The Representation of Social Processes by Markov Models[1]

Burton Singer
Columbia University

Seymour Spilerman
University of Wisconsin—Madison

In this paper we consider a class of issues which are central to modeling social phenomena by continuous-time Markov structures. In particular, we discuss (*a*) *embeddability,* or how to determine whether observations on an empirical process could have arisen via the evolution of a continuous-time Markov structure; and (*b*) *identification,* or what to do if the observations are consistent with more than one continuous-time Markov structure. With respect to the latter topic, we discuss how to select the specific structure from the list of alternatives which should be associated with the empirical process. We point out that the issues of *embeddability* and *identification* are especially pertinent to modeling empirical processes when one has available only fragmentary data and when the observations contain "noise" or other sources of error. These characteristics, of course, describe the typical work situation of sociologists. Finally, we note the type of situation in which a continuous-time model is the proper structure to employ and indicate that issues analogous to the ones we describe here apply to modeling social processes with discrete-time structures.

1. INTRODUCTION

Markov models provide a convenient framework for analyzing the structural mechanisms which underlie social change and for extrapolating shifts in the state distribution of a population. For reviews of applications and discussions of some pertinent mathematical issues, the reader is referred to Boudon (1973), Bartholomew (1973), and Singer and Spilerman (1974). Although most commonly employed in the study of mobility, Markov models have been applied to diverse substantative topics; they

[1] The work reported here was supported by grants NSF-GP-31505X and NSF-GS-38574 at Columbia University and by funds granted to the Institute for Research on Poverty of the University of Wisconsin by the Office of Economic Opportunity, pursuant to the Economic Opportunity Act of 1964. We also acknowledge computation funds from National Institute of Child Health and Human Development (NICHD) grant 1-PO1-HDO5876. Earl Kinmonth aided with the computations. Comments by Ken Land and the assistance of Harrison White and Hal Winsborough are gratefully acknowledged.

have been used, for instance, to study the influence of group norms on conformity (Cohen 1963), to measure distance in social networks (Beshers and Laumann 1967), and to analyze recidivism among delinquent juveniles (Wolfgang, Figlio, and Sellin 1972). The attractiveness of this mathematical formulation derives from the fact that it permits a researcher to focus upon the dynamic properties of a social process and ascertain the long-range consequences of particular institutional arrangements. An instructive example of this sort of inquiry is provided by Lieberson and Fuguitt (1967).

Several technical issues relating to the sensitive use of Markov models have begun only recently to receive an amount of attention that is commensurate with their importance. One matter concerns the phenomenon of population heterogeneity. In the initial attempts at modeling mobility processes by time-stationary Markov chains, socially heterogeneous populations were treated as though a single transition rule governed the movements of all individuals. Special kinds of discrepancies observed between the empirical data and predictions from these one-type Markov models were suggestive about the form of stochastic process which might provide a more realistic theoretical framework in which to view mobility (Blumen, Kogan, and McCarthy 1955). The main attempts at modifying the Markov model so it would provide a suitable description of movements by a heterogeneous population have involved viewing the population as consisting of a mixture of independent Markov processes, one for each individual or each distinct social group (McFarland 1970; Ginsberg 1971; Spilerman *1972a, 1972b;* Singer and Spilerman 1974).

A second issue concerns strategies for testing whether empirical observations are, in fact, compatible with an assumed class of models, such as general finite-state Markov processes, or a subset of them, such as birth and death processes. An example of this sort of inquiry is presented in Singer and Spilerman (1974, pp. 360–63), where an observed two-step matrix[2] $\hat{P}(2)$—representing occupational change between grandfathers' and respondents' generations—was examined for compatibility with a stationary discrete-time Markov structure. Conceptually, the problem is to decide whether the empirically determined matrix could have arisen via the evolution of the postulated process. Stated technically, it is to ascertain whether there exists a one-step transition matrix $P(1)$—which would be identified with grandfather-to-father transitions or, equivalently, with father-to-son transitions—such that $\hat{P}(2) = [P(1)]^2$. Where the answer is negative, it would be incorrect to predict future population dis-

[2] The symbol "∧" over a stochastic matrix or over an element in a matrix will mean that the quantity should be thought of as estimated directly from data. Matrices without this symbol should be viewed as obtained from a mathematical model.

tributions from a Markov model,[3] such as by raising the observed matrix to integer powers.

The same kind of issue must be faced with respect to compatibility of observed data with other model structures, and it is this fundamental sort of inquiry that we address in the present paper. We will concentrate on conditions for compatibility with a finite-state continuous-time Markov process, a mathematical structure which holds special interest for two reasons. First, although discrete-time formulations have been used in most applications of Markov models, the empirical processes under consideration commonly evolve continuously, and the appropriate technical apparatus would be a continuous-time model (Coleman 1964*a*, p. 129). The reason for the greater popularity of the discrete-time structure stems from its simpler mathematical nature, not from considerations of verisimilitude. Second, continuous-time Markov processes provide the underlying mathematical framework for James S. Coleman's (1964*a*) influential volume in mathematical sociology as well as for a number of more recent publications (Coleman 1968; Mayer 1972; Bartholomew 1973). Because of a neglect of the representation considerations that are discussed here, serious deficiencies exist with the estimation procedures used in these works. An additional reason for concentrating on compatibility with a continuous-time Markov framework is that the conceptual issues which must be addressed with more complicated mathematical structures, such as models of heterogeneous processes, already reveal themselves in this comparatively simple setting.

Representation becomes an issue when we have available only fragmentary data on population movements. Unfortunately, in the study of social phenomena, the common situation is to have very incomplete information about the evolution of an empirical process; frequently, observations have been taken at only two time points, $t = 0$ and $t = t_1$, yielding a single transition matrix[4] $\hat{P}(0,t_1)$. What we wish to determine, then, are the conditions which this observed matrix must satisfy in order for it to be viewed as an outcome of a continuous-time Markov model. For matrices satisfying the requisite criteria, we wish further to recover the parameters of the particular Markov structure that underlies the empirical process. These issues can be posed most effectively in terms of two sequential considerations—embeddability tests and the identification problem. In practice, a single calculation is usually informative on both matters.

Embeddability.—This issue refers to whether an observed transition

[3] In most applications of Markov models, tests of this sort are not made. Hodge (1966) provides an exception.

[4] Where it is understood that the initial observation is at $t = 0$, we will simplify our notation and write $\hat{P}(t_1)$, or even $\hat{P}$, in place of $\hat{P}(0,t_1)$.

matrix $\hat{P}(t_1)$ could have arisen via the evolution of a stationary continuous-time Markov process. It is well known that certain stochastic matrices are not compatible with such a formulation; this is the case, for instance, if $\hat{P}(t_1)$ has an element $\hat{p}_{ij}(t_1) = 0$, but some power of the matrix, say $\hat{P}(t_1)^n$, has a nonzero entry in the same position, that is, $\hat{p}_{ij}^{(n)}(t_1) \neq 0$ (Chung 1967, p. 126). Also, according to Coleman (1964*a*, p. 179; 1964*b*, p. 4), a stochastic matrix in which some main diagonal element is less than another entry in its column could not have been generated by a continuous-time Markov process. We shall show that Coleman's claim is in error.[5] For the present discussion, however, the essential point is that, while it is recognized that certain transition matrices cannot be represented by this mathematical structure, there is confusion over the full scope of the requirements for embeddability. Our first task, then, is to devise tests for ascertaining compatibility of an empirically determined matrix with a continuous-time Markov formulation.

Identification.—If the embeddability tests are passed, then we are guaranteed that $\hat{P}(t_1)$ could have been generated by at least one continuous-time Markov process. The identification problem refers to the possibility that the matrix could have originated from the operation of more than one Markov process. Consequently, our second task is to delineate the conditions under which the solution for the parameters of the Markov model will be unique. Also, for instances in which these conditions are not satisfied, we will require procedures for recovering the several Markov structures that could have produced the observed matrix and identifying the particular model from this list which should be associated with the data.

Sampling error and data-collection design.—Overlying the questions of embeddability and multiple solutions is the issue of sampling error. In most applications, an empirically determined transition matrix $\hat{P}(t_1)$ will have been constructed from a population sample. Repeated surveys of the population would produce somewhat different transition arrays, so we would be well advised to investigate the *sensitivity* of our estimate of the underlying Markov structure to sampling error. In particular, with respect to the matter of embeddability, we might wish to inquire whether a nonembeddable $\hat{P}(t_1)$ is "within error distance" of some embeddable matrix $\tilde{P}$. If it is, we could choose to carry out an analysis in which Markov methods are employed using the adjusted (embeddable) matrix $\tilde{P}$ instead of the observed array $\hat{P}(t_1)$.

[5] We wish to emphasize at the outset that our extensive criticism of estimation procedures used in Coleman's work in no way detracts from the utility of the mathematical formulations he employs or from his strategies in translating sociological theory into mathematical statements. Indeed, his work has been a source of inspiration to both of us.

The question of data error leads to more intriguing considerations with respect to the phenomenon of multiple solutions. Even if $\hat{P}(t_1)$ is compatible with a unique Markov process, it is possible that a slightly modified matrix $\tilde{P}$—within error distance of the original array—will produce a very different Markov structure from the one that has been identified. As a result, if the data derive from a population *sample,* then because of sampling variability we may have recovered the wrong Markov structure for the *population-level* process! We therefore discuss strategies for treating an empirically determined matrix as data containing considerable "noise" and identifying from it the particular Markov model to be associated with the substantive process.

Finally, there are crucial considerations regarding when to survey a population in order to facilitate model identification and parameter estimation. It is widely known, for instance, that if the interval between successive observations is very large (with respect to the rate of evolution of the empirical process), $\hat{P}(t_1)$ will resemble the equilibrium matrix, and the parameters of the continuous-time Markov model which produced the observed array cannot be recovered (Coleman 1968, p. 472). Yet the issue of data-collection design is considerably more complex than this simple remark conveys and involves decisions concerning the number of observations to be taken, the spacing between them, and interactions between these considerations.

2. MATHEMATICAL PRELIMINARIES AND EXAMPLES

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

$$\frac{dP(t)}{dt} = QP(t), \qquad P(0) = I, \tag{2.1}$$

where $P(t)$ and Q are $r \times r$ matrices. It is well known (Coleman 1964*a*, pp. 127–30; Chung 1967, pp. 251–57) that if Q has the structure

$$q_{ij} \geqslant 0 \text{ for } i \neq j, \quad q_{ii} \leqslant 0, \sum_{j=1}^{r} q_{ij} = 0 \text{ for } i = 1, \ldots, r, \tag{2.2}$$

then the functions $P(t)$, $t > 0$, which are solutions of (2.1) comprise the transition matrices of continuous-time stationary Markov chains. A typical element, $p_{ij}(t)$, of $P(t)$ has the interpretation:

$p_{ij}(t)$ = probability that an individual starting in state i at time 0 will be in state j at time t.

The Q-arrays, which are known as "intensity matrices," provide structural information about the population:

i) $q_{ij}/-q_{ii}$ = probability that an individual in state i will move to state j, given the occurrence of a transition.

ii) $1/-q_{ii}$ = expected length of time for an individual in state i to remain in that state.

We will denote the class of intensity matrices (arrays of the form [2.2]) by the symbol $\underline{\underline{Q}}$.

Solutions of (2.1) are given by the exponential formula

$$P(t) = e^{Qt}, \qquad t > 0, \tag{2.3}$$

where the matrix exponential e^A (A being an arbitrary $r \times r$ matrix) is defined by

$$e^A = \sum_{k=0}^{\infty} A^k/k!$$

The problem of finding simple test criteria on the entries of an observed stochastic matrix $\hat{P}(t_1)$, $t_1 < \infty$, which will guarantee that it can be written in the form (2.3) with $Q \in \underline{\underline{Q}}$, was first posed by Elfving (1937). It has come to be known as the *embedding problem* for continuous-time Markov chains.

An obvious description of the subclass $\underline{\underline{Z}}$ of stochastic matrices that are embeddable is given by

$$\underline{\underline{Z}} = \{P \text{ such that } \log P \in \underline{\underline{Q}}\}.$$

Attempts to develop practical test criteria or computer programs to determine membership in $\underline{\underline{Z}}$ are reported in Coleman (1964*a*, pp. 177–82), Mayer (1972, pp. 327–28), and Zahl (1955, p. 97). However, all these investigations suffer from a confusion about the full scope of the embedding problem, as well as from using an incomplete description of the logarithm function of matrix argument. This situation has resulted in a number of erroneous statements about the conditions under which an empirically determined matrix $\hat{P}(t_1)$ is, or is not, compatible with a continuous-time Markov process.

Example 1. Coleman (1964*a*, p. 179) has asserted that "the most obvious incompatibility is one in which for some state i, $n_{ii}/n_{i.}$ is less than some $n_{ji}/n_{j.}$ for some state j."[6] This statement suggests that a Markov

[6] n_{ij} = number of persons starting in state i at a reference time $t = 0$ who are in state j at a later time $t = 1$; $n_{i.} = \sum_{j=1}^{r} n_{ij}$. In our notation, $n_{ji}/n_{j.} = \hat{p}_{ji}$. Actually,

structure would not be a suitable model for a large class of mobility matrices (e.g., Prais 1955, table 1; Coleman 1964*a,* table 14.8) or, indeed, for any array in which some off-diagonal element exceeds the main diagonal entry in its column. That this assertion is incorrect can be seen from the matrix

$$\hat{P} = \begin{bmatrix} .260 & .169 & .248 & .323 \\ .327 & .275 & .146 & .252 \\ .269 & .346 & .232 & .153 \\ .162 & .285 & .305 & .248 \end{bmatrix} \qquad (2.4)$$

In every column there is a violation of Coleman's necessary criterion, yet this matrix can be represented as e^Q with

$$Q = \begin{bmatrix} -1.700 & 0.034 & 0.025 & 1.641 \\ 1.573 & -1.657 & 0.059 & 0.025 \\ 0.051 & 1.785 & -1.853 & 0.017 \\ 0.017 & 0.085 & 1.649 & -1.751 \end{bmatrix}$$

Example 2. Elsewhere, Coleman (1973, p. 21) has written, "It is not the case that any discrete-time Markov chain can be generated by an appropriate continuous-time process. Heuristically, those discrete-time chains that cannot be generated by a continuous-time process are those in which the equilibrium distribution is approached through a damped wave, rather than approached asymptotically."[7] Coleman's statement characterizing nonembeddable matrices is incorrect, as the following computations illustrate.

By exponentiating the intensity matrix Q from example 1 with $t = 1$, $P(1.00) = e^{1.00Q}$, the transition array (2.4) is reproduced. At time $t =$ 1.41,

$$P(1.41) = e^{1.41Q} = \begin{bmatrix} .231 & .233 & .261 & .275 \\ .284 & .244 & .201 & .271 \\ .285 & .296 & .206 & .213 \\ .224 & .299 & .257 & .220 \end{bmatrix},$$

and, at time $t = 2.24$,

$$P(2.24) = e^{2.24Q} = \begin{bmatrix} .248 & .271 & .239 & .242 \\ .252 & .259 & .235 & .254 \\ .265 & .262 & .223 & .250 \\ .261 & .275 & .226 & .238 \end{bmatrix}.$$

Coleman wrote $n_{ij}/n_{j.}$ in place of $n_{ji}/n_{j.}$. This is obviously in error, and elsewhere (Coleman 1964*b*, p. 4) he makes clear his intention.

[7] From the context, we interpret the word "asymptotically" to mean monotone, rather than oscillatory, convergence.

Note that each main diagonal entry $p_{ii}(t)$, observed over the three matrices, has the property $p_{ii}(1.00) > p_{ii}(1.41) < p_{ii}(2.24)$. This means that $p_{ii}(t)$ approaches an equilibrium value as $t \rightarrow \infty$ through damped oscillations and not asymptotically. Yet, because of the manner by which the sequence of P-matrices was constructed, they depict the evolution of a continuous-time Markov process.

Example 3. In attempting to represent an observed matrix $\hat{P}(t)$ in the form (2.3), Zahl (1955, p. 97) states that "the estimate of Q is taken to be

$$\frac{1}{t} \log \hat{P}(t) = \frac{1}{t} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}[\hat{P}(t) - I]^k}{k} \tag{2.5}$$

provided the series converges." Coleman (1968, p. 472) makes essentially the same claim. Yet, although convergence of (2.5) does provide a representation of $\log \hat{P}$, it *does not* guarantee that[8] $\log \hat{P} \in \underline{\underline{Q}}$. In particular, consider

$$\hat{P} = \begin{bmatrix} .600 & .330 & .070 \\ .302 & .560 & .138 \\ .380 & .040 & .580 \end{bmatrix}$$

The series representation for $\log \hat{P}$ converges to

$$\log \hat{P} = \begin{bmatrix} -.692 & .639 & .053 \\ .496 & -.733 & .237 \\ .707 & -.144 & -.563 \end{bmatrix}$$

which is not in $\underline{\underline{Q}}$, since $(\log \hat{P})_{32} = -.144 < 0$.

Example 4. In possibly the most serious of the misunderstandings, Coleman (1968, p. 472) has asserted that, "when [(2.5)] does not converge, this means that the data are not compatible with the assumptions of a continuous-time Markov process, or that the moves of the panel are too widely spaced." Mayer (1972, p. 328) makes essentially the same point: "The failure of [(2.5)] to converge for all transition matrices $P(t)$ reflects the fact that not all such matrices can arise from a continuous-time stationary Markov chain." These statements are in error. Equation (2.5) may fail to converge for matrices P, not resembling the equilibrium matrix, which nonetheless can be represented in the form e^Q with $Q \in \underline{\underline{Q}}$.

[8] In different contexts, we speak of checking whether $Q = (1/t) \log \hat{P} \in \underline{\underline{Q}}$ or whether $\log \hat{P} \in \underline{\underline{Q}}$. Because multiplication of a matrix by a real-valued quantity does not alter its character with respect to satisfying conditions (2.2), the two tests are equivalent.

Consider

$$\hat{P} = \begin{bmatrix} .3654 & .3762 & .2584 \\ .3292 & .3567 & .3141 \\ .4040 & .3188 & .2772 \end{bmatrix}$$

The series representation (2.5) converges if and only if $|\lambda_i - 1| < 1$ for all eigenvalues λ_i of $\hat{P}$. The matrix above has eigenvalues $\lambda_1 = 1$, $\lambda_2 = .053i$, $\lambda_3 = -.053i$. Thus $|\lambda_2 - 1| = |\lambda_3 - 1| > 1$ and (2.5) diverges. Nevertheless, $\hat{P} = e^Q$ for

$$Q = \begin{bmatrix} -1.805 & 1.718 & 0.087 \\ 0.044 & -1.784 & 1.740 \\ 2.262 & 0.017 & -2.279 \end{bmatrix},$$

and it is therefore embeddable.

The preceding examples highlight the confusions that exist concerning which transition matrices can be represented as outcomes of the evolution of a continuous-time Markov process. In particular, we have indicated that the standard recipe for estimating Q (the matrix of structural parameters which govern population movements)—via the power-series representation (2.5)—is highly deficient. The series does not provide a complete description of the logarithm of a matrix; as a result, it fails to converge for transition arrays that are compatible with a Markov formulation.

In fact, the inadequacy of equation (2.5) as a procedure for estimating the intensity matrix Q is even more fundamental than the illustrations above suggest. While the power series will converge to at most one version of log $\hat{P} \in \underline{\underline{Q}}$, *the equation* $\hat{P} = e^Q$ *can have multiple solutions* $Q \in \underline{\underline{Q}}$. This is a matter of great importance in sociological investigations, because the conventional strategy in using Markov models for theory construction emphasizes decomposing the q_{ij} elements of Q among theoretically postulated effect parameters (Coleman 1964*a*, chap. 5; 1964*b*, chap. 2; McDill and Coleman 1963). Clearly, one can hardly begin this task without ensuring that the *correct* Q has been recovered for the substantive process under study. Before considering the issues of multiple solutions and model identification, we address the conceptually prior question of embeddability of $\hat{P}$; that is, we seek to determine which transition matrices are compatible with a continuous-time Markov process.

3. EMBEDDABILITY OF $\hat{P}$

In the case of 2×2 matrices, a complete and practical solution to the question of embeddability was given by D. G. Kendall (see Kingman 1962, p. 15), who proved that

$$\hat{P} = \begin{bmatrix} \hat{p}_{11} & \hat{p}_{12} \\ \hat{p}_{21} & \hat{p}_{22} \end{bmatrix}, \qquad \hat{p}_{ij} \geqslant 0, \qquad \sum_j \hat{p}_{ij} = 1$$

is in $\underline{\underline{Z}}$ (or, equivalently, can be represented as e^Q, $Q \epsilon \underline{\underline{Q}}$) if and only if $\hat{p}_{11} + \hat{p}_{22} > 1$.

A solution to the embedding problem for stochastic matrices with an arbitrary finite number of states was provided by Kingman (1962). In particular, he proved that $\hat{P}$ can be written in the form e^Q, with $Q \epsilon \underline{\underline{Q}}$, if and only if (i) det $\hat{P} > 0$, *and* (ii) for every positive integer n, there is a stochastic matrix P_n such that $(P_n)^n = \hat{P}$. Unfortunately, condition (ii) does not lead to *practical* test procedures to be applied to $\hat{P}$, and Kingman pointed out the impossibility of obtaining general tests as simple as those in the 2×2 case for matrices of order greater than or equal to 3. A further mathematically interesting solution to the embedding problem has recently been given by Johansen (1973, p. 180); however, in keeping with Kingman's remarks, it too is not useful for practical computation.

This impasse has led to the development of a considerable number of easily applicable *necessary* conditions for an $r \times r$ stochastic matrix $\hat{P}$ to be in $\underline{\underline{Z}}$. These conditions are presented in Section 3.1, with illustrations of their use. A common feature of the tests is that they can be used only to assert that a particular matrix is *not* compatible with a Markov model. An empirically determined matrix which passes all the tests in Section 3.1 must still be subject to an examination based on *sufficiency* conditions for embeddability, if one hopes to pass on to the stage of model identification. With the results of Kingman (1962) and Johansen (1973) at hand, our only recourse is to develop simple computational procedures for obtaining all branches of log $\hat{P}$ compatible with the criteria in Section 3.1 and test these versions of the logarithm for membership in $\underline{\underline{Q}}$. This seemingly straightforward program leads to some surprisingly subtle phenomena, which are delineated in Section 3.2. General practical recommendations for testing an observed matrix $\hat{P}$ for embeddability are outlined in Section 3.3.

3.1 Necessary Conditions

Test criteria which empirically determined matrices must satisfy to be compatible with a family of mathematical models can be viewed usefully as devices for isolating matrices generated by these models from the class of all stochastic arrays. The necessary conditions listed below are the simplest such tests for distinguishing the subclass of matrices generated by continuous-time Markov models.

Condition 1.—(Austin and Ornstein; see Chung [1967, p. 126] for details.) If $\hat{p}_{ij}(t_1) = 0$, then $\hat{p}_{ij}^{(n)}(t_1) = 0$ for every integer n. If $\hat{p}_{ij}(t_1) \neq 0$, then $\hat{p}_{ij}^{(n)}(t_1) \neq 0$ for any integer n.

Condition 2.—(Kingman 1962) det $\hat{P} > 0$.

Condition 3.—(Elfving 1937) No eigenvalue λ_i of $\hat{P}$ can satisfy $|\lambda_i| = 1$ other than $\lambda_i = 1$. In addition, any negative eigenvalue must have even (algebraic) multiplicity.

Condition 4.—(Runnenberg 1962) All eigenvalues of $\hat{P}$ must lie inside a heart-shaped region H_r in the complex plane whose boundary is the curve $x(v) + iy(v)$, where

$$\left.\begin{aligned} x(v) &= \left[\exp\left(-v + v\cos\frac{2\pi}{r}\right)\right]\cos\left(v\sin\frac{2\pi}{r}\right) \\ y(v) &= \left[\exp\left(-v + v\cos\frac{2\pi}{r}\right)\right]\sin\left(v\sin\frac{2\pi}{r}\right) \end{aligned}\right\} \tag{3.1}$$

together with its symmetric image with respect to the real axis. In this parametrized formulation, $r =$ order of the matrix $\hat{P}$, and v is restricted by $0 \leqslant v \leqslant \pi/\sin(2\pi/r)$. The regions H_3, H_6, and H_{12} are displayed in figures 1, 2, and 3. The larger cone-shaped zones K_3, K_6, and K_{12} show the

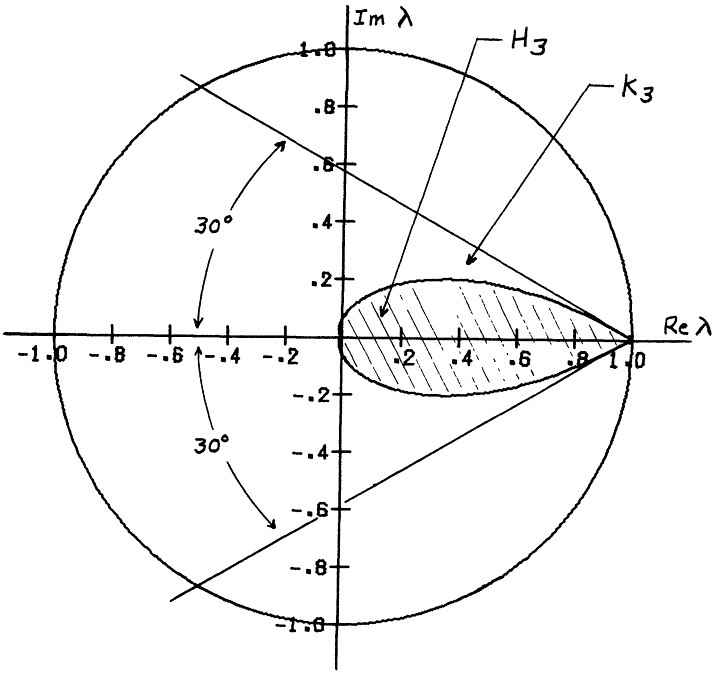

FIG. 1.—Eigenvalue regions for 3 × 3 stochastic matrices (K_3) and for the subset of them which is in $\underline{\underline{Z}}$ (H_3). A necessary condition for $\hat{P}$ to be embeddable is that all its eigenvalues lie in the shaded zone.

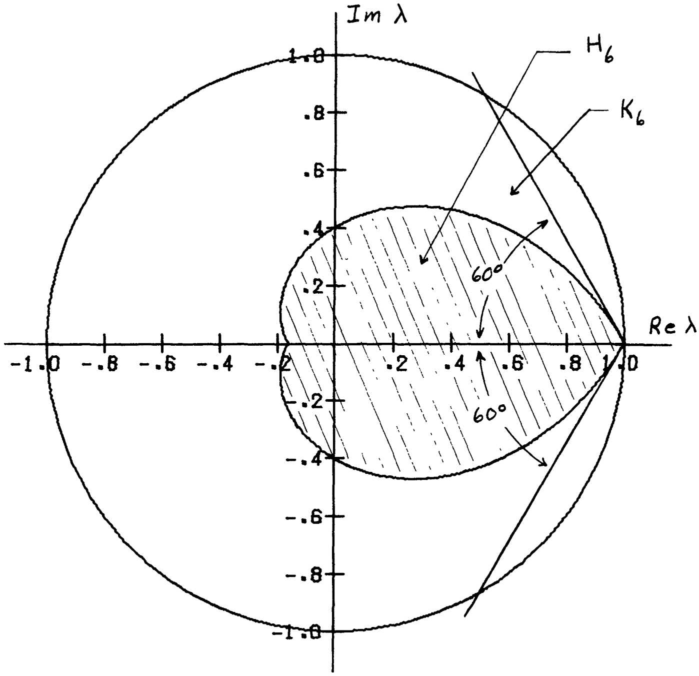

FIG. 2.—Eigenvalue regions for 6 × 6 stochastic matrices (K_6) and for the subset of them which is in $\underline{\underline{Z}}$ (H_6). A necessary condition for $\hat{P}$ to be embeddable is that all its eigenvalues lie in the shaded zone.

bounds on the eigenvalues of arbitrary 3 × 3, 6 × 6, and 12 × 12 stochastic matrices.

The cone-shaped zones arise from the requirement that the eigenvalues of an arbitrary stochastic matrix $\hat{P}$ must satisfy[9]

$$\left(\frac{1}{2}+\frac{1}{r}\right)\pi \leqslant \arg(\lambda - 1) \leqslant \left(\frac{3}{2}-\frac{1}{r}\right)\pi \qquad (3.2)$$

(where the argument[10] is in radians), together with the condition $|\lambda| \leqslant 1$. The additional limitation to the heart-shaped set H_r contained in K_r arises from the continuous-time Markov assumptions. This restriction can also be described by saying that the eigenvalues of $\hat{P}$ must satisfy (3.2) *and*

$$\left(\frac{1}{2}+\frac{1}{r}\right)\pi \leqslant \arg(\log \lambda) \leqslant \left(\frac{3}{2}-\frac{1}{r}\right)\pi \qquad (3.3)$$

9 These inequalities were established by Karpelewitsch (1951); they represent a considerable strengthening of the well-known restriction that all eigenvalues of a stochastic matrix must lie inside the unit circle.

10 For a complex number $\mu = a + bi$, we define $\arg(\mu) = tan^{-1}(b/a)$.

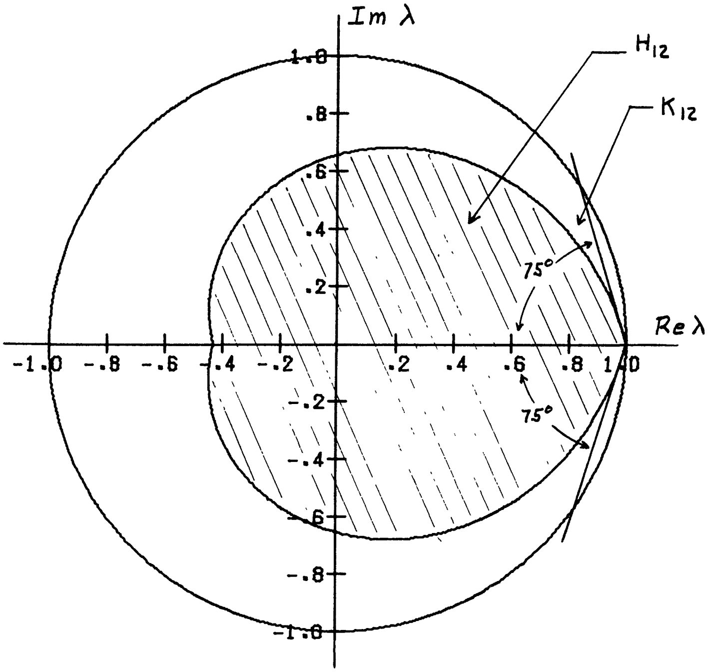

FIG. 3.—Eigenvalue regions for 12 × 12 stochastic matrices (K_{12}) and for the subset of them which is in $\underline{\underline{Z}}$ (H_{12}). A necessary condition for $\hat{P}$ to be embeddable is that all its eigenvalues lie in the shaded zone.

Examination of H_3 explains why failure of the series (2.5) to converge in example 4 did not rule out compatibility of $\hat{P}$ with a continuous-time Markov process. The region of convergence of (2.5) is $|\lambda_i - 1| < 1$, i.e., the unit circle centered at (1, 0), and the complex eigenvalues of the matrix in that example, while exterior to this region, are inside H_3.

Example 5. Suppose you observe the matrix

$$\hat{P} = \begin{bmatrix} .15 & .35 & .50 \\ .37 & .45 & .18 \\ .20 & .60 & .20 \end{bmatrix}.$$

Since det $\hat{P} = .05 > 0$, condition 2 is satisfied. However, $\hat{P}$ has eigenvalues $\lambda_1 = 1$, $\lambda_2 = -.1 + .2i$, $\lambda_3 = -.1 - .2i$ which, by (3.2), lie inside the cone K_3, but they are outside the heart-shaped zone H_3. Thus $\hat{P}$ cannot be represented as e^Q for any $Q \in \underline{\underline{Q}}$; in other words, it is not compatible with a continuous-time Markov model.

Example 6. Consider the matrix

$$\hat{P} = \begin{bmatrix} .20 & .40 & .40 \\ .35 & .20 & .45 \\ .40 & .40 & .20 \end{bmatrix}.$$

Here, det $\hat{P} = .04 > 0$, satisfying condition 2. The eigenvalues of $\hat{P}$ are $\lambda_1 = 1$, $\lambda_2 = \lambda_3 = -.2$, so that condition 3 applies and is satisfied. Nevertheless, λ_2 and λ_3 are outside the zone H_3. Thus $\hat{P}$ is not compatible with a continuous-time Markov model.

Example 7. Recall the matrix of example 3,

$$\hat{P} = \begin{bmatrix} .600 & .330 & .070 \\ .302 & .560 & .138 \\ .380 & .040 & .580 \end{bmatrix}.$$

This matrix satisfies the *necessary* conditions 1–4; however, it is still not representable as e^Q for any $Q \in \underline{\underline{Q}}$. This assertion is based on an examination of all versions of log $\hat{P}$ which are candidates for membership in $\underline{\underline{Q}}$. An understanding of these tests requires a complete description of log $\hat{P}$. This is the subject of the next section.

3.2 *The Matrix Equation* $\hat{P} = e^Q$

We require a definition of a function of matrix argument[11] which is sufficiently general to include analytic functions such as e^x and log x. It is useful to motivate the definition by an important property of polynomial functions $g(x)$. In particular, if

$$g(x) = a_0 + a_1x + a_2x^2 + \ldots a_nx^n$$

and A is an arbitrary square matrix, a natural definition of $g(A)$ is given by

$$g(A) = a_0I + a_1A + a_2A^2 + \ldots a_nA^n.$$

In addition, A can always be reduced to Jordan form J by some nonsingular matrix H, that is,

$$A = HJH^{-1}. \tag{3.4}$$

Finally, it is readily verified that

$$g(A) = Hg(J)H^{-1}. \tag{3.5}$$

Every Jordan matrix J has the following block structure:

[11] For a lucid and detailed mathematical exposition, the reader should consult Gantmacher (1960, chap. 5).

$$J = \begin{bmatrix} J_1 & & & & \mathbf{0} \\ & J_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ \mathbf{0} & & & \cdot & J_k \end{bmatrix}, \qquad J_i = \begin{bmatrix} \lambda_i & 1 & & & \mathbf{0} \\ & \lambda_i & 1 & & \\ & & \cdot & \cdot & \\ & & & \cdot & \cdot \\ & & & \cdot & 1 \\ \mathbf{0} & & & & \lambda_i \end{bmatrix}, \tag{3.6}$$

where λ_i is the ith eigenvalue of matrix A and occurs in J_i with multiplicity v_i, the order of J_i. (The λ_i appearing in different blocks J_i are not necessarily distinct.) Also, $\sum_{1}^{k} v_i = r$, the order of A.

The expression (3.5) will be useful in a wider context[12] than just polynomials, provided that we have a representation of $g(J)$ for arbitrary Jordan matrices J, which generalizes to analytic functions[13] $f(J)$. Then our program will be to define $f(A)$ according to (3.5), with g replaced by f, adding appropriate conventions for multiple-valued functions. For a polynomial function $g(x)$, we introduce its Taylor series expansion about $x = \lambda_i$ and write

$$g(J) = \begin{bmatrix} g(J_1) & & & & \mathbf{0} \\ & g(J_2) & & & \\ & & \cdot & & \\ & & & \cdot & \\ \mathbf{0} & & & \cdot & g(J_k) \end{bmatrix}, \tag{3.7}$$

where[14]

$$g(J_i) = g(\lambda_i)I + g'(\lambda_i)(J_i - \lambda_i I) + \dots + \frac{g^{(v_i-1)}(\lambda_i)}{(v_i-1)!}(J_i - \lambda_i I)^{v_i-1}$$

$$= \begin{bmatrix} g(\lambda_i) & g'(\lambda_i) & \frac{g''(\lambda_i)}{2!} & \cdots & \frac{g^{(v_i-1)}(\lambda_i)}{(v_i-1)!} \\ 0 & g(\lambda_i) & g'(\lambda_i) & \cdots & \frac{g^{(v_i-2)}(\lambda_i)}{(v_i-2)!} \\ \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & g(\lambda_i) \end{bmatrix}.$$

[12] The remainder of this section is more difficult mathematically and can be skipped at a first reading. Continue with Section 3.2a, "Distinct Eigenvalues."

[13] A function is said to be analytic at x if it has a derivative in a neighborhood containing the point.

[14] Although the Taylor series expansion has an infinite number of terms, $(J_i - \lambda_i I)^n = 0$ for all values of $n \geqslant v_i$.

Formula (3.7) has meaning for any function f which is analytic in a neighborhood of the eigenvalue λ_i. Thus, if f is *single valued* and analytic in a region of the complex plane containing the eigenvalues of A (e.g., $f(x) = e^x$), we define

$$f(A) = Hf(J)H^{-1} \tag{3.8}$$

where $f(J)$ is specified by (3.7) with g replaced by f.

If f is *multiple valued* (e.g., $f(x) = \sqrt{x}$, or $f(x) = \log x$), we define a *branch* of $f(A)$ corresponding to the similarity transformation H by

$$f_a(A) = Hf_a(J)H^{-1} \tag{3.9}$$

where

$$f_a(J) = \begin{bmatrix} f_{a_1}(J_1) & & & & \mathbf{0} \\ & f_{a_2}(J_2) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ \mathbf{0} & & & & f_{a_k}(J_k) \end{bmatrix}$$

and $f_{a_i}(x)$ is *any* single-valued branch of $f(x)$. Notice that different branches of $f(x)$ may be used with distinct Jordan blocks J_i and that each combination of $(f_{a_1}, f_{a_2}, \ldots, f_{a_k})$ will generate a different version of $f(A)$. Furthermore, the value of $f(A)$ may depend on the choice of H, a point to which we will have cause to return.[15] This definition was introduced by Cipolla (1932)—see also Rinehart (1955)—and represents the necessary level of generality for a discussion of solutions of the matrix equation $e^Q = \hat{P}$ ($\hat{P}$ is identified with A in the preceding discussion). We now specialize to the case where the eigenvalues of A are distinct. The repeated eigenvalue condition, while crucial to a complete understanding of embeddability, is more involved mathematically and will be considered separately.

3.2a. Distinct Eigenvalues

In this case, the Jordan matrix J reduces to a diagonal matrix D, in which the nonzero entries are the eigenvalues of A. Analogous to (3.4), we have

$$A = HDH^{-1} \tag{3.10}$$

[15] This matter is discussed in proposition 2.

where

$$D = \begin{bmatrix} \lambda_1 & & & & \mathbf{0} \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ \mathbf{0} & & & & \lambda_r \end{bmatrix} .$$

Also, the *eigenvector* corresponding to λ_i is contained in the ith column of H. The foregoing discussion regarding analytic functions of matrix argument carries over in its entirety, with the functions of Jordan blocks $f(J_i)$ replaced by functions of eigenvalues $f(\lambda_i)$. In particular, when f is multiple valued, (3.9) reduces to

$$f_a(A) = Hf_a(D)H^{-1} \tag{3.11}$$

where

$$f_a(D) = \begin{bmatrix} f_{a_1}(\lambda_1) & & & & \mathbf{0} \\ & f_{a_2}(\lambda_2) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ \mathbf{0} & & & & f_{a_r}(\lambda_r) \end{bmatrix} .$$

A different version of $f(A)$ is obtained from each combination of branches of $(f_{a_1}, f_{a_2}, \ldots, f_{a_r})$.

This discussion is relevant in the following way to the determination of embeddability. Ascertaining compatibility of an observed matrix $\hat{P}$ with a continuous-time Markov process requires investigating whether there exists an array $Q \in \underline{\underline{Q}}$ such that $\hat{P} = e^Q$. Lacking readily computable *sufficiency* conditions for general $r \times r$ stochastic matrices, our strategy must be to compute $\log \hat{P}$ and examine it for membership in $\underline{\underline{Q}}$. Now, the logarithm function is multiple valued,[16]

$$\log_k z = \log|z| + i(\theta + 2\pi k), \qquad k = 0, \pm 1, \pm 2, \ldots, \tag{3.12}$$

where z is an arbitrary complex number, $z = a + bi$; $|z| = \sqrt{a^2 + b^2}$; and $\theta = \tan^{-1} b/a$. Each value of k generates a different version of $\log z$, called a ***branch*** of the logarithm. In general, an infinity of branches will exist.

[16] The simplest way to appreciate the multiple-valued character of the logarithm is to begin with the definition: $x = \log y$ if x is a solution of the equation $e^X = y$ for a given y. Suppose x is such a solution. Then, for any integer k, $e^{x+2\pi ki} = e^x e^{2\pi ki} = e^x = y$ (since $e^{2\pi ki} = \cos 2\pi k + i \sin 2\pi k = 1$). Therefore, $\log y$ takes on the values x, $x \pm 2\pi i$, $x \pm 4\pi i$, etc.

From equations (3.11) and (3.12), we have

$$\log_K \hat{P} = H \log_K D\, H^{-1} \tag{3.13}$$

where[17]

$$\log_K D = \begin{bmatrix} \log_{k_1}\lambda_1 & & & & \mathbf{0} \\ & \log_{k_2}\lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ \mathbf{0} & & & & \log_{k_r}\lambda_r \end{bmatrix}.$$

Every combination of values of $(\log_{k_1}\lambda_1, \log_{k_2}\lambda_2, \ldots, \log_{k_r}\lambda_r)$ in (3.13) will yield a version of $\log \hat{P}$, so to determine embeddability one must check whether at least one branch is in $\underline{\underline{Q}}$. An important implication of *necessary condition 4* in Section 3.1 is that only *finitely* many branches of $\log \hat{P}$ need be checked for membership in $\underline{\underline{Q}}$. It is this feature which makes the computational tests described in detail in Section 3.3 feasible. Furthermore, in many applications, the number of branches which must be computed is quite small.

Sylvester's formula.—If A is an $r \times r$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$, and if f is single valued in a neighborhood of each of the eigenvalues, then equation (3.8) is equivalent to (Sylvester 1883)

$$f(A) = \sum_{i=1}^{r} f(\lambda_i) \prod_{j \neq i} \frac{(A - \lambda_j I)}{(\lambda_i - \lambda_j)}. \tag{3.14}$$

In addition, if f is multiple valued, then (3.14), with $f(\lambda_i)$ replaced by $f_{a_i}(\lambda_i)$, defines a version of $f(A)$ for each combination of branches of $(f_{a_1}, f_{a_2}, \ldots, f_{a_r})$; in other words, this equation is equivalent to (3.11).

Example 8. Consider the matrix

$$\hat{P} = \begin{bmatrix} .3654 & .3762 & .2584 \\ .3292 & .3567 & .3141 \\ .4040 & .3188 & .2772 \end{bmatrix},$$

which also appeared in example 4, and identify $\hat{P}$ with A in the discussion above. In order to solve the equation $\hat{P} = e^Q$, observe that $\hat{P}$ has distinct eigenvalues $\lambda_1 = 1$, $\lambda_2 = .053i$, $\lambda_3 = -.053i$. Setting $f(x) = \log x$ in Sylvester's formula, we obtain

[17] All logarithms are to base e. The subscript k denotes the branch number of the logarithm of a scalar quantity and takes on the values $k = 0, \pm 1, \pm 2, \ldots$. The subscript K denotes a version of the logarithm of a matrix and specifies a combination of branches of the logarithm of the eigenvalues.

$$\log \hat{P} = \log(\lambda_1)\frac{(\hat{P}-\lambda_2 I)(\hat{P}-\lambda_3 I)}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)} + \log(\lambda_2)\frac{(\hat{P}-\lambda_1 I)(\hat{P}-\lambda_3 I)}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)}$$

$$+ \log(\lambda_3)\frac{(\hat{P}-\lambda_1 I)(\hat{P}-\lambda_2 I)}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)}$$

$$= \begin{bmatrix} -1.805 & 1.718 & 0.087 \\ 0.044 & -1.784 & 1.740 \\ 2.262 & 0.017 & -2.279 \end{bmatrix},$$

which satisfies criterion (2.2) for membership in $\underline{\underline{Q}}$. In this calculation we used the principal branches of log λ_2 and log λ_3; namely, $\log \lambda_2 = \log(.053) + i\pi/2$ and $\log \lambda_3 = \log(.053) - i\pi/2$. Any other branch, e.g., $\log \lambda_2 = \log(.053) + i(\frac{\pi}{2} + 2\pi k)$ for an integer $k \neq 0$, would yield a version of $\log \hat{P}$ which is not in $\underline{\underline{Q}}$. For a similar reason, we use the principal branch of $\log(\lambda_1) = \log(1) = 0$.

An important feature of this example and of Sylvester's formula in general is that the logarithm of a matrix is well defined even when the power series (2.5) diverges, as it does here. For matrices with distinct eigenvalues λ_i satisfying $|\lambda_i - 1| < 1$, the series (2.5) is equivalent to the principal branch solution of (3.14)—$k = 0$ in equation (3.12). However, Sylvester's formula is more general, in that it will generate all branches of $\log \hat{P}$ as k is varied.[18] Furthermore, it leads to an evaluation of analytic functions of matrix argument as *finite* polynomials in the original matrix $\hat{P}$. The transcendental nature of $f(\hat{P})$ is incorporated entirely in the coefficients of this polynomial and involves only functions of eigenvalues. In particular, by rearranging terms, Sylvester's formula for general $r \times r$ matrices (3.14) can be written in the form

$$f(\hat{P}) = c_0 I + c_1 \hat{P} + c_2 \hat{P}^2 + \ldots + c_{r-1}\hat{P}^{r-1}$$

in which the c_i's are scalar functions of the eigenvalues of $\hat{P}$.

3.2b. Repeated Eigenvalues

When $\hat{P}$ has one or more sets of equal eigenvalues, the computations to determine embeddability can be considerably more involved. Unfortunately, even though the occurrence of repeated eigenvalues in an observed matrix $\hat{P}$ would be a rare event, we will have reason to consider adjustment

[18] Sylvester's formula has been effectively employed by Johansen (1974) in a recent study of the embedding problem. His results, however, are less general than the ones presented here, because Sylvester's formula also provides a less than complete description of the logarithm of a matrix. This point is elaborated in proposition 2.

strategies which make use of this condition. We therefore outline the main issues and analytic procedures at this point; some elaborations are found in Section 4.2 and in the Appendix.

It is useful to categorize matrices with repeated eigenvalues according to whether or not their elementary divisors[19] are distinct. Elementary divisors are said to be distinct if each eigenvalue λ_i appears in exactly one Jordan block $J_i(\lambda_i)$ in equation (3.6). They are said to be *nondistinct* if a repeated eigenvalue λ_i can serve as the diagonal element in more than one Jordan block. The importance of this distinction derives from the fact that the eigenvalues in a block are constrained to be on the same branch of a multiple-valued function—that is, they must have the same value of k in expression (3.12). The presence of nondistinct elementary divisors therefore permits different branches of log λ_i to be present simultaneously in log J, via the presence of λ_i in more than one Jordan block. It is this condition which creates exceptional difficulties in the calculation of log $\hat{P}$. The following propositions and examples outline the computations for the two multiple-eigenvalue cases:

Proposition 1.—If A is an $r \times r$ matrix with m different eigenvalues $\lambda_1, \ldots, \lambda_m$ having multiplicities $r_1, \ldots, r_m$ and elementary divisors $(\lambda - \lambda_1)^{r_1}, \ldots, (\lambda - \lambda_m)^{r_m}$—i.e., distinct elementary divisors—and if f is a functon that is single valued and analytic in a neighborhood of each of the eigenvalues, then $f(A)$ may be computed via (3.8) or by using the equivalent but computationally often simpler formula[20]

$$f(A) = \sum_{k=1}^{m} \sum_{s=1}^{r_k} c_{ks} \Big[f(\lambda_k) + (A - \lambda_k I) f'(\lambda_k) + \ldots + \frac{(A - \lambda_k I)^{s-1}}{(s-1)!} f^{(s-1)}(\lambda_k) \Big] \prod_{j \neq k} (A - \lambda_j I)^{r_j} (A - \lambda_k I)^{r_k - s} \tag{3.15}$$

where the terms c_{ks} are the coefficients in the partial-fraction expression

$$\frac{1}{\prod_{k=1}^{m} (\lambda - \lambda_k)^{r_k}} = \sum_{k=1}^{m} \sum_{s=1}^{r_k} \frac{c_{ks}}{(\lambda - \lambda_k)^s}.$$

When f is multiple valued, the various branches $f_a(A)$ may be found by computing (3.15) for all combinations of branches of $(f_{a_1}, f_{a_2}, \ldots, f_{a_r})$—that is, $f_{a_i}{}^{(v)}(\lambda_i)$ replaces $f^{(v)}(\lambda_i)$, $v = 0, 1, \ldots, s-1$, in equation

[19] On computing the elementary divisors of a matrix, consult Gantmacher (1960, pp. 139–45).

[20] When $r_k = 1$ for $k = 1, \ldots, m$, then (3.15) reduces to Sylvester's formula (3.14).

(3.15). With respect to determining embeddability of $\hat{P}$, the number of versions of $\log \hat{P} = f(\hat{P})$ which need to be examined is discussed in Section 3.3.

Example 9. Consider the matrix

$$\hat{P} = \begin{bmatrix} .1600 & .5300 & .3100 \\ .0527 & .4900 & .4577 \\ .1100 & .1400 & .7500 \end{bmatrix},$$

and identify $\hat{P}$ with A in the preceding discussion. The eigenvalues of $\hat{P}$ are $\lambda_1 = 1$ and $\lambda_2 = .2$, with multiplicities $r_1 = 1$ and $r_2 = 2$, respectively. First note that both eigenvalues lie in H_r (fig. 1). It is also the case that the elementary divisors of $\hat{P}$ are distinct; they are $(\lambda - 1)$ and $(\lambda - .2)^2$. We may therefore solve for all solutions to $\hat{P} = e^Q$ by using equation (3.15) and setting $f(\lambda_i) = \log \lambda_i$:

$$\log \hat{P} = c_{11}(\log 1)(\hat{P} - \lambda_2 I)^2 + c_{21}(\log \lambda_2)(\hat{P} - \lambda_1 I)(\hat{P} - \lambda_2 I) + c_{22}\left[(\log \lambda_2)I + \frac{1}{\lambda_2}(\hat{P} - \lambda_2 I)\right](\hat{P} - \lambda_1 I). \quad (3.16)$$

Selecting the principal branch of the logarithm for each eigenvalue, the first term in expression (3.16) disappears, since $\log 1 = 0$. From the remaining terms, we obtain

$$Q = \log \hat{P} = \begin{bmatrix} -2.046 & 1.993 & 0.053 \\ 0.024 & -0.818 & 0.794 \\ 0.315 & 0.043 & -0.358 \end{bmatrix}.$$

As in the previous example, we could have chosen some other branch of the logarithm function, $\log .2 \pm 2\pi k i$, for an integer $k \neq 0$. However, (3.16) would then produce matrices with complex entries, and these have no meaning in the context of Markov models (i.e., they are not in $\underline{\underline{Q}}$).

Proposition 2.[21]—All solutions of the equation $e^Q = A$ are called branches of the logarithm function of A, and they are given by (Gantmacher 1960, pp. 239–41)

$$Q = \log A = HB \log J\, B^{-1}H^{-1} \quad (3.17)$$

where

i) H is any nonsingular matrix which reduces A to Jordan form, $A = HJH^{-1}$.

ii) B is an arbitrary nonsingular matrix that commutes with J; that is, $BJ - JB = 0$.

[21] The remainder of this section is more difficult mathematically and can be skipped at a first reading. Continue with Section 3.3.

iii)

$$\log J = \begin{bmatrix} \log J_1 & & & & \mathbf{0} \\ & \log J_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ \mathbf{0} & & & & \log J_k \end{bmatrix},$$

where

$$\log J_j = \begin{bmatrix} \log \lambda_j & \frac{-1}{\lambda_j} & \frac{1}{\lambda_j^2} & \cdots & \frac{(-1)^{v_j-1}}{\lambda_j^{v_j-1}(v_j-1)!} \\ 0 & \log \lambda_j & \frac{-1}{\lambda_j} & \cdots & \frac{(-1)^{v_j-2}}{\lambda_j^{v_j-2}(v_j-2)!} \\ \cdot & 0 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \log \lambda_j \end{bmatrix},$$

$\log \lambda_j = \log |\lambda_j| + i(\arg \lambda_j + 2\pi k)$, k is an integer, and $v_j =$ multiplicity of λ_j in the elementary divisor $(\lambda - \lambda_j)^{v_j}$.

If the elementary divisors of A are *distinct,* B may be replaced by the identity in (3.17), and log A is independent of the choice of H. It is this property which permits the simpler representations (3.14) and (3.16). When the elementary divisors of A are *nondistinct,* computation of all versions of log A requires a knowledge of the matrices B which satisfy $BJ - JB = 0$. These matrices contain a finite number of parameters, each of which can be an arbitrary complex number. Every product HB represents a similarity transformation which reduces A to Jordan form and, at the same time, generates a distinct version of log A. This leads to uncountably many versions of log A; and there may, in fact, be a continuum of such matrices, all or part of which is in $\underline{\underline{Q}}$. It is precisely these matrices with nondistinct elementary divisors which prevent the development of simple general solutions to the embedding problem. In any other situation, a researcher need only compute polynomials in $\hat{P}$ to evaluate log $\hat{P}$, and test a finite number of branches of the logarithm for membership in $\underline{\underline{Q}}$.

Example 10. Consider the matrix

$$\hat{P} = \frac{1}{3}\begin{bmatrix} 1+2X & 1-X & 1-X \\ 1-X & 1+2X & 1-X \\ 1-X & 1-X & 1+2X \end{bmatrix}, \qquad (3.18)$$

where $X = -e^{-2\sqrt{3}\pi}$, and identify $\hat{P}$ with A in the preceding discussion. The eigenvalues of $\hat{P}$ are $\lambda_1 = 1$, and $\lambda_2 = \lambda_3 = X$; the elementary divisors are $(\lambda - 1)$, $(\lambda - X)$, $(\lambda - X)$, which are nondistinct. Consequently, the Jordan matrix associated with $\hat{P}$ is

$$J = \begin{bmatrix} J_1 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{bmatrix}.$$

Also, a similarity transformation H, such that $\hat{P} = HJH^{-1}$, is given by

$$H = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \frac{1}{2}(-1 + i\sqrt{3}) & \frac{1}{2}(-1 - i\sqrt{3}) \\ 1 & \frac{1}{2}(-1 - i\sqrt{3}) & \frac{1}{2}(-1 + i\sqrt{3}) \end{bmatrix}. \quad (3.19)$$

In computing $\log \hat{P} = Q$, choose $\log J_1 = \log 1 = 0$; $\log J_2 = \log X = -2\sqrt{3}\pi + i\pi$; and $\log J_3 = \log X = -2\sqrt{3}\pi - i\pi$. Now, formula (3.17) with $B = I$, the identity matrix, yields

$$\log \hat{P} = 2\pi\sqrt{3} \begin{bmatrix} -\frac{2}{3} & \frac{1}{2} & \frac{1}{6} \\ \frac{1}{6} & -\frac{2}{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{6} & -\frac{2}{3} \end{bmatrix}, \quad (3.20)$$

which belongs to $\underline{\underline{Q}}$.

To manufacture other versions of $\log \hat{P}$ which are also in $\underline{\underline{Q}}$, observe that the matrices which commute with J are all of the form

$$B = \begin{bmatrix} a & 0 & 0 \\ 0 & c_{11} & c_{12} \\ 0 & c_{21} & c_{22} \end{bmatrix}, \quad (3.21)$$

where $\{c_{ij}\}$ and a are arbitrary complex numbers subject only to the restriction that B be invertible. For $\log \hat{P}$ to be in $\underline{\underline{Q}}$, we may limit consideration to matrices B with entries satisfying,

$$\left.\begin{array}{ll} \text{i)} & c_{11}c_{12} - c_{21}c_{22} = 0 \\ \text{ii)} & u \text{ is real, where } u = \dfrac{c_{11}c_{22} + c_{12}c_{21} + 2c_{21}c_{22}}{c_{11}c_{22} - c_{12}c_{21}} \\ & v \text{ is real, where } v = \dfrac{c_{11}c_{22} + c_{12}c_{21} - 2c_{21}c_{22}}{c_{11}c_{22} - c_{12}c_{21}} \\ \text{and} & \\ \text{iii)} & |u| \leqslant 2 \text{ and } |v| \leqslant 2. \end{array}\right\} \quad (3.22)$$

Conditions (i) and (ii) guarantee that log $\hat{P}$ will be real valued, while (iii) ensures that the entries will satisfy criteria (2.2). Each choice of $\{c_{ij}\}$ then yields a version of log $\hat{P}$ which is a member of $\underline{\underline{Q}}$, and they are all given by

$$\log \hat{P} = HB \log J\, B^{-1}H^{-1}$$

$$= 2\pi\sqrt{3}\begin{bmatrix} -\frac{2}{3} & \frac{1}{3}+\frac{u}{6} & \frac{1}{3}-\frac{u}{6} \\ \frac{1}{3}-\frac{v}{6} & -\frac{2}{3}-\frac{(u-v)}{12} & \frac{1}{3}+\frac{(u+v)}{12} \\ \frac{1}{3}+\frac{v}{6} & \frac{1}{3}-\frac{(u+v)}{12} & -\frac{2}{3}+\frac{(u-v)}{12} \end{bmatrix}. \tag{3.23}$$

The matrix (3.20) arises in the special case where $c_{11} = c_{22} = 1$, $c_{12} = c_{21} = 0$, and thus $u = v = 1$. The nonzero constant a in matrix B does not enter into the formula for log $\hat{P}$, because it can only multiply the first row of log J, all of whose entries are 0.

With this example at hand,[22] some remarks concerning the role of such matrices in social mobility studies are in order (these comments will be elaborated upon in Section 4). If the primary purpose of an investigation is to obtain structural information about the propensity of individuals in a population to move between particular states, then our major concern must center on the possible values of $q_{ij}/-q_{ii}$ for $i \neq j$ in branches of log $\hat{P} = Q$ which are in $\underline{\underline{Q}}$. These ratios have the interpretation "propensity to move from state i to state j when a change in state occurs." The continuum of branches of log $\hat{P}$ which are given by (3.23) represents a continuum of propensities to move between states, all compatible with the observed matrix $\hat{P}$. Focusing on mobility out of state 1 in (3.23), we see that

$$0 \leqslant \frac{q_{12}}{-q_{11}} = \frac{1/3 + u/6}{2/3} \leqslant 1$$

and

$$0 \leqslant \frac{q_{13}}{-q_{11}} = \frac{1/3 - u/6}{2/3} \leqslant 1.$$

Thus, on the basis of observations at two time points which give rise to $\hat{P}$

[22] The matrix (3.18) was introduced by Cuthbert (1973) in order to exhibit an example of a stochastic matrix compatible with a continuum of Markov models. Cuthbert's continuum arises when you choose $c_{11} = c_{22} = 1$ and $c_{12} = c_{21}$, real. Then the constraints on u and v entail that $|c_{12}| = |c_{21}| \leqslant 1/3$. This choice does not, however, lead to all of the branches of log $\hat{P}$ given in (3.23), which represents an exhaustive list in $\underline{\underline{Q}}$.

given by (3.18), we cannot even say whether individuals who start out in state 1 tend to favor state 2 or state 3 when they move. Clearly, this situation is totally uninformative about the underlying mobility mechanism, and the present example thereby serves to highlight the unusual difficulties which can arise in the case of repeated eigenvalues with non-distinct elementary divisors.

3.3 *Summary of Steps to Determine Embeddability*

3.3a. Distinct Eigenvalues

The most common eigenvalue configuration for $\hat{P}$, an empirically determined stochastic matrix, is one in which the roots are distinct. In testing $\log \hat{P}$ for membership in $\underline{\underline{Q}}$, we therefore start with this case.

Step 1.—Check that the necessary conditions (2) and (3) in Section 3.1 are satisfied.

Step 2.—Check the eigenvalues of $\hat{P}$ for membership in the heart-shaped zone H_r described in Section 3.1. If this test is passed, proceed to step 3 or 4.

Step 3.—If the eigenvalues of $\hat{P}$ are all real and positive, compute $\log \hat{P}$ using either the power series (2.5), Sylvester's formula (3.14), or the diagonalization transformation (3.11). Only the principal branch of the logarithm ($k = 0$ in equation [3.12]) will be real valued, and any of the procedures will yield the *unique* version of $\log \hat{P}$ that can possibly be in $\underline{\underline{Q}}$.

Step 4.—If $\hat{P}$ has complex eigenvalues, they must occur in conjugate pairs. For each such pair $(\lambda, \bar{\lambda})$, determine all branches of their logarithms which satisfy Runnenberg's condition,

$$\pi\left(\frac{1}{2} + \frac{1}{r}\right) \leqslant \arg(\log_k \lambda) \leqslant \pi\left(\frac{3}{2} - \frac{1}{r}\right), \qquad (3.24)$$

where $r =$ order of matrix $\hat{P}$, $\arg(\log_k \lambda) = \tan^{-1}\left(\frac{\theta + 2\pi k}{\log \rho}\right)$, and k specifies a branch of $\log_k \lambda$ according to[23]

$$\log_k \lambda = \log \rho + i(\theta + 2\pi k); \quad k = 0, \pm 1, \pm 2, \ldots, \ 0 < \theta < \pi. \qquad (3.25)$$

Now select one of the branches for each complex conjugate pair, and compute $\log \hat{P}$ via (3.11) or by using Sylvester's formula (3.14). Check the resulting matrix for membership in $\underline{\underline{Q}}$. Repeat this calculation for all

[23] If $\lambda = a + bi$, then $\rho = |\lambda| = \sqrt{a^2 + b^2}$ and $\theta = \tan^{-1}(b/a)$.

branches satisfying (3.24). Clearly, there are only a finite number of such computations to be performed, and they will yield all versions of $\log \hat{P} \in \underline{\underline{Q}}$.

In particular, if we represent a pair of complex conjugate eigenvalues $(\lambda, \bar{\lambda})$ by $(\rho e^{i\theta}, \rho e^{-i\theta})$, $0 < \rho < 1$ and $0 < \theta < \pi$, then the number of branches of $\log \lambda$ which need to be examined in testing $\log \hat{P}$ for membership in $\underline{\underline{Q}}$ is $U(r) + L(r) + 1$, where

$$\left.\begin{aligned} U(r) &= \text{integer part of} \left| \frac{(\log \rho)\tan\left[\pi\left(\frac{1}{2}+\frac{1}{r}\right)\right] - \theta}{2\pi} \right| \\ L(r) &= \text{integer part of} \left| \frac{(\log \rho)\tan\left[\pi\left(\frac{3}{2}-\frac{1}{r}\right)\right] - \theta}{2\pi} \right| \end{aligned}\right\} \quad (3.26)$$

and r is the order of the matrix.[24] Here $U(r)$ specifies the upper bound to $+k$, and $L(r)$ the lower bound to $-k$, with respect to the multiple-valued logarithm function (3.25). Since the computation of $U(r)$ and $L(r)$ is to be performed for each pair of complex conjugate eigenvalues of $\hat{P}$, the number of versions of $\log \hat{P}$ that must be examined is

$$\prod_{j=1}^{v} [U_j(r) + L_j(r) + 1],$$

where v, the upper limit, denotes $\hat{P}$'s number of complex conjugate eigenvalue pairs. The value of this product will usually be small (frequently $U_j(r) = L_j(r) = 0$ for most j's). In Section 4.2 we indicate why it is especially rare for a branch other than the principal branch of $\log \hat{P}$ to be in $\underline{\underline{Q}}$ when the matrix is of low order ($r \leqslant 3$). In larger arrays, however, one might have to examine multiple versions of $\log \hat{P}$ to determine embeddability.

3.3b. Data Noise and Repeated Eigenvalues

Because our data are commonly contaminated by the effects of sampling variability and measurement error, one cannot be certain that an empirically determined matrix $\hat{P}$ is the correct transition matrix for the population of interest. As a consequence, if the preceding calculations indicate that $\hat{P}$ is not embeddable, but the violations in $\log \hat{P}$ are not severe, a researcher should consider adjusting the observed matrix to a nearby $\tilde{P}$ which is embeddable and continuing his analysis with the modified matrix.

24 These formulas were computed from (3.24) by solving for k (in the arc tangent) at each bound.

Strategies for making such an adjustment usually operate on log $\hat{P}$, perturbing it to a matrix $Q_o \in \underset{=}{Q}$, and then estimate $\tilde{P}$, the modified array, via $e^{Q_o} = \tilde{P}$.

There are several procedures for altering log $\hat{P}$ so it will satisfy a priori chosen conditions, such as membership in $\underset{=}{Q}$. Zahl (1955, p. 98) suggests setting the offending elements (negative q_{ij}'s, $i \neq j$, in the present context) to zero and modifying the main diagonal entries so that the row sum condition, $\sum_j q_{ij} = 0$, will be satisfied. Coleman (1964*a*, pp. 178–80) uses an iterative routine which forces selected q_{ij} elements to zero in the computation of log $\hat{P}$, thereby smearing the compensatory adjustments over the remaining nonzero entries. In example 11, we illustrate the adjustment process using yet another procedure, one which minimizes the sum of squared differences between log $\hat{P}$ and $Q \in \underset{=}{Q}$. General recommendations regarding which of the techniques is advantageous in a particular problem are currently being prepared.

Example 11. Suppose you observe the matrix

$$\hat{P} = \begin{bmatrix} .600 & .330 & .070 \\ .302 & .560 & .138 \\ .380 & .040 & .580 \end{bmatrix},$$

which also appeared in example 3. This matrix has eigenvalues $\lambda_1 = 1$, $\lambda_2 = .370 + .011i$, $\lambda_3 = .370 - .011i$. Applying Runnenberg's condition in the form (3.26), we find that $U = L = 0$; hence only the principal branch of the logarithm needs to be examined for membership in $\underset{=}{Q}$. Calculating this branch,

$$\log \hat{P} = \begin{bmatrix} -.692 & .639 & .053 \\ .496 & -.733 & .237 \\ .707 & -.144 & -.563 \end{bmatrix},$$

which is not in $\underset{=}{Q}$ since $(\log \hat{P})_{32} = -.144 < 0$. This raises the question of whether a small perturbation of $\hat{P}$ would yield a logarithm in $\underset{=}{Q}$. To this end, we determine the nearest intensity matrix Q_o to log $\hat{P}$, and check whether or not e^{Q_o} represents a "small perturbation" of $\hat{P}$. The notion of "nearest" will be defined by $\min_{Q \in \underset{=}{Q}} ||\log \hat{P} - Q||$, where $||A - B|| = \sqrt{\sum_{i,j} (a_{ij} - b_{ij})^2}$.

In the present example, the minimum is obtained for

$$Q_o = \begin{bmatrix} -.692 & .639 & .053 \\ .496 & -.733 & .237 \\ .635 & 0 & -.635 \end{bmatrix}.$$

Calculation of $e^{Q_o} = \tilde{P}$ yields

$$\hat{P} + (\text{small perturbation}) = \tilde{P} = \begin{bmatrix} .598 & .334 & .068 \\ .298 & .568 & .134 \\ .349 & .104 & .547 \end{bmatrix},$$

and a case might now be made that $\hat{P}$ was not embeddable only because of sampling error or other data noise. To *conclude* that the substantive process actually is Markovian with Q_o as the governing intensity matrix, tests of the sort described in Section 5, based on three or more time points, must be passed.

Repeated eigenvalues.—From a computational point of view, the notion of repeated eigenvalues means that they agree to within a prescribed finite number of digits. If you take a large random sample of stochastic matrices, then those matrices with repeated eigenvalues tend to occur with a frequency close to zero. On the other hand, the entries in $\hat{P}$ which arise in mobility investigations are often subject to considerable sampling variability and other sources of error. Our concern, therefore, is to know whether a small perturbation in $\hat{P}$, call it $\tilde{P}$, would lead to branches of log $\tilde{P}$ radically different from those of log $\hat{P}$. These radical differences can occur in passing from a distinct to a repeated eigenvalue matrix, which in turn can be viewed as being "within error distance" of the original distinct eigenvalue matrix. This suggests that a distinct eigenvalue matrix $\hat{P}$ which is compatible with a Markov model *and* has a pair of eigenvalues within a prescribed number of digits of each other should be perturbed to a $\tilde{P}$ with repeated eigenvalues. Then the structure of the continuum should be displayed as in example 10. If the branches of log $\tilde{P}$ which are in $\underline{\underline{Q}}$ are sufficiently varied, this would lead us to report that our observations $\hat{P}$ based on data collected at two time points are uninformative about the underlying mobility mechanism.

The additional tasks to be undertaken, then, in a situation where $\hat{P}$ has eigenvalues which are close to being repeated, consist of carrying out the following procedures:

Step 5.—Adjust the observed stochastic matrix $\hat{P}$ so that it will have repeated eigenvalues.

Step 6.—Determine the structure of the continuum using the simulation strategy described in the Appendix, and check whether some part of the continuum is in $\underline{\underline{Q}}$.

The task of adjusting $\hat{P}$ so that it will have repeated eigenvalues is not difficult when the eigenvalues close together are complex conjugates. Fortunately, it is this situation which is of primary practical interest. If we represent these eigenvalues in polar form, $(\lambda, \bar{\lambda}) = (\rho e^{i\theta}, \rho e^{-i\theta})$, $0 < \theta < \pi$, where $\theta \approx 0$ or $\theta \approx \pi$, then the corresponding eigenvalues in log $\hat{P}$ are log $\rho \pm i(\theta + 2\pi k)$. We now want to alter $\hat{P}$ so that one of the approximate equalities is replaced by an *exact* equality. For a scalar t, $t \log \hat{P} = Ht \log DH^{-1}$ will have among its eigenvalues $t \log \rho \pm i(t\theta + 2\pi kt)$. Therefore, if we choose $t = t_1 = 2\pi k/(2\pi k + \theta)$ or $t = t_2 = (\pi + 2\pi k)/(\theta + 2\pi k)$, where k is the largest branch number[25] that satisfies (3.24), the matrix $\tilde{P} = e^{t \log \hat{P}}$ will have repeated real eigenvalues, either $(\rho^{t_1}e^{i2\pi k}, \rho^{t_1}e^{-i2\pi k}) = (\rho^{t_1}, \rho^{t_1})$ or $(\rho^{t_2}e^{i\pi(2k+1)}, \rho^{t_2}e^{-i\pi(2k+1)}) = (-\rho^{t_2}, -\rho^{t_2})$. This technique is called "riding log $\hat{P}$." It was employed in example 2, and it is applied again in Section 4.2.

4. MULTIPLE SOLUTIONS OF $\hat{P} = e^Q$

4.1 *Conceptual Overview*

The tests outlined in the preceding section permit a researcher to ascertain whether or not an empirically determined matrix $\hat{P}(t_1)$, constructed from observations at times $t = 0$ and $t = t_1$, is compatible with a continuous-time Markov process. When the answer is affirmative, at least one version of log $\hat{P}(t_1)$ will be in $\underline{\underline{Q}}$. In general, as we have observed, it may be necessary to examine several branches of log $\hat{P}(t_1)$ to resolve the question of embeddability. For instance, when $\hat{P}(t_1)$ has complex eigenvalues, each complex conjugate pair will generate $U + L + 1$ candidates for membership in $\underline{\underline{Q}}$.

In discussing the tests in Section 3, our objective was to investigate embeddability; we sought to determine whether *any* of the log $\hat{P}(t_1)$ candidates was, in fact, a bona fide member of $\underline{\underline{Q}}$. In the present section, we shift emphasis and inquire into *how many* versions of log $\hat{P}(t_1)$ can belong to $\underline{\underline{Q}}$. Stated technically, we wish to compute the number of different solutions Q to the equation $e^{Qt_1} = \hat{P}(t_1)$ which have the required structure (2.2). In the discussion that follows, we shall assume $\hat{P}(t_1)$ is embeddable; that is, at least one version of the logarithm is in $\underline{\underline{Q}}$.

Under certain conditions, it is possible to guarantee that this solution $Q \in \underline{\underline{Q}}$ will be *unique*. In particular, this is so whenever one of the following sufficiency conditions is satisfied:

i) The eigenvalues of $\hat{P}(t_1)$ are distinct, real, and positive.

ii) $\min_i\{\hat{p}_{ii}(t_1)\} > 1/2$, where $\hat{p}_{ii}(t_1)$ is the diagonal element in the ith row of $\hat{P}(t_1)$.

iii) det $\hat{P}(t_1) > e^{-\pi} = .0432$.

The first criterion derives from the fact that only the principal branch of log $\hat{P}(t_1)$ is real valued under the indicated eigenvalue constraints.

[25] k may be positive or negative. The sign is chosen according to whether one wants to "move backward" to a repeated eigenvalue situation ($+k$) or "move forward" ($-k$). Note also that $k = 0$ will not generate a continuum at $\theta = 0$. These issues are addressed in greater detail in Sections 4.2 and 4.3.

Also, in this circumstance, the assessment that $Q = (1/t_1)\log \hat{P}(t_1) \in \underline{\underline{Q}}$ will be unique is independent of the choice of t_1, since the eigenvalues of $P(t)$ generated by such a Q retain the specified properties for all times t. Additionally, when the eigenvalues satisfy (i), the series formula (2.5) will converge to the unique version of the logarithm in $\underline{\underline{Q}}$.

The second and third criteria were established by Cuthbert (1972, 1973) and refer to the specific times t in the evolution of $P(t) = e^{Qt}$ at which the solution $Q \in \underline{\underline{Q}}$ will be unique. For the purpose of model identification, conditions (ii) and (iii) reveal that *every* Markov chain (identified by a matrix $Q \in \underline{\underline{Q}}$ via the relation $P(t) = e^{Qt}$) has an interval of time $[0, T]$ during which only one version of $\log P(t_1)$, $0 < t_1 < T$, is in $\underline{\underline{Q}}$. [The location of the uniqueness interval at the origin follows from the fact that $L(t)$, the number of branches of $\log P(t)$ in $\underline{\underline{Q}}$, is a nondecreasing function of time—see fig. 6, Section 4.2.]

These comments suggest that, in planning an observational study where Markov models are to be utilized for identifying non–directly observable mobility mechanisms (Q-matrices), it is advisable to take the first two observations as close together as possible, while still allowing a representative amount of movement to occur. The question of what constitutes an appropriate time interval is clearly tied to the nature of the particular substantive process. The point to be highlighted here is that, because of the complications which arise when there are multiple solutions, this sort of consideration is consequential in developing sampling strategies for situations where the number of time points at which data can be collected is very restricted.

Except when one of the special conditions (i), (ii), or (iii) is satisfied, it is possible for several branches of $\log \hat{P}(t_1)$ to be in $\underline{\underline{Q}}$. This nonuniqueness phenomenon, illustrated in the examples below, has received very little attention in scientific disciplines (physics, engineering, sociology) in which Markov processes are frequently utilized. Nonetheless, the existence of multiple solutions $Q \in \underline{\underline{Q}}$ to the equation $e^{Qt_1} = \hat{P}(t_1)$ is not at all uncommon.

Example 12. Consider the empirically determined matrix[26]

$$\hat{P}(t_1) = \begin{bmatrix} .234 & .252 & .264 & .250 \\ .252 & .237 & .245 & .266 \\ .268 & .255 & .230 & .247 \\ .248 & .271 & .248 & .233 \end{bmatrix}.$$

[26] The relative closeness of this array to the equilibrium matrix is not a requirement for the existence of multiple branches, except for small order arrays such as 3×3 and 4×4. See n. 29 below on this point.

This array can be represented in the form e^{Qt_1}, $Q \in \underline{\underline{Q}}$ and $t_1 = 1$, with either of the matrices

$$Q_1 = \begin{bmatrix} -3.350 & 0.134 & 0.067 & 3.149 \\ 3.132 & -3.306 & 0.144 & 0.030 \\ 0.035 & 3.233 & -3.395 & 0.127 \\ 0.137 & 0.033 & 3.149 & -3.319 \end{bmatrix}$$

or

$$Q_2 = \begin{bmatrix} -3.329 & 3.312 & 0.005 & 0.012 \\ 0.033 & -3.337 & 3.209 & 0.095 \\ 0.016 & 0.023 & -3.334 & 3.295 \\ 3.294 & 0.050 & 0.027 & -3.371 \end{bmatrix}.$$

From the perspective of uncovering structural mechanisms, the matter of identifying the "correct" Q for an empirical process must be a central consideration, because the alternative intensity matrices consistent with the mathematical formalism $\hat{P}(t_1) = e^{Qt_1}$ will lead to different substantive conclusions. If only the branch $(1/t_1)\log \hat{P}(t_1) = Q_1$ were recovered, one would assert that the most frequent transitions are $S_1 \rightarrow S_4$, $S_2 \rightarrow S_1$, $S_3 \rightarrow S_2$, and $S_4 \rightarrow S_3$. In contrast, if only the branch $(1/t_1)\log \hat{P}(t_1) = Q_2$ were computed, one would contend that the process evolves principally through the following pattern of movements: $S_1 \rightarrow S_2$, $S_2 \rightarrow S_3$, $S_3 \rightarrow S_4$, and $S_4 \rightarrow S_1$. Since, in applications of continuous-time Markov processes, attention has been directed to the relative magnitudes of the q_{ij} entries and to apportioning these elements among theoretically specified effect parameters (e.g., Coleman 1964*a*, chap. 6; McDill and Coleman 1963; Bartholomew 1973, chap. 5), identification of the appropriate intensity matrix would appear to be a necessary initial step in this sort of analysis.

This task may be divided into two component issues: (a) recovery of all matrices $Q \in \underline{\underline{Q}}$ that are compatible with the representation $e^{Qt_1} = \hat{P}(t_1)$ and (b) selection from this list of alternative Q-matrices the correct one for the empirical process at hand. Procedures for accomplishing the first task are presented in the current section. The second issue can be resolved by bringing additional substantive information to bear on the nature of the process to aid in choosing among the alternative Q-matrices, by collecting data at more than two time points, or by sampling the population over a briefer time interval (e.g., within the region of uniqueness). These matters will be considered in Section 5.

4.2 *How Multiple Versions of log $\hat{P}(t) \in \underline{\underline{Q}}$ Arise*

The simplest way to describe how multiple matrices $Q \in \underline{\underline{Q}}$ originate is to consider the case of a general 3×3 stochastic matrix $P(t)$ which has complex eigenvalues. Expressing this matrix in diagonal form, we have

$P(t) = HD(t)H^{-1}$. For convenience, we write the complex eigenvalues of $P(t)$ as exponentials,

$$D(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda(t) & 0 \\ 0 & 0 & \bar{\lambda}(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{t(a+bi)} & 0 \\ 0 & 0 & e^{t(a-bi)} \end{bmatrix}, \quad (4.1)$$

where $\bar{\lambda}(t)$ denotes the complex conjugate of the eigenvalue $\lambda(t)$. Then $\log \hat{P}(t) = H \log D(t)H^{-1}$, in which

$$\log D(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & at + i(bt + 2\pi k) & 0 \\ 0 & 0 & at - i(bt + 2\pi k) \end{bmatrix},$$

$$\mathrm{k} = 0, \pm 1, \pm 2, \ldots . \quad (4.2)$$

We specify $b > 0$. Also note, for reference, that because $|\lambda(t)| < 1$ for all t, $at = \log |\lambda(t)| < 0$.

Applying Runnenberg's necessary condition for embeddability (3.24), we have

$$\frac{5\pi}{6} \leqslant \tan^{-1}\left(\frac{bt + 2\pi k}{at}\right) \leqslant \frac{7\pi}{6}, \quad k = 0, \pm 1, \pm 2, \ldots, \quad (4.3)$$

where the inverse tangent specifies arg(log $\lambda(t)$) in (3.24). For a fixed t, we therefore have a series of tests, one for each integer (branch) k. The point to be emphasized here is that, since every branch of log $P(t)$ whose eigenvalues satisfy (4.3) is a candidate for membership in $\underline{\underline{Q}}$, more than one version of the logarithm may, in fact, be in $\underline{\underline{Q}}$. It is also the case that, as t increases and $P(t)$ evolves to the equilibrium matrix of the process, the number of branches of log $P(t)$ that are potentially in $\underline{\underline{Q}}$ becomes larger. These phenomena are illustrated in figures 4 and 5.

Figure 4 displays the locations of various branches[27] of the eigenvalues $\log \lambda(t) = (a \pm bi)t + 2\pi k$, $t = 1$, in relation to Runnenberg's criterion. The wedge-shaped region (solid lines) defines the boundaries of this necessary condition for embeddability—all eigenvalues of log $P(t)$ must lie in the zone. In this illustration, only the principal branch ($k = 0$) is located in the wedge-shaped region; other branches of the logarithm, which differ by multiples of 2π in their imaginary parts, lie outside the wedge.

Now consider the effect of letting t increase. With respect to the principal branch of log $\lambda(t)$, $\tan^{-1}(bt/at) = \tan^{-1}(b/a)$, and hence the

[27] For brevity in the discussion, we focus on positive branches ($k > 0$). An analogous description can be presented for negative branches of the logarithm.

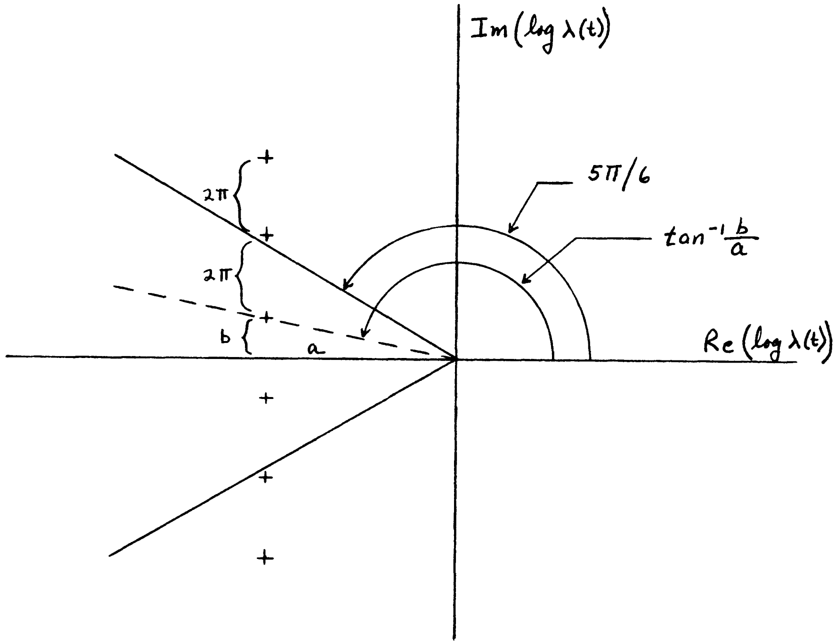

FIG. 4.—Eigenvalues of log $P(t)$, for $t = 1$. The pair of crosses closest to the negative real axis depicts the principal branch of the logarithm, log $\lambda = a \pm bi$. The pair next further out represents the branch for $k = 1$, i.e., log $\lambda = a \pm i(b + 2\pi)$, and so forth.

argument of the logarithm is unchanged. With regard to any other branch $k > 0$, since

$$\frac{bt + 2\pi k}{at} = \frac{b + 2\pi k/t}{a} > \frac{b + 2\pi k}{a}$$

(the inequality follows because $a < 0$) and since $\tan^{-1}x$ is an increasing function of x in the second quadrant, we have

$$\begin{aligned} \arg(at + i[bt + 2\pi k]) &= \tan^{-1}([bt + 2\pi k]/at) \\ &> \tan^{-1}([b + 2\pi k]/a) = \arg(a + i[b + 2\pi k]). \end{aligned}$$

This calculation shows that the angle made by a branch of the logarithm $(k > 0)$, with respect to the positive real axis, enlarges with time. As a result, additional branches enter the wedge, and the number of versions of log $P(t)$ that are candidates for membership in $\underline{\underline{Q}}$ increases. This phenomenon is illustrated in figure 5.

If we let $L(t) = \{$number of branches of log $P(t) \in \underline{\underline{Q}} |$given $t\}$, the next relevant considerations are: (i) $L(t)$ itself is a monotone nondecreas-

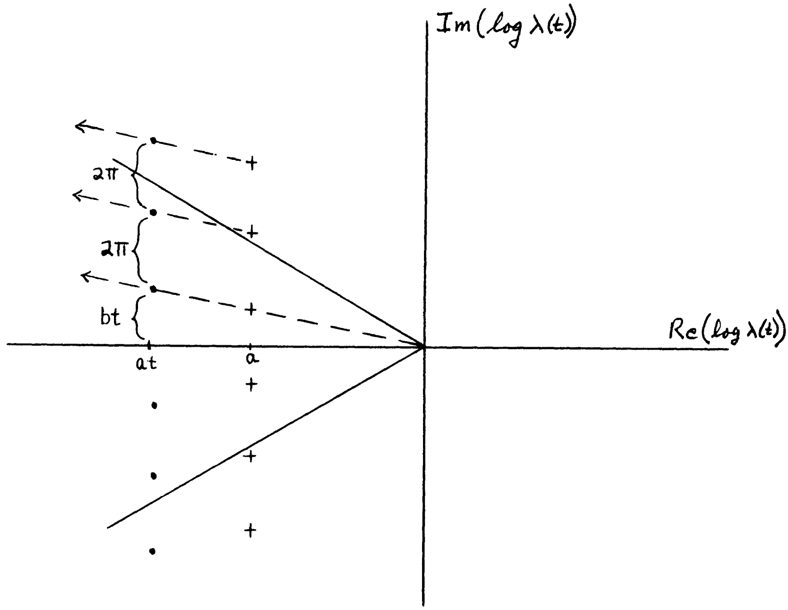

FIG. 5.—Trajectories of the eigenvalues of log $P(t)$, as a function of time. The dashed lines with arrowheads show the trajectories of the branches of log $\lambda(t)$, as a function of t.

ing function of t (except, possibly, for isolated time points), and (ii) if $L(t) > 1$ at some time t (other than one of the isolated time points), then $L(t) \rightarrow \infty$ as $t \rightarrow \infty$ (Cuthbert 1972, 1973). The graph of $L(t)$ in figure 6 is the prototype for the evolution of any Markov chain where Q

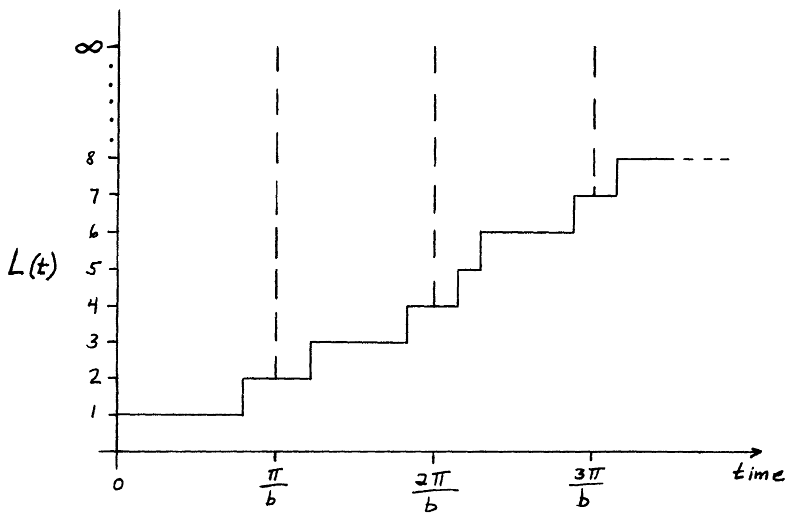

FIG. 6.—Number of branches of log $P(t)$ in $\underline{\underline{Q}}$, as a function of time

has distinct eigenvalues and at least one complex conjugate pair. At times $t = \pi/b, 2\pi/b, 3\pi/b, \ldots, n\pi/b, \ldots$, the complex conjugate eigenvalues of $P(t)$ will equal $\exp[an\pi/b \pm i(n\pi + 2\pi k)]$, $n = 1, 2, 3, \ldots$, and $k = 0, \pm 1, \pm 2, \ldots$. This expression reduces to one of the multiple real-root conditions, either $\lambda_2(t) = \lambda_3(t) = \exp(an\pi/b)$ or $\lambda_2(t) = \lambda_3(t) = -\exp(an\pi/b)$, according to whether n is even or odd. The point to be stressed is that, at these times, $P(t) = e^{Qt}$ has repeated eigenvalues with nondistinct elementary divisors, which will give rise to a continuum of branches of $\log P(n\pi/b)$.

From the point of view of model identification—determining the correct $Q \in \underline{\underline{Q}}$ for a substantive process—these times are a source of difficulty, because their locations are a priori unknown. Knowledge of $\log \hat{P}(t_1)$, where $\hat{P}(t_1)$ has the same structure as $P(n\pi/b)$ in the preceding illustration, can be useless for making statements about the propensity of individuals to move between particular states.[28] If many observations in time were to be allowed in a particular study, we could prepare sampling plans for model identification which would be relatively uninfluenced by this phenomenon. With observations at only two, three, or four time points being a constraint in most studies, however, a single uninformative matrix $\hat{P}(t_i)$ can make a considerable difference in the available information for identifying the Q-matrix underlying a substantive process.

With general r-state matrices, the preceding discussion is complicated by the possible presence of more than one pair of complex conjugate eigenvalues. The graph of $L(t)$ (fig. 6) would then be altered in two ways: first, there would be additional isolated time points at which $L(t) = +\infty$. These correspond to the instants at which the added complex eigenvalues have zero imaginary parts and become repeated real roots. Second, the rise in the step function can be much steeper. This is because the wedge-shaped region (fig. 4), which determines the number of branches of $\log \lambda(t)$ that can generate candidates for membership in $\underline{\underline{Q}}$, widens as a function of r, the order of the matrix. This phenomenon is illustrated in figures 7, 8, and 9.

Figure 7 displays the wedge-shaped zones for general 3-state and 6-state matrices; the respective angles made with the positive real axis are determined by the inequalities (4.3). From the illustrative representation of an eigenvalue of $\log \hat{P}$ and its complex conjugate, we see that, while only the principal branch lies in the wedge for 3×3 matrices, two additional branches would be candidates for membership in $\underline{\underline{Q}}$ if this same eigenvalue belonged to the larger array. It is this fact, together with the presence of additional complex conjugate eigenvalues to generate candidates for membership in $\underline{\underline{Q}}$, which prompted our remark in Section 3.3 to

[28] The same remark holds for a $\hat{P}$ which is considered to be within error distance of $P(n\pi/b)$.

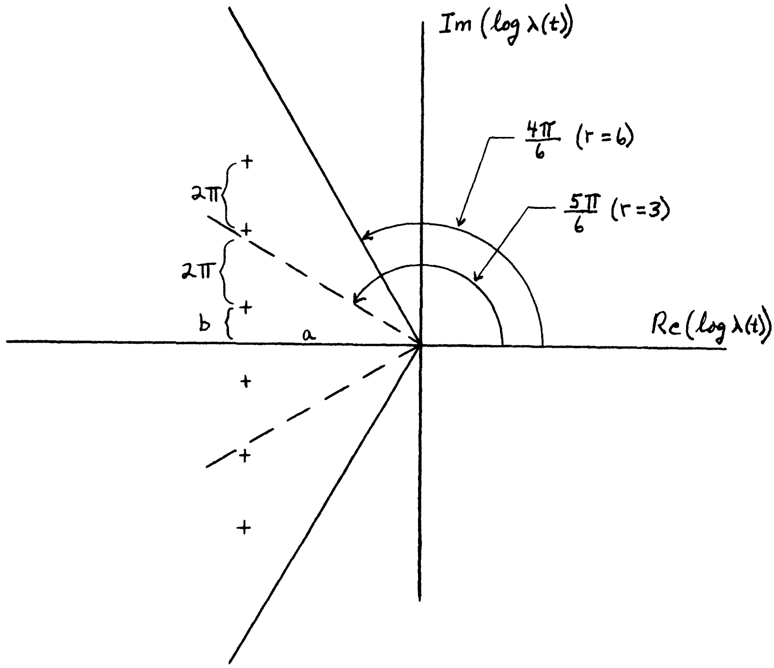

FIG. 7.—Runnenberg's wedge criterion, illustrated for 3×3 and 6×6 matrices, for $t = 1$.

the effect that the number of branches which must be checked for embeddability increases directly with the order of $\hat{P}$. In the context of the present discussion, we emphasize that the computations are more likely to produce multiple versions of log $\hat{P} \in \underline{\underline{Q}}$ in large-order arrays.

Figure 8 presents the same information as figure 7, but from a different perspective. The preceding plot depicted the constraints on the branches of the eigenvalues of *log* $\hat{P}$, as they relate to eligibility for membership in $\underline{\underline{Q}}$. In figure 8 we display the conditions on the eigenvalues of $\hat{P}$, in the case of 4×4, 6×6, 12×12, and 20×20 matrices, for it to generate at least two candidates for membership in $\underline{\underline{Q}}$. We thereby see in a more direct fashion how the constraints are relaxed as the matrix size is increased.[29] Finally, in figure 9 we show the restrictions for different numbers of logarithms to be eligible for membership in $\underline{\underline{Q}}$, in the particular instance of a 20×20 array. The outer, heart-shaped region, labeled H_{20}, is a graph of Runnenberg's necessary conditions: all eigenvalues of $\hat{P}$

[29] In connection with this point, we refer the reader to the 4×4 matrix $\hat{P}(t_1)$ in example 12. The reason why it is reasonably similar to the equilibrium matrix for the process can now be appreciated; namely, the complex conjugate eigenvalues are close to zero in magnitude.

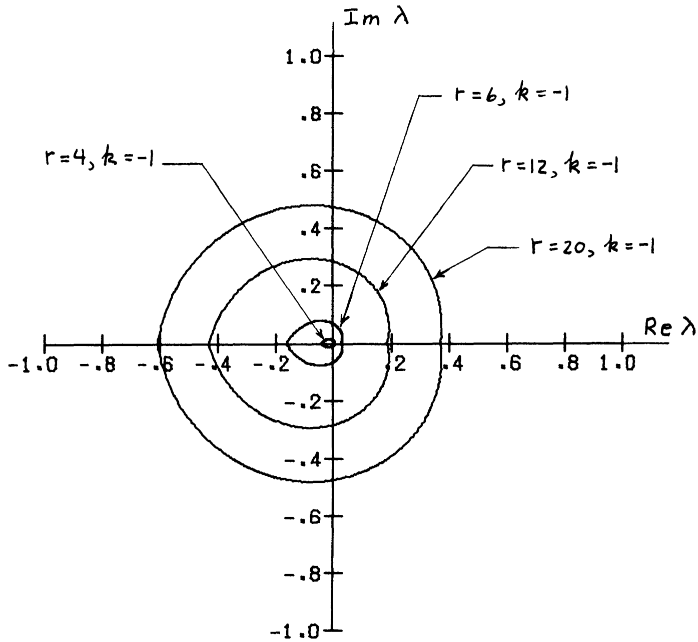

FIG. 8.—Eigenvalue regions of $\hat{P}$ in which two versions of log $\hat{P}$ are candidates for membership in $\underline{\underline{Q}}$, for 4 × 4, 6 × 6, 12 × 12, and 20 × 20 matrices. Each eigenvalue of $\hat{P}$ interior to the curve relevant to its size will generate at least two logarithm candidates for membership in $\underline{\underline{Q}}$.

must lie in this zone for the matrix to be embeddable. The interior curves delineate the regions in which an eigenvalue of $\hat{P}$ will generate multiple branches of log $\hat{P}$ that can be in $\underline{\underline{Q}}$; for instance, if some eigenvalue λ_j lies interior to the curve labeled "$k = -1$," then each of the two branches of its logarithm,

$$\log \lambda_j = a + bi \quad \text{and} \quad \log \lambda_j = a + i(b - 2\pi),$$

will generate versions of log $\hat{P}$ which must be examined for membership in $\underline{\underline{Q}}$.

The most severe form of nonuniqueness of log $\hat{P}(t_1)$ occurs for Markov chains $P(t) = e^{Qt}$ having real eigenvalues which remain repeated for *all* $t > 0$, instead of separating into complex conjugates, as was the case in the preceding discussion. The transition mechanisms associated with such chains are by no means pathological from a substantive point of view, and the prototype of this phenomenon is illustrated in the following example.[30]

[30] The remainder of this section is more difficult mathematically and can be skipped at a first reading. Continue with Section 4.3.

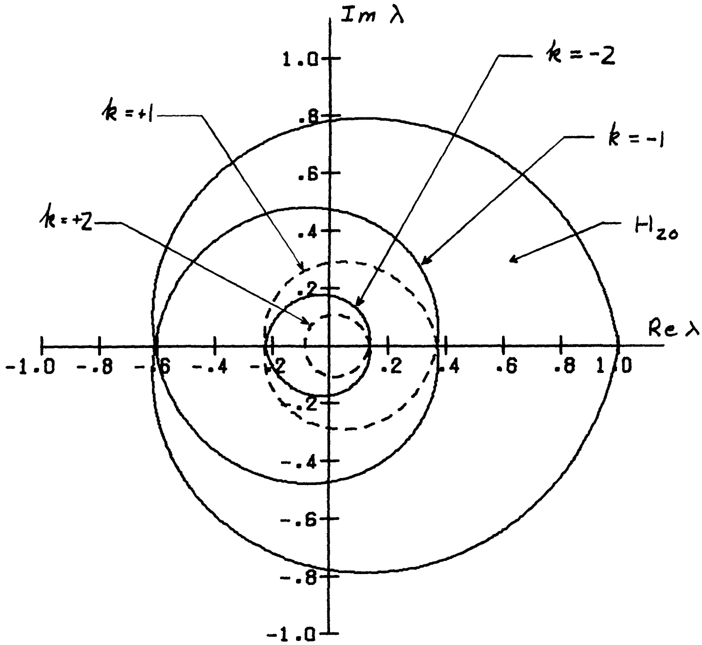

FIG. 9.—Eigenvalue regions of $\hat{P}$ in which multiple versions of log $\hat{P}$ are candidates for membership in $\underline{\underline{Q}}$, for 20 × 20 matrices. All eigenvalues of $\hat{P}$ must lie in the region H_{20} for $\hat{P}$ to be embeddable. If an eigenvalue is interior to a k-curve, it generates $|k| + 1$ versions of log $\hat{P}$ which may be in $\underline{\underline{Q}}$.

Example 13. Consider the matrix

$$P(t) = \frac{1}{3}\begin{bmatrix} 1 + 2e^{-3t/2} & 1 - e^{-3t/2} & 1 - e^{-3t/2} \\ 1 - e^{-3t/2} & 1 + 2e^{-3t/2} & 1 - e^{-3t/2} \\ 1 - e^{-3t/2} & 1 - e^{-3t/2} & 1 + 2e^{-3t/2} \end{bmatrix}, \quad (4.4)$$

where $t > 0$. $P(t)$ has eigenvalues 1, $e^{-3t/2}$, $e^{-3t/2}$ (which are repeated irrespective of the choice of t) and nondistinct elementary divisors $(\lambda - 1)$, $(\lambda - e^{-3t/2})$, $(\lambda - e^{-3t/2})$. Note that this is the matrix of example 10 with $X = e^{-3t/2}$.

From the discussion of repeated eigenvalues with nondistinct elementary divisors (Section 3.2b), we know that all branches of $(1/t) \log P(t)$ may be computed via

$$\frac{1}{t} \log P(t) = \frac{1}{t} HB \log J(t) B^{-1} H^{-1} \quad (4.5)$$

where H is any similarity transformation that reduces $P(t)$ to diagonal form (e.g., eq. [3.19]), B is a matrix with complex entries (3.21) which commutes with $J(t)$, and log $J(t)$ has the form

$$\log J(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{-3t}{2} + 2\pi ki & 0 \\ 0 & 0 & \frac{-3t}{2} - 2\pi ki \end{bmatrix}, \quad (4.6)$$

in which $k = 0, \pm 1, \pm 2, \ldots$ specifies branches of the logarithm.

We now describe how a continuum arises in this eigenvalue condition. The first time that the complex eigenvalues in (4.6) satisfy Runnenberg's condition (4.3) with $k \neq 0$ occurs at $t^* = 4\pi/\sqrt{3}$. Before this time only the branch $k = 0$ of log $J(t)$ will be in the wedge-shaped zone (fig. 4). It can be checked that, when $k = 0$, $B \log J(t) = \log J(t) B$, and therefore equation (4.5) reduces to $(1/t)\, H \log J(t)\, H^{-1}$ for every matrix B. This means that at most one version of $(1/t) \log P(t)$ can be in $\underset{=}{Q}$. Indeed,

$$Q = \frac{1}{t} \log P(t) = \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -1 \end{bmatrix} \text{for } 0 < t < \frac{4\pi}{\sqrt{3}}. \quad (4.7)$$

When $t > t^* = \frac{4\pi}{\sqrt{3}}$, a second branch of log $J(t)$ in (4.6) enters the wedge-shaped zone (see fig. 5). In this circumstance, it is no longer the case that $B \log J(t) = \log J(t)\, B$, and a continuum of versions of $(1/t) \log P(t)$ will be generated, each version corresponding to a choice of $\{c_{ij}\}$ in B (eq. [3.21]). A bit of computation will show that, if $\{c_{ij}\}$ are restricted according to

i) $c_{11}c_{12} - c_{21}c_{22} = 0$

ii) u is real, $u = \dfrac{c_{11}c_{22} + c_{12}c_{21} + 2c_{21}c_{22}}{c_{11}c_{22} - c_{12}c_{21}}$

v is real, $v = \dfrac{c_{11}c_{22} + c_{12}c_{21} - 2c_{21}c_{22}}{c_{11}c_{22} - c_{12}c_{21}}$

iii) $|ku| < \sqrt{3}t/4\pi$ and $|kv| < \sqrt{3}t/4\pi$

where k is an integer (the branch number), then all choices of $\{c_{ij}\}$ will yield matrices $Q \in \underline{\underline{Q}}$, and they are summarized by

$$Q = \frac{1}{t} \log P(t) =$$

$$\begin{bmatrix} -1 & \frac{1}{2} + \frac{2k\pi}{\sqrt{3}}\frac{u}{t} & \frac{1}{2} - \frac{2k\pi}{\sqrt{3}}\frac{u}{t} \\ \frac{1}{2} - \frac{2k\pi}{\sqrt{3}}\frac{v}{t} & -1 - \frac{k\pi}{\sqrt{3}}\frac{(u-v)}{t} & \frac{1}{2} + \frac{k\pi}{\sqrt{3}}\frac{(u+v)}{t} \\ \frac{1}{2} + \frac{2k\pi}{\sqrt{3}}\frac{v}{t} & \frac{1}{2} - \frac{k\pi}{\sqrt{3}}\frac{(u+v)}{t} & -1 + \frac{k\pi}{\sqrt{3}}\frac{(u-v)}{t} \end{bmatrix}$$

$$\text{for } t > \frac{4\pi}{\sqrt{3}}. \qquad (4.8)$$

The graph of $L(t)$ versus t for this Markov chain is shown in figure 10. From the point of view of model identification, the second observation t_1 must be taken before $t^* = 4\pi/\sqrt{3}$. After this time, a repeated observation will yield the matrix (4.8), which is completely uninformative about the propensity to move between different states. The fundamental difficulty illustrated by this example is that empirically determined matrices with nondistinct elementary divisors *in which this property is retained through time* may be associated with a continuum of intensity matrices for all times t_1 greater than some threshold t^*. To distinguish this "essential"

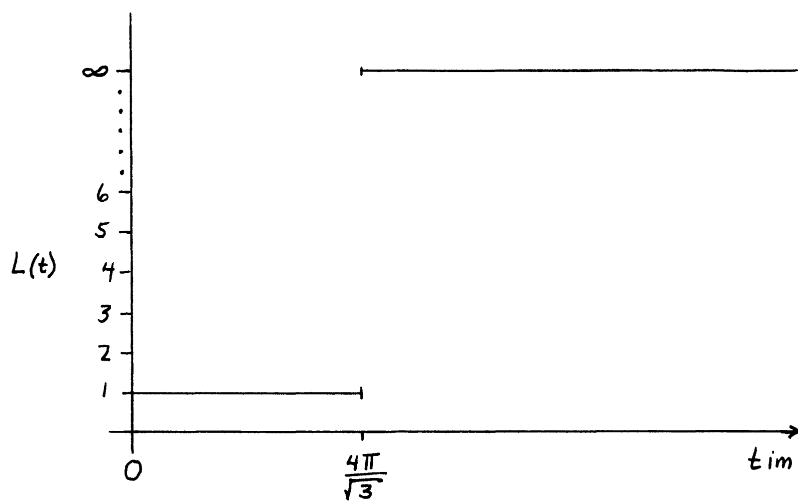

FIG. 10.—Number of branches of log $\hat{P}(t) \in \underline{\underline{Q}}$ as a function of time, for $\hat{P}(t)$ in example 13.

continuum case from the chance occurrence of an "isolated" continuum (viz. the points π/b, $2\pi/b$, . . , $n\pi/b$, . . . in Section 4.2), a researcher should check whether the eigenvalues of $\hat{P}(t_1 + \Delta t)$, some $\Delta t > 0$, are repeated with nondistinct elementary divisors when his initial matrix $\hat{P}(t_1)$ has these properties.

4.3 *Summary of Correspondence between Eigenvalue Characteristics and Number of Matrices $Q \in \underline{\underline{Q}}$*

The number of versions of log $\hat{P}$ that can possibly be in $\underline{\underline{Q}}$, as this relates to the eigenvalue characteristics of $\hat{P}$, is summarized in table 1. The left

TABLE 1

EIGENVALUES OF $\hat{P}$ AND THE NUMBER OF MATRICES $Q \in \underline{\underline{Q}}$

Eigenvalue Characteristics	Embeddable?	How Many Q's?
1. Positive, distinct	Possibly	One
2. Positive, repeated, distinct elementary divisors	Possibly	One
3. Positive, repeated, nondistinct elementary divisors	Possibly	One or continuum
4. Negative, distinct	Never	. . .
5. Negative, repeated, odd multiplicity	Never	. . .
6. Negative, repeated, even multiplicity	Possibly	Continuum
7. Complex, distinct, member of a conjugate pair	Possibly	One or multiple
8. Complex conjugate, repeated	Possibly	One, multiple, or continuum
9. Mixture of the types above	Possibly	The most extreme form of nonuniqueness present in any component of the mixture

tab of the table refers to a single eigenvalue of $\hat{P}$ or to a set of eigenvalues sharing a common property (e.g., complex conjugates). The evaluation in the farthest right column assumes that embeddability is met; in other words, that at least one version of log $\hat{P}$ is in $\underline{\underline{Q}}$. In making this evaluation, it is also presumed that the remaining eigenvalues of $\hat{P}$ do not satisfy a condition which is compatible with a greater number of candidates for membership in $\underline{\underline{Q}}$; for instance, all eigenvalues must belong to categories (1) and (2) in order to conclude, on the basis of an examination of eigenvalues alone, that at most one version of log $\hat{P}$ is in $\underline{\underline{Q}}$.

A second point to be noted in connection with the table is that the eigenvalue conditions which rule out embeddability do so by not being compatible with a real-valued version of log $\hat{P}$. For example, if $\hat{P}$ has a unique negative eigenvalue, $\lambda = -a$ ($a > 0$), its logarithm will be log $a + ik\pi$, $k = 0, \pm 1, \pm 2, \ldots$, which always has a nonzero imaginary part.

The corresponding eigenvector h in the similarity transformation $\hat{P} = HJH^{-1}$ will be *real* valued (since $\lambda = -a$ is distinct and real), and $\log \hat{P} = H \log JH^{-1}$ will have the identical eigenvector corresponding to its *complex* eigenvalue. There is no way in which $\log \hat{P}$ can be real valued in this circumstance. What alters the situation in the case of repeated negative eigenvalues with even multiplicity is that, when the elementary divisors are not distinct, the eigenvectors corresponding to the repeated eigenvalues will be complex conjugates, and real versions of $\log \hat{P}$ can result. In particular, this will occur when different branches of the logarithm of $-a$ are present simultaneously in $\log D$.

Finally, we emphasize that the eigenvalue configurations most commonly found in empirically determined matrices involve combinations of distinct positive and distinct complex conjugates (categories [1] and [7]).

5. TESTING STRATEGIES

5.1 *Identification of Structural Parameters*

We assume first that the process under observation is time stationary, that the data are free of measurement and classification error, and that the entire population has been surveyed, so sampling variability is not a concern. These assumptions have also been made, though without being noted explicitly, in the preceding sections. In this environment, the identification problem arises when observations are taken at only two time points ($t = 0$, $t = t_1$), and the matrix $\hat{P}(t_1)$ constructed from these observations can be represented in the form $\hat{P}(t_1) = e^{Qt_1}$ for multiple arrays $Q \in \underline{\underline{Q}}$. A researcher then has the following options:

i) He may bring to bear other information about the substantive process. For instance, if $\hat{P}(t_1)$ were the matrix in example 12, a researcher might have reason to believe that $q_{12} > q_{14}$ and therefore Q_2, not Q_1, governs the evolution of the process. Clearly, such a choice can be made only when there is a finite list of intensity matrices, and not when a continuum is present.

ii) If an opportunity exists to collect data at a third time point, it should be selected so as not to be an integer multiple of the initial interval $(0, t_1)$. The reason is that, at multiples of an observation interval, the same list of Q-matrices can reappear; this was the case, for instance, with the times π/b, $2\pi/b$, etc. in figure 6. If, however, the third observation is taken at $t_2 \neq kt_1$, k an integer,[31] then even in the presence of multiple

[31] This recommendation assumes that we have observed the first appearance of a continuum, which will be the most common situation. If we have observed the second occurrence, the times $t_2 = \frac{kt_1}{2}$ should be avoided. If it is a third occurrence, omit the

branches of $\log \hat{P}(t_1, t_2) \in \underline{\underline{Q}}$, only one version of the logarithm, Q_0, will have the property

$$Q_0 = \frac{1}{t_1} \log \hat{P}(0, t_1) = \frac{1}{t_2 - t_1} \hat{P}(t_1, t_2). \tag{5.1}$$

This correspondence will identify the *unique* Q that can be associated with the empirical process.

There is an additional virtue in collecting data at three or more time points. The embeddability problem concerns only the question of compatibility of a *single* stochastic matrix $\hat{P}(t_1)$—that is, observations at two time points—with a continuous-time Markov process. We have seen that, on the basis of this information alone, it is frequently possible to rule out a Markov structure. However, when data are available from more than two time points, a direct test can also be made of the fundamental dynamic assumption of a first-order Markov process, namely that the future state of the system depends only on current state, not on its history. These additional necessary conditions are specified by tests of the sort

$$\hat{P}(t_i, t_k) = \hat{P}(t_i, t_j)\, \hat{P}(t_j, t_k), \quad 0 < t_i < t_j < t_k. \tag{5.2}$$

The availability of data at three time points provides the most rudimentary opportunity to check this assumption. Formal statistical tests of the validity of the Markov property are described in Anderson and Goodman (1957) and Billingsley (1961).

Study design considerations.—The potential for nonuniqueness can be minimized at the study design stage. If the use of Markov models is contemplated, the survey times should be chosen close together in time, while still permitting a representative amount of movement to take place. When the number of states is small (say, $r \leqslant 5$), it should be possible to select t_1 so that $\min_i\{\hat{p}_{ii}(t_1)\} > 1/2$. If $\hat{P}(t_1)$ is embeddable, this condition on the diagonal elements ensures that $\log \hat{P}(t_1) \in \underline{\underline{Q}}$ will be unique (see Section 4.1). When the number of system states is large, it may not be possible to satisfy this condition and still retain an adequate amount of population movement to estimate $\log \hat{P}(t_1)$ accurately. Even in this circumstance, however, t_1 should be selected reasonably close in time to the initial observation, since the degree of nonuniqueness of $Q \in \underline{\underline{Q}}$ is a monotone increasing function of time (fig. 6), except for isolated instants such as $\{k\pi/b\}$.

In most data-gathering situations, one has neither a priori information concerning the rate of movement (to assist in selecting the second observation) nor an opportunity to schedule the second wave of a survey accord-

---

times $t_2 = \frac{kt_1}{3}$, etc. As a practical guide, if a researcher avoids the two sets of time points cited in this footnote, he is unlikely to encounter a second continuum.

ing to these considerations. A more pragmatic suggestion would be to collect detailed retrospective information about the process. Ideally, this should consist of "sample path" data; that is, complete information about a respondent's duration in each system state over the time interval of interest. When such data are deemed too costly to collect, a respondent should be queried regarding his system state at several prechosen time points in the past (e.g., one year ago, two years ago, etc.). Having gathered such information, the researcher may utilize the estimation procedures and model tests that require more than two observations in time.

5.2 Sampling Error and Data Noise

The data available to researchers are commonly contaminated by errors of various sorts. While we may wish to make statements about a *population-level* process, information is usually collected for a population *sample*. Similarly, errors of measurement can result in the misclassification of individuals with respect to system state.

Ordinarily, these are not very serious problems. In many sampling situations, the inference made about a population parameter, using standard statistical procedures, tends to be incorrect to a degree that varies *continuously* with the magnitude of the measurement error. By using distributional statistics, one can put confidence bounds around an estimate and describe the interval in which the population-level parameter lies. However, measurement error and sampling variability carry greater consequence when we seek to identify the non–directly observable structural mechanisms (Q-matrices) that underlie Markov processes. In particular, when an empirically determined matrix $\hat{P}(t_1)$ is in the vicinity of a second stochastic matrix $\tilde{P}$ which can be expressed in the form $\tilde{P} = e^Q$ for multiple versions of $\log \tilde{P} \in \underline{\underline{Q}}$, then a small error in the estimate of $\hat{P}(t_1)$ can result in the recovery of a matrix $Q \in \underline{\underline{Q}}$ which, while unique, is *the wrong intensity matrix for the substantive process.*

Example 14. Suppose you observe

$$\hat{P}_1(t_1) = \begin{bmatrix} .232 & .249 & .266 & .253 \\ .254 & .236 & .242 & .268 \\ .270 & .258 & .228 & .244 \\ .245 & .274 & .250 & .231 \end{bmatrix}.$$

This matrix can be written in the form e^{Qt_1}, $t_1 = 1$, for a *unique* version of $\log \hat{P}(t_1) \in \underline{\underline{Q}}$,

$$Q_1 = \begin{bmatrix} -3.216 & 0.129 & 0.064 & 3.023 \\ 3.007 & -3.174 & 0.138 & 0.029 \\ 0.034 & 3.104 & -3.260 & 0.122 \\ 0.132 & 0.032 & 3.023 & -3.186 \end{bmatrix}.$$

If one believed $\hat{P}_1(t_1)$ to be error free, it would be reasonable to conclude that Q_1 describes the evolution of the dynamic process. However, in a fallible environment, a second survey of the same population would produce a slightly different observed matrix. Consider

$$\hat{P}_2(t_1) = \begin{bmatrix} .231 & .255 & .266 & .248 \\ .250 & .234 & .247 & .269 \\ .271 & .252 & .227 & .250 \\ .251 & .275 & .245 & .229 \end{bmatrix}.$$

No element of this matrix differs from its counterpart in $\hat{P}_1(t_1)$ by an amount in excess of .006 in magnitude, so it is not unreasonable to suggest that the two matrices represent different samples from a single parent population. However, while it is the case that $\hat{P}_2(t_1)$ is also compatible with a continuous-time Markov process for a *unique* $Q \in \underline{\underline{Q}}$, this intensity matrix is given by

$$Q_2 = \begin{bmatrix} -3.164 & 3.148 & 0.005 & 0.011 \\ 0.031 & -3.170 & 3.049 & 0.090 \\ 0.015 & 0.022 & -3.167 & 3.130 \\ 3.130 & 0.048 & 0.026 & -3.204 \end{bmatrix}.$$

Matrices Q_1 and Q_2 represent very different structural mechanisms and would lead to contrary conclusions about the nature of the substantive process. What has happened is that, while $\hat{P}_1(t_1)$ and $\hat{P}_2(t_1)$ are each compatible with the representation e^{Qt_1} and have unique logarithms in $\underline{\underline{Q}}$, the two empirically determined P-matrices lie in the vicinity of a third, $\tilde{P}$, which in turn can be represented as a Markov process for multiple matrices $Q \in \underline{\underline{Q}}$. Indeed,

$$\tilde{P} = \hat{P}_1(t_1 + .05) = e^{(t_1+.05)Q_1}$$

$$\tilde{P} = \hat{P}_2(t_1 + .05) = e^{(t_1+.05)Q_2}$$

and this common P-array is the same one presented in example 12 to illustrate the phenomenon of multiple intensity matrices.[32]

Specific error structures.—In the context of sampling variability or measurement error, then, a researcher cannot assume that, because $\hat{P}(t_1) = e^{Qt_1}$ for a unique $Q \in \underline{\underline{Q}}$, this intensity matrix describes the evolution of the substantive process. He must either remove the error from the observed matrix and use the "purged" array for estimating structural parameters or examine the intensity matrices of other P's that are within "error distance" of his empirically determined matrix.

[32] The Q-matrices in example 12 are the ones in this illustration multiplied by $t = 1.05$.

Misclassification error can be incorporated formally in a description of observed transition matrices by introducing the representation

$$\hat{P}(t_i,t_j) = \bar{P}(t_i,t_j) \; o \; E(t_i,t_j) \qquad 0 \leqslant t_i < t_j, \tag{5.3}$$

where $\hat{P}(t_i,t_j)$ is an empirically determined $r \times r$ matrix of transition probabilities based on observations at times t_i and t_j; $\bar{P}(t_i,t_j)$ is a fitted $r \times r$ matrix of transition probabilities representing the error-free or purged mobility structure; $E(t_i,t_j)$ is an $r \times r$ matrix of residuals interpreted as errors due to misclassification; and the symbol o denotes either *addition* or *multiplication*. Motivating the representation (5.3) is the view that matrix $\bar{P}$, rather than $\hat{P}$, should be tested for compatibility with a Markov process and Q should be estimated from the equation $\bar{P} = e^Q$.

Calculation of $\bar{P}$ and E must be based on an assumed model of the error structure, together with independent estimates of the parameters. For example, if the states are occupational categories and there is a natural ordering among them (e.g., on the basis of a prestige scale), an individual who actually moves from state i at time t_1 to state j at time t_2 may have probability c_1 of being recorded in state $j-1$ at time t_2, probability c_2 of being recorded in state $j+1$ at time t_2, and probability $1-c_1-c_2$ of being recorded correctly. If this kind of measurement error is believed to operate, it implies a representation of the form

$$\hat{P}(t_1,t_2) = \bar{P}(t_1,t_2)E \tag{5.4}$$

where

$$E = \begin{bmatrix} 1-c_1 & c_2 & 0 & 0 & \mathbf{0} \\ c_1 & 1-c_1-c_2 & c_2 & 0 & \\ 0 & c_1 & 1-c_1-c_2 & c_2 & \\ & & \ddots & \ddots & \ddots \\ & & & & c_2 \\ \mathbf{0} & & & c_1 & 1-c_2 \end{bmatrix}.$$

Given c_1 and c_2 based on independent misclassification estimates, we could solve the matrix equation (5.4) for $\bar{P}(t_1,t_2)$. See Coleman (1964*b*) for approaches of this sort to the study of change in a fallible environment.

Random error.—In general, a formal model of the error structure will not be available, yet we may wish to make allowance for the effect of "noise" in the data. We recommend a strategy of "exploring" a neighborhood of the observed matrix $\hat{P}(t_1)$, to ascertain whether nearby P-arrays are compatible with intensity matrices that are very different from the initial Q-matrix.

A reasonable procedure for exploring a neighborhood of $\hat{P}(t_1)$ would be to "ride" its associated intensity matrix Q_o. By this is meant computing

$P(t)$ from the representation $P(t) = e^{Q_o t}$, using for t the values $t_1 - \Delta t$, $t_1 - 2\Delta t, \ldots, t_1 - h\Delta t$, and $t_1 + \Delta t$, $t_1 + 2\Delta t, \ldots, t_1 + k\Delta t$, where the termination points h and k are the last times that $P(t_1 - j\Delta t)$ and $P(t_1 + l\Delta t)$ can be considered "within sampling or measurement error" of the observed matrix. Next, examine the eigenvalues in the sequence of matrices:

a) If there is a complex conjugate pair $(\lambda, \bar{\lambda}) = (a \pm bi)$ whose imaginary part passes through zero, then $\hat{P}(t_1)$ is in a neighborhood of some matrix $\tilde{P}$ which has repeated real eigenvalues. Associated with this array, a continuum of matrices $Q \in \underline{\underline{Q}}$ will satisfy the relation $\tilde{P} = e^Q$. Strategies for exploring the structure of a continuum are discussed in the Appendix.

b) If a continuum does not occur within error distance, recover all matrices $Q \in \underline{\underline{Q}}$ that are compatible with the representation $P(t_1 + k\Delta t) = e^{Q_o(t_1 + k\Delta t)}$, where k was chosen as the forward stopping point of the sequence of P-matrices.[33] The complete solution to the problems of determining the number of candidates for membership in $\underline{\underline{Q}}$ and computing all versions of $\log P \in \underline{\underline{Q}}$ was presented in Section 3.

If it is the case that $\log \hat{P} \in \underline{\underline{Q}}$ is unique under the perturbations of $\hat{P}(t_1)$, this intensity matrix can be viewed as the sole mobility structure compatible with a Markov formulation of the substantive process. Stated more transparently, additional samples from the same population can be expected to produce similar Q-matrices. In contrast, if multiple mobility mechanisms $Q \in \underline{\underline{Q}}$ are found for matrices $\tilde{P}$ within error distance of the observed array $\hat{P}(t_1)$, one of the procedures described in Section 5.1 for selecting among alternative intensity matrices must be utilized.

In an environment containing error, the advantages of collecting data at three or more points in time are especially apparent. We noted earlier (Section 5.1) that three time points are the minimum number for a direct test of the dynamic assumption underlying a first-order Markov process, that is, for checking that

$$\hat{P}(0,t_2) = \hat{P}(0,t_1)\,\hat{P}(t_1,t_2), \qquad 0 < t_1 < t_2. \tag{5.5}$$

In practice, this entails evaluating whether $||\hat{P}(0,t_2) - \hat{P}(0,t_1)\,\hat{P}(t_1,t_2)|| < \epsilon$, for $\epsilon > 0$, and some suitably chosen norm (e.g., $||A|| = \sqrt{\sum_{i,j} a_{ij}^2}$). When (5.5) is satisfied *and* it is also the case that

$$\frac{1}{t_1}\log \hat{P}(0,t_1) \approx \frac{1}{t_2 - t_1}\log \hat{P}(t_1,t_2) \approx \frac{1}{t_2}\log \hat{P}(0,t_2), \tag{5.6}$$

[33] Because $L(t) = \{$the number of branches of $\log P(t) \in \underline{\underline{Q}}\}$ is a nondecreasing function of time (except for isolated occurrences of continua), it is not necessary to examine points earlier than t_1.

we would define Q, the common intensity matrix for the process, as an average of these three estimates. In the presence of sampling or measurement error, then, collection of data at three or more time points permits a test of the fundamental Markov assumption and also facilitates an accurate calculation of Q, through the pooling of several estimates.

In an instance where (5.5) is satisfied but equation (5.6) is not, the process will still be Markovian, though it is no longer time stationary. This leads to the problem of testing observed matrices for compatibility with a time-homogeneous Markov model (the null hypothesis) against special non-time-homogeneous alternatives. We hope to discuss this important issue in a future publication.

As a final comment on analytic strategy in the context of data noise, we emphasize that, while the occurrence of multiple matrices $Q \epsilon \underline{\underline{Q}}$ may not be very common in an error-free environment, it characterizes the normal work situation when data are fallible. This is because we advise a researcher to examine a neighborhood of an observed $\hat{P}(t_1)$ for the presence of additional intensity matrices and to consider each recovered $Q \epsilon \underline{\underline{Q}}$ as possibly governing the evolution of the empirical process. Due to data noise, then, we suggest *creating* a multiple $Q \epsilon \underline{\underline{Q}}$ situation when an observed transition matrix has associated with it a unique logarithm in $\underline{\underline{Q}}$. For this reason, collection of data at three or more time points should be a routine requirement when the use of Markov models is contemplated.

6. CONCLUSIONS

The point of departure for this study was the gross misunderstanding among researchers concerning which stochastic matrices are compatible with a continuous-time Markov process having stationary transition probabilities. We noted that the power-series representation of the logarithm of a matrix (eq. [2.5])—the principal formula used in estimating the structural parameters that govern the evolution of a Markov process—permits an intensity matrix to be recovered only for a subset of this class of stochastic models. By resorting, instead, to the spectral-decomposition representation, we were able to estimate intensity matrices for Markov models in instances where (2.5) does not converge; that is, in cases of transition arrays which Coleman and others have considered not to be compatible with this mathematical structure. In the course of the investigation, we also raised new issues which a researcher must consider; these include, principally, the possibility that multiple intensity matrices may be compatible with an empirically determined transition array and the fact that, as a result of data "noise," recovery of a unique $Q \epsilon \underline{\underline{Q}}$ does not preclude the possibility that the observed process is governed by an entirely different intensity matrix.

In subsequent papers, we intend to address two additional issues which a researcher desiring to use Markov models in a flexible and creative manner must entertain: (*a*) How should a priori restrictions be placed on the elements of a Q-matrix, and (*b*) how can a researcher discriminate among the alternative mathematical models which, on substantive grounds, provide reasonable descriptions of his data? The first topic was mentioned, in passing, in Section 3.3, when we sought to adjust a nonembeddable $\log \hat{P}$ to a neighboring $Q \in \underline{\underline{Q}}$. More generally, we may wish to estimate the parameters of a sociological theory which specifies that certain instantaneous transitions are prohibited (see Coleman 1964*a*, chaps. 4 and 5, for examples). The second topic refers to testing data for compatibility with a subset of Markov models (such as birth and death processes) versus general finite-state Markov processes and to comparing the fit of Markov models with that of other mathematical structures, such as mixtures of Markov processes or semi-Markov processes.

As a final point, we emphasize that the problems addressed in this paper cannot be avoided by employing a *discrete-time* Markov framework in place of a continuous-time formulation. In the discrete-time model, the counterpart to the task of estimating $Q \in \underline{\underline{Q}}$ entails recovering the one-step transition matrix for an empirical process; that is, taking the appropriate root of the observed matrix $\hat{P}$. Like a logarithm, a root is a multiple-valued function, so the problem of nonuniqueness which we have discussed here arises also in that formulation. Conceptually, the discrete-time model embodies a further difficulty: because most social processes evolve continuously, there usually isn't a compelling reason for preferring one specification of the unit time interval to another. (For instance, in studying *intra*generational occupational mobility, should the unit time interval be five years or three years or six months?) Yet this is a question of great consequence, because an empirically determined matrix (estimated, let us say for this illustration, from observations ten years apart) may be consistent with a discrete-time Markov structure *for some choices of the unit time interval but not for other choices* (see Singer and Spilerman 1974, pp. 360–63, for an example). Where no substantive meaning can be attached to a particular interval length,[34] this does not imply that the unit time interval can be specified at the convenience of the researcher or that tests of the sort described here can be ignored. Rather, it suggests that the appropriate mathematical structure is a continuous-time formulation, the procedures for which have been discussed in this paper.

[34] Examples of instances where a substantive meaning *can* be attached to an interval length and for which discrete-time Markov chains provide an appropriate analytic structure are (i) the popular preference in presidential elections—interval length equals four years—and (ii) school grades for a cohort or an individual—interval length equals one semester.

APPENDIX

Exploring a Continuum[35]

In the case where $\hat{P}$ has repeated eigenvalues and nondistinct elementary divisors, the value of log $\hat{P}$ depends on the choice of similarity transformation that is used to reduce $\hat{P}$ to Jordan form. A computer-based strategy to test a representative collection of branches of log $\hat{P}$ for membership in $\underline{\underline{Q}}$ is the most direct approach we can currently recommend for deciding on compatibility of $\hat{P}$ with a continuous-time Markov model. If a branch of log $\hat{P}$ which belongs to $\underline{\underline{Q}}$ is discovered during the computer tests, then it can be shown that there is in fact a *continuum* of branches which are in $\underline{\underline{Q}}$. The testing strategy outlined below and illustrated in a simple example is also designed to give some indication of the extent of the continuum of branches which are in $\underline{\underline{Q}}$.

Step 1.—Compute one similarity transformation H which reduces $\hat{P}$ to Jordan form. The method of computation is entirely at the discretion of the researcher (see Gantmacher 1960, chap. 6, for suggestions).

Step 2.—Take a random sample of points in an 8-dimensional square region with center at the origin.[36] For each sample of 8 numbers, use them as the real and imaginary parts of the parameters in the matrices B which commute with $J = H^{-1}\hat{P}H$. Then evaluate

$$\log \hat{P} = HB \log J \, B^{-1}H^{-1}$$

where

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{11} & c_{12} \\ 0 & c_{21} & c_{22} \end{bmatrix}, \qquad \log J = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \log \lambda_2 & 0 \\ 0 & 0 & \overline{\log \lambda_2} \end{bmatrix},$$

the $\{c_{ij}\}$ are given by

$$c_{11} = x_{11} + iy_{11} \qquad c_{12} = x_{12} + iy_{12}$$
$$c_{21} = x_{21} + iy_{21} \qquad c_{22} = x_{22} + iy_{22},$$

$\{x_{ij}\},\{y_{ij}\}$ are the 8 numbers associated with each sample point, and $\overline{\log \lambda_2}$ denotes the complex conjugate of $\log \lambda_2$. Note whether this branch of log $\hat{P}$ is in $\underline{\underline{Q}}$. Several hundred such evaluations may be necessary in order to identify those matrices B, if any, which yield versions of log $\hat{P} \in \underline{\underline{Q}}$.

[35] The Appendix is more difficult mathematically and can be skipped at a first reading.

[36] We recommend beginning this search in the 2-dimensional subspace defined by the conditions $c_{11} = c_{22}$, $c_{12} = c_{21}$, real. Then extend the search to the 4-dimensional space defined by the restriction that $\{c_{ij}\}$ be real, and finally introduce complex numbers in the full 8-dimensional space. Improved strategies for exploring this kind of continuum are currently in the preliminary development stage.

The preceding computations do not increase in complexity for $r \times r$ matrices having a single pair of repeated real roots, which is the situation most likely to arise. In this general case, B will have the form

$$B = \begin{bmatrix} 1 & 0 & & & & & & & & \mathbf{0} \\ 0 & a_{11} & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & a_{vv} & 0 & 0 & & & & \\ & & & 0 & c_{11} & c_{12} & & & & \\ & & & 0 & c_{21} & c_{22} & & & & \\ & & & 0 & 0 & 0 & a_{v+3,\, v+3} & & & \\ & & & & & & & \ddots & & \\ \mathbf{0} & & & & & & & & & a_{rr} \end{bmatrix}.$$

However, since only the $\{c_{ij}\}$ generate a continuum, the same simulation as before is involved. In carrying out these computations, the reader is reminded that, if λ_2 is a repeated *negative* root, the simulation must be performed for all branches of $\log \lambda_2 = \log |\lambda_2| + i(\pi \pm 2\pi k)$ which satisfy Runnenberg's necessary condition for embeddability (eq. [4.3]). If λ_2 is a repeated *positive* root, the calculations must be carried out for all branches, except $k = 0$, of $\log \lambda_2 = \log |\lambda_2| \pm 2\pi ki$ which satisfy Runnenberg's condition. (The case $k = 0$ can produce at most one version of $\log \hat{P} \in \underline{\underline{Q}}$—see example 13.)

Example 15. Recall the matrix of example 10,

$$\hat{P} = \frac{1}{3}\begin{bmatrix} 1 + 2x & 1 - x & 1 - x \\ 1 - x & 1 + 2x & 1 - x \\ 1 - x & 1 - x & 1 + 2x \end{bmatrix},$$

with $x = -e^{-2\sqrt{3}\pi}$. This array is reduced to Jordan form by the similarity transformation

$$H = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \frac{1}{2}(-1 + \sqrt{3}i) & \frac{1}{2}(-1 - \sqrt{3}i) \\ 1 & \frac{1}{2}(-1 - \sqrt{3}i) & \frac{1}{2}(-1 + \sqrt{3}i) \end{bmatrix}.$$

Our problem is to indicate how a random sampling scheme of the type mentioned above could give some insight into the variety of branches of log $\hat{P}$ which are in $\underline{\underline{Q}}$. To illustrate the ideas, we restrict our consideration to the subset of matrices B of the form

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \beta & \alpha \\ 0 & \alpha & \beta \end{bmatrix},$$

where α and β are arbitrary *real* numbers.

A computing strategy designed to identify matrices B yielding branches of log $\hat{P} \in \underline{\underline{Q}}$ would begin by generating uniformly distributed (α, β) values within a square centered at the origin—the boundaries $|\alpha| = |\beta| = 10$ are chosen here for illustrative purposes. Each generated value represents a point on the α–β plane for which $\log \hat{P} = HB \log J B^{-1}H^{-1}$ is to be computed. If the resulting matrix is in $\underline{\underline{Q}}$, a "+" is recorded at the point; otherwise, a dot is recorded. In the present example, the flared pattern shown in figure 11, known as the "Iron Cross of the Red Baron (2d class)," would result.

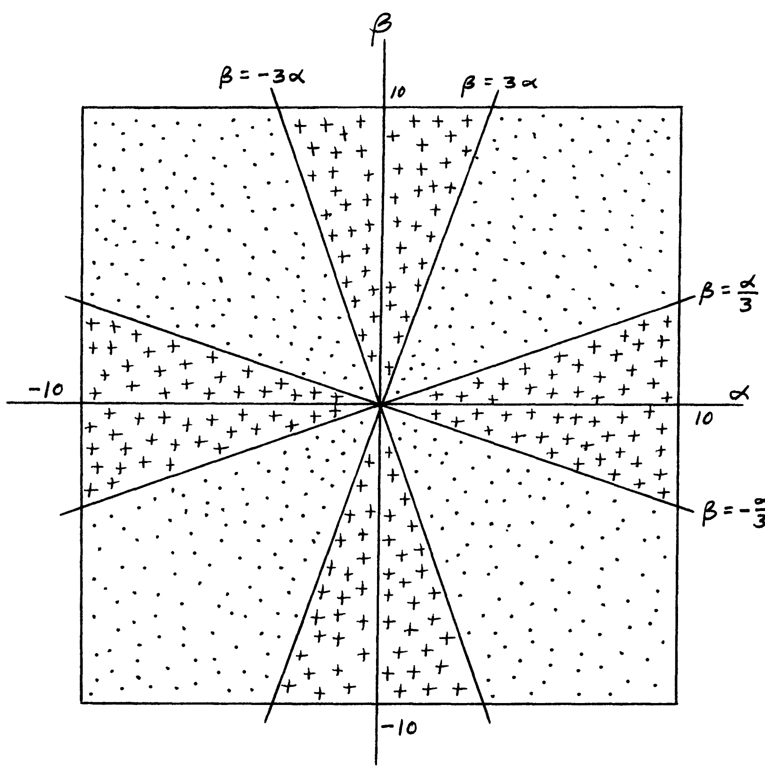

FIG. 11.—Restrictions on matrix B to generate branches of log $\hat{P} \in \underline{\underline{Q}}$. The (α, β) values for which log $\hat{P} \in \underline{\underline{Q}}$ are indicated by the symbol "+." The contours of constant values of log $\hat{P}$ are the straight lines $\beta = \left(\frac{1+x}{1-x}\right)\alpha$, where $1/2 \leqslant |x| \leqslant 2$.

The restrictions on α and β can be summarized by the inequality $\frac{1}{2} \leqslant \left|\frac{\beta - \alpha}{\beta + \alpha}\right| \leqslant 2$. What is more to the point, the structure of the continuum is identical to the one reported in equation (3.23). (This may be verified by replacing $\{c_{ij}\}$ in [3.21] with the appropriate α and β values and computing the restrictions [3.22].) In general, however, by limiting $\{c_{ij}\}$ to real values, only a portion of the continuum will be produced.

REFERENCES

Anderson, T. W., and L. A. Goodman. 1957. "Statistical Inference about Markov Chains." *Annals of Mathematical Statistics* 28 (1): 89–109.

Bartholomew, D. J. 1973. *Stochastic Models for Social Processes.* 2d ed. New York: Wiley.

Beshers, J. M., and E. O. Laumann. 1967. "Social Distance: A Network Approach." *American Sociological Review* 32 (April): 225–36.

Billingsley, P. 1961. *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press.

Blumen, I., M. Kogan, and P. J. McCarthy. 1955. *The Industrial Mobility of Labor as a Probability Process.* Cornell Studies of Industrial and Labor Relations, vol. 6. Ithaca, N.Y.: Cornell University Press.

Boudon, R. 1973. *Mathematical Structures of Social Mobility.* New York: American Elsevier.

Chung, K. L. 1967. *Markov Chains with Stationary Transition Probabilities.* Berlin: Springer.

Cipolla, M. 1932. "Sulle matrice espressione analitiche di un'altra." *Rendiconti del Circolo Matematico di Palermo,* no. 56, pp. 144–54.

Cohen, B. 1963. *Conflict and Conformity: A Probability Model and Its Application.* Cambridge, Mass.: M.I.T. Press.

Coleman, James S. 1964*a*. *Introduction to Mathematical Sociology.* New York: Free Press.

———. 1964*b*. *Models of Change and Response Uncertainty.* Englewood Cliffs, N.J.: Prentice-Hall.

———. 1968. "The Mathematical Study of Change." Pp. 428–78 in *Methodology in Social Research,* edited by Hubert M. Blalock and Ann B. Blalock. New York: McGraw-Hill.

———. 1973. *The Mathematics of Collective Action.* Chicago: Aldine-Atherton.

Cuthbert, J. R. 1972. "On Uniqueness of the Logarithm for Markov Semi-Groups." *Journal of the London Mathematical Society* 4 (May): 623–30.

———. 1973. "The Logarithm Function for Finite-State Markov Semi-Groups." *Journal of the London Mathematical Society* 6 (May): 524–32.

Elfving, G. 1937. "Zur Theorie der Markoffschen Ketten." *Acta Social Science Fennicae n.,* series A.2, no. 8, pp. 1–17.

Gantmacher, F. R. 1960. *The Theory of Matrices.* Vol. 1. New York: Chelsea.

Ginsberg, R. 1971. "Semi-Markov Processes and Mobility." *Journal of Mathematical Sociology* 1 (July): 233–63.

Hodge, R. W. 1966. "Occupational Mobility as a Probability Process." *Demography* 3 (1): 19–34.

Johansen, S. 1973. "A Central Limit Theorem for Finite Semi-Groups and Its Application to the Imbedding Problem for Finite-State Markov Chains." *Zeitschrift für Wahrscheinlichkeitstheorie,* no. 26, pp. 171–90.

———. 1974. "Some Results on the Imbedding Problem for Finite Markov Chains." *Journal of the London Mathematical Society* 8 (July): 345–51.

Karpelewitsch, F. I. 1951. "On the Characteristic Roots of a Matrix with Non-negative Elements." *Isvestija,* série mathématique, no. 15, pp. 361–83.

Kingman, J. F. C. 1962. "The Imbedding Problem for Finite Markov Chains." *Zeitschrift für Wahrscheinlichkeitstheorie,* no. 1, pp. 14–24.

Lieberson, S., and G. V. Fuguitt. 1967. "Negro-White Occupational Differences in the Absence of Discrimination." *American Journal of Sociology* 73 (September): 188–200.

Mayer, Thomas F. 1972. "Models of Intra-generational Mobility." Pp. 308–57 in *Sociological Theories in Progress,* edited by Joseph Berger, Morris Zeldich, Jr., and Bo Anderson. New York: Houghton Mifflin.

McDill, Edward L., and James S. Coleman. 1963. "High School Social Status, College Plans, and Interest in Academic Achievement: A Panel Analysis." *American Sociological Review* 28 (December): 905–18.

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 (June): 463–76.

Prais, S. J. 1955. "Measuring Social Mobility." *Journal of the Royal Statistical Society* 118 (ser. A): 55–66.

Rinehart, R. F. 1955. "The Equivalence of Definitions of a Matric Function." *American Mathematical Monthly* 62 (June): 395–414.

Runnenberg, J. Th. 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): 536–41.

Singer, Burton, and Seymour Spilerman. 1974. "Social Mobility Models for Heterogeneous Populations." Pp. 356–401 in *Sociological Methodology, 1973–1974,* edited by Herbert L. Costner. San Francisco: Jossey-Bass.

Spilerman, Seymour. 1972*a*. "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain." *American Sociological Review* 37 (June): 277–94.

———. 1972*b*. "Extensions of the Mover-Stayer Model." *American Journal of Sociology* 78 (November): 599–626.

Sylvester, J. J. 1883. "On the Equation to the Secular Inequalities in the Planetary Theory." *Philosophical Magazine* 16 (5): 267–69.

Wolfgang, Marvin E., Robert M. Figlio, and Thorsten Sellin. 1972. *Delinquency in a Birth Cohort.* Chicago: University of Chicago Press.

Zahl, Samuel. 1955. "A Markov Process Model for Follow-up Studies." *Human Biology* 27 (May): 90–120.