INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

CLUSTERING ON THE MAIN DIAGONAL

Burton Singer
Seymour Spilerman

UNIVERSITY OF WISCONSIN - MADISON

Clustering on the Main Diagonal

Burton Singer
Columbia University

Seymour Spilerman
Russell Sage Foundation

April 1978

ABSTRACT

In this paper we discuss the well-known empirical regularity
commonly referred to as "clustering on the main diagonal." We
show that this phenomenon is to be associated with certain
model types, in particular, with heterogeneity formulations. In
contrast, the reverse situation--underprediction of the main
diagonal elements of a Markov model--will arise when the under-
lying process is a certain common semi-Markov model. We use these
main diagonal relations as an introduction to the general topic
of model discrimination.

CLUSTERING ON THE MAIN DIAGONAL

I. INTRODUCTION

An empirical regularity in mobility matrices which has given rise to a considerable volume of research concerns the phenomenon of "clustering on the main diagonal." The reference model in terms of which this regularity has been identified is the discrete-time Markov chain with stationary transition probabilities. By this model we mean the formal relation between stochastic matrices $\{P(0,k)\}_{k=1,2,\ldots}$ given by

$$P(0,k) = P(0,1)^k, \qquad k = 1,2,3,\ldots . \qquad (1)$$

In practice $P(0,1)$ is usually estimated by $\hat{P}(0,1)$, a transition matrix constructed in the usual way from observations on a population[1] at times $t = 0,1$. $P(0,k)$ then is the k-step transition matrix predicted via the formulation (1) and obtained by raising the observed matrix to the k-th power.

The imagery consistent with the Markov model is one of a homogenous population changing system states (occupations, industries, income categories, etc.) in a manner such that knowledge of an individual's current state conveys complete information about his subsequent movements. This simple formulation has found extensive use, both as a base-line model against which to compare more complex formulations (Hodge 1966; McFarland 1970; McCall 1971; Coleman 1974a), and as a method of forecasting the evolution of social processes (Rogers 1966; Tarver and Gurley 1965;

Lieberson and Fuguitt 1967). For further details on the mathematics of Markov chains, the reader is referred to Feller (1968, chapter 17).

The phenomenon of clustering on the main diagonal refers to the fact that where observations can be taken on a population at time k, k = 2,3,..., as well as at times 0 and 1, the main diagonal entries of the observed k-step transition matrix $\hat{P}(0,k)$ frequently bear the following relation to the main diagonal entries predicted by the Markov model:

$$\left(\hat{P}(0,k)\right)_{ii} > [\hat{P}(0,1)^k]_{ii}, \qquad i = 1,2,...r, \qquad (2)$$

where r equals the number of system states. In words, the Markov model tends to underpredict the observed main diagonal entries.

This fact has stimulated a considerable amount of research since Blumen, Kogan and McCarthy (1955)--hereafter referred to as BKM--first commented upon the phenomenon of clustering. BKM, and most subsequent authors (Goodman 1961; Mayer 1972; McFarland 1970; Bartholemew 1967; Spilerman 1972a, 1972b; Singer and Spilerman 1974, 1976b,1977), have interpreted the presence of clustering as evidence for heterogeneity--the fact that the population being observed consists of individuals who differ in rate of movement, in proclivity to make certain transitions, or in both ways. It has also been remarked that the phenomenon (2) can arise in a homogeneous population in which the length of

stay by an individual in a system state is not exponentially distributed (McGinnis 1968; Ginsberg 1971; Singer and Spilerman 1976b).[2] Finally, Coleman (1964b) has pointed out that clustering can occur as a consequence of "response uncertainty;" for our purposes, a form of measurement error.

Given the volume of discussion about clustering on the main diagonal and the diverse explanations that have been invoked to account for it, it seems reasonable to pose several questions in regard to the inequality (2). In particular, we wish to ascertain which sorts of stochastic processes can produce this phenomenon and which sorts of processes cannot. A second question pertains to whether the reverse of inequality (2)--overprediction of the observed main diagonal elements by a Markov model--can occur, and which sorts of processes will generate that regularity. The value in posing these questions is that they speak to the much neglected task of model discrimination; the construction of simple tests for choosing among competing explanations of a social process.

The particular model types that we shall compare against simple Markov are mixtures of Markov processes and a special parametric family of semi-Markov processes. We restrict attention to these processes because they provide the simplest setting in which to discuss clustering phenomena while still exhibiting the subtle behavior of more complicated models such as those incorporating response uncertainty (Coleman 1964b). It is

important to emphasize that we are discussing the behavior of models per se. Hence, from an empirical point of view, our analysis is relevant to stochastic matrices generated by a large number of observations. Although clustering as in (2) can occur under a Markov model just due to sampling variability, we do not discuss this important aspect of the clustering phenomenon in the present paper.

To address the question of which model types satisfy inequality (2) and which do not, it is necessary to consider explicit formulations of each kind of process. In the next section we therefore present two versions of the mover-stayer model, a more general formulation of population heterogeneity, and an example of a semi-Markov process. All of these models have appeared in the sociological literature or otherwise make sense for social processes. The discussions will be brief and the reader is referred elsewhere for fine details on the mathematics and for estimation procedures. In sections III-V we report our main results on the phenomenon of clustering in relation to each of the model types. In the final pages we provide additional comments on the topic of data collection design as it relates to the task of model discrimination.

5

II. MODEL SPECIFICATION

In this section we outline the structures of four models which later are compared with a Markov chain in regard to the evolution of the main diagonal elements of the transition matrix P(0,t). The initial three models are formulations of population heterogeneity; these are appropriate in the many sociological contexts in which it would be simplistic to treat a population as homogeneous with respect to the behavior under study. In each case, the underlying model for an individual in the population is Markovian; heterogeneity is expressed in the way the individual-level processes are aggregated. The final model constitutes a semi-Markov process; it refers to situations in which one believes an assumption of exponential waiting times (no effect of duration in current state on an individual's departure rate) to be unrealistic. McGinnis' (1968) formulation of "cumulative inertia" is a classic instance in which the theory involved refers to the shape of the waiting time distribution, i.e., to whether or not it is exponential.

Heterogeneity formulations. The heterogeneity models we consider may all be represented by the following generic formalism. Denote by $P_\lambda(0,t)$ the transition matrix for an individual with expected rate of movement equal to λ. (For each individual $P_\lambda(0,t)$ is assumed to evolve according to a first-order Markov chain with stationary transition probabilities.) The observed population-level process may then be written as

$$P_{\mu}(0,t) = \int_{\Lambda} P_{\lambda}(0,t)d\mu(\lambda), \qquad (3)$$

where $d\mu(\lambda)$ is a distribution function which describes the
proportion of type-λ individuals in the population and the
integral sign indicates that we sum the $P_{\lambda}(0,t)$ arrays, each
weighted by the proportion of type-λ individuals in the popula-
tion, over all person-types $\lambda \in \Lambda$. Different formulations of
population heterogeneity then amount to different specifications
of $d\mu(\lambda)$.

It should be noted that the heterogeneity models we consider
all involve mixtures of Markov processes in which mixing is on
the rate of movement parameter λ. This amounts to specifying
that all individuals in the population have the same proclivities
of transferring to the various destination states when they move
(we denote this common propensity by the transition matrix M),
and that population heterogeneity can be expressed entirely in
terms of individual differences in the expected rate of movement.
While restrictive, this formulation is still sufficiently general
to permit the observations we wish to make about model discrimin-
ation and the phenomenon of clustering. There is also precedent
for this formulation in that it underlies BKM's mover-stayer
model as well as Spilerman's (1972a) extension of the mover-stayer
model.

A second restriction, implicit in (3), on the diversity of
population heterogeneity is the requirement that the rate of
movement parameter, λ, be the same for an individual in all

system states. This specification was not made by BKM in their initial presentation of the mover-stayer model, though they did propose it to make the mathematics of more complex heterogeneity models tractable (BKM 1956, pp. 138-46); further, Spilerman's extension (1972a) does utilize this simplification. For ease in comparing the three heterogeneity formulations which we outline, the common generic structure (3) is retained throughout.

A. <u>Blumen, Kogan and McCarthy's (1956) mover-stayer model.</u> This simple formulation of heterogeneity consists of a discrete-time process in which it is assumed that the population is built up from two types of persons--stayers, who never leave their origin states, and movers, who evolve in accordance with a first-order Markov chain with stationary transition probabilities. Despite its simplicity--perhaps, because of this--the mover-stayer model has been used widely to accommodate heterogeneity: in studies of industry change (BKM 1956) and income evolution (McCall 1971), to cite but two examples.

Formally, let Δ = the sampling interval of the process (3 months in BKM's study). The mover-stayer model then is defined for the discrete time sequence $T = \{t : t = k\Delta, k = 0, 1, 2,...\}$. Let $\Lambda = \{\lambda_1 = 0, \lambda_2 = 1\} = \{\text{stayer, mover}\}$ and introduce the mixing distribution

$$
d\mu(\lambda) = \begin{cases} s & \text{if } \lambda = \lambda_1 \\ 1-s & \text{if } \lambda = \lambda_2 \\ 0 & \text{otherwise} \end{cases}
$$

where s is a scalar, $0 < s < 1$. Finally, define a k-step transition matrix for each subpopulation according to

$$
P_{\lambda_1}(0,k\Delta) = \left[P_{\lambda_1}(0,\Delta)^0 \right]^k = I^k = I \quad \text{(Stayers)} \tag{4a}
$$

$$
P_{\lambda_2}(0,k\Delta) = \left[P_{\lambda_2}(0,\Delta)^1 \right]^k \overset{\text{def}}{=} M^k \quad \text{(Movers)}. \tag{4b}
$$

With this specification we obtain for the population-level process

$$
P_{\mu}(0,k\Delta) = sI + (1-s)M^k, \qquad k = 1, 2, \ldots, \tag{5}
$$

which is the familiar mover-stayer model subject to the additional requirement that a common fraction of stayers, s, is present in all origin states.

B. <u>A continuous-time version of the mover-stayer model.</u> As a second specification of heterogeneity, we consider a continuous-time version of BKM's model. This formulation will be useful in highlighting some implications of continuous-time versus discrete-time processes in modeling social phenomena. Moreover, as we have argued elsewhere (Singer and Spilerman 1976b), for processes which evolve continuously in time, a model having this same character has advantages with respect to identification of the

structure of the evolutionary process. Indeed, for the subject BKM studied--industrial mobility--we suggest that it would have been advantageous for them to employ a continuous-time formulation.

The continuous-time analog of the mixture (5) can be specified by defining T = {non-negative real numbers}, with Λ and μ exactly as before. Then stayers still evolve according to the identity matrix

$$P_{\lambda_1}(0,t) = e^{0t(M-I)} = I \qquad (6a)$$

while movers evolve according to the continuous-time Markov process,[3]

$$P_{\lambda_2}(0,t) = e^{\lambda_2 t(M-I)} . \qquad (6b)$$

In equation (6b) the rate of movement parameter λ_2 has the interpretation, $1/\lambda_2$ = expected waiting time between transitions; and M is a stochastic matrix which specifies movement proclivities when a transition occurs.

If observations on a continuous-time mover-stayer mixture are taken at times $t_k = k\Delta$, $k = 0, 1, 2,...$ and if you identify Δ with $1/\lambda$ (as BKM did implicitly), then the observed transition

matrices can be represented by the recipe,

$$P_\mu(0,k\Delta) = sI + (1-s)e^{k(M-I)} \tag{7}$$

which is the continuous-time analog to (5).

C. A continuous distribution of types of individuals. We now specialize equation (3) to mixtures of a continuous type,

$$P_\mu(0,t) = \int_0^\infty P_\lambda(0,t)d\mu(\lambda) . \tag{8}$$

As in the preceding model, each individual in the population is assumed to evolve according to a continous-time Markov process

$$P_\lambda(0,t) = e^{\lambda t(M-I)}, \qquad t \geq 0, \tag{9}$$

where M is a stochastic matrix which describes movement probabilities when a transition occurs.

In contrast with the earlier models we assume now that instead of two types of persons, or n types, there is a continuous distribution of individuals identified by their expected rate of movement. In particular, we specify heterogeneity according to the two parameter family of gamma distributions

$$d\mu(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda}{\Gamma(\alpha)} \qquad \alpha, \beta, \lambda > 0. \tag{10}$$

This function is a very general one and is able to accommodate

a variety of unimodal shapes.

If observations on this continuous-time process are taken at the instants $t_k = k\Delta$, $k = 0, 1, 2, \ldots$, and if Δ is identified with $1/\lambda$ for ease of comparison with the preceding models, the present heterogeneity formulation may be written as (Spilerman 1972a),

$$P_\mu(0, k\Delta) = \left(\frac{\beta}{\beta + k\Delta}\right)^\alpha \left[I - \left(\frac{k\Delta}{\beta + k\Delta}\right) M\right]^{-\alpha} . \tag{11}$$

Semi-Markov processes. In the preceding models, each individual-level process was first-order Markov. This specifies, first of all, that an individual's past locations are immaterial to understanding his future moves; only current state is pertinent. It is this feature of Markov chains which is usually articulated when the assumptions underlying the process are explained (e.g., Tarver and Gurley 1965; Hodge 1966; Lieberson and Fuguitt 1967). Yet, there is a further strong assumption in the Markov model, namely that duration time τ in state i follows an exponential distribution (and is independent of duration times in prior states):

$$\text{Prob}_i(\tau < t) = F_i(t) = 1 - e^{-\lambda t}, \quad i = 1, 2, \ldots, r. \tag{12}$$

Use of the exponential distribution[4] amounts to stating that the probability of departing from state i during the infinitesimal interval t + dt, conditional on being in state i at time t, equals

$$r_i(t)dt = \frac{f_i(t)dt}{1-F_i(t)} = \frac{\lambda e^{-\lambda t}dt}{1-(1-e^{-\lambda t})} = \lambda dt, \quad i = 1, 2, \ldots, r \quad (13)$$

where $f_i(t)$ is the density function corresponding to $F_i(t)$. Substantively, this result says that there is no impact of duration in a system state on the probability of leaving. Individuals neither settle down (in an occupation, residence location) nor grow weary of the setting.

It has been pointed out (Morrison 1967; McGinnis 1968; Land 1969) that this assumption of a constant departure rate is untenable for many social processes. Indeed, McGinnis' (1968) "axiom of cumulative inertia" speaks directly to this point; he contends that individuals are more likely to remain at a residence location, or in a job, the longer they have been there. To model such a process, we would need to have $r_i(t)$ in equation (13) appearing as a decreasing function of time, rather than as a constant.

Semi-Markov models constitute a class of stochastic processes which maintain the assumption that past locations are immaterial to future states given current location, while permitting the duration-time distribution to be more general than exponential. We require the following terminology for this model class. Let M be a stochastic matrix of transition probabilities which describe the propensity to move to particular states when a transition occurs. Let P(0,t) be the transition matrix constructed from observations on a population at times 0 and t. Finally, let $F_i(t)$, $i = 1, 2, \ldots, r$, be a distribution function which has the interpretation, "probability that a move has occurred by time t;" we assume it has a density function $f_i(t)$.

The mathematical formalism of semi-Markov processes that we use in the final two models derives from the system of integral equations

$$p_{ij}(0,t) = \delta_{ij}[1 - F_i(t)] + \sum_{k=1}^{r} \int_0^t f_i(s)m_{ik}p_{kj}(0,t-s)ds, \quad (14)$$

where $\delta_{ij} = 1$ if $i = j$, 0 if $i \neq j$; and $1 \leq i, j \leq r$. These expressions, known as the backward equations for a continuous-time semi-Markov process (Feller 1971, pp. 483-497), are amenable to the following interpretation: (1) When $i \neq j$, $p_{ij}(0,t)$ consists of the sum of products of three factors: the probability of a first departure from state i at time s, the probability of a state i to state k transition at that instant, and the probability of transferring to stage j by some combination of moves in the interval $t-s$. The summation is over all intermediate states k and over all time divisions s in the interval $(0,t)$. (2) When $i = j$, in addition to the above term, there is the possibility of not transferring out of state i during $(0,t)$. This probability is given by the first term.

Although this is not the most general formulation of semi-Markov processes (see Pyke 1961a, 1961b; Ginsberg 1971; Singer and Spilerman 1974 for more detailed statements), it is a convenient starting point for our purposes. If we make the further assumption that $F_i(t) = F(t)$, $i = 1, 2,\ldots, r$--in keeping with our earlier discussion concerning state independence of the waiting-time distribution--then the semi-Markov model has a simple representation

in matrix form,

$$P(0,t) = \sum_{n=0}^{\infty} [F_n(t) - F_{n+1}(t)]M^n , \qquad (15)$$

where $F_n(t)$ denotes the n-fold convolution of F(t) with itself.[5]
This formulation often permits tractable representations to be
constructed once F(t) is specified. Note, incidentally, that
if F(t) is specified by equation (12), the semi-Markov model
(15) reduces to the familiar continuous-time Markov chain
$P(0,t) = e^{\lambda t(M-I)}$ [see note 3].

D. A semi-Markov model with increasing departure rate. This
formulation would pertain to processes in which the probability
of a move increases with duration in the state. Substantive contexts
in which this specification makes sense are the stage models of
developmental psychologists (e.g., Kohlberg 1973; Piaget 1954;
Loevinger 1966). The basic notion here is that an individual
passes through a more or less ordered sequence of states (develop-
mental stages), whose onsets are age dependent. For some behaviors
or abilities (e.g., psychosexual stages), duration at a particular
developmental level may be programmed genetically into the organism.
For other phenomena (e.g., cognitive stages) experience through
interaction with the environment plays a more central role in pre-
paring an individual for entering the next developmental level.

In stage theories, then, there is a basic notion that the
probability of departure increases with duration in a state. We can
formalize this assumption by specifying F(t) to be a gamma distribu-
tion (equation 10), with $\alpha = 2$ and $\beta = \lambda > 0$ arbitrary:

$$\text{Prob}(\tau < t) = F(t) = \int_0^t \lambda^2 u e^{-\lambda u} du = 1 - (1+\lambda t)e^{-\lambda t} . \qquad (16)$$

Here, the rate of movement out of state i, i = 1, 2, ...,r, is given by $r(t) = \lambda^2 t/[1 + \lambda t]$, which increases monotonically to an asymptote $r(\infty) = \lambda$ (Figure 1). Specific theories as to stage sequences could be built directly into the M matrix of equation (15); see Singer and Spilerman (1978) for details.

With the specification of F(t) in (16), the difference between the n-fold and (n+1)-fold convolution is

$$F_n(t) - F_{n+1}(t) = e^{-\beta t}\left[\frac{(\beta t)^{2n}}{(2n)!} + \frac{(\beta t)^{2n+1}}{(2n+1)!}\right]. \tag{17}$$

Substituting (17) into (15) and evaluating the resulting power series yields[6]

$$P(0,t) = e^{-\beta k\Delta}\left[\cosh \beta k\Delta M^{1/2} + M^{-1/2}\sinh \beta k\Delta M^{1/2}\right], \tag{18}$$

in which cosh and sinh are hyperbolic functions of matrix argument. Equation (18) is the transition matrix for a semi-Markov process in which the probability of a state change increases with duration in the state.

III. MOVER-STAYER MODELS AND CLUSTERING

Eigenvalue conditions. We will have at our disposal some powerful mathematical machinery if we alter the definition of clustering very slightly. In particular, the specification
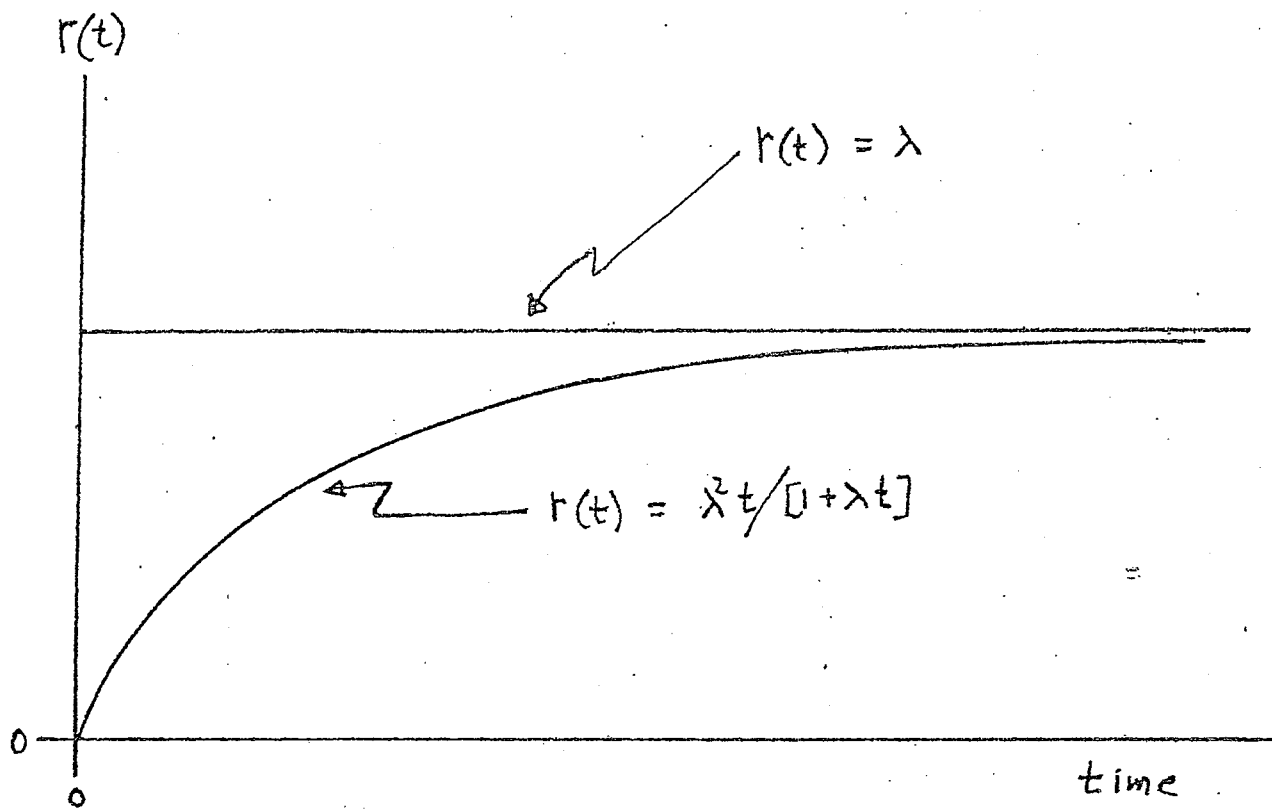
Figure 1. Rate of Movement Function r(t) for Different Waiting-Time Distributions.

[a]$r(t) = \lambda$ for exponentially distributed waiting times. $r(t) = \lambda^2 t/[1 + \lambda t]$ for gamma distributed, waiting times.   –

introduced by BKM (1956, pp. 79-95) and used by subsequent researchers is[7]

$$(P^k(0,\Delta))_{ii} < (P(0,k\Delta))_{ii}, \quad i = 1, 2, \ldots, r. \quad (19)$$

In words, it is required that each diagonal element of the Markov model underpredict the corresponding entry of the observed process. The formulation of clustering we propose instead is

$$\sum_{i=1}^{r} (P^k(0,\Delta))_{ii} < \sum_{i=1}^{r} (P(0,k\Delta))_{ii} ; \quad (20)$$

equivalently,

$$\text{trace } P_\mu^{\,k}(0,\Delta) - \text{trace } P_\mu(0,k\Delta) < 0. \quad (21)$$

Thus the inequality is assumed to hold between the sums of main diagonal elements, though not necessarily for each pair individually. We will indicate momentarily the mathematical advantage of this specification. First, we show that it is a reasonable definition of clustering.

If equation (19) is satisfied for a mobility process, then (21) will also hold. Thus, in this most common situation, the two formulations will yield identical results. Similarly, if the reverse of inequality (19) is satisfied (overprediction by the Markov model), then the reverse of (21) will hold. Only in instances where some main diagonal elements satisfy (19), while others do not, are the two

formulations in potential disagreement. Yet, in this case, under (19) there is no explicit criterion as to when an inference of clustering can be made; for example, do one or two reversals among main diagonal entries in a large matrix invalidate such a conclusion? BKM (1956, pp. 60-64), incidentally exhibited as evidence of clustering matrices in which one reversal often was present. In summary, inequality (21) does not alter the character of the traditional formulation and has the advantage of providing explicit criteria under which clustering can be asserted when (19) is an excessively stringent requirement.

For mathematical purposes the attractiveness of inequality (21) is that it enables a relation between clustering and the eigenvalue properties of matrices to be established. In particular, let

$$D = \begin{bmatrix} \delta_1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \delta_r \end{bmatrix} \tag{22}$$

be the array of eigenvalues corresponding to some matrix P. It is well known (e.g., Bellman, 1970, p. 96) that

$$\text{trace } P \overset{\text{def}}{=\!=} \sum_{i=1}^{r} (P)_{ii} = \sum_{i=1}^{r} \delta_i \; ;$$

that is, the sum of the main diagonal elements of a matrix is equal

to the sum of its eigenvalues. Thus if δ_i is an eigenvalue of $P_\mu(0,k\Delta)$, and δ_i^* an eigenvalue of $P_\mu^k(0,\Delta)$, an equivalent expression to (21) is

$$\sum_{i=1}^{r} \delta_i^* - \sum_{i=1}^{r} \delta_i < 0. \tag{23}$$

Mover-stayer mixtures. Now consider a transition matrix $P_\mu(0,k\Delta)$, constructed from observations at $t = 0$ and $t = k\Delta$ on BKM's mover-stayer model (5),

$$P_\mu(0,k\Delta) = sI + (1-s)M^k . \tag{24}$$

If the eigenvalues of $P_\mu(0,k\Delta)$ are represented by matrix D in (22) and the eigenvalues of M by an analogous diagonal matrix $V = \{v_i\}$, then each eigenvalue δ_i may be expressed as

$$\delta_i = s + (1-s)v_i^k , \quad i = 1, 2,\ldots,r. \tag{25}$$

To see this assume that H is a matrix whose columns are eigenvectors of $P_\mu(0,k\Delta)$ in (24). Then,

$$D = H^{-1}P_\mu(0,k\Delta)H$$

$$= H^{-1}[sI + (1-s)M^k]H = sI + (1-s)H^{-1}M^kH$$

$$= sI + (1-s)[H^{-1}MH]^k . \tag{26}$$

Since D is a diagonal matrix, $H^{-1}MH$ must also be a diagonal matrix--

call it V--and its entries are the eigenvalues of M. Equation (25) follows from this argument.

By similar reasoning it can be shown that each eigenvalue δ_i' of $P_\mu^{\ k}(0,\Delta)$ has the form

$$\delta_i' = [s + (1-s)v_i]^k \ .$$ (27)

With the above mathematical equipment at hand, the clustering criterion (20) may be written as (using [23], [25] and [27]),

$$\sum_{i=1}^{r} \left[\left(s + (1-s)v_i \right)^k - \left(s + (1-s)v_i^{\ k} \right) \right] < 0.$$ (28)

We have therefore transformed an inequality between main diagonal elements of mover-stayer and Markov matrices into an inequality between their eigenvalues. Further, the eigenvalues v_i in (28) pertain to the matrix of structural parameters M, which is constant over time even though $P_\mu(0,t)$ is not.

To ascertain the conditions under which the inequality (28) will be satisfied, we first determine whether or not all terms in the sum

$$\sum_{i=1}^{r} \left[\left(s + (1-s)v_i \right)^k - \left(s + (1-s)v_i^{\ k} \right) \right]$$ (29)

necessarily have the same sign. If v_i is a positive real eigenvalue other than $v_i = 1$, then by Jensen's inequality (Feller 1971, pp. 153-4),

$$[s + (1-s)v_i]^k - [s + (1-s)v_i^{\ k}] < 0 \qquad (30)$$

for $k = 2, 3, 4,...$ and all $s \in (0,1)$. Our first conclusion, therefore is that in a comparison between BKM's mover-stayer model and a Markov chain, <u>if all eigenvalues of $P_\mu(0,\Delta)$ are distinct, positive, and real, the inequality (21) will hold and clustering on the main diagonal will be observed.</u>

We further see that if the difference (29) is ever to be positive and a reversal of (28)--equivalently (21)--observed, it must occur for matrices having complex conjugate or negative real

eigenvalues for which Jensen's inequality is reversed. We therefore ask whether there are regions inside the unit disk in the complex plane for which

$$\text{Re}[s + (1-s)v]^k - \text{Re}[s + (1-s)v^k] > 0 \tag{31}$$

where Re(.) denotes the real part of the eigenvalue. The unit disk is the relevant region to examine because all eigenvalues of stochastic matrices are restricted to this region. Further, it is sufficient to examine the real parts of the eigenvalues because complex eigenvalues of matrices with real entries come in conjugate pairs (e.g., $x + iy$ and $x - iy$), so that for matrix P with eigenvalues δ_i, $i = 1, 2, \ldots, r$,

$$\text{trace } P = \sum_{i=1}^{r} \delta_i = \sum_{i=1}^{r} \text{Re } \delta_i .$$

Case $k = 2$ in equation 31. Writing v in the form $v = x + iy$, we obtain

$$\text{Re}[s + (1-s)v]^2 - \text{Re}[s + (1-s)\bar{v}^2] = s(s-1)[(1-x)^2 - y^2] > 0$$

which is satisfied by eigenvalues v with components such that

$$(1 - x - y)(1 - x + y) < 0. \tag{32}$$

The eigenvalues in the unit disk for which this inequality holds are shown in the shaded region of Figure 2.[8]

An important consequence of (32) is that

$$Re[s + (1-s)v]^2 - Re[s + (1-s)v^2] < 0$$

for _any_ eigenvalue v of a stochastic matrix M of order $r \leq 4$ and for every $s \in (0,1)$. This is a consequence of the inequalities of Karpelewitsch (1951) [also see Singer and Spilerman (1976), pp. 10-13] which restrict the eigenvalues of general $r \times r$ stochastic matrices according to

$$\left[\frac{1}{2} + \frac{1}{r}\right]\pi \leq arg(v-1) \leq \left[\frac{3}{2} - \frac{1}{r}\right]\pi \ . \tag{33}$$

These inequalities define the cone-shaped regions K_3 and K_4 drawn in dashed lines in Figure 2. We observe that when $r \leq 4$ all eigenvalues of a stochastic matrix must lie in the complement of the shaded region in Figure 2 and reversal of (21) cannot occur.

Case k = 3 in equation (31). The region for which the inequality holds is now more complicated geometrically than the shaded region of Figure 2. Furthermore, the sign of the difference
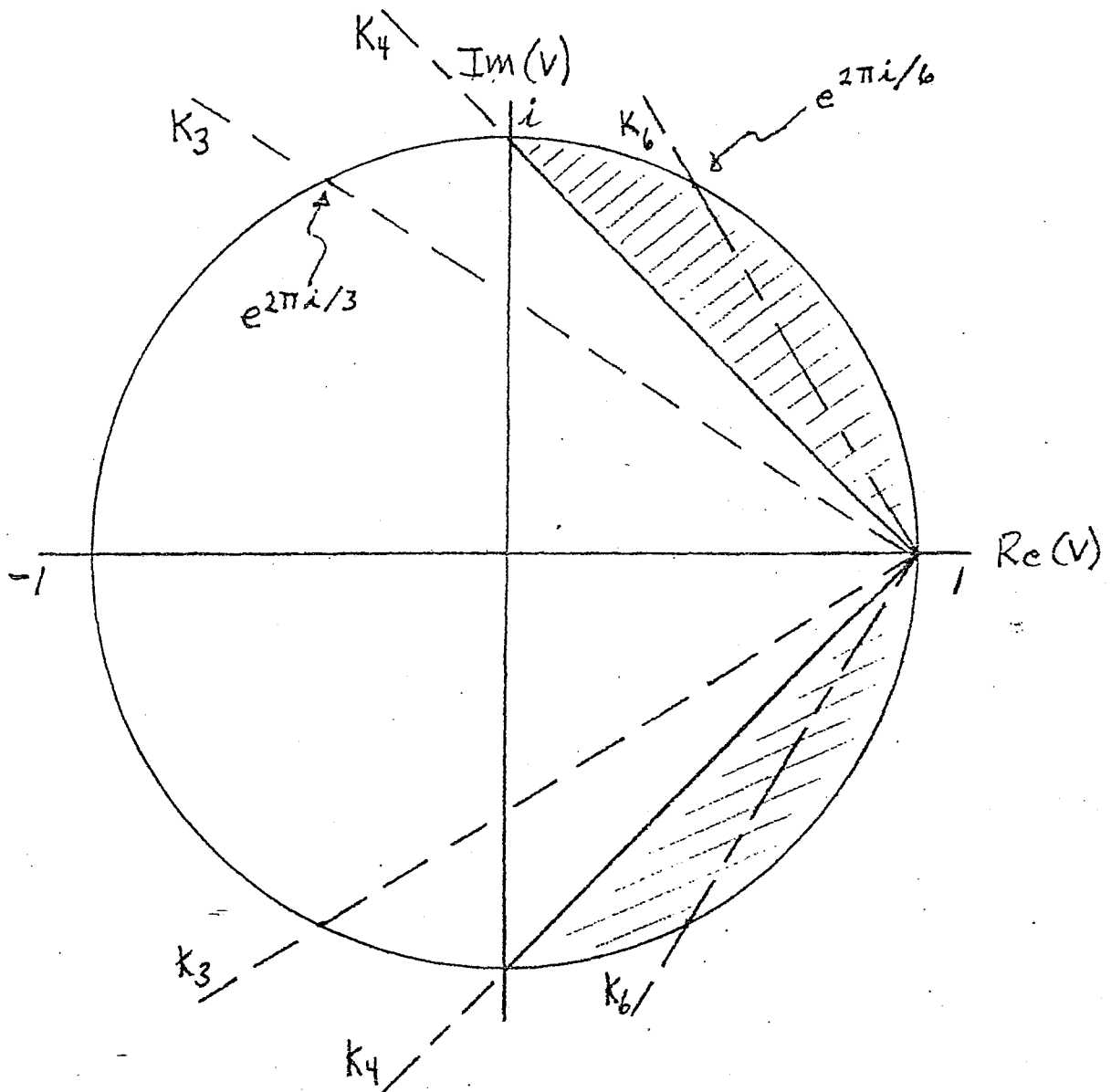
$$Re[s + (1-s)v]^3 - Re[s + (1-s)v^3]$$

Figure 2. Eigenvalue Region for a Reversal of Equation 28 with k = 2
(Shaded area), Together with Cone-shaped Restrictions on
Eigenvalue Locations of 3 x 3, 4 x 4, and 6 x 6 Matrices[a].

[a]Each cone-shaped region (dashed lines) is symmetric with respect to the
x-axis.

depends on s, unlike in the situation for k = 2. By a calculation similar to that in the previous case we obtain the shaded regions in Figure 3 as the location of eigenvalues of matrices M for which a reversal of (21) may occur.[9]

The above restrictions have rather far-reaching implications as will now be explained. To see this consider the more refined restrictions on eigenvalues of stochastic matrices of orders 3, 4, and 6 due to Karpelewitsch (1951). In particular, the shaded areas in Figures 4, 5, and 6 display the regions within the unit circle of the complex plane in which all eigenvalues of matrices of orders 3,4, and 6, respectively, must lie. These areas should be considered in conjunction with the shaded region in Figure 3, which reports eigenvalue conditions <u>necessary</u> for a reversal of (21)--equivalently (28). It is evident that there is an overlap between the regions, even for stochastic matrices of order 3. Further, the area of overlap widens as the matrix order is increased.

For 3 × 3 stochastic matrices the region of overlap (Figures 3 and 4) is restricted to negative real values of v, in particular to the region $v \in [-1, -\frac{(1+s)}{(2-s)}]$. Note, incidentally, that this constrains s to the interval $(0,\frac{1}{2}]$. By a judicious choice of M and s we can therefore illustrate mover-stayer mixtures in which this heterogeneity formulation underpredicts the diagonal elements of the corresponding Markov process.
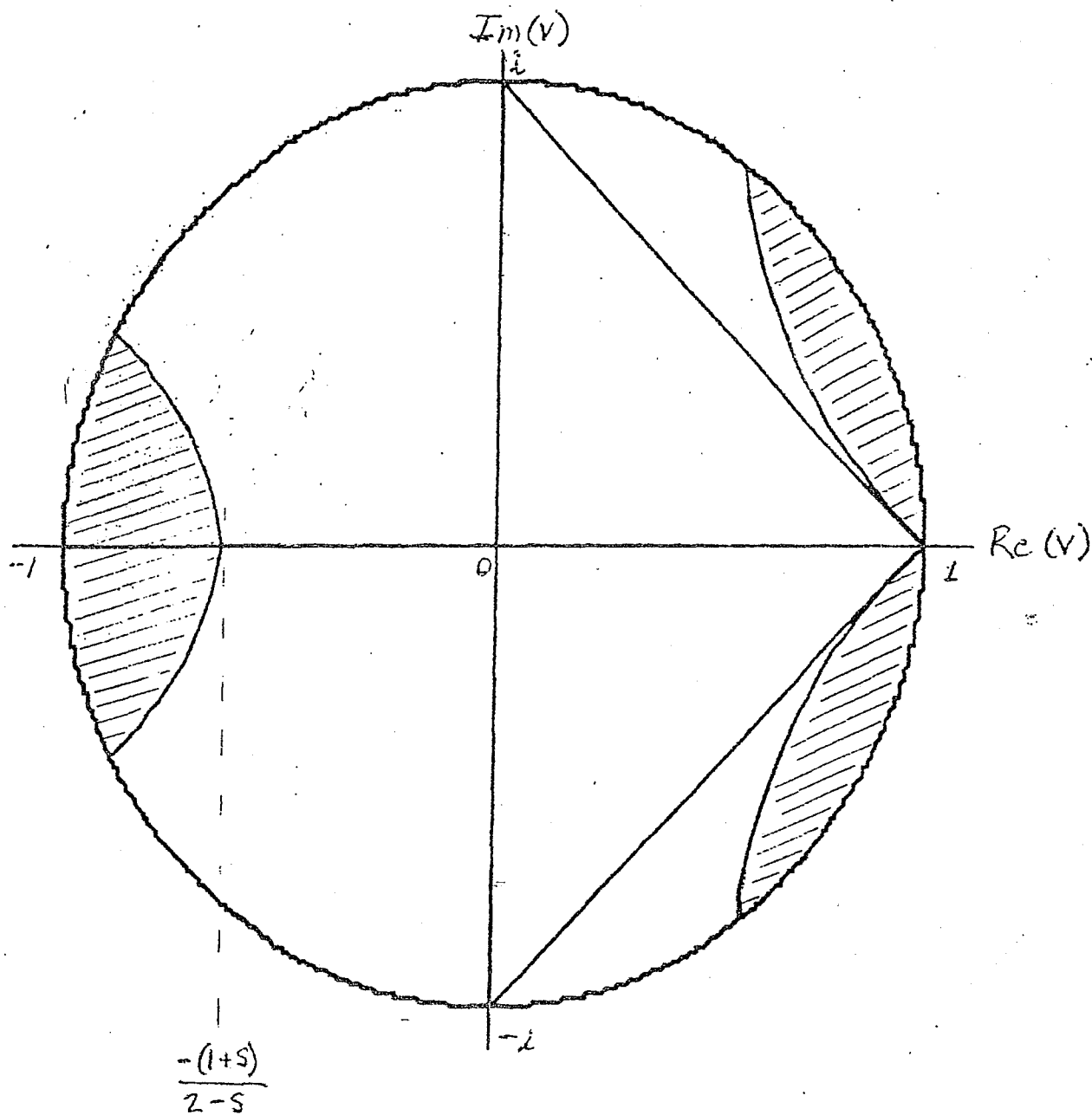
Figure 3.   Eigenvalue Regions for a Reversal of Equation (28) with k = 3
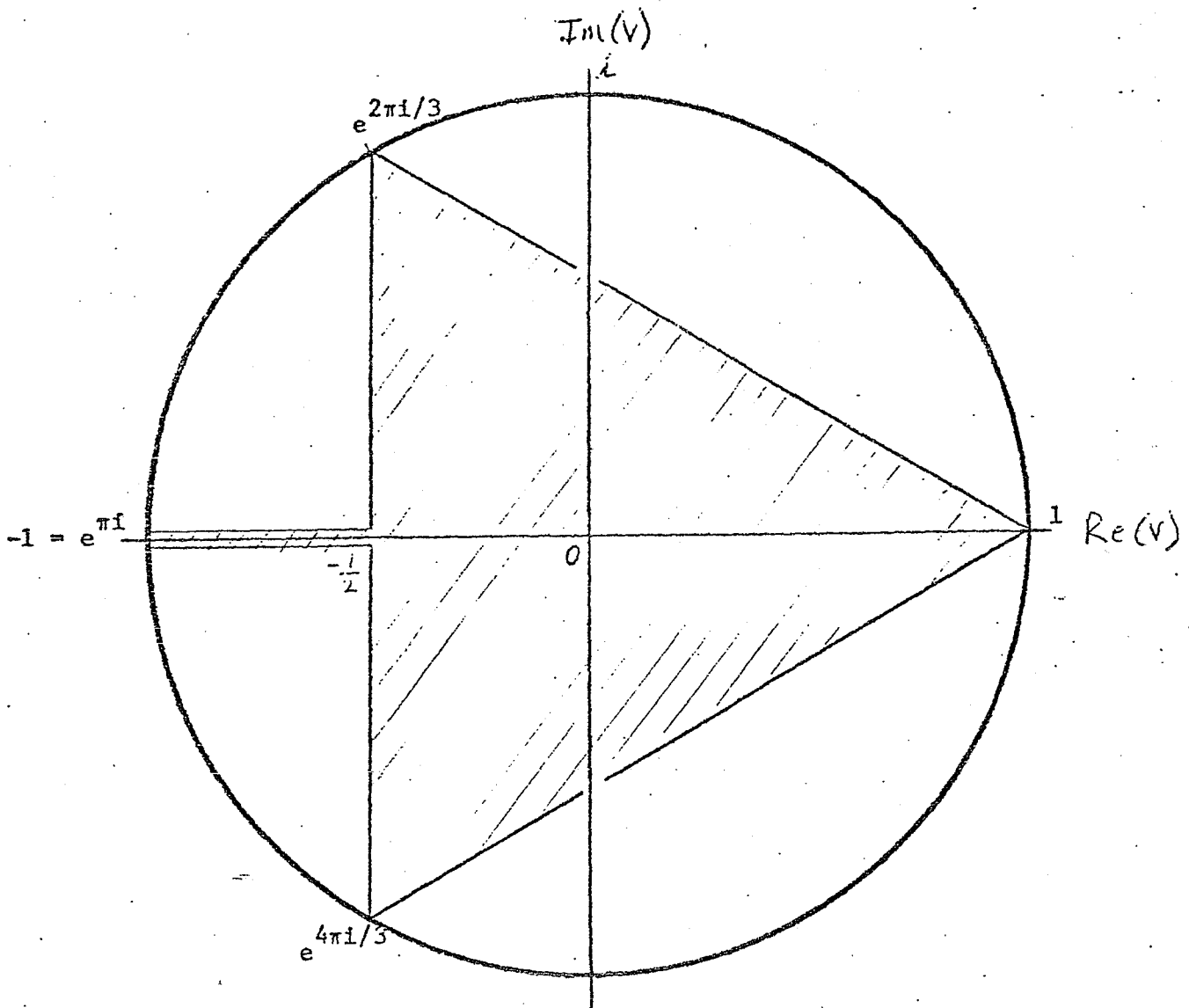   (Shaded area)[a].

[a]Value of s in this graph = 0.18.

Figure 4. Restrictions on Eigenvalues of 3 x 3 Stochastic Matrices[a].

———————————

[a]All eigenvalues of a 3 x 3 stochastic matrix must lie in the triangle or on the negative real axis.
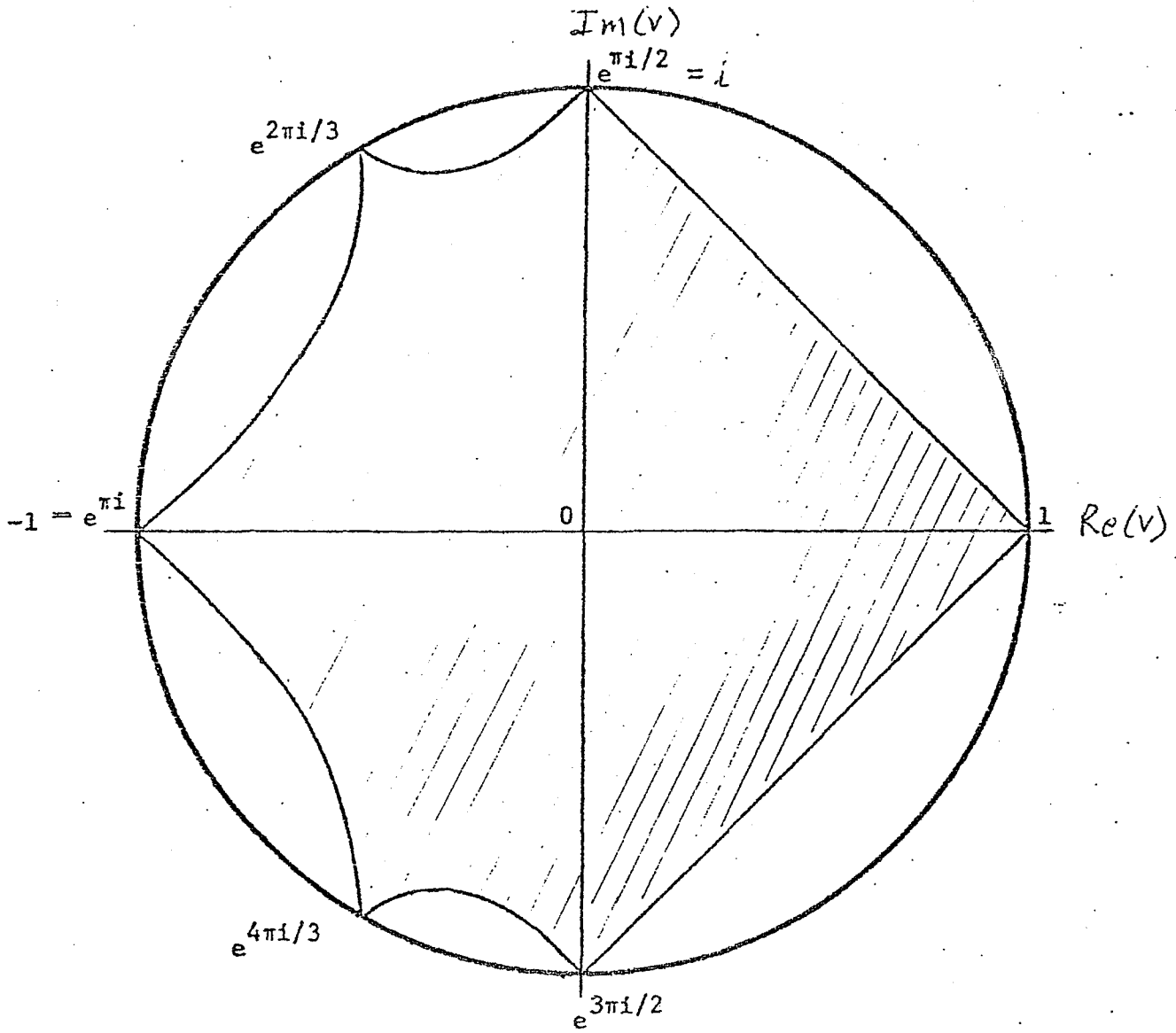
Figure 5. Restrictions on Eigenvalues of 4 x 4 Stochastic Matrices[a].

----

[a]All Eigenvalues of a 4 x 4 stochastic matrix must lie within Robin's cape or on its boundary.
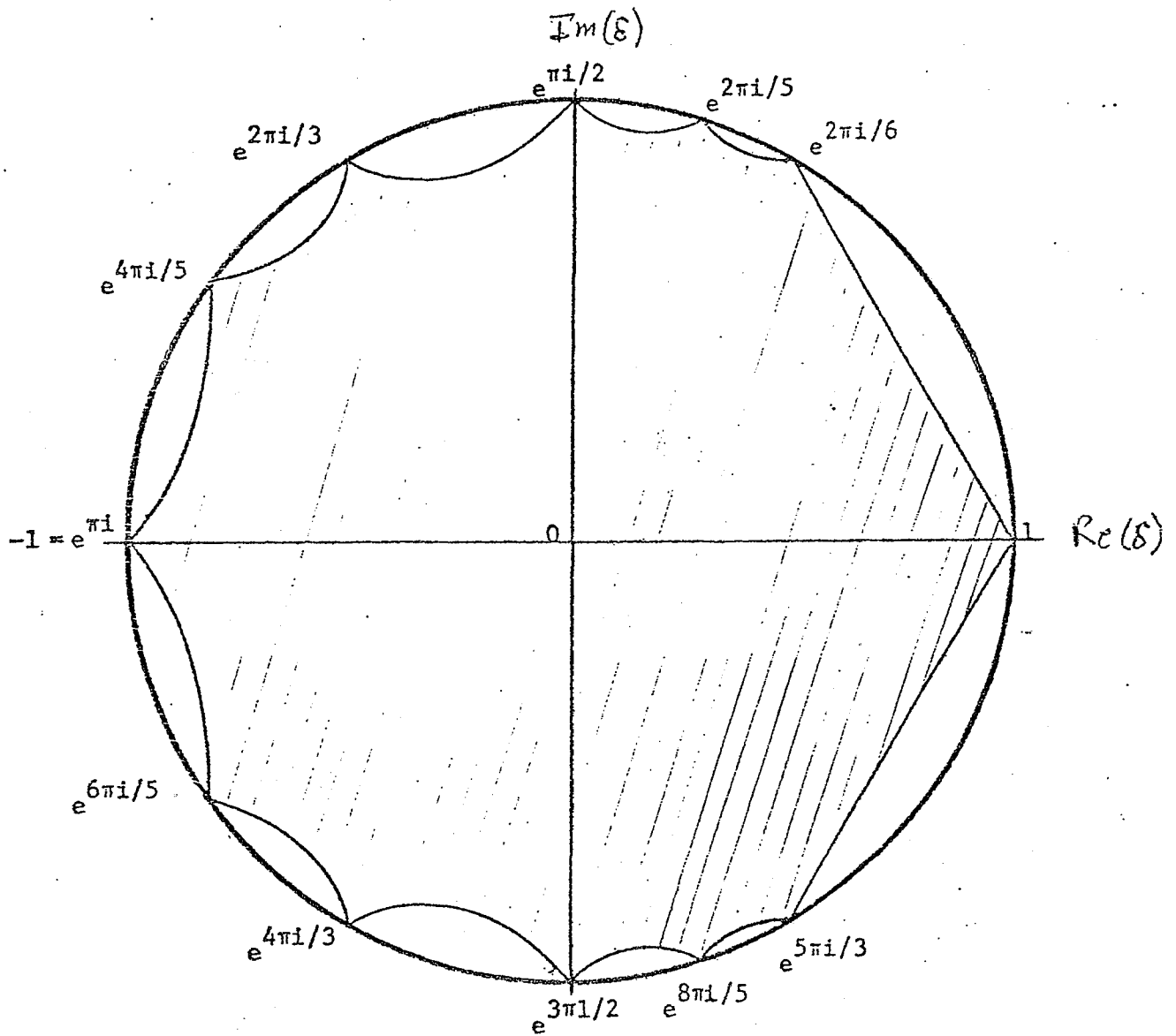
Figure 6. Restrictions on Eigenvalues of 6 x 6 Stochastic Matrices[a].

---

[a]All eigenvalues of 6 x 6 stochastic matrices must lie within Batman's cape or on its boundary.

Example 1. Suppose an empirical process evolves according
to the mover-stayer formulation,

$$P_{\mu}(0,\Delta) = sI + (1-s)M$$

$$= \begin{pmatrix} .1302 & .0326 & .8372 \\ .0352 & .0963 & .8685 \\ .8685 & .0013 & .1302 \end{pmatrix}$$

where s = .0625 and

$$M = \begin{pmatrix} .0722 & .0347 & .8931 \\ .0375 & .0361 & .9264 \\ .9264 & .0014 & .0722 \end{pmatrix} .$$

Comparing the mixture model $P_{\mu}(0,k\Delta)$--i.e., the observed process--
with the Markov model $P_{\mu}^{k}(0,\Delta)$ which uses $P_{\mu}(0,\Delta)$ as a one-step
transition matrix, we find

$$\text{trace } P_{\mu}^{2}(0,\Delta) = 1.502 < \text{trace } P(0,2\Delta) = 1.755$$

but

$$\text{trace } P_{\mu}^{3}(0,\Delta) = .6486 > \text{trace } P_{\mu}(0,3\Delta) = .6089 .$$

Continuing these calculations,

$$\text{trace } P_{\mu}^{4}(0,\Delta) = 1.248 < \text{trace } P_{\mu}(0,4\Delta) = 1.548$$

$$\text{trace } P_{\mu}^{5}(0,\Delta) = .8248 > \text{trace } P_{\mu}(0,4\Delta) = .7784$$

and so forth.

Example 2. As part of a study of interpersonal relationships among American high school youth in the 1950's, J. Coleman (1961) asked students in Northern Illinois high schools in October, 1957 and again in May, 1958 whether or not:

1) they perceived themselves to be members of the leading crowd in their school;

2) they can maintain their principles and be a member of the leading crowd.

Affirmative answers to each question were scored + and negative answers were scored -. Thus, an individual can respond to the above questions in one of four possible ways at each observation time: (response to (1), response to (2)) = (+,+), or (+,-), or (-,+), or (-,-). We then identify these responses as possible states of a stochastic process. In connection with this survey, Coleman proposed a theory about attitude changes in an adolescent population in which individuals could alter their views on either issue (1) or (2) at any one time but could not change their attitude on both issues simultaneously. This theoretical proposal implies that transition matrices M describing attitude changes when they occur should be of the special form

$$M = \begin{pmatrix} 0 & \alpha & 1-\alpha & 0 \\ \beta & 0 & 0 & 1-\beta \\ \gamma & 0 & 0 & 1-\gamma \\ 0 & \delta & 1-\delta & \theta \end{pmatrix}$$

where $0 < \alpha, \beta, \gamma, \delta < 1$.

These matrices have eigenvalues $v_1 = 1$, $v_2 = -1$,

$v_3 = \{\frac{1}{2}[A - (A^2 - 4B)^{1/2}]\}^{1/2}$ and $v_4 = -\{\frac{1}{2}[A - (A^2 - 4B)^{1/2}]\}^{1/2}$

where

$$A = \delta(1-\beta) + (1-\gamma)(1-\delta) + \alpha\beta + \gamma(1-\alpha)$$

$$B = \alpha\beta(1-\delta)(1-\gamma) - \beta\delta(1-\alpha)(1-\gamma)$$
$$+ \gamma\delta(1-\alpha)(1-\beta) - \alpha\gamma(1-\delta)(1-\beta) .$$

A somewhat tedious calculation verifies that if any M, as defined above, is incorporated in a discrete-time mover-stayer mixture, then there is an <u>interval</u> of stayer fractions $(0,s^*)$ such that

$$\text{trace } \left(sI + (1-s)M\right)^3 > \text{trace } \left(sI + (1-s)M^3\right)$$

for every $s \in (0,s^*)$, if and only if

$$-2 + 3\sum_{i=2}^{4} v_i^2 - 2\sum_{i=2}^{4} v_i^3 = 3[1 + A - \sqrt{A^2 - 4B}\,] > 0 .$$

Furthermore, for such M, s^* is given by the formula

$$s^* = \frac{2 - 3\sum_{i=2}^{4} v_i^2 + 2\sum_{i=2}^{4} v_i^3}{\sum_{i=2}^{4} (v_i - 1)^3} .$$

To illustrate the calculations, suppose that $\alpha = \beta = \frac{1}{4}$ and $\gamma = \delta = \frac{1}{2}$. Then $v_1 = 1$, $v_2 = -1$, $v_3 = \frac{1}{4}$, $v_4 = -\frac{1}{4}$ and $s* = 0.342$. This means that for any value of $s\varepsilon(0,.342)$, a mover-stayer mixture having the above M-matrix will show under-prediction of the main diagonal entries of the corresponding Markov process.

Despite the fact that reversals can take place for discrete-time mover-stayer matrices of order 3 and 4, the analogous inequality cannot occur for any 3×3 or 4×4 continuous-time mover-stayer mixture. This arises because every eigenvalue v of $e^{(M-I)}$, where M is a stochastic matrix, must lie inside or on the boundary of the heart-shaped region H_r in the complex plane (Figure 7). [See Singer and Spilerman (1976, pp. 10-13) for details; see Runnenberg (1962) for a proof.] Further, if $v \varepsilon H_r$, then $s + (1-s)v$ is also in H_r, all $s \varepsilon (0,1)$. Since for $r \leq 4$ the hearts H_3 and H_4 are properly contained in the complement of the shaded region in Figure 2, reversals of (21) cannot occur for continuous-time models of these orders.

With respect to mover-stayer mixtures we therefore conclude:

(i) A reversal of (21)--equivalently (28)--cannot occur when all eigenvalues of $P_\mu(0,\Delta)$ are distinct, real, and positive. This statement holds irrespective of whether the model is a discrete or a continuous-time formulation; for all integer values of r and k, and for all spacings $\Delta > 0$.
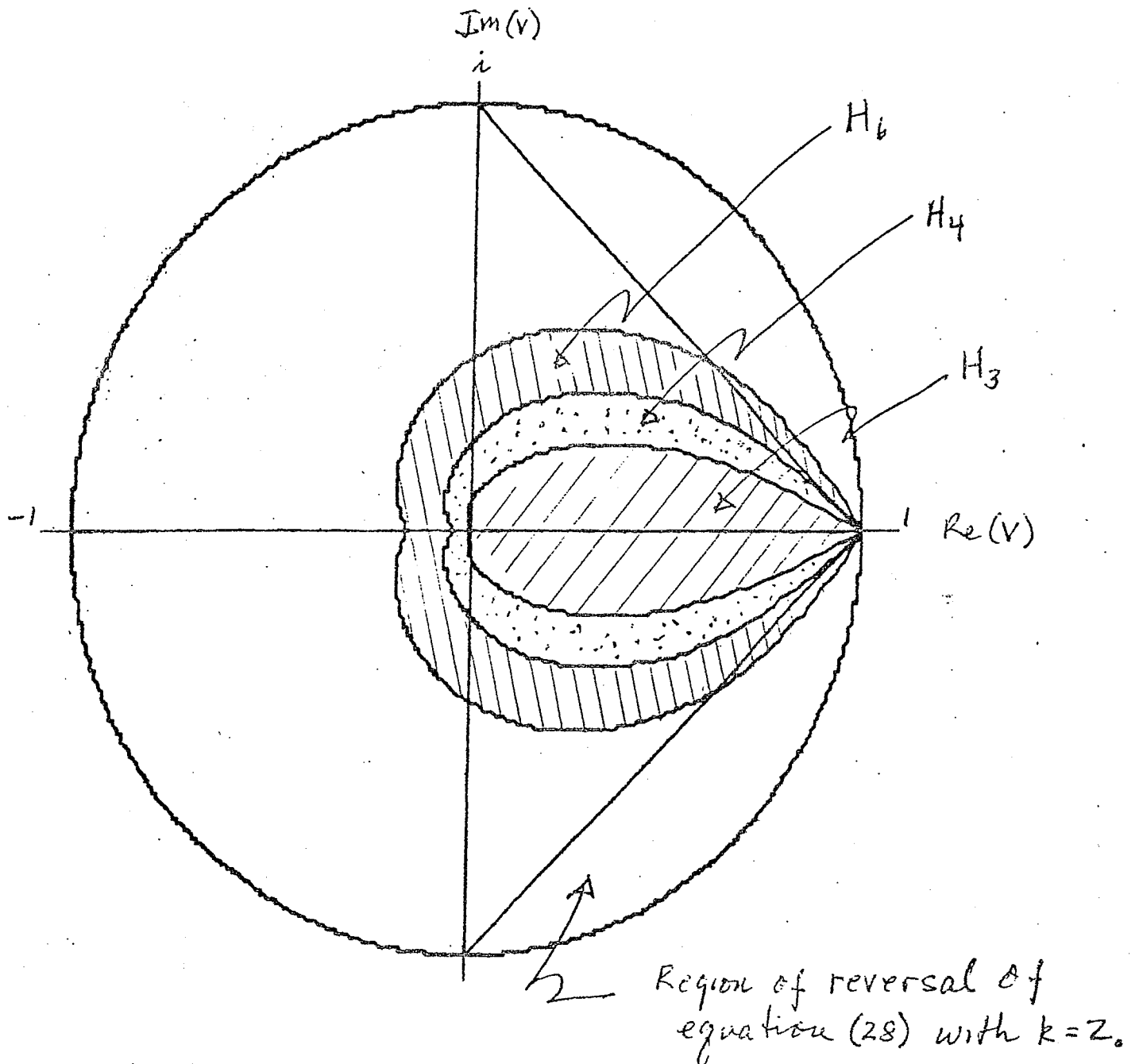
Figure 7. Restrictions on Eigenvalues of 3 x 3, 4 x 4, and 6 x 6
            Continuous-Time Mover-Stayer Mixtures[a].

_____

[a]Each heart-shaped region includes the ones of smaller size. Thus, the
eigenvalues of a 3 x 3 continuous-time mover-stayer mixture must lie within
the inner-most region (H_3), the eigenvalues of a 4 x 4 continuous-time mover-
stayer mixture must lie within the intermediate region (H_4), etc.

(ii) For k = 2 in (21) we have the further restriction that a reversal cannot occur for negative real eigenvalues (see Figure 2); complex eigenvalues are necessary.

(iii) As the order of matrix P is increased the eigenvalue region in which reversals can occur becomes more extensive. It is possible, however, to exhibit reversals for k = 3 in a discrete-time mover-stayer model even for matrices as small as 3 × 3 and 4 × 4.

(iv) Finally, a general <u>sufficient</u> condition for a reversal of (21) is that all eigenvalues lie in the shaded region. In large-order matrices a not uncommon situation would be for some to fall in this region while others lie outside it. In this circumstance the eigenvalue conditions constitute only <u>necessary</u> criteria and a direct calculation of equation (29) is necessary.

What has been demonstrated of importance in this section is that mover-stayer mixtures do not necessarily imply clustering on the main diagonal; in fact, under-prediction of the corresponding elements in a Markov chain can occur. Thus, a failure to observe clustering in an empirical process does not rule out the possibility that the process was generated by a mover-stayer mixture, as long as some eigenvalue of $\hat{P}(0,\Delta)$ lies in the region of reversal (Figures 2,3). However, a failure to observe clustering when no eigenvalue lies in this region would rule out the possibility that the underlying model is mover-stayer. In particular, when all eigenvalues are distinct, real, and positive, clustering must occur when the empirical process is mover-stayer.

IV. A CONTINUOUS MASS DISTRIBUTION AND CLUSTERING

Consider the class of mixtures of continuous-time Markov chains having transition probabilities given by (8) and (9), i.e., by

$$P_\mu(0,t) = \int_0^\infty e^{\lambda t(M-I)} d\mu(\lambda) \tag{34}$$

where the rate parameter λ is distributed in the population according to the probability measure $\mu(\lambda)$ via the relation,

$$\text{Prob}(\lambda < c) = \int_0^c d\mu(\lambda) . \tag{35}$$

Even without specifying $\mu(\lambda)$ further we can make a rather general statement about clustering for this mixture process. In particular, if $P_\mu(0,t)$ has distinct, real, positive eigenvalues[10], then for any k and any spacing interval Δ we have

$$\text{trace } P_\mu{}^k(0,\Delta) - \text{trace } P_\mu(0,k\Delta) < 0 . \tag{36}$$

Thus, inequality (21) holds and clustering must occur (see Appendix 1). A particularly important consequence of this result is that if an observed matrix $\hat{P}(0,\Delta)$ satisfies the noted eigenvalue condition but inequality (36) is reversed, no model of the form (34) could have generated the matrix.

For arrays $P_\mu(0,\Delta)$ with complex eigenvalues, a reversal of (36) can occur but the conditions for this are complicated and vary with the parametric specification of $d\mu(\lambda)$, the order of the matrix, and the eigenvalues themselves. To illustrate a reversal and report some conditions which make it possible, we consider the particular

case of order $r \leq 4$ and gamma distributed mixtures, i.e., $d\mu(\lambda)$ defined by (10).

To clarify what is at issue in producing a reversal it is convenient to express the left-hand side of (36) in terms of the eigenvalues of M-I and the mixing measure $d\mu(\lambda)$. To this end let $\{\delta_i(t)\}$, $1 \leq i \leq r$ be the eigenvalue of $P_\mu(0,t)$. We then have

$$\text{trace } P_\mu^k(0,\Delta) - \text{trace } P_\mu(0,k\Delta)$$

$$= \sum_{i=1}^{4} \delta_i^{k}(\Delta) - \sum_{i=1}^{4} \delta_i(k\Delta) \quad . \tag{37}$$

Since $P_\mu(0,t)$ is defined by (34), its eigenvalues are mixtures of the eigenvalues $e^{\lambda t z_i}$ of $e^{\lambda t(M-I)}$. In particular

$$\delta_i(t) = \int_0^\infty e^{\lambda t z_i} \, d\mu(\lambda) \quad , \quad i = 1,\ldots,r \tag{38}$$

where $\{z_i\}$, $1 \leq i \leq r$, are the eigenvalues of M-I. With $d\mu(\lambda)$ specified by (10) we obtain for (38)

$$\delta_i(t) = \left(\frac{\beta}{\beta - tz_i}\right)^\alpha \quad , \quad 1 \leq i \leq r \tag{39}$$
$$\alpha > 0 \quad \beta > 0 \quad .$$

Thus equation (37) reduces to

$$\sum_{i=1}^{4} \left[\left(\frac{\beta}{\beta - \Delta z_i}\right)^{k\alpha} - \left(\frac{\beta}{\beta - k\Delta z_i}\right)^\alpha \right] . \tag{40}$$

Now, a matrix of order $r = 4$ can have only a single pair of complex conjugate eigenvalues; thus, necessary conditions on α, β, Δ, and $\{z_i\}$ which determine the sign of expression (40) can be presented as conditions which determine the sign of

$$\left(\frac{\beta}{\beta - \Delta z_i}\right)^{k\alpha} - \left(\frac{\beta}{\beta - k\Delta z_i}\right)^{\alpha} \tag{41}$$

for the complex eigenvalue[11] z_i. A numerical evaluation of (41) was performed for $k = 2, 3,$ and 4; for (α, β) combinations defined by the list

$$\alpha = .5, \ 1.5, \ 2, \ 3, \ 4, \ 6$$

$$\beta = .2, \ .5, \ .75, \ 1, \ 1.5, \ 2, \ 3, \ 4, \ 6, \ 8;$$

and for $\Delta z = x + iy$ taking on values on the grid defined by

$$x = -.7 - k \ (.1), \qquad k = 0, 1, 2, \ldots, 10$$

$$y = .1 + j \ (.1), \qquad j = 0, 1, 2, \ldots, 14.$$

These calculations reveal the (α, β) pairs for which it is possible for (41) to be positive for some complex number that is an eigenvalue of a 4×4 matrix $M-I$, with M stochastic. The results of this exercise are reported in Figure 8. In particular, a necessary

condition for (41) to be positive is that (α,β) lie in the shaded region of the figure.

For a given (α,β) pair such that (41) can be positive we wish to characterize the eigenvalue region Δz = x + iy for which (41) is, in fact, positive. Recall that the restriction

$$\pi(\frac{1}{2} + \frac{1}{r}) \leq \arg (\Delta z) \leq \pi \quad , \tag{42}$$

together with its symmetric counterpart below the x-axis, is a necessary condition for Δz to be an eigenvalue of an r × r intensity matrix (Singer and Spilerman 1976a, pp. 10-12). Inequality (42) with r=4 restricts the eigenvalue region to that portion of the complex plane below the 45 degree line in Figure 9. The bent line which defines the lower boundary derives from two further considerations. First, the segment of this line closest to the origin arises from the simulation calculations discussed above. The steeper rise, to the left of the bend, is due to an asymptotic argument described in Singer and Spilerman (1977, p. 30). The slope of this line is -tan(π/12) which means that if M, and hence M-I, has complex conjugate eigenvalues with a sufficiently small imaginary part, (41), and thereby (40), will be negative for all Δ > 0. In particular, this will be the case if
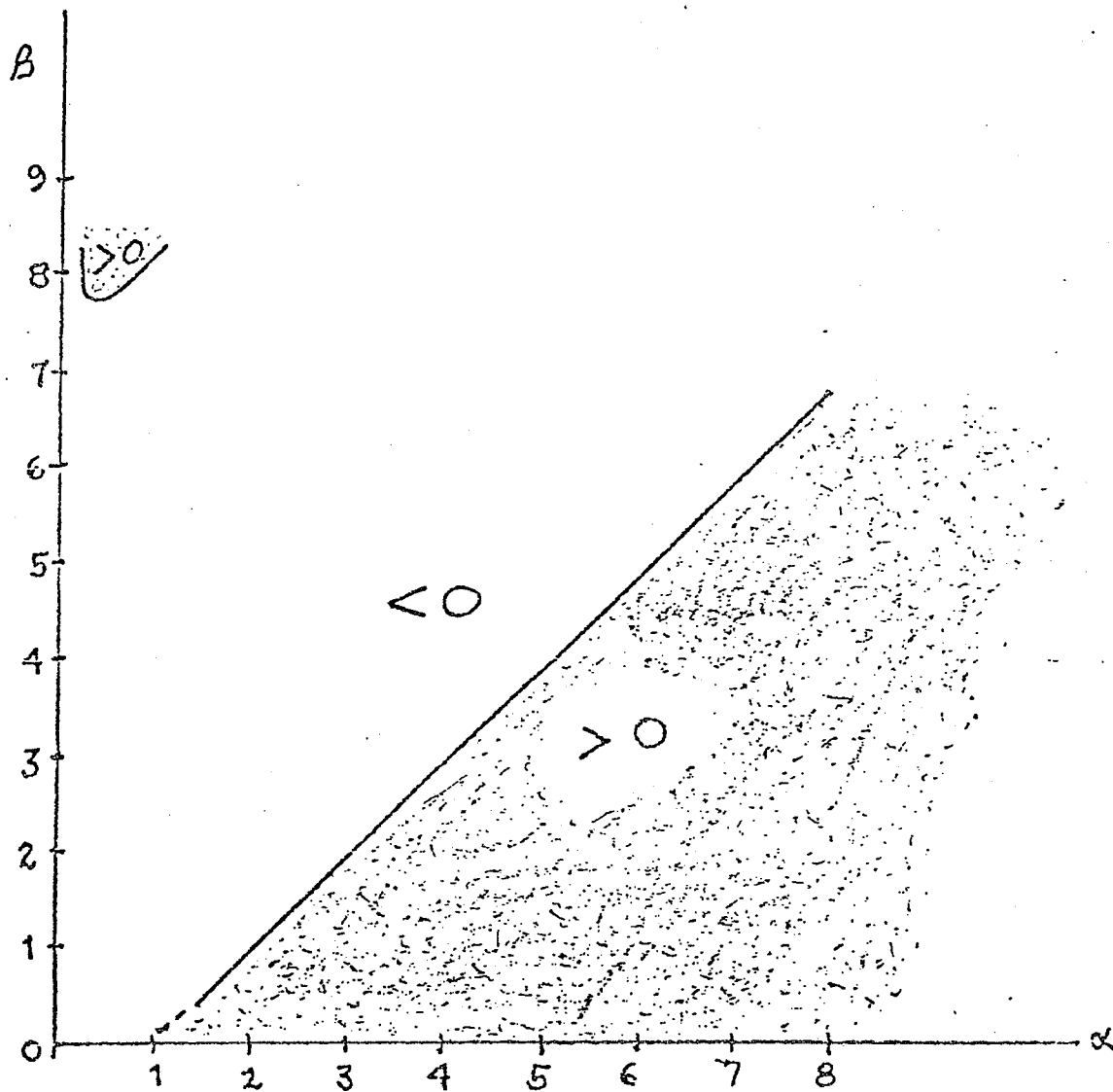
Figure 8.   Region for which Equation (41) can be Positive[a].

---

[a]Shaded area shows (α, β) values for which equation (41) can be positive. Calculations are for complex eigenvalues of 4 x 4 stochastic matrices where μ is a gamma mixture with parameters (α, β) and k = 2,3,4.

$$\tan^{-1}\left|\frac{y_j}{x_j}\right| \leq \frac{\pi}{12} \tag{43}$$

where $z_j = x_j \pm iy_j$ are eigenvalues of M-I.

With this characterization of the region where (41) will be positive, it is a straightforward matter to exhibit concrete gamma mixtures of Markov chains for which, for example,

$$\text{trace } P_\mu^2(0,\Delta) - \text{trace } P_\mu(0,2\Delta) > 0$$

all $\Delta \geq 1$.

<u>Example 3.</u> Suppose M is a 3 × 3 circulant matrix

$$M = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_3 & a_1 & a_2 \\ a_2 & a_3 & a_1 \end{pmatrix}$$

with $a_1 = .3332$, $a_2 = -.6620$, $a_3 = .0448$. Then M-I has eigenvalues $z_1 = 0$, $z_2 = -1 + .5i$, $z_3 = -1 - .5i$. The numerical calculations described above reveal that for $\Delta = 1$ and $\mu(\lambda)$ a gamma measure with $\alpha = 6$, $\beta = 1$,

$$\text{trace } P_\mu^2(0,1) - \text{trace } P_\mu(0,2) = 3.8 \times 10^{-4} > 0.$$

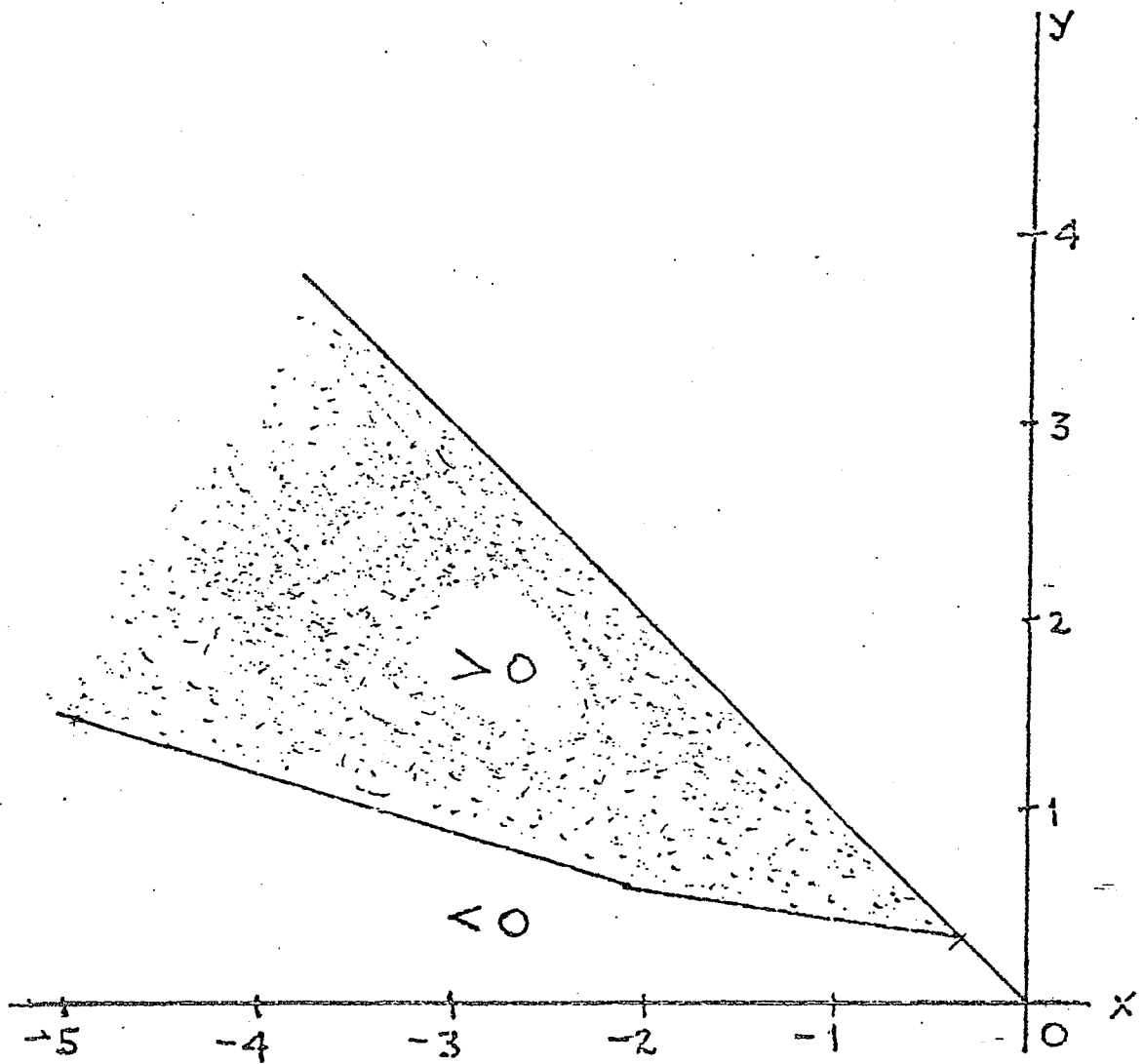Figure 9. Region for which Equation (41) is Positive.[a]

---

[a]Shaded area shows eigenvalue region for which equation (41) will be positive. Calculations are for μ a gamma mixture with $\alpha = 6$, $\beta = 1$.

Additional calculations and reference to (43) indicate that for
$\Delta > .86$

$$\text{trace } P_\mu^{\ 2}(0,\Delta) - \text{trace } P_\mu(0,2\Delta) > 0$$

while for spacings $\Delta < .86$ the reverse of this inequality holds.

V. A SEMI-MARKOV MODEL AND CLUSTERING

In each of the preceding mixture models if
$P_\mu(0,\Delta)$ contains only distinct, real, positive eigenvalues,
clustering on the main diagonal is guaranteed; i.e., a reversal
of (21) cannot occur. Moreover, this result is rather general,
holding for matrices $P_\mu(0,\Delta)$ of any order, all values of k in (21),
all spacing intervals Δ, and any heterogeneity formulation based on
a distribution of rates of movement.[12] Overprediction of the main
diagonal elements by a Markov model can occur in a mixture formulation
only if matrix $P_\mu(0,\Delta)$ has negative real or complex eigenvalues.
Further, in a continuous-time mixture overprediction can occur only
if there are complex eigenvalues.

For practical purposes it would be useful to know what
proportion of transition matrices constructed from observations on
a social process contain only distinct, real, positive eigenvalues.
While we have not addressed this question in a systematic fashion,
our experience from examining published matrices suggests that upward
of 90% of arrays constructed from sociological data have distinct,
real, positive eigenvalues. Our results associating clustering with
heterogeneity are, therefore, strong statements with respect to data
analysis.

We hasten to add that there is no obvious substantive reason as to why phenomena such as job turnover and geographic migration, for example, should be associated with the general class of matrices having distinct, real, positive eigenvalues. However, a proper subclass of these matrices for which there is some meaningful interpretation is the class of totally positive matrices. Such matrices, in addition to having distinct, positive, real eigenvalues, satisfy the determinant inequalities

$$\det \begin{pmatrix} p_{i_1 j_1} & \cdots & p_{i_1 j_k} \\ \vdots & & \vdots \\ p_{i_k j_1} & \cdots & p_{i_k j_k} \end{pmatrix} > 0$$

for any ordered row indices $i_1 < i_2 < \ldots < i_k$ and column indices $j_1 < j_2 < \ldots < j_k$, $k = 2, 3, \ldots, 4$. These matrices arise from birth and death processes--i.e., movement at a single transition can only take place to nearest neighbor states, such as in an ordered hierarchy where states are identified with prestige rankings. Thus M in a birth and death process must satisfy $m_{ij} = 0$ if $|i-j| > 1$. In this circumstance,

$$P_\mu(0,\Delta) = \int_\Lambda e^{\Delta\lambda(M-I)} d\mu(\lambda)$$

is totally positive for every mixing distribution μ and for every
Δ > 0.

A specific setting for which this class of models forms
a natural initial baseline is the study of careers in hierarchical
organizations such as a civil service bureaucracy or an industrial
firm. In this connection see S. Stewman's (1975) description of
promotion patterns in the Michigan State Police.

We underscore the importance of our results about eigenvalue
conditions by pointing out that when $P(0,\Delta)$, having distinct, real,
positive eigenvalues, arises from a semi-Markov process, the
reverse of equation (21) can occur. Further, if (21) holds,
this rules out a semi-Markov process with gamma distributed
waiting times, i.e., equation (18), and constitutes evidence for
heterogeneity.[13] If the reverse of (21) is found, this rules out
any mixture model--mover-stayer or continuous mass distribution--
and constitutes evidence for a semi-Markov process (among the
classes of models we have considered in this paper).

To provide the reader with the flavor of the argument regarding
over-prediction of the main diagonal elements of the semi-Markov
process (18) by a Markov model, we sketch the proof of

$$\text{trace } P^k(0,\Delta) > \text{trace } P(0,k\Delta) \qquad (44)$$

(i.e., a reversal of [21]) for the case $k = 2$ with $P(0,\Delta)$ a matrix

having distinct, real, positive eigenvalues. Define

$$u(t, v_i) = \sum_{n=0}^{\infty} \left[F_n(t) - F_{n+1}(t) \right] v_i^n$$

where $F_n(t) - F_{n+1}(t)$ is given by (17) and v_i is an eigenvalue of matrix M. $u(t, v_i)$ is then an eigenvalue of $P(0,t)$. Thus,

$$\text{trace } P^2(0, \Delta) = \sum_{i=1}^{r} u^2(\Delta, v_i),$$

and

$$\text{trace } P(0, 2\Delta) = \sum_{i=1}^{r} u(2\Delta, v_i).$$

Now, for any real number $v \, \varepsilon \, (0,1)$ we have (see note 6)

$$u^2(\Delta, v) - u(2\Delta, v) = e^{-2\beta\Delta} \left[\cosh \beta\Delta\sqrt{v} + \frac{1}{\sqrt{v}} \sinh \beta\Delta\sqrt{v} \right]^2$$

$$- e^{-2\beta\Delta} \left[\cosh 2\beta\Delta\sqrt{v} + \frac{1}{\sqrt{v}} \sinh 2\beta\Delta\sqrt{v} \right]$$

which, after some algebra and use of double angle formulas, reduces to

$$u^2(\Delta,v) - u(2\Delta,v) = e^{-2\beta\Delta} \sinh^2 \beta\Delta\sqrt{v} \left[\frac{1-v}{v}\right].$$

This expression is positive for all $v \in (0,1)$, $\beta > 0$, $\Delta > 0$. Finally, observe that for the eigenvalue $v_1 = 1$, $u^2(\Delta,v_1) - u(2\Delta,v_1) = 1 - 1 = 0$. Thus

$$\text{trace } P^2(0,\Delta) - \text{trace } P(0,2\Delta) = \sum_{i=1}^{r} \left[u^2(\Delta,v_i) - u(2\Delta,v_i)\right]$$

$$= 0 + \sum_{i=2}^{r} \left[u^2(\Delta,v_i) - u(2\Delta,v_i)\right] > 0.$$

A proof of (44) may be constructed along similar lines for $k = 3$, 4 and any matrix M having distinct, real eigenvalues. Thus, we obtain rather general results; namely that for a semi-Markov process with gamma distributed waiting times, there always will be a reversal of equation (21)--for $k = 2$, 3, 4,--if $P(0,\Delta)$ has distinct, real, positive eigenvalues. With complex eigenvalues it is possible for (21) and a reversal of (21) to both occur, the particular outcome depending on k, β, and

the region of the complex plane in which the eigenvalues lay.
Determination of these conditions requires a numerical exercise
of the sort illustrated in the preceding section with the continuous
mass distribution function.

VI. CONCLUSIONS

We have sought to demonstrate in this paper that the
familiar phenomenon of clustering on the main diagonal is an
inherent consequence of the structure of certain model types.
Further, while it has not to our knowledge been discussed
previously, underprediction of the main diagonal elements of a
Markov model should be associated with other kinds of stochastic
models that are also applicable to social processes. Our results
are strongest for matrices $\hat{P}(0,\Delta)$ having distinct, real, positive
eigenvalues, which happens to be the most common situation with
respect to data from a social process.

In particular, if a matrix $\hat{P}(0,\Delta)$ satisfies this eigenvalue
condition, then if it arises from a heterogeneity formulation--
whether mover-stayer or continuous mass distribution--a reversal of
(21) cannot occur; that is "clustering on the main diagonal" is a
consequence of the model's structure. If a matrix $\hat{P}(0,\Delta)$ arises
from a semi-Markov process with gamma-distributed waiting times,
then a reversal of (21) will always occur. (Even in the case of
distinct, real, positive eigenvalues we cannot make a general
statement about all semi-Markov formulations, as could be stated
for heterogeneity formulations; it is possible that a semi-Markov
model with a different waiting time distribution function would be

consistent with (21). Lacking appropriate general theorems about semi-Markov processes, one must attempt to carry out the sort of proof outlined in the preceding section for different parametric formulations of this model type.)

Main diagonal conditions--whether clustering or underprediction in relation to a Markov process--constitute one sort of empirical regularity that can be used to advantage in the activity of model discrimination. Stated generally, by model discrimination we mean a set of procedures which would allow a researcher to choose among competing model types on the basis of a few strategic observations on an empirical process. Our findings in this paper contribute to model discrimination techniques in that, under appropriate eigenvalue conditions, we have shown that main diagonal relationships can be used to reject a heterogeneity formulation or to reject a particular semi-Markov model.

This paper does not, however, constitute a concerted effort to develop model discrimination procedures. That task is a difficult one, and different criteria may have to be developed depending on the sort of data one can collect on an empirical process. (One immediate consequence of our concern with model discrimination is therefore a heightened interest in questionnaire and data collection design, to ensure that the proper information for differentiating among competing models of an empirical process will be available.) It is our intention to address this constellation of issues in the near future.

APPENDIX I

__Theorem:__ If z is a real eigenvalue of matrix M-I, where
M is stochastic, and $\mu(\lambda)$ is an arbitrary probability measure on
$[0,\infty)$, then

$$\delta^k(\Delta) - \delta(k\Delta) < 0$$

for k = 2, 3,... and all $\Delta > 0$, where

$$\delta(\Delta) = \int_0^\infty e^{\lambda \Delta z} \, d\mu(\lambda).$$

__Proof:__ Let Λ be a random variable with distribution function
$\text{Prob}(\Lambda < c) = \int_0^c d\mu(\lambda)$. Consider the random variable Y, defined
in terms of Λ according to $Y = e^{\Delta z \Lambda}$, where $\Delta > 0$ and $z < 0$.[14]
With these random variables we can write

$$EY = E(e^{\Delta z \Lambda}) = \int_0^\infty e^{\Delta z \lambda} d\mu(\lambda) = \delta(\Delta).$$

Now introduce the convex functions $f_k(x) = x^k$, k = 2, 3,..., where
$x \geq 0$, and observe that a direct application of Jensen's inequality
(Feller 1971, pp. 153-4) yields

$$\delta^k(\Delta) = (EY)^k \leq E(Y^k) = E(e^{k\Delta z \Lambda}) = \delta(k\Delta).$$

NOTES

[1]By the symbol "^" over a matrix or over entries in a matrix we shall mean estimates made directly from observations. Entries without this symbol will refer to calculations from a mathematical model.

[2]With the discrete-time Markov chain, the waiting time to a move follows a geometric distribution. The exponential distribution is the continuous-time analog of the geometric, in the sense that the conditional probability of an event occurring in time interval $(t, t+dt)$, given that no event occurred prior to time t, is independent of the value of t.

[3]This formulation arises as follows. Assume that the duration time in a state is exponentially distributed with average waiting time equal to $1/\lambda$. An equivalent statement is that moves occur in time according to a Poisson process in which the probability of making exactly k moves during $(0,t)$ is given by $g_k(0,t) = (\lambda t)^k e^{-\lambda t}/k!$, $k = 0, 1, 2,\dots$. If M is the transition matrix followed at each move, then $P(0,t)$, the transition matrix which represents population movements between times 0 and t, can be viewed as a weighted average of the terms M^k, $k = 0, 1, 2,\dots$, in which the weights are given by $g_k(0,t)$. That is

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

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

[4] The parameter λ is not indexed by system state since we have assumed that $\lambda_i = \lambda$, $i = 1, 2, \ldots, r$.

[5] $F_n(t)$ represents the probabiltiy that the waiting time for n events to occur is less than t; that is, $F_n(t) = \sum_{k=n}^{\infty} g_k(0,t)$ where $g_k(0,t) =$ probability of exactly k events in the interval $(0,t)$. Thus $F_n(t) - F_{n+1}(t) = \sum_{n}^{\infty} g_k(0,t) - \sum_{n+1}^{\infty} g_k(0,t) = g_n(0,t)$, and equation 15 can be viewed as a sum of probabilities for making $k = 0, 1, 2, \ldots$ transition events, each weighted by M^k, the transition matrix followed when making k moves.

[6] The power series is initially constructed in scalar argument,

$$u(t,\delta) = \sum \left[F_n(t) - F_{n+1}(t) \right] \delta^n$$

$$= e^{-\beta t} \left[\sum_{n=0}^{\infty} \frac{(\beta t \sqrt{\delta})^{2n}}{(2n)!} + \frac{1}{\sqrt{\delta}} \sum_{n=0}^{\infty} \frac{(\beta t \sqrt{\delta})^{2n+1}}{(2n+1)!} \right]$$

$$= e^{-\beta t} \left[\cosh \beta t \sqrt{\delta} + \frac{1}{\sqrt{\delta}} \sinh \beta t \sqrt{\delta} \right].$$

This scalar valued analytic function can be extended to an analytic function of matrix argument by any of the procedures in Singer and

Spilerman (1976).

[7]The matrices $P(0,k\Delta)$, $k = 1,2,\ldots$ in (19)-(21) refer to transition matrices constructed from an arbitrary discrete-state stochastic process. In particular, the notions of clustering defined in (19)-(21) make sense for general discrete-state processes. The matrices defined by (3) and (18) are specializations to mixtures of Markov chains and to a special parametric family of semi-Markov processes, respectively.

[8]Because each section of the shaded region is convex and contains the eigenvalues 1 and v (or 1 and $\overline{v}$, the complex conjugate of v), $s \cdot 1 + (1-s)v$ [and $s \cdot 1 + (1-s)\overline{v}$] are also in the shaded region for any $s\epsilon(0,1)$. Thus the eigenvalue condition on M for a reversal of (21)--i.e., presence in the shaded region of Figure 2--pertains as well to the eigenvalues of $P_\mu(0,\Delta)$.

[9]Because the shaded regions now depend on s they pertain to the eigenvalues of M, not $P_\mu(0,\Delta)$.

[10]Stated in terms of matrix M we need require only distinct, real eigenvalues. This is because each eigenvalue z_i of M-I is transformed into an eigenvalue $\delta_i(t)$ of $P_\mu(0,t)$ via the relation $\delta_i(t) = \int_0^\infty e^{\lambda t z_i} d\mu(\lambda)$. Since e^x takes on positive values for any real argument x, and since the integration may be viewed as a weighted average of different exponentials, $\delta_i(t)$ will be positive for real, negative roots $z_i + 1$ of matrix M.

[11] A matrix of order r = 4 with a pair of complex conjugate eigenvalues will have two real roots. By Jensen's inequality (Appendix I) the real roots add negative terms to the sum (40). A necessary condition for (40) to be positive is that (41) be positive. A sufficient condition is that the sum of positive terms in (40)--which correspond to complex conjugate roots--exceed the sum contributed to (40) by the real valued roots.

[12] We remind the reader that we have considered only forms of heterogeneity that can be expressed in terms of individual differences in the rate of movement.

[13] Our conclusions with respect to heterogeneity formulations are general, not dependent on the parametric specification of heterogeneity (other than the fact that mixing is on the rate of movement term). Our results with regard to semi-Markov processes are specific to the formulation we have analyzed, namely, gamma distributed waiting times.

[14] If matrix M is stochastic then every real eigenvalue z of M-I will be negative or zero.

REFERENCES

Bartholomew, D. J.

1973   Stochastic Models for Social Processes, 2nd Edition, New York: John Wiley and Sons.

Bellman, Richard

1970   Introduction to Matrix Analysis, 2nd Edition, New York: McGraw-Hill.

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

1955   The Industrial Mobility of Labor as a Probability Process, Cornell Studies of Industrial and Labor Relations, vol. 6. Ithaca, New York: Cornell University.

Coleman, James S.

1961   The Adolescent Society, New York: The Free Press.

1964a   Introduction to Mathematical Sociology, New York: The Free Press.

1964b   Models of Change and Response Uncertainty, Englewood Cliffs, New Jersey: Prentice-Hall.

Feller, William

1968   Introduction to Probability Theory and its Applications vol. I, New York: John Wiley, 3rd Edition.

1971   Introduction to Probability Theory and its Applications vol. II, New York: John Wiley and Sons.

Ginsberg, R.

1971   "Semi-Markov processess and mobility," Journal of Mathematical Sociology, 1:233-263.

Goodman, L.A.

    1961   "Statistical methods for the mover-stayer model,"

                Journal of the American Statistical Association

                56:841-868.

Hodge, R.W.

    1966   "Occupational mobility as a probability process,"

                Demography 3:19-34.

Karpelewitsch, F.I.

    1951   "On the characteristic roots of a matrix with

                non-negative elements," Isvestija, Serie Mathematique,

                15:361-383.

Kohlberg, L.

    1973   "Continuities in childhood and adult moral development

                revisited," in Paul Baltes and K. W. Schaie (eds.)

                Life-Span Developmental Psychology. New York: Academic

                Press.

Land, K.

    1969   "Duration of Residence and Prospective Migration,"

                Demography 6:133-140.

Lieberson, S. and G. Fugiutt

    1967   "Negro-white occupational differences in the absence

                of discrimination," American Journal of Sociology

                73:188-200.

Loevinger, J.

    1966   "The meaning and measurement of ego development,"

                American Psychologist 21:195-206.

Mayer, T.S.

    1972   "Models of intra-generational mobility," pp. 308-57 in _Sociological Theories in Progress_, edited by Joseph Berger, Morris Zelditch, Jr., and Bo Anderson. New York: Houghton-Mifflin.

McCall, J. J.

    1973   _Income Mobility, Racial Discrimination and Economic Growth_. Lexington, Mass.: Heath.

McFarland, D. 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): 712-722.

Piaget, J.

    1954   - _The Construction of Reality in the Child_, New York: Basic Books.

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.

Runnenberg, J. T.

1962   "On Elfving's problem of imbedding a time-discrete Markov
        chain in a continuous time one for finitely many states,"
        Proceedings Koninklijke Nederlandse Akademie van
        Wetenschappen, ser. A, Mathematical Sciences 65(5):
        536-541.

Singer, B. and Spilerman, S.

1974   "Social mobility models for heterogeneous populations,"
        in Herbert Costner (ed.) Sociological Methodology
        1973-74. San Francisco: Jossey-Bass, pp. 356-401.

1976a   "The representation of social processes by Markov
        models," American Journal of Sociology 82, July:1-54.

1976b   "Some methodological issues in the analysis of
        longitudinal surveys," Annals of Economic and Social
        Measurement, vol. 5, no. 4, pp. 447-474.

1977   "Trace inequalities for mixtures of Markov chains,"
        Advances in Applied Probability, December, (in press).

1978   "Mathematical representations of development theories,"
        in Paul Baltes and John Nesselroade (eds.) Manual of
        Longitudinal Research Methodology, New York: Academic
        Press (in press).

Spilerman, S.

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

1972b   "The analysis of mobility processes by the introduction
        of independent variables into a Markov chain,"
        American Sociological Review 37(June):277-294.

Stewman, S.

1975   "Two Markov models of open system occupational mobility:

underlying conceptualization and empirical tests,"

American Sociological Review 40:298-321.

Swift, Joanne

1972   The Localization of Characteristic Roots of Stochastic

Matrices, unpublished M.Sc. Thesis, Department of

Mathematics, McGill University, Montreal.

Tarver, J. O. and Gurley, W. R.

1965   "A stochastic analysis of geographic mobility and

population projections of the census divisions in the

United States," Demography 2:134-139.