INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

162-73

EXTENSIONS OF MARKOV MODELS OF SOCIAL MOBILITY TO HETEROGENEOUS POPULATIONS

Burton Singer and Seymour Spilerman

UNIVERSITY OF WISCONSIN ~ MADISON

EXTENSIONS OF MARKOV MODELS OF SOCIAL MOBILITY TO HETEROGENEOUS POPULATIONS*

Burton Singer

Columbia University and
Princeton University

Seymour Spilerman

University of Wisconsin

---

*The work reported here was supported by grants N00014-67-A-0151-0017 at Princeton University, NSF-GP-31505X at Columbia University, and by the Russell Sage Foundation and the Institute for Research on Poverty at the University of Wisconsin-Madison. The Institute's funds were provided by the Office of Economic Opportunity pursuant to the Economic Opportunity Act of 1964.

CONTENTS

EXTENSIONS OF MARKOV MODELS OF SOCIAL MOBILITY TO HETEROGENEOUS POPULATIONS

1. INTRODUCTION

There is an extensive and diverse literature on the application of discrete state Markov models to social processes. This formal structure has been an important analytic device in the study of occupational and industrial mobility (Matras 1960; Hodge 1966; Blumen, Kogan, and McCarthy 1955), income dynamics (Smith and Cain 1967; McCall 1971), and geographic migration (Rogers 1966; Tarver and Gurley 1965; Brown 1970). The key features of a problem which suggest the use of Markov models as a baseline or for projection are (a) a specified list of system states which may be occupations, industries, income categories, or geographic regions, (b) the availability of repeated observations on population movements among the states, and (c) an interest in the dynamics of the transition process.

In applications of Markov models it is frequently assumed (often implicitly) that the population can be considered to be homogeneous, and therefore is representable by a single Markov process. However, where investigators have obtained data at several time points so that n-step matrices could be observed and compared with the Markov predictions (i.e., checking the validity of $\hat{P}(n) = [\hat{P}(1)]^n$), it has often been found that the projections deviate from the observed values, and do so in a characteristic manner. Blumen, Kogan, and McCarthy (1955) were the first to suggest that the tendency of the Markov model to underpredict the main diagonal entries of an observed matrix can be attributed, in many research contexts, to population heterogeneity.

Blumen, Kogan, and McCarthy introduced the "mover-stayer" model to contend with this phenomenon. They postulated the presence of two types of persons: movers, who transfer according to a single Markov chain, and stayers, who remain permanently rooted in their origin states. Using this extremely restrictive form of the notion of heterogeneity, they constructed a simple mathematical model for the evolution of the total population and devised estimation procedures for its parameters. Recent work on extending the Markov framework to incorporate population heterogeneity in a more flexible manner has proceeded in two directions. The difference between them derives from the particular strategy adopted to accommodate heterogeneity; what they share in common is an assumption that the individual level process can be considered Markovian.

In one approach (McFarland 1970; Spilerman 1971a) each person is assumed to move according to a Markov chain, but follows a transition matrix which is unique to him. Population heterogeneity is therefore attributed to individual differences in the tendency to select particular destination states at a move. In the second approach (Ginsberg 1971; Spilerman 1972b), heterogeneity is accommodated by permitting individual differences in the rate at which transition events occur. At each transition, however, it is assumed that a single matrix common to all persons governs choice of destination. Thus, in this formulation, the burden of explaining heterogeneity is cast entirely upon variations in the rate of movement.

With regard to the utility of these extensions,[1] each provides a suitable framework for analyzing certain processes. Investigations into job mobility (Palmer 1954, p. 50) and geographic migration (Taeuber, Chiazze, and Haenszel 1968) have concluded that substantial individual differences exist in rate of movement. The second formulation would be appropriate for studying these processes and forecasting changes in their state distributions. In contrast, there are social phenomena for which population heterogeneity is primarily a consequence of individual differences in the probability of making particular transitions. Intergenerational occupational mobility is the most apparent instance; indeed, the very notion of different rates of movement seems inappropriate so that all population heterogeneity would have to arise from individual proclivities for certain transitions. A more detailed comparison of these complementary perspectives is presented in Spilerman (1972a).

In the present paper we present a unified framework in which to view the models mentioned above as well as more intricate social mobility models. The essence of our conceptual apparatus involves a formal distinction between the individual level or microscopic process, which is usually unobserved, and the population level or macroscopic process. An individual will be identified by a collection of rates which describe the average times he stays in particular states before moving, or by a stochastic matrix whose entries can be interpreted as propensities to favor transitions to certain states. It is also conceivable that an individual be classified by specifying <u>both</u> a matrix of rates of movement and a stochastic matrix listing probabilities of making particular transitions when a move occurs.

From a description of individual level behavior and a specification of the form which heterogeneity takes, we show how the population level process can be constituted. From a data analysis perspective, however, our situation is usually the reverse; we generally lack sufficient information to identify the individual level process. Instead, a researcher is constrained to sample the population at a few points in time and obtain counts of the number of persons making particular transitions, as well as other statistics concerning the population level process. From these observations we wish to infer the parameters of the unobserved, individual level process. One reason for recovering these parameters is that they can be used to reject a model by showing the implied, individual level description to be unrealistic for the problem at hand. Another reason is that they provide the basic ingredients for making statements about future trends in the population.

Our program is to first describe a class of Markov and semi-Markov processes which will serve as models for the evolution of individual behavior. In addition, the ideas involved in identification of non-directly observable parameters are illustrated by examples with the simplest Markov chain models (section 2). In section 3 we again proceed via a sequence of examples to show how the above mentioned population processes can be described mathematically. The basic mathematical structures characterizing the observable macroscopic level processes are mixtures of Markov and semi-Markov models. This notion is explained from the point of view of weighted averages of stochastic processes and from the alternative perspective of observable histories.

Finally we present some examples in section 4 of an "inverse problem" where gross macroscopic level information is used to obtain partial, and in a few instances complete, information about the mechanics of the individual level process. This aspect of our study involves an independent mathematical development which will be presented elsewhere in a joint paper by the authors. Our purpose here is simply to illustrate the ideas involved and communicate their relevance for the study of social mobility.

2. MARKOV CHAINS -- TRANSITION PROBABILITIES

2.1 Discrete time processes

The simplest mathematical caricature that we will employ to describe the evolution of an individual (or a homogeneous population) is a discrete time Markov chain. This stochastic process $\{X(k),\ k = 1,2,3,\ldots\}$ should be viewed as detailing state transitions by an individual, where the system states might be geographic regions, occupations, industries, or income categories, depending upon the particular substantive problem. Probability statements about the process are governed by the analytical recipe,

$$\text{Prob}\left(X(k+n) = j \,\middle|\, X(k) = i\right) = m_{ij}^{(n)}$$

for $k = 0,1,2,\ldots$, and $n = 0,1,2,\ldots$. The element $m_{ij}^{(n)}$ is the (i,j) entry in the stochastic matrix M^n (n-fold matrix multiplication of M). M is itself a stochastic matrix and describes single transitions by an individual; its (i,j) entry m_{ij} has the interpretation, "probability of moving from state i to state j in one step."

The typical empirical setting in which mobility data are gathered does not allow the matrix M to be estimated directly from the movements of an individual. An investigator usually observes the locations of many persons at a few time points $n_0 = 0$, $n_1, n_2, \ldots$, and estimates stochastic matrices $\hat{P}(n_1)$, $\hat{P}(n_2), \ldots$, where the (i,j) entry in $\hat{P}(n_k)$ denotes[2] the proportion of individuals from among those in state i at time n_0 who are in state j at time n_k.

One assumption which is frequently made is that the population is homogeneous in its movements. This permits the matrix $\hat{P}(n_k)$, estimated from the observed locations of the population at times n_0 and n_k, to be associated with the evolution of a single individual. We also make this assumption in the present section. A second specification which is often adopted is to identify the smallest observational time interval (n_0,n_1) with the unit interval of the process. Thus, we might define $M = \hat{P}(n_1)$ in which case, if $n_k = hn_1$ for some integer h, we have

$$P(n_k) = \hat{P}(n_1)^h = M^h .$$

The above identification is often used when a discrete time mathematical model is desired for a process evolving continuously in time but having no natural unit time interval which can be associated with it. On the other hand, when a discrete sequence of times can be identified as spaced at substantively meaningful intervals, it sometimes happens that the unit time interval is smaller than the minimum observation interval. That is, while we observe a population at times n_0 and n_1, the unit interval of the process could be $(n_0,[1/h]n_1)$ for some integer $h > 1$. A problem of this sort might arise if we collected data on respondent's occupation $(n_1 = 2)$ and grandfather's occupation $(n_0 = 0)$, but neglected to obtain information on father's occupation. Given the observed $\hat{P}(2)$ matrix we might inquire into whether it can be represented in the form $\hat{P}(2) = M^2$ for some stochastic matrix M. A question of this very nature was posed by Robert W. Hodge (1966) in an inquiry into the extent of inheritance of occupational status beyond a single generation.

The problem of deciding whether a particular observed matrix $\hat{P}(n)$ may be represented in the form $\hat{P}(n) = M^n$ for some stochastic matrix M is the simplest version of what we will refer to as the <u>embeddability problem</u>. This is equivalent to asking whether or not the observed data are compatible with a discrete time Markov model. If the answer to this question is affirmative then we require a procedure for recovering the one-step transition matrix M. This is a statement of the <u>inverse problem</u>. In the present context it entails tabulating the n-th roots of $\hat{P}(n)$ in order to determine all transition matrices M which are compatible with our observations.

To illustrate these ideas in a simple mathematical setting suppose we are in a substantive situation which allows observations to be made only at times $n_0 = 0$ and $n_1 = 3$, and that we estimate the stochastic matrix

$$\hat{P}(3) = \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix} . \qquad (2.1)$$

Computing cube roots of this matrix we find that

$$[\hat{P}(3)]^{1/3} = \begin{pmatrix} \frac{1}{2}[1 - \frac{1}{\sqrt[3]{3}}] & \frac{1}{2}[1 + \frac{1}{\sqrt[3]{3}}] \\ \frac{1}{2}[1 + \frac{1}{\sqrt[3]{3}}] & \frac{1}{2}[1 - \frac{1}{\sqrt[3]{3}}] \end{pmatrix} = M \qquad (2.2)$$

is the unique stochastic matrix which is a cube root of $\hat{P}(3)$. Thus the empirically determined matrix $\hat{P}(3)$ is embeddable in a discrete time Markov chain with one-step transition matrix M given by (2.2).

In general, we will be able to ascertain embeddability without having to compute the n-th roots of $\hat{P}(n)$ [see section 4]. If the matrix is found to be embeddable in a discrete time Markov process then we should be able to identify at least one stochastic matrix M such that $\hat{P}(n) = M^n$. In the above example, the inverse problem reduced to calculating all cube roots of $\hat{P}(3)$ which are stochastic matrices. An explicit analytical recipe for calculating the roots of an arbitrary stochastic matrix is described in Appendix I.

If the natural time scale of the problem is such that the matrix (2.1) corresponds to observations at times n = 0 and n = 2, then the process is incompatible with a discrete time Markov chain theory.[3] This is because (2.1) has no square roots which are stochastic matrices; thus the observations in this new time scale cannot be represented in the form $\hat{P}(2) = M^2$, and the process is not embeddable in a discrete time Markov chain structure.

A further illustration of the complexity of inverse problems can be seen from the following mathematical example. Suppose you initially take observations at times n = 0 and n = 2 and that you estimate

$$\hat{P}(2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

On the surface it might appear as though you are observing a population in which there is no mobility between states. However, $\hat{P}(2)$ has three distinct square roots all of which are stochastic matrices:

$$\sqrt{\hat{P}(2)} = \left\{ \begin{array}{l} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \right.$$

Each square root has a different substantive interpretation, and only the first of these (i.e., the identity matrix) corresponds to no mobility. The second matrix in the above list corresponds to a situation in which individuals starting in either state 1 or state 3 cycle back and forth between these states while an individual starting in state 2 never moves. The third matrix may be identified with a population in which individuals either cycle back and forth between states 1 and 2 or remain stationary in state 3. Discrimination between these alternatives may be either on substantive grounds or on the basis of a further set of observations.

If additional observations are possible, they should be taken at one of the times 3,5,7,..., etc. and <u>not</u> at an even time period. In particular, if $\hat{P}(2)$ is the identity and a Markov model adequately describes the process then the pair of matrices $(\hat{P}(2), \hat{P}(3))$ must be one of the following distinct sets:

$$S_1: \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$S_2: \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$S_3: \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

while the pair $(\hat{P}(2), \hat{P}(4))$ must be

$$S_4: \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The set S_1 is compatible with a unique Markov chain model with $M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Similarly, S_2 is only compatible with

$$M = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

and S_3 is only compatible with

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

On the other hand, S_4 is consistent with all three models. This indicates that it is often desirable to take observations at time points which are not evenly spaced if you want to discriminate between substantively distinct stochastic models all compatible with data from a few periods.

A further point is that if $(\hat{P}(2), \hat{P}(3))$ with $\hat{P}(2)$ equal to the identity deviates from all of the three sets S_1, S_2, S_3 then the data are not consistent with a discrete time Markov structure. You can also rule out a Markov model by obtaining empirical matrices $(\hat{P}(2), \hat{P}(4))$ with $\hat{P}(2)$ equal to the identity but $\hat{P}(4)$ being a stochastic matrix not equal to the identity. These examples illustrate the importance of not regarding compatibility of observations with a particular model during a small number of time periods as strong evidence for validity of the model as a description of the underlying process. Nevertheless, examination of alternative solutions of the inverse problem based on observations at a few time points can be a useful exploratory tool for calling attention to possibly unsuspected mobility mechanisms M and for suggesting more realistic models.

2.2. Continuous time processes

The natural time scale for many mobility processes is not a discrete sequence of intervals such as generations or decades, but a continuum of time points. In particular, geographic migration (Brown 1972; Spilerman 1972b) and occupational mobility (Blumen, Kogan and McCarthy 1955) can be viewed more realistically as processes in which state changes occur at random time points, and probabilities of moves between particular states are governed by Markov transition matrices. Several extensions of this formulation which are appropriate for heterogeneous environments appear in section 3. In the present discussion we establish a framework for the extensions by indicating three alternative descriptions of continuous time Markov chains; the level of generality and substantive utility of each is delineated.

I. Q-Matrices

Consider a stochastic process with a finite number of states whose transition probabilities are governed by the system of ordinary differential equations

$$\frac{dP(t)}{dt} = QP(t), \qquad P(0) = I \tag{2.3}$$

where Q is an r x r matrix whose entries satisfy $-\infty < q_{ii} < 0$, $q_{ij} \geq 0$ for $i \neq j$, $\sum_{j=1}^{r} q_{ij} = 0$, and r = number of states. The system (2.3) has a unique solution given by the exponential formula

$$P(t) = e^{tQ}, \qquad t > 0 \tag{2.4}$$

and the matrices Q and P(t) have the following substantive interpretations:

$$\begin{cases} -q_{ii}dt = \text{probability that an individual in state } i \text{ at time } t \text{ exits from that state during the time interval } (t, t+dt) \\ q_{ij}dt = \text{probability that an individual in state } i \text{ at time } t \text{ moves to state } j\ (j \neq i) \text{ during the interval } (t, t+dt) \end{cases}$$

$p_{ij}(t)$ = probability that an individual starting in state i at time 0 is in state j at time t [(i,j) entry in P(t)].

Mobility processes whose transition probabilities are governed by (2.3) have the property that their state at time t+s given the complete history of the process up to time t is only dependent on the last observation, namely the state at time t. This is a statement of the Markov property. Furthermore, the holding time until exiting from a particular state i is exponentially distributed with parameter $-q_{ii}$. This is the most general formulation of continuous-time finite-state Markov chains arising in social mobility studies.

When continuous histories are available on all population movements during the time interval (0,t), the matrix Q can be estimated directly from the observed transitions. A maximum likelihood procedure has been reported by Meier (1955) and involves the following calculations:

$$\begin{cases} \hat{q}_{ij} = \dfrac{N_{ij}}{A_{i.}} & \text{for } j \neq i \\ \hat{q}_{ii} = -\sum\limits_{j\neq i} \hat{q}_{ij} & \text{for } j = i \end{cases}$$

where $A_{i.}$ = total occupation time in state i during (0,t) by all individuals in the population

N_{ij} = total number of transitions from state i to state j during (0,t).

In addition, availability of individual histories allows the suitability of a continuous time Markov model to be examined on several different grounds.

(i) Estimate Q from individual histories up to time t, compute $e^{(t+s)\hat{Q}}$, and compare this theoretically-based transition matrix with the observed matrix $\hat{P}(t+s)$. This is a check on compatibility of the data with the mathematical structure (2.3) and (2.4).

(ii) Check the goodness of fit of an exponential distribution to the holding time distribution in each state.

(iii) Assess the strength of longer range dependence on past history if there are sufficient time series records.

The above general formulation of continuous time Markov transition matrices has been used in numerous sociological contexts (e.g., Coleman 1964, pp. 177-182; Bartholomew 1967, pp. 77-78). However, the analysis of social mobility in a heterogeneous population is greatly facilitated by the alternative formulations presented in the next two sections. In particular, they provide the basis for a classification scheme which allows individuals (or sub-populations) to be characterized according to either their rate of movement, their propensity to move to particular states, or both simultaneously. This kind of classification also leads to a straightforward mathematical caricature of mobility processes in a heterogeneous context when individual histories are not directly observable. Nevertheless, even in the present formulation, substantial information about individual level behavior can be inferred by an extension of the inverse problem arguments already presented and solving (2.4) for Q. This analysis is described in section 4.

II. Subordination

A starting point for the development of mobility models appropriate to a heterogeneous population is to consider Q-matrices of the special form $Q = \lambda(M-I)$, where λ is a positive constant and M is a stochastic matrix. Populations in which transition probabilities evolve according to $e^{t\lambda(M-I)}$ may be given the following substantive interpretation. An individual starting in state i at time 0 stays there for an exponentially distributed length of time τ_0 with

$$\text{Prob } (\tau_0 \geq t) = e^{-\lambda t}, \qquad t > 0.$$

At the end of this period he makes a transition to state j with probability m_{ij}, the (i,j) entry in the stochastic matrix M. It is not assumed that $m_{ii} = 0$; hence, the individual may have a positive probability of staying in the same state[4] after time τ_0. Once an individual has moved according to M, he stays in his new state for another exponentially distributed length of time τ_1 which is independent of τ_0 and of the state he is in, but again satisfies

$$\text{Prob } (\tau_1 \geq t) = e^{-\lambda t}, \qquad t > 0.$$

Now he makes another move according to m_{jh}, and the previous process of waiting an exponentially distributed length of time and moving according to the entries in M is repeated.

In this imagery the constant $\frac{1}{\lambda}$ describes an individual's mean waiting time before moving (or before making a decision to possibly move), while M characterizes his propensity to transfer to particular states. An alternative formulation of the process derives from an interpretation of λ as measuring rate of movement. The random variables $\{Y(t), t > 0\}$ which describe an individual's history may be written in the special form

$$Y(t) = X(T_\lambda(t)) \qquad (2.5)$$

where Prob $(Y(t) = j | Y(0) = i)$ is the (i,j) entry in $P(t) = e^{t\lambda(M-I)}$, $T_\lambda(t)$ is a Poisson process with parameter λ, and $X(k)$, $k = 0,1,2,\ldots$ is a discrete time Markov chain with one-step transition probabilities governed by M.

Representation of stochastic processes by a recipe such as (2.5) is known in the mathematics literature as subordination (Feller 1971, pp. 345-349); more precisely, the process $Y(t)$ is said to be subordinated to $X(k)$ using $T_\lambda(t)$ as an operational or intrinsic clock. Elaborating further, expression (2.5) says that you can also think of individuals whose transition probabilities are governed by $e^{t\lambda(M-I)}$ as evolving according to the prescription:

(i) Wait in an initial state i until the first jump time of a Poisson process.

(ii) At this instant, change state once according to the laws of a discrete time Markov chain whose one-step transition matrix is M.

(iii) Wait in the new state j until the Poisson process jumps for the second time.

(iv) Now change state again according to M.

(v) Repeat the above procedure for successive Poisson jump times.

In this setting, the constant λ governing rate of movement appears only in the description of the Poisson process or, if you like, the intrinsic clock. Specifically,

$$\begin{aligned} &\text{Prob (exactly } v \text{ moves up to time } t) \\ &= \text{Prob (exactly } v \text{ jumps in a Poisson process with parameter } \lambda \text{ during } (0,t)) \\ &= \text{Prob } (T_\lambda(t) = v) = \frac{(\lambda t)^v e^{-\lambda t}}{v!} . \end{aligned}$$

Likewise, the matrix M governing the propensity to move to particular states appears only in the description of the discrete time Markov chain $X(k)$ according to

$$\text{Prob}\left(X(k+1) = j \mid X(k) = i\right) = m_{ij} .$$

This matrix provides a static picture of the population at an instant of movement. The dynamics are regulated by the intrinsic clock $T_\lambda(t)$.

A final point is that the interpretation (2.5) gives rise to a series representation for $e^{t\lambda(M-I)}$ in the form

$$e^{t\lambda(M-I)} = \sum_{v=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^v}{v!} M^v . \tag{2.6}$$

The individual terms in this series indicate that rate of movement and propensity to move to particular states are factors regulated by two independent sociologically identifiable quantities, λ and M. It is useful to contrast this isolation of rate of movement and propensity to move in the factored matrix λ(M-I) with the more general formulation involving Q-matrices. The two formulations are related according to

$$\begin{cases} -q_{ii}dt = -\lambda(m_{ii}-1)dt = \text{probability of leaving state } i \text{ during the time interval } (t,\ t+dt) \\ q_{ij}dt = \lambda m_{ij}dt = \text{probability of a move from state } i \text{ to state } j\ [j \neq i] \text{ during } (t,\ t+dt). \end{cases}$$

We wish to emphasize that mobility processes in a stationary environment are most usefully described by waiting times in states together with transition probabilities which are independent of the waiting times, but allow for the possibility of remaining in the current state at a move. Q-matrices permit descriptions in terms of transition mechanisms and waiting time distributions as independent quantities, but this would naturally take the form

$$Q = \begin{pmatrix} -q_{11} & & \bigcirc \\ & \ddots & \\ \bigcirc & & -q_{rr} \end{pmatrix} \left[\begin{pmatrix} 0 & \ddots & \frac{q_{ij}}{-q_{ii}} \\ & \ddots & \quad i \neq j \\ & & 0 \end{pmatrix} - \mathbf{I} \right].$$

The matrix $\left\| \frac{q_{ij}}{-q_{ii}} \right\|_{i \neq j}$ describes jumps without allowing for further holding in the origin state; in particular,

$$\frac{q_{ij}}{-q_{ii}} = \{\text{probability of a move from state } i \text{ to state } j \text{ given a departure}\}.$$

The diagonal matrix $\left\| -q_{ii} \right\|$ has entries interpretable as holding time rates in the sense that

$$\text{Prob}\left(Y(u) = i,\ s < u < s + t \middle| Y(s) = i\right) = e^{q_{ii}t}.$$

Nevertheless, this formulation seems less sociologically meaningful than expressing Q in the factored form $\lambda(M-I)$. If you start with a fixed Q, which means that you are viewing Q as the basic ingredient in the model, infinitely many such factorizations are possible. However, our attitude here is that the natural starting point for a description of mobility processes is λ and M, with Q defined in terms of these basic ingredients.

The present description of mobility in which an individual is characterized by a rate λ, a transition matrix M, and evolves according to probabilities $e^{t\lambda(M-I)}$ was the starting point for Spilerman's (1972b) extension of the mover-stayer model to a population with a continuum of types. A limitation of this extension is that the waiting time distributions until transition according to M do not depend on a person's state; rather, an individual is compelled to move at the same rate irrespective of his location (Spilerman 1972b, p. 609). The description of section III provides a basis for eliminating this restriction and presents the most general continuous time Markov chain formulation of mobility by an individual.

III. A General Factored Representation of Q

Consider Q-matrices of the form $\Lambda(M-I)$ where

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \bigcirc \\ & \lambda_2 & & \\ & & \ddots & \\ \bigcirc & & & \lambda_r \end{pmatrix}, \quad \lambda_i > 0 \text{ for } i = 1,2,\ldots,r$$

and M is a stochastic matrix. The evolution of individuals governed by the transition matrix $e^{t\Lambda(M-I)}$ may be described more concretely as follows:

(i) An individual starting in state i at time 0 stays there for an exponentially distributed length of time τ_0 with

$$\text{Prob}\ (\tau_0 \geq t | X(0) = i) = e^{-\lambda_i t}, \qquad t > 0.$$

(ii) At the end of this time period he makes a decision to move to state j with probability m_{ij}. In general, $m_{ii} \neq 0$.

(iii) Now he waits in state j for an exponentially distributed length of time τ_1 with

$$\text{Prob}\ (\tau_1 \geq t | X(1) = j) = e^{-\lambda_j t}.$$

(iv) Then he makes another decision to move to state h with probability m_{jh}.

(v) The above sequence is repeated.

The waiting times τ_k only depend on the current state X(k), not on the past history[5] X(0), X(1)...,X(k-1) and $\tau_0, \tau_1, \ldots, \tau_{k-1}$. However an individual's rate of movement is now characterized by the diagonal matrix Λ rather than a single constant λ as in the previous section. The present description reduces to that of section II when $\lambda_1 = \lambda_2 = \ldots. = \lambda_r$.

Again, it should be observed that a given Q-matrix has infinitely many factorizations of the form Λ(M-I) with

$$\begin{cases} q_{ii} = \lambda_i(m_{ii}-1) & \\ q_{ij} = \lambda_i m_{ij} & \text{for } i \neq j \end{cases}$$

However, the basic ingredients of mobility models with individuals evolving according to continuous time Markov chains are the matrices Λ and M, with Q defined in terms of them. When individual histories are available, M and Λ can be estimated from observed movements according to the recipe

$$\begin{cases} \hat{m}_{ij} = \dfrac{N_{ij}}{N_{i.}} \\ \hat{\lambda}_i = \dfrac{N_{i.}}{A_{i.}} \end{cases} \qquad (2.7)$$

where N_{ij} = number of transitions from i to j during the observation period (0,t)

$N_{i.} = \sum_j N_{ij}$ = total number of moves from state i (including within-state moves) during (0,t)

$A_{i.}$ = total occupation time in state i by all individuals during (0,t).

Our representation of the random variables Y(t) in section II, which describe individual histories by

$$Y(t) = X(T_\lambda(t))$$

with $T_\lambda(t)$ a Poisson process, does not carry over to the more general formulation indicated here. We could, in principal, write

$$Y(t) = X(T^*(t))$$

where

$$T^*(t) = \{\text{number of transitions up to time } t \text{ in a Markov chain governed by } e^{t\Lambda(M-I)}\};$$

however, $T^*(t)$ does not have a simple family of formulas analogous to the Poisson distribution describing its evolution. Thus we will retain our first interpretation of alternating exponential holding times and decisions to possibly move as the simplest generic caricature of mobility for an individual.

From an analytical point of view the simplicity of the Poisson series representation

$$e^{t\lambda(M-I)} = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} M^k$$

with the terms describing rate of movement and those describing transitions appearing as separate multiplicative factors does not carry over to our more general formulation. In particular

$$e^{t\Lambda(M-I)} \neq e^{t\Lambda M} e^{-t\Lambda} \neq \sum_{k=0}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} M^k .$$

The nonequivalence of the three expressions is due to the fact that Λ and M do <u>not</u> commute, i.e., $\Lambda M \neq M\Lambda$, and it is this algebraic point which makes computations with $e^{t\Lambda(M-I)}$ considerably more difficult than with the model of section 2.II. We will return to this issue again in section 3 when we compare this description of individual level mobility in a heterogeneous population with a description where $\lambda_1 = \lambda_2 = \ldots = \lambda_r$.

3. HETEROGENEITY AND MIXTURES OF STOCHASTIC PROCESSES

3.1. Mixtures of Markov processes and inversion formulas

Thus far we have discussed Markov chain models as they pertain to repeated moves by a single individual or to the movements of a homogeneous population. In the context of social mobility, observable populations are rarely homogeneous with respect to the frequency with which individuals move or their propensity to transfer to particular states. However, in early studies (Prais 1955; Matras 1960; Tarver and Gurley 1965), it was tacitly assumed that the population under consideration could be viewed as a homogeneous unit and that histories associated with a single Markov process could be thought of as typical of all segments of the population.

In using discrete time Markov transition matrices as a baseline for comparison with particular data sets, a standard strategy is to estimate an n-step transition matrix $\hat{P}(n)$ from the data, calculate the n-th roots[6] $[\hat{P}(n)]^{1/n} = M$ viewing these as one-step transition matrices, and then compare M^{n+k} with the observed matrix $\hat{P}(n+k)$. For many social phenomena a substantial discrepancy was noted between M^{n+k} and $\hat{P}(n+k)$, and it was suggested that this is because the population should really be viewed as heterogeneous with different stochastic processes describing the evolution of different sub-populations. The first detailed discussion of this kind of inadequacy of Markov models to depict social mobility was by Blumen, Kogan, and McCarthy (1955). They documented the phenomenon of "lumping on the main diagonal" (the presence of more individuals in these cells of an n-step transition matrix than predicted by a Markov model), and showed that it can derive from treating a heterogeneous population as though it was homogeneous. Furthermore, they constructed a discrete time model in which two types of persons were distinguished, each evolving independently according to a different Markov process. Efficient estimation procedures for the parameters of this model were subsequently developed by Leo Goodman (1961).

The analytic context for which the mover-stayer model was developed is one where evolution of each distinct type of individual is not directly observable. An investigator is constrained to sample the total pooled population (also referred to as a macroscopic level description) at a few time points and obtain counts of the number of individuals starting in a particular state who are in any other state at the end of a sampling interval. This situation is typical of all mobility environments discussed in the present work. A key step in understanding the underlying mobility process and the appropriateness of particular models, then, is an identification of the non-directly observable quantities (one-step transition matrices and rates of movement for continuous time processes) by a mathematical analysis relating information about the pooled population back to the behavior of individuals. This is another instance of an inverse problem, analogous to the discussion of section 2.1 but complicated by the fact that we are treating several types of individuals simultaneously.

In order to clarify these ideas, we present four examples which form the simplest mathematical caricatures of the notion of a mixture of stochastic processes and which are also substantively meaningful in the context of social mobility.

Example 1:

Consider a population consisting of two kinds of individuals. Persons of each type evolve independently according to a discrete time

Markov chain. We denote by $X_1(k)$ and $X_2(k)$, $k = 0,1,2,\ldots$, the random variables describing the movements among states by persons of each type. Probability statements about $X_1(k)$ and $X_2(k)$ are assumed to follow the theoretical recipe

$$\text{Prob } (X_\lambda(k) = j \mid X_\lambda(0) = i) = m_{ij;\lambda}^{(k)}$$

for $k = 0,1,2,\ldots$, and $\lambda = 1$ or 2. The element $m_{ij;\lambda}^{(k)}$ is the (i,j) entry in the stochastic matrix M_λ^k (k-fold matrix multiplication of M_λ). We will refer to the bivariate process $(X_1(k), X_2(k))$ as a microscopic or individual-level description of a mobility process.

In empirical situations we usually observe values of a random variable $Y(k)$ which are possible states of either the process $X_1(k)$ or $X_2(k)$; that is, we can observe how an individual sampled from the population evolves through time although we cannot assign him to a particular person type. We also assume that we can estimate

$$s_i = \{\text{proportion of individuals from among those in state } i \text{ at time } 0 \text{ who are classified as type } 1\}.$$

Procedures for estimating s_i have been reported by Blumen, Kogan, and McCarthy (1955) and Goodman (1961).

Evolution of the stochastic process $Y(k)$, $k = 0,1,2,\ldots$, is described by the transition probabilities

$$\text{Prob } (Y(k) = j \mid Y(0) = i) = s_i m_{ij;1}^{(k)} + (1-s_i) m_{ij;2}^{(k)}$$

or in matrix form

$$P(k) = SM_1^k + (I-S)M_2^k \tag{3.1}$$

with

$$S = \begin{pmatrix} s_1 & & & \bigcirc \\ & s_2 & & \\ & & \ddots & \\ \bigcirc & & & s_r \end{pmatrix}.$$

The univariate process Y(k) is referred to as a macroscopic or population-level description of a mobility process. Its interpretation in this sense arises from the fact that an observer who can only see histories Y(0), Y(1), Y(2),..., and the family of matrices P(1), P(2), P(3),..., cannot discover that in fact Y(k) is generated by a composite of two types of individuals evolving according to $X_1(k)$ and $X_2(k)$ respectively. However, once the interpretation of a heterogeneous population is brought in, you can formulate a theory of evolution of a mixture of two types of individuals as in equation (3.1), with Y(k) describing the composite or pooled population. Hence the term, "mixture of stochastic processes."

The particular theory (3.1) describes a population in which the mixing distribution remains constant through time and is identified with the proportions of individuals in each state (which may be a job category or a geographic region) who are of type 1 at the reference time k = 0. This theoretical description reduces to the classical mover-stayer model when $M_1 = I = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ & & \ddots \\ 0 & & & 1 \end{pmatrix}$. It should also be emphasized that the macroscopic level process Y(k) is not Markovian even though the components of the pooled population are assumed to evolve according to discrete time Markov processes.

With this theoretical picture at hand, we illustrate the notions of embeddability and inverse problem for mixtures of stochastic processes within the context of the mover-stayer model. For the simplest mathematical structure consider a 2-state mobility process where you are constrained to observe the pooled population at times $n = 0$ and $n = 2$. Denote the 2×2 stochastic matrix estimated from the data by

$$\hat{P}(2) = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} \qquad 0 < a,b < 1 .$$

The entry in row i, column j (i,j = 1 or 2) has the interpretation, "proportion of individuals in state i at time 0 who are in state j at time 2." Our first task is to determine necessary and sufficient conditions so that $\hat{P}(2)$ is compatible with the theoretical description

$$P(2) = S + (I-S)M^2 , \qquad (3.2)$$

To further simplify the calculations while still retaining a substantively meaningful description of a pooled population of movers and stayers, suppose that $s_1 = s_2 =$ some $s \geq 0$. Then solve the matrix equation (3.2) for M and obtain

$$M = [(I-S)^{-1} (P(2)-S)]^{1/2} , \qquad (3.3)$$

Replacing the theoretical P(2) by the empirically determined $\hat{P}(2)$ we can check, using the calculations in Appendix I, that the inversion formula (3.3) (which is the solution of the inverse problem) will yield a legitimate stochastic matrix M if and only if $a + b > 1 + s$. Putting this another way, an empirically determined 2×2 matrix $\hat{P}(2)$ is said to be embeddable in a mover-stayer framework if and only if the inequality is satisfied. It is also important to notice that if you alter what you regard as the natural time scale of the mobility process so that the matrix $\begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}$ is thought of as $\hat{P}(n)$ for <u>any</u> even number n, then the condition $a + b > 1 + s$ is still necessary and sufficient for the data to be compatible with the theoretical framework of (3.1). In fact, in the 2 x 2 case this condition also ensures a unique inverse.

The criterion $a + b > 1 + s$ becomes more meaningful if you recall that the original data sets examined by Blumen, Kogan, and McCarthy gave rise to $\hat{P}(n)$ having diagonal elements larger than those predicted by a simple Markov chain model. For a two state process a criterion for $\hat{P}(n)$, n even,[7] to be compatible with the Markov structure M^n is just $a + b > 1$; however, once you postulate two types of individuals, stayers and movers, in proportions s and 1-s respectively, you are describing evolution via transition matrices whose diagonal elements must be larger than the corresponding one-type Markov model by precisely the mixing fraction s. An analogous condition also holds for processes with more than 2 states and for rather general matrices $S = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_r \end{pmatrix}$. A full mathematical discussion, however, is somewhat intricate and will appear in Singer and Spilerman (1973). A final point with regard to the criterion $a + b > 1 + s$ is that even if you cannot estimate the mixing fraction s directly, you can still indicate the largest possible value of s which allows the matrix $\hat{P}(n)$ to be compatible with a mover-stayer theoretical framework; namely maximum s such that $1 + s < a + b$.

Example 2:

Let $X_1(k),\ldots,X_N(k)$ be independent discrete time Markov chains describing the movements of N types of individuals each evolving according to a distinct stochastic matrix M_λ, $\lambda = 1,\ldots,N$. For individuals of each type,

$$\text{Prob}\ (X_\lambda(k) = j\,|\,X_\lambda(0) = i) \;=\; m_{ij;\lambda}^{(k)}$$

where $m^{(k)}_{ij;\lambda}$ is the (i,j) entry in M^k_λ. Let

$$S_\lambda = \begin{pmatrix} s_{\lambda 1} & & \bigcirc \\ & \ddots & \\ \bigcirc & & s_{\lambda r} \end{pmatrix}, \qquad \sum_{\lambda=1}^{N} S_\lambda = I,$$

where $s_{\lambda i}$ = {proportion of the population in state i at time 0 which consists of type λ persons}. Analogous to example 1 the individual or microscopic level vector process $(X_1(k),\ldots,X_N(k))$, k = 0,1,2,..., is not directly observable; however, the pooled population is observable at a few time points and it is from this data that information about the matrices $M_1,\ldots,M_N$, postulated by the theoretical structure, must be inferred. The population level process may be described by a family of random variables Y(0), Y(1),..., whose values are the possible states of each of the individual sub-populations. The evolution of Y(k) is governed by the stochastic matrices

$$P(k) = \sum_{\lambda=1}^{N} S_\lambda M^k_\lambda, \qquad k = 0,1,2,\ldots .$$

This formulation of a mixture of Markov processes can be extended to the case where each individual in the population has his own M_λ matrix. This approach was suggested by McFarland (1970) in recognition of the fact that heterogeneity in social mobility is attributable to an assortment of individual differences--in race, ethnicity, parental SES, educational attainment, and so forth. Spilerman (1972a) has presented a regression method for estimating the individual M_λ matrices from an observed population level matrix $\hat{P}(1)$ and data on the determinants of heterogeneity. Because of its complexity, a discussion of embeddability and inversion methods for this model will be deferred to the companion mathematical paper (Singer and Spilerman 1973).

Example 3:

Let $\{X_\lambda(t)\}_{\lambda>0}$ be a continuum of independent continuous time Markov chains whose transition probabilities are governed by the exponential formula

$$e^{t\lambda(M-I)} , \qquad \lambda > 0,\ t \geq 0 . \qquad (3.4)$$

The processes $\{X_\lambda(t)\}_{\lambda>0}$ should be thought of as describing the evolution of infinitely many different types of individuals, each type being identified by a number λ which specifies its rate of movement. For a fixed value of λ (one type of person) this is just the continuous time Markov chain model described in (2.II). Now, however, we envision a heterogeneous population where a type-λ individual has waiting times between moves which are exponentially distributed with parameter λ, independent of his previous state. All types of individuals are treated as having the same propensity to move among the states, prescribed by the matrix M.

As in our previous examples, the vector process $\{X_\lambda(t)\}_{\lambda>0}$ is not directly observable, but we postulate that type-λ individuals occur in the total population with a frequency described by a probability density $g(\lambda)$. Then the observable macroscopic level process, which consists of the mixture (or pooling) of all types of individuals, can be described by random variables $Y(t)$, $t \geq 0$ whose values are the possible states of the component types $\{X_\lambda(t)\}_{\lambda>0}$, and whose transition probabilities are governed by the mixture of Markov transition matrices

$$P(t) = \int_0^\infty e^{t\lambda(M-I)} g(\lambda)d\lambda. \qquad (3.5)$$

The entries of $P(t)$ have the usual interpretation,

$$p_{ij}(t) = \text{Prob}\ (Y(t) = j | Y(0) = i).$$

This formulation may be viewed as an extension of the mover-stayer model, and it was developed in that light by Spilerman (1972b). In applications, $g(\lambda)$ is commonly specified as a gamma density

$$g(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta\lambda}, \qquad \alpha,\beta > 0 \qquad (3.6)$$

because of the ability of this functional form to describe a variety of unimodal curves, unimodality being a reasonable characterization of the frequency of occurrence of different types of persons (with respect to rate of mobility) in heterogeneous populations (Palmer 1954, p. 50; Taeuber, Chiazze and Haenszel 1968, p. 46).

Subject to the hypothesis (3.6), the integral (3.5) may be evaluated as

$$P(t) = \left(\frac{\beta}{\beta+t}\right)^\alpha \left[I - \left(\frac{t}{\beta+t}\right)M\right]^{-\alpha} \qquad (3.7)$$

where the $-\alpha^{\text{th}}$ root of the above matrix is defined by the power series in M,

$$P(t) = \lim_{n \to \infty} \sum_{k=0}^{n} \left(\frac{\beta}{\beta+t}\right)^\alpha \left(\frac{t}{\beta+t}\right)^k M^k . \qquad (3.8)$$

Equation (3.7) is amenable to two substantive interpretations depending upon the role assigned to the parameters α and β:

Formulation 1. View the population as heterogeneous with the gamma family of distributions describing the proportion of individuals of type-λ in the total population. Then the macroscopic level process $Y(t)$ may be represented as

$$Y(t) \quad = \quad X(T_{(\alpha,\beta)}(t)) \tag{3.9}$$

where $X(k)$, $k = 0,1,2,\ldots,$ is a discrete time Markov chain with one-step transition matrix M and $T_{(\alpha,\beta)}(t)$ is a negative binomial process acting as the intrinsic clock for the pooled population. $T_{(\alpha,\beta)}(t)$ may be thought of as a Poisson process with gamma distributed parameter. It has the probability distribution

$$\text{Prob } (T_{(\alpha,\beta)}(t) = v) \quad = \quad \binom{a+v-1}{v} \left(\frac{t}{\beta+t}\right)^{v} \left(\frac{\beta}{\beta+t}\right)^{\alpha} \tag{3.10}$$

in which the term for v denotes the proportion of the population making exactly v transitions in the time interval $(0,t)$. This is just another instance of the notion of subordination already discussed in section 2.II, except that now $X(T_{(\alpha,\beta)}(t))$ is no longer a Markov process.

From the point of view of available data, information on $T_{(\alpha,\beta)}(t)$ often can be obtained without collecting individual histories. For instance, data on the distribution of number of moves in mobility processes are reported in Palmer (1954, p. 50) and Lipset and Bendix (1959, p. 158). Using the mean and variance of the negative binomial variate $T_{(\alpha,\beta)}(t_1)$, where t_1 is an arbitrary observation time, estimates of the gamma parameters α and β can be computed from these observed population distributions. This procedure was employed by Spilerman (1972b, p. 614) in an example using data which was artificial, but constructed to simulate the nature of heterogeneity in occupational mobility. He reports the values $\alpha = 1.37$, $\beta = 0.92$ whose ramifications we will now discuss.

The value $\alpha = 1.37$ is especially interesting[8] if you consider the following alternative description of the negative binomial process.

(i) Consider a special clock starting in position 0 and remaining there for a random length of time τ_0 governed by[9]

$$\text{Prob } (\tau_0 \geq t) = \left(\frac{\beta}{\beta+t}\right)^{\alpha}, \qquad t > 0.$$

At the end of the epoch τ_0 the special clock moves to position 1.

(ii) The special clock remains in position 1 for a random length of time τ_1 independent of τ_0 but having the same distribution:

$$\text{Prob } (\tau_1 \geq t) = \left(\frac{\beta}{\beta+t}\right)^{\alpha}, \qquad t > 0.$$

At the end of the epoch τ_1, the special clock moves to position 2.

(iii) The above steps are repeated and at any particular time, the position of the special clock should be identified with the number of moves up to time t of a negative binomial process $T_{(\alpha,\beta)}(t)$.

The important point about the waiting time distribution for $\tau_0, \tau_1, \ldots,$ is that with $1 < \alpha < 2$ it has an infinite variance (and finite mean). This means that a substantial portion of the pooled population moves very rarely or not at all, which is another way of saying that there is considerable heterogeneity.

<u>Formulation 2</u>. View the population as <u>homogeneous</u> but evolving according to the following recipe.

(i) An individual starting in state i remains there for a length of time τ_0 governed by the 2-parameter family of distributions

$$\text{Prob}\ (\tau_0 \geq t) = \left(\frac{\beta}{\beta+t}\right)^{\alpha} .$$

At the end of the epoch τ_0 he makes a decision to move to state j (j may be equal to i) according to the stochastic matrix M.

(ii) The individual remains in state j for a new random length of time τ_1 independent of τ_0 but having the same distribution. At the end of the epoch τ_1 he again makes a decision to move according to the stochastic matrix M.

(iii) The above steps are repeated, and individuals in the homogeneous population evolve according to a continuous time stochastic process X(t) with transition probabilities given by (3.7). This is a special form of semi-Markov process; a more general treatment of this class of processes in the context of social mobility models appears in section 3.3.

A key point about this formulation is that it is suitable for describing processes where individual histories are not available to the observer and a more flexible class of waiting time distributions than just exponential is desired. As we remarked earlier, exponential waiting times guarantee that your mathematical model is a Markov process; however this requirement seems unnecessarily stringent for describing mobility.[10] The 2-parameter family of distributions

$$F_{(\alpha,\beta)}(t) = 1 - \left(\frac{\beta}{\beta+t}\right)^{\alpha}, \qquad \alpha,\beta > 0,\ t > 0 \qquad (3.11)$$

which arose in the above discussion describes a more general family of densities $f_{(\alpha,\beta)}(t)$ all of which have the same form as the exponential density (see Figure 1). Indeed, the exponential with parameter λ arises as a limiting case of $F_{(\alpha,\beta)}(t)$ when[11] $\alpha \to \infty$, $\beta \to \infty$, and $\frac{\alpha}{\beta} = \lambda$.

Figure 1 about here

The simplest mathematical caricature of embeddability and an inverse problem for the model (3.7) arises again for a 2-state process. For this special situation an empirically determined stochastic matrix

$$\hat{P}(t_1) = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} \qquad 0 < a,b < 1$$

corresponding to observations at times 0 and t_1 is compatible with the theoretical framework (3.7) if and only if $a + b > 1$. For a fixed α and β in (3.7) this condition also ensures a unique solution to the inverse problem which is given by

$$M = \left(\frac{\beta+t_1}{t_1}\right)\left[I - \left(\frac{\beta}{\beta+t_1}\right)\left(\hat{P}(t_1)\right)^{-1/\alpha}\right]. \qquad (3.12)$$

Necessary and sufficient conditions for an observed r x r stochastic matrix $\hat{P}(t_1)$ to be representable in the form (3.7) can be determined in principle. However, the criteria become very complicated as r increases and computational algorithms to test for embeddability are needed. In section 4.1 we present some general embeddability criteria for r = 3 and indicate conditions valid for arbitrary r when the sociological context permits stronger assumptions about the structure of M than just requiring it to be a stochastic matrix. A full discussion of the computational problems arising from embeddability tests will appear in Singer and Spilerman (1973).

In concluding our discussion of the present example, it should be pointed out that a simple strategy to check for embeddability and uniqueness of M is to calculate an r x r matrix according to (3.12) and check whether or not the computation process yields a stochastic matrix as opposed to a complex valued one. If this matrix is stochastic and if the observed $\hat{P}(t_1)$ satisfies the condition $\inf_i[\hat{p}_{ii}(t_1)] > 1/2$, you can verify that the computed matrix M is in fact the <u>unique</u> stochastic M compatible with the data. This test was used by Spilerman (1972b, p. 607). It should be emphasized that, in general, embeddable matrices $\hat{P}(t_1)$ can give rise to drastically different M arrays when $\inf_i[\hat{p}_{ii}(t_1)] \leq 1/2$. Examples of this situation are presented in section 4.1.

One interpretation of empirically determined matrices for which the diagonal elements are all greater than 1/2 is that on the natural time scale of the mobility process the observations are sufficiently close to time 0 so that many moves away from the origin state have not yet occurred. In a mathematical context $\inf_i[\hat{p}_{ii}(t_1)] > 1/2$ is a condition which guarantees that in computing $(\hat{P}(t_1))^{-1/\alpha}$ you are on the principal branch of the $-1/\alpha$-th root of the matrix, thereby ruling out complex matrices as well as other real stochastic matrices M in the inversion formula (3.12). A complete mathematical treatment of these issues will appear in Singer and Spilerman (1973).

Example 4:

From a substantive point of view, a principal defect of the individual level description in example 3 is the requirement that a person's waiting time distribution be the same in every state. We should like to eliminate this constraint and permit a full Markov model to characterize the movements of an individual. This is desirable since there are many instances in which rate of movement is a function of system state; for example, industries differ in their rates of employee separation (Blauner 1964, pp. 198-203).

We therefore classify a person according to the diagonal matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_r \end{pmatrix} \qquad \lambda_i \geq 0,\ i = 1,2,\ldots,r$$

where $1/\lambda_i$ has the interpretation, "average waiting time in state i." Then let $\{X_\Lambda(t)\}_{\Lambda>0}$ be a continuum of independent continuous time Markov chains whose transition probabilities are governed by

$$e^{t\Lambda(M-I)}, \qquad t \geq 0. \tag{3.13}$$

This is just the formulation of section (2.III) except that now the family $\{X_\Lambda(t)\}_{\Lambda>0}$ is thought of as describing the evolution of infinitely many different types of individuals, each type being identified by a distinct positive diagonal matrix Λ. Individuals of type Λ are viewed as occurring in the total population with a proportion specified by a joint probability density $g(\lambda_1,\ldots,\lambda_r)$. The macroscopic level (pooled population) is then described by random variables $Y(t)$, $t \geq 0$ whose values are the possible states of the component types $\{X_\Lambda(t)\}$ and whose transition probabilities are governed by the mixture of Markov transition matrices

$$P(t) = \int_0^\infty \cdots \int_0^\infty e^{t\Lambda(M-I)} g(\lambda_1,\ldots,\lambda_r)d\lambda_1\ldots d\lambda_r \ . \qquad (3.14)$$

A flexible 2r-parameter family of distributions analogous to (3.6) and useful for describing heterogeneity in the full population is

$$g(\lambda_1,\ldots,\lambda_r) = \prod_{i=1}^{r} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i} e^{-\beta_i\lambda_i} \qquad (3.15)$$

where $\alpha_i, \beta_i > 0$, and $i = 1,2,\ldots,r$.

A major analytical difficulty arises in dealing with the representation (3.14), even for specializations such as (3.15), because simple evaluations of the integrals in terms of rational functions of M or finite linear combinations of exponentials cannot be carried out. The source of this mathematical difficulty is the fact that the matrices Λ and M are non-commutative, i.e., $\Lambda M \neq M\Lambda$. A discussion of numerical methods for evaluating expressions such as (3.14) will appear in Singer and Spilerman (1973).

Although we cannot obtain a convenient expression for P(t), analogous to (3.7) in the case where the rate of movement parameter was specified by a scalar, or even an efficient computational algorithm, we can evaluate P(t) numerically for illustrative purposes. In particular, equation (3.14) may be written

$$P(t) = \int_0^\infty \cdots \int_0^\infty \sum_{k=0}^{\infty} \frac{(\Lambda(M-I))^k}{k!} t^k g(\lambda_1,\ldots,\lambda_r)d\lambda_1\ldots d\lambda_r$$

$$= \sum_{k=0}^{\infty} \frac{t^k}{k!} \begin{pmatrix} \lambda_1 u_{11} & \cdots & \lambda_1 u_{1r} \\ \vdots & & \vdots \\ \lambda_r u_{r1} & \cdots & \lambda_r u_{rr} \end{pmatrix}^k g(\lambda_1,\ldots,\lambda_r)d\lambda_1\ldots d\lambda_r$$

where $U = M-I$, i.e., $u_{ij} = m_{ij} - \delta_{ij}$. Specifying $g(\lambda_1,\ldots,\lambda_r)$ as a product of gamma densities (3.15), we obtain

$$P(t) = I + t\left(\cdot\cdot u_{ij}\frac{\alpha_i}{\beta_i}\cdot\cdot\right) + \frac{t^2}{2!}\left(\cdot\cdot\frac{(\alpha_i+1)\alpha_i}{\beta_i^2}u_{ii}u_{ij} + \frac{\alpha_i}{\beta_i}\sum_{k\neq i}\frac{\alpha_k}{\beta_k}u_{ik}u_{kj}\cdot\cdot\right) + \ldots \quad (3.16)$$

where the entries in matrices represent the (i,j) terms. Although the corresponding terms of higher order arrays increase rapidly in complexity, the calculations can be carried out by computer for a few terms of the series.

Artificial data were prepared in order to compare this model with ones in the earlier examples. The underlying structure of the constructed data is revealed in Table 1. Panel A shows the individual-level matrix M which was assumed to govern the movements of all persons. The waiting time distributions are displayed in panel B; they were constructed by assuming that a gamma density with parameters (α_i,β_i) describes the population heterogeneity in state i with respect to rate of movement. By varying these parameters over the system states we have built into the data the full range of generality consistent with the present model.

<u>Table 1 about here</u>

Using the information in Table 1 in conjunction with equation (3.16) the matrices $\hat{P}(1)$, $\hat{P}(2)$, and $\hat{P}(3)$ were constructed for the process. These arrays are reported in row 1 of Table 2. We will interpret them as "observed data;" they depict a mobility process in which there is population heterogeneity with respect to rate of movement, and an individual's rate can depend on the state he is in.

<u>Table 2 about here</u>

How good a fit would the mover-stayer extension or the Markov chain model provide to these data? To investigate this matter the $\hat{P}(1)$ matrix in Table 2 together with $(\hat{\alpha},\hat{\beta})$, the parameters of the waiting time distribution for the pooled population, were used to estimate M via (3.12). Equation (3.7) was then employed to calculate P(2) and P(3), the matrices predicted by the mover-stayer model. These arrays are presented in the second row of Table 2. Markov chain estimates were obtained by raising $\hat{P}(1)$ to the 2-nd and 3-rd powers (which provides identical results to projection from (2.4)). These matrices are reported in row 3 of Table 2.

The two models produce different kinds of errors when compared with the "observed data." The Markov model permits the waiting time distributions to vary by state but constrains them to be exponential. This produces an underestimation of the proportions on the main diagonal when population heterogeneity in a state is considerable (as it is for state 1),[12] but accurate results where the heterogeneity is small (state 2). The mover-stayer extension permits the waiting times to be other than exponential but constrains them to be represented by a single distribution. With the present data the mover-stayer projections overestimate both main diagonal entries.

It should be noted that the mover-stayer model is not completely specified in the example. Both the general model of this section (equation 3.16) and the Markov model are insensitive to the proportion of the total population in an origin state. This is not true for the mover-stayer extension when the assumption concerning state independent waiting times is violated, as it is here. Since the parameters $(\hat{\alpha},\hat{\beta})$ are calculated from the movements of all individuals, the estimated values will differ according to the origin state distribution of the population. What this means is that there are a variety of mobility situations, all consistent with the data in Table 1 and with the matrices in rows 1 and 3 of Table 2, which will produce different arrays with the mover-stayer model. The particular matrices reported in row 2 are based on the additional assumption that the population was evenly distributed between states 1 and 2 at time 0 (footnote b of Table 1).

Rather than pursue computational details our intention here is simply to point out that the theoretical framework (3.14) provides the most general macroscopic level description of a heterogeneous population with individuals classified in terms of their rates of movement, the rates being state dependent, and evolving according to independent Markov chains. As in example 3, the observable process is a particular case of a semi-Markov process and this provides a second interpretation for (3.16), as a homogeneous population with non-exponential waiting times. Here the waiting time distributions depend on the state according to

$$\text{Prob } (\tau_k \geq t \,|\, X(k) = i) = \left(\frac{\beta_i}{\beta_i + t} \right)^{\alpha_i} . \qquad (3.17)$$

The particular formula (3.17) arises when the family (3.15) is used to describe the proportion of type-Λ individuals in the pooled population. The variable $X(k)$, $k = 0,1,2,\ldots$, denotes a discrete time Markov chain governed by the stochastic matrices M, M^2, $M^3,\ldots$, and only describes the jumps of $Y(t)$, not its waiting times in particular states.

In applying this model it should be noted that the data requirements are more extensive than was the case previously. First, until inverse procedures are developed, we must have available $\hat{M}$ rather than $\hat{P}(t)$, although the latter is the more commonly published datum. Second, we demand a separate waiting time distribution for each system state in order to estimate the parameters (α_i,β_i), $i = 1,\ldots,r$. This is in contrast to the estimation of α and β in example 3 using the subordination representation $Y(t) = X(T_{(\alpha,\beta)}(t))$. There we required either the waiting time distribution for the entire population or, what is more generally available, the distribution of number of moves in the population during $(0,t)$. A similar readily computable description of $Y(t)$ in the present case, governed by (3.14) and (3.15), is not possible due to the state dependence of the waiting times. In principle we could write $Y(t) = X(T^*(t))$ where $T^*(t) = \{$number of transitions in $Y(t)$ up to time $t\}$; however there is no simple 2r-parameter family of processes, analogous to the negative binomial process, which enables us to solve for (α_i,β_i), $i = 1,\ldots,r$ in terms of number of transitions in a sampling interval. For these reasons, in contrast with the mover-stayer extension (example 3), individual-level data files will be necessary to exploit this model. They could derive from either retrospective histories (e.g., the Taeuber data file on residence change (Taeuber, Chiazze and Haenszel 1968)) or from panel studies (e.g., the New Jersey Negative Income Tax Experiment conducted by the Institute for Research on Poverty of the University of Wisconsin).

Embeddability and inverse problems for the present model, given observations at several time points $\hat{P}(t_1)$, $\hat{P}(t_2),\ldots,\hat{P}(t_k)$ and varying degrees of information about (α_i,β_i), $i = 1,\ldots,r$ (ranging from estimation of all 2r constants down to rough inequalities about their range), are also complicated by the fact that Λ and M are non-commutative. In particular, no simple representation for M in terms of logarithms and rational functions of $P(t_i)$ analogous to (3.12) is available. Numerical inversion methods involving special multi-dimensional extensions of the classical Lagrange inversion formula for scalar valued analytic functions will be discussed in the context of mobility models in Singer and Spilerman (1973). It should also be pointed out that a complete discussion of the inverse problem and non-uniqueness of M in the present setting poses substantial mathematical difficulties which are unresolved as of this writing. The interested reader should consult Singer and Spilerman (1973) for precise statements of these questions.

3.2. Identification of the determinants of population heterogeneity

Let $\{X_{\Lambda,M}(t)\}_{\Lambda>0,M>0}$ be a collection of independent continuous time Markov chains whose transition probabilities are governed by the prescription

$$e^{t\Lambda(M-I)} .$$

We therefore classify a person in terms of a diagonal matrix of movement rates and a stochastic matrix which specifies his transition propensities at a move. For convenience we will subscript individual q's parameters and write (Λ_q, M_q). We make no assumption regarding particular distributions for $\{\Lambda_q\}$ and $\{M_q\}$ in the population; however, we do require the availability of individual-level attribute data. Our intention is to discuss a method for ascertaining the determinants of population heterogeneity with respect to both rate of movement and propensity to favor transitions to certain states.

(a) Heterogeneity in the rate of movement: Consider the regression equation,

$$A_{ic} = a_i + \sum_{k=1}^{K} b_{ik} X_{kc} + e_i , \qquad i = 1,\dots,r \tag{3.18}$$

where A_{ic} is the c-th waiting time interval in state i during $(0,t_1)$, $(X_1,\ldots,X_K)$ are variables which are expected to explain individual differences in rate of movement, and the error terms are assumed to be independently distributed. The observations in this regression are the C waiting times in state i. A person will contribute more than one observation if he made several moves during $(0,t_1)$ which originated in state i; if he was in this state throughout the interval (and failed to move) he will appear once with $A_{ic} = t_1$.

This specification is intimately related to the rate of movement parameter of the continuous time Markov model according to

$$\frac{1}{N_{i.}} \sum_{c}^{C} \hat{A}_{ic} = \frac{1}{N_{i.}} \sum_{c}^{C} A_{ic} = \frac{1}{\hat{\lambda}_i} \tag{3.19}$$

where the sums are taken over all waiting time intervals in state i during $(0,t_1)$, and $N_{i.}$ denotes the number of moves originating in state i. The first equality results from the least squares procedure of fitting a regression plane ($\hat{A}_{ic}$ is the predicted c-th waiting time, the prediction having been made from the attribute profile $(X_{1c},\ldots,X_{Kc})$ of the individual associated with this waiting time). The second equality is just equation (2.7) restated in terms of waiting times. The $\hat{\lambda}_i$ value pertains to the single Markov chain that would be estimated if heterogeneity were ignored; it provides a suitable reference in terms of which population heterogeneity may be described. Combining (3.18) and (3.19),

$$\frac{1}{\hat{\lambda}_i} = \hat{a}_i \frac{C}{N_{i.}} + \sum_{k=1}^{K} \hat{b}_{ik} \left(\frac{\sum_{c}^{C} X_{kc}}{N_{i.}} \right), \qquad i = 1,\ldots,r \tag{3.20}$$

and this indicates how the regression produces a decomposition of the

Markov parameter $\hat{\lambda}_i$. The term in parentheses in (3.20), incidently, can be interpreted as the "typical" individual profile associated with a waiting time interval.

(b) Heterogeneity with respect to choice of destination state: Define a variable y_{ijc} which equals 1 if the c-th move originating in state i during $(0,t_1)$ resulted in a transition to state j, and 0 if it did not. Now consider the equation

$$y_{ijc} = a_{ij} + \sum_{h=1}^{H} b_{ijh}X_{hc} + e_{ij}, \qquad i,j = 1,\ldots,r \quad (3.21)$$

where $(X_1,\ldots,X_H)$ are variables which are expected to relate to choice of destination at a move, and the error terms e_{ij} are independently distributed. The observations for this regression are all C <u>moves</u> which originated in state i. An individual will appear more than once if he made several moves from state i during $(0,t_1)$; he will not contribute an observation if he failed to move.

The relation between this equation and the corresponding Markov parameters is given by

$$\frac{1}{N_{i.}} \sum_{c}^{N_{i.}} \hat{y}_{ijc} = \frac{1}{N_{i.}} \sum_{c}^{N_{i.}} y_{ic} = \frac{N_{ij}}{N_{i.}} = \hat{m}_{ij} \qquad (3.22)$$

where N_{ij} equals the number of state i to state j transitions. The equality between the first two terms follows from the regression procedure; the second equality derives from the definition of y_{ijc}, and the third from equation (2.7). Again, the value $\hat{m}_{ij}$ refers to the single Markov chain that would result from treating the population as though it were homogeneous; it provides a useful benchmark from which to characterize heterogeneity. Combining (3.21) and (3.22) the decomposition of the Markov parameter $\hat{m}_{ij}$ may be expressed as

$$\hat{m}_{ij} = \hat{a}_{ij} + \sum_{h}^{H} \hat{b}_{ijk} \left(\frac{\sum\limits_{c}^{N_{i.}} X_{hc}}{N_{i.}} \right), \qquad i,j = 1,\ldots,r. \quad (3.23)$$

This equation describes the population heterogeneity with respect to choice of destination state at a move. The term in parentheses in (3.23) depicts the typical individual profile associated with a move.

The two regression equations (3.18) and (3.21) therefore lead to a decomposition of population heterogeneity in a way that is intimately related to the continuous time Markov chain formulation. Further elaboration of this procedure, in the context of a discrete time Markov model, may be found in Spilerman (1972a).

In theory these regressions could be used to construct a Λ_q and M_q for each individual in the population as was done in Spilerman (1972a). If this is carried out the population level transition matrix would be written

$$P(t) = V^{-1} \sum_{q} V_q e^{t\Lambda_q(M_q - I)} \qquad (3.24)$$

where V_q is a matrix with entry 1 on the main diagonal of the i-th row and zero in all other cells (i denoting individual q's location at time 0) and $V = \sum_q V_q$. Expression (3.24)[13] describes the population level process when each individual q evolves independently according to a continuous time Markov chain with parameters (Λ_q, M_q). It is the continuous time analog of McFarland's (1970) formulation to accommodate heterogeneity, which was summarized in example 2 of section 3.1.

In practice, the estimates for this construction are likely to be poor since we would be computing M_q matrices for non-movers during $(0,t_1)$ using only information on choice of destination state by movers (3.21). Consequently, the utility of this formulation lies mainly in its contribution to analyzing heterogeneity, rather than to estimating individual level parameters for projection.

3.3. Semi-Markov processes and their mixtures

When the multivariate density $g(\lambda_1,\ldots,\lambda_r)$ in equation (3.14) is specified as a product of univariate density functions, $\prod g_i(\lambda_i)$, then the same mathematical formalism (3.14) applies to (a) a heterogeneous population in which each individual moves according to a Markov process with transition matrix M, the heterogeneity in rate of movement being described by $g_i(\lambda_i)$ in state i; and (b) a homogeneous population in which an individual waits in state i according to the distribution function

$$F_i(t) = \int_0^t \int_0^\infty \lambda e^{-\lambda t} g_i(\lambda) d\lambda dt$$

before transferring according to M.

For the macroscopic level process Y(t) of example 4, $\{g_i(\lambda_i)\}$ were specified as gamma densities (3.15), and the corresponding waiting time distributions were given by

$$F_i(t) = \text{Prob}\left(\begin{array}{r} \text{waiting time until a transition} \\ \text{is less than } t \end{array} \middle| \begin{array}{c} \text{present state} \\ \text{is } i \end{array}\right)$$

$$= 1 - \left(\frac{\beta_i}{\beta_i + t}\right)^{\alpha_i}, \qquad \alpha_i, \beta_i > 0, \quad i = 1,2,\ldots,r.$$

In the Y(t) process of example 3 $g_i(\lambda_i) = g(\lambda)$ [equation 3.6] and $F_i(t) = F(t)$ [equation 3.11]; that is, the description of population heterogeneity in the first perspective, and the waiting time to a move in the second, are independent of the state of the process.

These macroscopic level processes are special cases of what are known as semi-Markov processes (Pyke 1961a, 1961b; Ginsberg 1971). This model is usually presented in the conceptual imagery of a homogeneous population with waiting time distributions which need not be exponential. For an explicit formulation consider a stochastic process $Z(t)$, $t \geq 0$, with a finite number of states which, again, may be occupational categories, geographic regions, or income levels. The transition probabilities for the semi-Markov processes treated here are the unique solutions of the system of integral equations

$$p_{ij}(t) = \delta_{ij}(1-F_i(t)) + \sum_{k=1}^{r} \int_0^t f_i(s)m_{ik}p_{kj}(t-s)ds \qquad (3.25)$$

where $p_{ij}(t) = \text{Prob}\ (Z(t) = j \mid Z(0) = i)$; $\delta_{ij} = 1$ if $i = j$, 0 if $i \neq j$; and $1 \leq i,j \leq r$.

$F_i(t)$ is a distribution function which has the interpretation, "probability that a move has occurred by time t;" we assume that it has a density $f_i(t)$. The stochastic matrix M with entries m_{ij} describes the propensity to move to particular states. Equation (3.25) is therefore amenable to the following interpretation: (a) When $i \neq j$, $p_{ij}(t)$ consists of the sum of products of three factors: the probability of a first transition out of state i at time s, the probability of a state i to state k transition at that move, and the probability of transferring to state j by some combination of moves during (s,t). The summation is over all intermediate states k and over all time points s in the interval (0,t). (b) When $i = j$, then, in addition to the above factor, there is a possibility of not transferring out of state i during (0,t). The associated probability is specified by the first term.

When $f_i(t) = \lambda_i e^{-\lambda_i t}$, then the system (3.25) is equivalent to the differential equations (2.3) with Q written in the factored form $Q = \Lambda(M-I)$. Thus the integral equation formulation (3.25) is a very transparent way of saying that a stochastic process governed by these equations behaves like a Markov process except that the waiting time distributions can be represented by general density functions $f_i(t)$.

With these preliminary notions at hand we now indicate two contexts in which semi-Markov processes are a natural description of social mobility.

Example 1:

R. McGinnis (1968) refers to a phenomena which he calls cumulative inertia and which has the interpretation that the longer a person remains in a particular state (occupation, geographic region, etc.) the less likely he is to move out of that state in the immediate future. Presumably, with increasing duration a person establishes social linkages and in other ways acclimates to his setting so that the attractiveness of remaining is increased. Ginsberg (1971) has pointed out that a semi-Markov process with decreasing event rate provides a formalization of this notion. A mathematical caricature of cumulative inertia can be stated in terms of the waiting time distributions $F_i(t)$ via the function

$$r_i(t) = \frac{f_i(t)}{1-F_i(t)} . \tag{3.26}$$

The expression $r_i(t)dt$ can be interpreted as the probability that a person known to be in state i at time t will exit from that state in the next dt units of time. Then cumulative inertia simply means that r(t) is a monotone decreasing function of t.[14] A simple 2-parameter family of waiting time distributions with monotone decreasing cumulative inertia is given by

$$F_i(t) = 1 - e^{-\lambda_i t^{\gamma_i}} \quad \text{with } \lambda_i > 0, \quad 0 < \gamma_i < 1. \qquad (3.27)$$

For this specification, $r_i(t) = \lambda_i \gamma_i t^{\gamma_i - 1}$. Now classification of an individual evolving according to a semi-Markov process would be to characterize him by the family of distributions $\mathbb{F} = \left\{F_1(t), \ldots, F_r(t)\right\}$ describing his waiting times in any state and the stochastic matrix M describing his propensity to move to particular states.

Example 2:

In the framework of mixtures of stochastic processes we can consider $\{X_{\mathbb{F}}(t)\}$ as a continuum of independent semi-Markov processes describing the mobility of individuals whose rates of movement are governed by (3.27). It is assumed that the individual rates are distributed in the total population with proportions governed by a probability density $g(\lambda_1, \ldots, \lambda_r; \gamma_1, \ldots, \gamma_r)$. Then the macroscopic level process Y(t) is defined as a stochastic process whose possible states coincide with those of $\{X_{\mathbb{F}}(t)\}$ but whose transition probabilities are governed by

$$P(t) = \int_0^\infty \cdots \int_0^\infty \cdots \int_0^1 \cdots \int_0^1 \Big[S(t;\, \lambda_1, \ldots, \lambda_r;\, \gamma_1, \ldots, \gamma_r; M)$$

$$g(\lambda_1, \ldots, \lambda_r;\, \gamma_1, \ldots, \gamma_r)\Big]\, d\lambda_1 \ldots d\lambda_r d\gamma_1 \ldots d\gamma_r \qquad (3.28)$$

where $S(t;\, \lambda_1, \ldots, \lambda_r;\, \gamma_1, \ldots, \gamma_r; M)$ denotes the stochastic matrix solution of (3.25) with $\{F_i(t)\}$ specified by (3.27). In general we must appeal to numerical integration methods to evaluate $S(t;\, \lambda_1, \ldots, \lambda_r;\, \gamma_1, \ldots, \gamma_r; M)$

as explicit simple representations analogous to (3.7) are a rarity for semi-Markov processes. Again, a useful 2r-parameter family of densities describing the composition of the pooled population is given by

$$g(\lambda_1,\ldots,\lambda_r) = \prod_{i=1}^{r} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i - 1} e^{-\beta_i \lambda_i}, \qquad \alpha_i, \beta_i > 0$$

In this simple setting we treat $\gamma_1,\ldots,\gamma_r$ as fixed; although this is certainly not an essential conceptual restriction. The rationale for using the gamma distributions in the present context remains the same as that presented for mobility processes where individuals evolved according to Markov rather than semi-Markov processes. The question of computationally effective solutions of the inverse problem for mixtures of semi-Markov processes is at present unresolved; an indication of the mathematical difficulties and some suggested lines of attack will be presented in Singer and Spilerman (1973).

This very general process (3.28), incidentally, provides us with a formulation in which <u>both</u> a duration of residence effect and population heterogeneity can be postulated. In the preceding models, and in the few other discussions of semi-Markov models as they pertain to social mobility (Ginsberg 1971, p. 254), one was compelled to specify individual-level behavior as Markovian if the heterogeneity perspective was adopted, and the population as homogeneous if a duration effect mechanism was postulated. Clearly, both processes could be operative and we should prefer a model in which they can be incorporated simultaneously.

A final point concerning semi-Markov processes which should be mentioned is that the formulation given by equation (3.25) does not describe the most general process of this kind as treated in the mathematics literature. In particular, the original semi-Markov framework allowed for waiting time distributions which could depend on the next future state as well as the current state of the process. This level of generality, however, does not seem appropriate for most social mobility situations; hence we have restricted our attention to a subclass of semi-Markov processes which requires the estimation of fewer parameters.

4. INVERSE PROBLEMS

In the previous sections we have indicated a few examples of inverse problems and their associated embeddability questions. This aspect of our study really involves an independent mathematical development which also seems to be of considerable importance outside the context of social mobility models, and which will be elaborated in a separate publication. In the present section we simply illustrate the flavor of inverse problems and give some indication of general diagnostic strategies for recovering partial information about the fine structure of a mobility process from information about its behavior at a few points in time.

Before proceeding to the examples, we would like to point out where the inverse problems of the present study fit into a larger mathematical framework. To fix the ideas, recall the matrix differential equation

$$\frac{dP(t)}{dt} = QP(t), \qquad\qquad P(0) = I$$

whose solutions are the transition probabilities for continuous time Markov chains. Rather than being given a particular differential equation (i.e., a fixed Q) and asked to compute a solution P(t) (a "direct problem"), an inverse problem has a class of differential equations and partial information about a solution (usually $\hat{P}(t_i)$ for a few values of i) as given ingredients. From this information the problem is to find the particular differential equation which is compatible with the observed solution.

The overall strategy of inverse problem formulation and interpretation in the context of social mobility will now be described entirely within the context of a homogeneous population evolving according to a continuous time Markov chain. The key point is that all of the issues which must be faced in the more complicated mixtures of Markov and semi-Markov formulations are already present in this setting.

Step 1--Embeddability

From an empirical point of view, the most primitive question to be asked about a stochastic matrix $\hat{P}(t_1)$, estimated from observations at times 0 and t_1, is whether or not it is compatible with the theoretical framework

$$\hat{P}(t_1) = e^{t_1 Q}$$

where Q is an r x r matrix satisfying

$$-\infty < q_{ii} < 0, \quad q_{ij} \geq 0 \text{ for } i \neq j, \quad \sum_{j=1}^{r} q_{ij} = 0 . \qquad (4.1)$$

This problem has a long history in the mathematics literature (see Singer and Spilerman 1973 for references), and our purpose here is to indicate its solution for 2 and some 3-state processes as well as general finite state birth and death processes.

Case 1:

If $\hat{P}(t_1)$ is a 2 x 2 stochastic matrix denoted by $\begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}$, $0 < a,b < 1$, then it can be represented as $e^{t_1 Q}$ with Q satisfying (4.1) if and only if $a + b > 1$. (See Appendix I for a proof.)

It is interesting to note that the matrix $\begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}$ which, as indicated in (2.1), is compatible with a discrete time Markov model for $t_1 = 3$ is not compatible with a continuous time Markov model for any positive time t_1. Another interesting feature of the condition $a + b > 1$ is that this automatically guarantees uniqueness of Q. In the 2 x 2 case we therefore have a single criterion which ensures both embeddability and uniqueness; note also that this is a weaker requirement than the general sufficiency condition for uniqueness, $\inf_i [\hat{p}_{ii}(t_1)] > 1/2$.

When the inequality $a + b > 1$ is satisfied the unique Q matrix governing the evolution of the continuous time Markov chain is given by

$$Q = \frac{1}{t_1} \log \hat{P} = \frac{\log(a+b-1)}{t_1(a+b-2)} \begin{pmatrix} a-1 & 1-a \\ 1-b & b-1 \end{pmatrix}. \qquad (4.2)$$

A further ramification here is that compatibility of the data with a continuous time Markov model and unique identification of Q can be checked by observations at time 0 and only one other time point; this time point may be chosen arbitrarily by the experimenter. As subsequent examples indicate, this simplicity of embeddability tests and identification of Q no longer holds even for 3-state processes.

Case 2:

(a) If $\hat{P}(t_1)$ is a 3 x 3 stochastic matrix with distinct real eigenvalues $1 > \lambda_1 > \lambda_2 > 0$ then it can be represented as $e^{t_1 Q}$ for at least one Q satisfying (4.1) if and only if

$$p_{ij}^{(2)} \leq \hat{p}_{ij} \frac{(\lambda_2^2 -1)\log \lambda_1 - (\lambda_1^2 -1)\log \lambda_2}{(\lambda_2-1)\log \lambda_1 - (\lambda_1-1)\log \lambda_2}, \qquad \text{all } i \neq j \quad (4.3)$$

where $p_{ij}^{(2)}$ is the (i,j) entry in $[\hat{P}(t_1)]^2$.

(b) If $\hat{P}(t_1)$ has eigenvalues 1, λ, λ where λ is real and $1 > \lambda > 0$, then it can be represented as $e^{t_1 Q}$ for at least one Q satisfying (4.1) if and only if

$$p_{ij}^{(2)} \leq \hat{p}_{ij} \frac{\lambda^2 \log \lambda^2 - \lambda^2 + 1}{\lambda \log \lambda - \lambda + 1}, \qquad \text{all } i \neq j. \quad (4.4)$$

Similar criteria can be given for 3 x 3-matrices having complex eigenvalues as well as for general r-state matrices. The explicit inequalities become very intricate as r increases; however, they are all established by observing that log P may always be evaluated in principle as a polynomial in P of degree at most r-1 [Caley-Hamilton Theorem (Stein 1967, p. 196)]. Then inequalities such as (4.3) and (4.4) arise by requiring that

$$\log P = C_0 I + C_1 P + \ldots + C_{r-1} P^{r-1}$$

be real and a matrix satisfying (4.1).

Case 3:

If a sociological context allows us to restrict consideration to continuous time models where the only allowable transitions are to nearest neighbor states, then we have a simpler criterion for $\hat{P}(t_1)$ to be representable as $e^{t_1 Q}$ for at least one Q such that (4.1) holds but with $q_{ij} = 0$ for $|i - j| > 1$ (i.e., Q is a Jacobi intensity matrix).

In particular, an r x r stochastic matrix $\hat{P}(t_1)$ is representable as $e^{t_1 Q}$ with Q a Jacobi intensity matrix if and only if all of its entries and the following 2 x 2 sub-determinants are strictly positive (Karlin and McGregor 1957): Choose $i_1 < i_2$, $j_1 < j_2$ arbitrarily and check

$$\det \begin{pmatrix} \hat{p}_{i_1,j_1} & \hat{p}_{i_1,j_2} \\ \hat{p}_{i_2,j_1} & \hat{p}_{i_2,j_2} \end{pmatrix} > 0 \qquad (4.5)$$

$$1 \leq i_1 < i_2 \leq r, \quad 1 \leq j_1 < j_2 \leq r.$$

It is important to notice that the class of models e^{tQ} with Q a Jacobi matrix has been widely used in sociological investigations even outside the context of social mobility (e.g., Coleman 1964, chapters 10, 11, and 14). In the mathematics literature these processes are referred to as finite state birth and death processes (W. Feller 1968, chapter 17), and they also have a long history of use as baseline models in particle physics, chemistry, and biology. The criteria (4.5) provide a simple, readily computable test for compatibility of a stochastic matrix with a birth and death model for an arbitrary finite number of states. The interested reader can check that in Coleman's (1964, pp. 462-465) application of a birth and death structure to English mobility data there are numerous violations of this embeddability condition.

With these examples at hand, we should emphasize that the above inequality tests can be considered as devices for isolating the class of stochastic matrices compatible with particular continuous time Markov structures[15] from the class of all stochastic matrices. Once it is concluded that a matrix is embeddable in a Markov structure, the next step is to identify all intensity matrices Q which could have given rise to the observed $\hat{P}(t_1)$.

Step 2--Identification of Q

If the observed matrix $\hat{P}(t_1)$ is embeddable and the condition $\inf_i[\hat{p}_{ii}(t_1)] > 1/2$ is satisfied, then we can calculate a unique Q,

$$Q = \frac{1}{t_1} \log \hat{P}(t_1). \tag{4.6}$$

In order to illustrate the methodological difficulties which might arise when dealing with matrices $\hat{P}(t_1)$ which are embeddable in a continuous time Markov structure, but do not satisfy the condition $\inf_i[\hat{p}_{ii}(t_1)] > 1/2$, we consider the following example.

Suppose you estimate[16]

$$\hat{P}(t_1) = \begin{pmatrix} \frac{1}{3} + \frac{2\varepsilon}{3} & \frac{1}{3} - \frac{\varepsilon}{3} & \frac{1}{3} - \frac{\varepsilon}{3} \\ \frac{1}{3} - \frac{\varepsilon}{3} & \frac{1}{3} + \frac{2\varepsilon}{3} & \frac{1}{3} - \frac{\varepsilon}{3} \\ \frac{1}{3} - \frac{\varepsilon}{3} & \frac{1}{3} - \frac{\varepsilon}{3} & \frac{1}{3} + \frac{2\varepsilon}{3} \end{pmatrix} \tag{4.7}$$

where $\varepsilon = e^{-2\pi\sqrt{3}}$ and $t_1 = \frac{4\pi}{\sqrt{3}} \approx 7.26$. This stochastic matrix has a representation $e^{t_1 Q}$ for the following intensity matrices:

$$Q_1 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \qquad Q_2 = \begin{pmatrix} -1 & 1/2 & 1/2 \\ 1/2 & -1 & 1/2 \\ 1/2 & 1/2 & -1 \end{pmatrix}.$$

If a researcher is constrained to estimating just this $\hat{P}(t_1)$ from data, then he will find these two substantively distinct matrices compatible with his observations and with a continuous time Markov model. In particular, Q_1 and Q_2 correspond to processes where the holding times between moves are exponentially distributed with parameter 1, regardless of state; however, a process governed by Q_1 only allows transitions through states in the cyclic pattern $1 \to 2 \to 3 \to 1 \to 2 \to 3$...etc. On the other hand, a process governed by Q_2 allows equally likely transitions from any one state to any other state. Sociological argument must decide which of these two alternatives is substantively meaningful if a single observation beyond $t = 0$ is a constraint on the study.

Alternatively if you compute e^{tQ_1} and e^{tQ_2} for general times t, you find that $P_1(t) = e^{tQ_1}$ is a 3 x 3 stochastic matrix with entries

$$p_{ii}^{(1)}(t) = 1/3 + 2/3\ e^{-3t/2} \cos \sqrt{3}t/2, \qquad i = 1,2,3$$

$$p_{12}^{(1)}(t) = p_{23}^{(1)}(t) = p_{31}^{(1)}(t) = 1/3 + 2/3\ e^{-3t/2} \cos\left(\frac{\sqrt{3}t}{2} - \frac{2\pi}{3}\right)$$

$$p_{13}^{(1)}(t) = p_{21}^{(1)}(t) = p_{32}^{(1)}(t) = 1/3 + 2/3\ e^{-3t/2} \cos\left(\frac{\sqrt{3}t}{2} + \frac{2\pi}{3}\right)$$

while $P_2(t) = e^{tQ_2}$ is a 3 x 3 stochastic matrix with entries

$$p_{ii}^{(2)}(t) = \frac{1}{3} + \frac{2}{3} e^{-3t/2}, \qquad i = 1,2,3$$

$$p_{ij}^{(2)}(t) = \frac{1}{3} - \frac{1}{3} e^{-3t/2} \qquad \text{for } i \neq j.$$

When $t = \frac{4k\pi}{\sqrt{3}}$ $(k = 0,1,2,\ldots)$, $P_1(t) = P_2(t)$ and you cannot discriminate between these two processes. The difficulty to be highlighted here is that the observation time dictated by the experiment turned out to be inconvenient for unique identification of Q. However, if one more observation is allowed, the above calculations indicate that if it is taken at time $\frac{6\pi}{\sqrt{3}}$ (i.e., $k = 3/2$), then $\log P\left(\frac{4\pi}{\sqrt{3}}\right)$ and $\log P\left(\frac{6\pi}{\sqrt{3}}\right)$ must be constant multiples of <u>either</u> Q_1 or Q_2. Whichever is observed will identify the unique Q-matrix compatible with a continuous time Markov model for the mobility process under examination.

The phenomenon described above arises due to non-uniqueness of the logarithm of a stochastic matrix. As indicated earlier, a sufficient condition for uniqueness in the Markov case is that $\inf_i [\hat{p}_{ii}(t_1)] > 1/2$; however, there is no apriori reason to believe that this condition will hold in environments where Markov models might be applied. Hence a more thorough understanding of the nature of non-uniqueness as illustrated by the above example is clearly needed. Some progress in this direction both for the Markov and semi-Markov models arising as mixtures according to the recipes of section 3 is described in Singer and Spilerman (1973); however, substantial mathematical difficulties are still unresolved with regard to these questions.

5. CONCLUSIONS

We conclude with an overview of diagnostic strategies for the social mobility models formulated in the previous sections. Many of our remarks should be viewed as suggestions for future research; however, a discussion of several mobility data sets from this point of view together with a presentation of appropriate data analytic techniques will appear in a separate publication of the authors.

(A) Select a class of Markov or semi-Markov models, such as those in section 3, which seem to correspond to prior evidence and theories about the nature of heterogeneity in the population you are observing.

(B) Check for embeddability. Necessary and sufficient conditions for an observed 2 x 2 matrix to be embeddable in a continuous time Markov process, and for an observed r-state matrix to be embeddable in a birth and death structure were reported in section 4.

(C) Check for a unique solution to the inverse problem. A sufficient condition for uniqueness in the continuous time models we have explored is the criterion $\inf_i [p_{ii}(t_1)] > 1/2$. The non-uniqueness phenomena illustrated in our previous example frequently occurs for logarithms, roots, and more general inverse formulas of matrices with repeated eigenvalues. Thus a useful strategy in dealing with empirically determined matrices $\hat{P}(t_1)$ having distinct eigenvalues some of which are within several significant digits of each other, is to adjust $\hat{P}(t_1)$ to force equality of the eigenvalues and compute all M-matrices or Q-matrices (depending on the context) compatible with both a repeated eigenvalue estimate and a distinct eigenvalue estimate. The point here is that much sociological data involves severe "noise," and you may miss an opportunity to examine and interpret a substantively meaningful matrix M by treating the observed matrix $\hat{P}(t_1)$ as though it was error free. Indeed, because of sampling error in the observed matrix $\hat{P}(t_1)$ you may have computed the <u>wrong</u> M for the process! The numerical analysis problems connected with these adjustments will also be indicated in Singer and Spilerman (1973).

(D) The calculation of compatible matrices M via numerical inversion algorithms should be followed by a determination of time points beyond that of the original observations t_0 and t_1 at which it would be possible to discriminate among competing candidates. An instance of this was the identification of the time $\frac{6\pi}{\sqrt{3}}$ in the previous section where an estimated $\hat{P}\left(\frac{6\pi}{\sqrt{3}}\right)$ could be used to discriminate between Q_1 and Q_2. This portion of the diagnostic process really falls under the framework of experimental designs for mobility processes, an area which to the best of our knowledge is completely unexplored.

APPENDIX I. COMPUTING FUNCTIONS OF MATRIX ARGUMENT

A key step in the production of inversion formulas is often the evaluation of an analytic function of matrix argument. In the continuous time Markov case we required a computation of Q from an observed $\hat{P}$ and a postulated structure e^{tQ}. This involves calculation of log $\hat{P}$ where $\hat{P}$ is a stochastic matrix. Analogously, in the discrete time Markov model we required a recipe for computing the n-th roots of $\hat{P}$.

A natural formulation of analytic functions f(z) with z replaced by a matrix is the contour integral definition

$$f(P) = \frac{1}{2\pi i} \int_{\Gamma} (\zeta I - P)^{-1} f(\zeta) d\zeta$$

where Γ is a smooth closed curve which encloses the eigenvalues of P and $f(\zeta)$ is single-valued and analytic. The components of $(\zeta I-P)^{-1}$ are of the form

$$(\zeta I-P)^{-1}_{ij} = \frac{(-1)^{i+j}\phi_{ij}}{\phi(\zeta)}$$

where $\phi(\zeta)$ = determinant of $(\zeta I-P)$, and

ϕ_{ij} = determinant of the (n-1) x (n-1) matrix obtained by deleting the j^{th} row and i^{th} column of $\zeta I-P$.

To illustrate the use of this formulation we calculate log P when P is the 2 x 2 stochastic matrix

$$P = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} \qquad 0 < a,b < 1.$$

The steps in the computation are

$$(\zeta I-P) = \begin{pmatrix} \zeta-a & a-1 \\ b-1 & \zeta-b \end{pmatrix}, \qquad \left|\left|(-1)^{i+j}\phi_{ij}\right|\right| = \begin{pmatrix} \zeta-b & 1-a \\ 1-b & \zeta-a \end{pmatrix},$$

$$\phi(\zeta) = \zeta^2 - \zeta(a+b) + (a+b-1).$$

The eigenvalues of P are the roots of $\phi(\zeta) = 0$ and are given by

$$\zeta_1 = 1, \quad \zeta_2 = a + b - 1.$$

Residue evaluation of $\frac{1}{2\pi i}\int_\Gamma \frac{(-1)^{i+j}\phi_{ij}\log(\zeta)d\zeta}{(\zeta-1)(\zeta-(a+b-1))}$ for $1 \leq i,j \leq 2$ yields

$$\log P = \frac{\log(a+b-1)}{a+b-2}\begin{pmatrix} a-1 & 1-a \\ 1-b & b-1 \end{pmatrix}.$$

For this to be a legitimate Q-matrix we require simply that log(a+b-1) is real. This will happen if and only if $a + b > 1$. Thus we obtain the condition that P is representable as e^Q if and only if $a + b > 1$.

For a second application of the contour integral we calculate the cube roots of

$$P(3) = \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}$$

which is (2.1) in the text. Here

$$(\zeta I-P) = \begin{pmatrix} \zeta-1/3 & -2/3 \\ -2/3 & \zeta-1/3 \end{pmatrix}, \qquad \left|\left|(-1)^{i+j}\phi_{ij}\right|\right| = \begin{pmatrix} \zeta-1/3 & 2/3 \\ 2/3 & \zeta-1/3 \end{pmatrix},$$

$$\phi(\zeta) = \zeta^2 - 2\zeta/3 - 1/3 = (\zeta-1)(\zeta+1/3) = 0.$$

Residue evaluation of $\frac{1}{2\pi i}\int_\Gamma \frac{(-1)^{i+j}\phi_{ij}\zeta^{1/3}d\zeta}{(\zeta-1)(\zeta+1/3)}$ for $1 \leq i,j \leq 2$ yields

$$[P(3)]^{1/3} = \begin{pmatrix} 1/2 - 1/2\sqrt[3]{3} & 1/2 + 1/2\sqrt[3]{3} \\ 1/2 + 1/2\sqrt[3]{3} & 1/2 - 1/2\sqrt[3]{3} \end{pmatrix},$$

as the only real valued root. This matrix is (2.2) in the text.

A more extensive discussion of the role of contour integral formulations in producing inversion formulas for the models of section 3 is given in F. John (1965, pp. 103-118).

FOOTNOTES

[1] We assume that the reader is familiar with the four papers cited in the preceding paragraph and with the rudiments of discrete and continuous time Markov processes.

[2] The symbol ^ over a stochastic matrix means that it should be thought of as a quantity estimated directly from data; matrices without ^ should be viewed as obtained from a mathematical model.

[3] Incompatibility with a discrete time Markov chain implies incompatibility with a continuous time Markov structure. The converse is not true.

[4] The utility of this formulation can be illustrated by an example. If the process concerns geographic migration and the system states are regions of the county, a state i to state i transition would represent change of residence within a region. Even if it is unreasonable conceptually to "move" and not change state (as in movements among marital statuses) we might still want to speak of "exposures to movement" or "decisions to possibly move."

[5] $\{X(k)\}$ are again the random variables of a discrete time Markov chain governed by M which describes moves when they occur.

[6] In most applications to mobility it has been assumed that the natural time scale of the process is such that $n = 1$, thus eliminating the need for computing roots of matrices. This assumption is tantamount to saying that the natural time scale has intervals which are the same length as a sampling interval, thereby obscuring consideration of

6 (cont.) alternative underlying time scales and transition mechanisms M which might be compatible with the data and substantively meaningful.

[7] It should be observed that a 2 x 2 stochastic matrix $\hat{P}(n)$ with n odd is always compatible with a discrete time Markov structure. Thus if you are restricted to taking observations only at time 0 and one other time, the additional measurement should be made at an even time to provide the most elementary test of compatibility with a Markov model. This same discussion also applies to 2 x 2 stochastic matrices $\hat{P}(n)$ thought of as observations generated by a mover-stayer model.

[8] The second parameter, β, is merely a scaling factor.

[9] This distribution is obtained by assuming a mixture of exponential waiting times,

$$P(\tau_0 \geq t) = \int_t^\infty f(t)dt = \int_t^\infty \int_0^\infty \lambda e^{-\lambda t} g(\lambda) d\lambda dt$$

where $g(\lambda)$ is a gamma density.

[10] In fact, sociological hypotheses which invoke the Markov property (independence of future state from past locations, given current state) to describe state changes at the occurrence of a move often place no requirement on the waiting time distributions. It would be incorrect to test such a thesis by fitting a Markov chain to the data.

[11] This can be seen by expanding both (3.11) and the cumulative distribution function for an exponential, $F_\lambda(t) = 1 - e^{-\lambda t}$, in power series.

[12] The variance of the gamma density, $\frac{\alpha_i}{\beta_i 2}$, provides a measure of the extent of heterogeneity in state i. For state 1 the variance equals 2.6, for state 2 it equals .039.

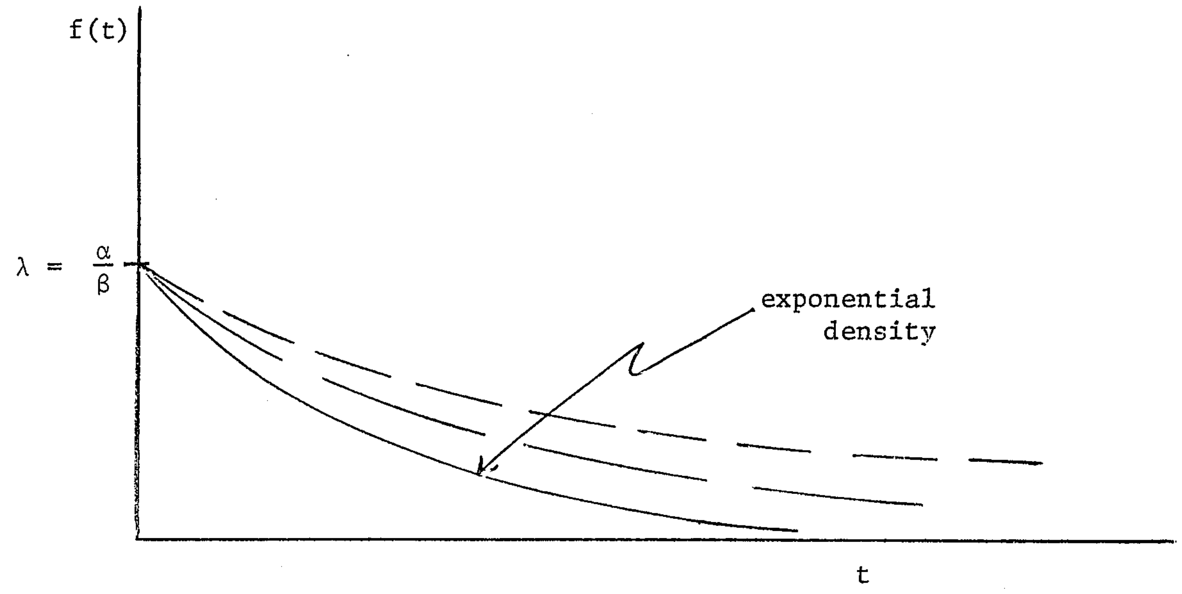
[13] The formula is defined only if V is non-singular. This condition will hold if one or more persons occupy each origin state at time zero.

[14] In the Markov case $r_i(t) = \frac{\lambda_i e^{-\lambda_i t}}{1-[1-e^{-\lambda_i t}]} = \lambda_i$ for i = 1,...,r. This says that the rate of movement is constant, irrespective of duration in the state.

[15] If a matrix is embeddable in a Markov structure this means that it could have been generated by a Markov process; further tests of the sort outlined in section 2.2-I are necessary to confirm this possibility.

[16] This example is originally due to J. Speakman (1967).

FIGURE 1. Family of Compound Exponential Curves*



---

*The family of densities $f(t) = \left(\frac{\alpha}{\beta+t}\right)\left(\frac{\beta}{\beta+t}\right)^{\alpha}$ was constructed by the integration

$$f(t) = \int_0^{\infty} \lambda e^{-\lambda t} g(\lambda)\, d\lambda$$

where $g(\lambda)$ is a gamma density with parameters (α,β).

Table 1. Structure of Heterogeneity in the Population, Simulated Data

A. Individual-level transition matrix

$$M = \begin{pmatrix} .83 & .17 \\ .20 & .80 \end{pmatrix}$$

B. Cumulative waiting time distribution, by state[a]

	(1)	(2)	(3)
Waiting time	State 1 (α=2.1, β=0.9)	State 2 (α=1.4, β=6.0)	Pooled population[b] ($\hat{\alpha}$=0.61, $\hat{\beta}$=0.49)
0	0	0	0
.2	.344	.044	.194
.4	.538	.086	.312
.6		.125	.391
.8		.161	.449
1.0		.194	.493
1.2		.225	.503
1.4		.254	
1.6		.282	
1.8		.307	
2.0		.332	
2.2		.354	
2.4		.376	
2.6		.396	
2.8		.415	

---

[a] The entries in columns (1) and (2) were generated from the cumulative distribution,

$$F_i(t) = 1 - \left(\frac{\beta_i}{\beta_i + t}\right)^{\alpha_i}, \qquad i = 1,2$$

using the indicated (α_i, β_i). Values are reported for $F_i(t) < .500$ in states 1 and combined states, for $F_2(t) < .400$, and for the first entries exceeding these figures.

[b] An identical number of persons was assumed to be present in each state at $t = 0$. The entries in this column were therefore obtained by summing across the states and dividing by 2. The parameters ($\hat{\alpha}$,$\hat{\beta}$) for the pooled population were estimated using the median and interquartile range of the empirical distribution in column 3 (Mood 1950, p. 387).

Table 2. Population-Level Transition Matrices
Estimated from the Simulated Data

	P(1)	P(2)	P(3)
Postulated population structure (equation 3.16)	$\begin{bmatrix} .702 & .298 \\ .037 & .963 \end{bmatrix}$	$\begin{bmatrix} .528 & .472 \\ .062 & .938 \end{bmatrix}$	$\begin{bmatrix} .422 & .578 \\ .079 & .921 \end{bmatrix}$
Projections from Mover-Stayer extension (equation 3.7)[a]	*	$\begin{bmatrix} .574 & .426 \\ .054 & .946 \end{bmatrix}$	$\begin{bmatrix} .501 & .499 \\ .063 & .937 \end{bmatrix}$
Projections from Markov model ($P(n) = [\hat{P}(1)]^n$)	*	$\begin{bmatrix} .504 & .496 \\ .062 & .938 \end{bmatrix}$	$\begin{bmatrix} .372 & .628 \\ .079 & .921 \end{bmatrix}$

---

[a] Procedure uses $\hat{P}(1)$ and $(\hat{\alpha},\hat{\beta})$ from column (3) of Table 1 as input data.

REFERENCES

Bartholomew, D. J.

1967 <u>Stochastic Models for Social Processes</u>. New York: Wiley.

Blauner, Robert.

1964 <u>Alienation and Freedom</u>. Chicago: University of Chicago Press.

Blumen, I., M. Kogan, and P. J. McCarthy.

1955 <u>The Industrial Mobility of Labor as a Probability Process</u>. Cornell Studies of Industrial and Labor Relations, Vol. 6. Ithaca, New York: Cornell University.

Brown, Lawrence.

1970 "On the use of Markov chains in movement research." <u>Economic Geography</u> 46 (June-supplement):393-403.

Coleman, James S.

1964 <u>Introduction to Mathematical Sociology</u>. New York: Free Press.

Feller, William.

1968 <u>Introduction to Probability Theory and its Applications</u>. Vol. I. New York: Wiley (3rd edition).

1971 <u>Introduction to Probability Theory and its Applications</u>. Vol. II. New York: Wiley.

Ginsberg, R.

1971 "Semi-Markov process and mobility." <u>Journal of Mathematical Sociology</u> 1:233-263.

Goodman, L. A.

1961 "Statistical methods for the mover-stayer model." <u>Journal of the American Statistical Association</u> 56:841-868.

Hodge, R. W.

1966 "Occupational mobility as a probability process." Demography 3:19-34.

John, Fritz.

1965 Ordinary Differential Equations. New York: Courant Institute Lecture Notes.

Karlin, S. and J. McGregor.

1957 "The differential equations of birth and death processes and the Stieltjes moment problem." Transactions of the American Mathematical Society 85:489-546.

Lipset, Seymour M., and R. Bendix.

1959 Social Mobility in Industrial Society. Berkeley: University of California Press.

Matras, J.

1960 "Comparison of intergenerational occupational mobility patterns." Population Studies 14:163-169.

McCall, John J.

1971 "An analysis of poverty: Some preliminary findings." Journal of Business 44 (April):125-147.

McFarland, David D.

1970 "Intra-generational social mobility as a Markov process: Including a time-stationary Markovian model that explains observed declines in mobility rates over time." American Sociological Review 35:463-476.

McGinnis, R.

1968 "A stochastic model of social mobility." American Sociological Review 33 (October):712-22.

Mood, Alexander M.

1950 Introduction to the Theory of Statistics. New York: McGraw-Hill.

Palmer, G.

1954 Labor Mobility in Six Cities. New York: Social Science Research Council.

Prais, S. J.

1955 "Measuring social mobility." Journal of the Royal Statistical Society 118 (Series A):55-66.

Pyke, R.

1961a "Markov renewal processes: Definitions and preliminary properties." Annals of Mathematical Statistics 32:1231-1242.

1961b "Markov renewal processes with finitely many states." Annals of Mathematical Statistics 32:1243-1259.

Rogers, Andrei.

1966 "A Markovian analysis of migration differentials." Proceedings American Statistical Association, Social Science Section, Washington, D. C.: American Statistical Association.

Singer, Burton and Seymour Spilerman.

1973 "Identification of a class of semi-Markov processes." To appear.

Smith, James D. and Glen Cain.

1967 "Markov chain applications to household income distribution." University of Wisconsin. Mimeo.

Speakman, J.

1967 "Two Markov chains with a common skeleton." Zeitschrifft für Wahrscheinlichkeitstheorie 7:224.

Spilerman, Seymour.

1972a "The analysis of mobility processes by the introduction of independent variables into a Markov chain." *American Sociological Review* 37 (June):277-94.

1972b "Extensions of the mover-stayer model." *American Journal of Sociology* 78:599-627.

Stein, F. Max.

1967 *Introduction to Matrices and Determinants*. Belmont, Calif.: Wadsworth.

Taeuber, Karl E., L. Chiazze, Jr., and W. Haenszel.

1968 *Migration in the United States*. Washington, D. C.: Government Printing Office.

Tarver, James D. and William R. Gurley.

1965 "A stochastic analysis of geographic mobility and population projections of the census divisions in the United States." *Demography* 2:134-139.