BIOMETRICS 39, 907–919
December 1983

Estimation in Markov Models from Aggregate Data

J. D. Kalbfleisch, J. F. Lawless and W. M. Vollmer

Department of Statistics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

SUMMARY

In this paper, situations in which individuals move through a finite set of states according to a continuous-time Markov process are considered. Only aggregate data are available: these consist of the number of individuals in each state at specified observation times. We develop conditional least squares and approximate maximum-likelihood-estimation procedures for time-homogeneous models, and extend the methods so that they can handle immigration of individuals into the system during observation. Asymptotic covariance estimates are presented, and some problems for future study are noted.

1. Introduction

In many biological and sociological investigations, observations are made as individuals or particles pass through a finite number of states or compartments. Examples in medicine include studies of individuals who transit through a series of healthy and diseased states until death is observed (Berlin, Brodsky and Clifford, 1979; Temkin, 1978; Fix and Neyman, 1951), and investigations of the flow of tracer particles through parts of the body (Carter, Matrone and Mendenhall, 1964; Shah, 1976). In entomology, the life cycles of insects are studied by observing insect populations as they pass through developmental stages (Read and Ashford, 1968; Dempster, 1961). Investigations of social systems and processes also provide many examples (Bartholomew, 1973; Tuma, Hannan and Groeneveld, 1979). Frequently, the observations consist of the numbers of individuals in a population that occupy the various states at specific time points. Such aggregate data are common in areas such as biomedicine (Matis and Hartley, 1971; Kodell and Matis, 1976), animal ecology (Read and Ashford, 1968; Manly, 1974) and the social sciences (Lee, Judge and Zellner, 1970; Bartholomew, 1973). Bartholomew, 1973).

Markov models are used in many of these areas. If separate individuals are observed either continuously or at discrete time points, it is straightforward to write down the likelihood and develop efficient inference procedures (Bartlett, 1955; Basawa and Rao, 1980), but this is not the case for aggregate data. This paper presents statistical methodology for handling aggregate data in terms of continuous-time Markov models. Previous work in this area has been done primarily in the context of compartmental-model analysis (see Matis and Hartley, 1971; Faddy, 1976; Kodell and Matis, 1976). Computational and other difficulties have thus far hindered a satisfactory general treatment: the main problems have been that the distributional structure of aggregate data collected over time makes exact maximum likelihood methods unfeasible, and that the least squares methods so far proposed have been computationally cumbersome. By making use of the essential Markov structure of the data, we are able to present greatly simplified procedures for least squares and approximate maximum likelihood estimation.

Section 2 reviews some properties of continuous-time Markov processes, and §3 gives estimation procedures for time-homogeneous processes. The methods are illustrated by

Key words: Aggregate data; Compartmental models; Conditional least squares; Estimation; Markov processes; Minimum chi square.

application to data from a compartmental model. Section 4 describes simplifications that are sometimes possible when observations are taken at equally spaced time points. The subsequent sections deal briefly with a number of related problems and an additional example.

2. Markov Processes with Finite State Space

Markov processes are extensively discussed in many texts on stochastic processes; see, for example, Cox and Miller (1965) for a thorough discussion of theory, and Bartholomew (1973) and Chiang (1968, 1980) for many examples of applications. In this section we give a number of pertinent results about continuous-time Markov processes with finite state space.

A Markov process $\{X(t)\colon 0 < t < \infty\}$ with state space $\{1, 2, \ldots, k\}$ can be specified in terms of the $k \times k$ transition probability matrices $\mathbf{P}(s, t)$, $0 \leqslant s \leqslant t$, with (i, j)th element

$$P_{ij}(s, t) = \mathrm{pr}\{X(t) = j \mid X(s) = i\}, \quad i, j = 1, \ldots, k.$$

Alternatively, the process can be specified in terms of the instantaneous transition intensities

$$q_{ij}(t) = \lim_{\Delta t \to 0} P_{ij}(t, t + \Delta t)/\Delta t, \quad i \neq j.$$

For convenience, we also define

$$\begin{aligned} q_{ii}(t) &= -\sum_{j \neq i} q_{ij}(t) \\ &= -\lim_{\Delta t \to 0}\{1 - P_{ii}(t, t + \Delta t)\}/\Delta t, \end{aligned} \tag{2.1}$$

and the $k \times k$ transition intensity matrix $\mathbf{Q}(t)$ with (i, j)th element $q_{ij}(t)$.

The time-homogeneous process, where $q_{ij}(t) = q_{ij}$, is a stationary process with $P_{ij}(s, t)$ being a function of $t - s$ only, and we write $\mathbf{P}(t) = \mathbf{P}(0, t)$. In this case, the forward Kolmogorov equations

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{Q},$$

where $\mathbf{Q} = (q_{ij})$, admit the unique solution

$$\begin{aligned} \mathbf{P}(t) &= \exp(\mathbf{Q}t) \\ &= \sum_{r=0}^{\infty} \mathbf{Q}^r t^r / r!, \end{aligned} \tag{2.2}$$

subject to the boundary condition $\mathbf{P}(0) = \mathbf{I}$ (cf. Cox and Miller, 1965, p. 182). As is well-known, $\mathbf{P}(t)$ can be determined from $\mathbf{Q}$ by simple computational techniques. In particular, if $\mathbf{Q}$ has distinct eigenvalues $d_1, d_2, \ldots, d_k$, and $\mathbf{A}$ is a $k \times k$ matrix whose jth column is a right eigenvector corresponding to d_j, then $\mathbf{Q} = \mathbf{A}\mathbf{D}\mathbf{A}^{-1}$, where $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_k)$ is a diagonal matrix with $d_1, \ldots, d_k$ on the main diagonal. Substitution, term by term, in (2.2) then shows that

$$\mathbf{P}(t) = \mathbf{A}\,\mathrm{diag}\{\exp(d_1 t), \ldots, \exp(d_k t)\}\mathbf{A}^{-1}. \tag{2.3}$$

When $\mathbf{Q}$ has repeated eigenvalues, an analogous decomposition of $\mathbf{Q}$ into Jordan canonical form can be used (see Cox and Miller, 1965). From a computational viewpoint, (2.3) is particularly convenient; once $\mathbf{A}$ and $d_1, \ldots, d_k$ are determined, $\mathbf{P}(t)$ can be quickly evaluated for any t.

Suppose now that the time-homogeneous intensities are differentiable functions of a vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$. Thus $q_{ij} = q_{ij}(\boldsymbol{\theta})$ and $\mathbf{Q} = \{q_{ij}(\boldsymbol{\theta})\}$. In what follows, we require the derivatives of the entries of $\mathbf{P}(t)$ with respect to $\theta_1, \ldots, \theta_r$. As is shown in the Appendix, these derivatives can be computed without obtaining explicit expressions for $P_{ij}(t)$ in terms of $\boldsymbol{\theta}$. Provided $\mathbf{Q}$ has distinct eigenvalues $d_1, \ldots, d_k$, the matrix with entries $\partial P_{ij}(t)/\partial\theta_h$ is

$$\frac{\partial \mathbf{P}(t)}{\partial \theta_h} = \mathbf{A}\mathbf{V}_h\mathbf{A}^{-1}, \quad h = 1, \ldots, r, \tag{2.4}$$

where $\mathbf{V}_h$ is a $k \times k$ matrix with (i, j)th element

$$\begin{array}{ll} g_{ij}^{(h)}\{\exp(d_i t) - \exp(d_j t)\}/(d_i - d_j), & i \neq j, \\ g_{ii}^{(h)} t \exp(d_i t), & i = j, \end{array}$$

and $g_{ij}^{(h)}$ is the (i, j)th element in $\mathbf{A}^{-1}(\partial\mathbf{Q}/\partial\theta_h)\mathbf{A}$. The derivatives $\partial\mathbf{Q}/\partial\theta_h$ are usually simple, and calculation of (2.4) is straightforward once $\mathbf{A}$ and $\mathbf{D}$ are obtained. The restriction requiring $\mathbf{Q}$ to have distinct eigenvalues is of little practical importance. Rare cases in which the postulated models have repeated eigenvalues could be handled by developing a similar algorithm based on the Jordan canonical decomposition.

3. Estimation with Aggregate Data

3.1 *Aggregate Data*

Suppose that observations are made on a group of individuals who act independently of one another, with each individual passing through states according to a time-homogeneous Markov process with state space $\{1, 2, \ldots, k\}$ and $k \times k$ transition intensity matrix $\mathbf{Q} = \{q_{ij}(\boldsymbol{\theta})\}$. Suppose further that observations are made at Times $t_0 < t_1 < \cdots < t_m$, and that the data consist only of the total numbers of individuals, $N_j(t_l)$, in State j at Time t_l, $j = 1, \ldots, k$, $l = 0, 1, \ldots, m$. For the present, we suppose that the system is closed so $\sum_{j=1}^{k} N_j(t_l) = N$, $l = 0, 1, \ldots, m$, and we also suppose that the numbers of individuals in States $1, \ldots, k$ at Time t_0 are known. Let $u_l = t_l - t_{l-1}$ for $l = 1, \ldots, m$, and define vectors

$$\begin{aligned} \mathbf{N}_l' &= \{N_1(t_l), \ldots, N_k(t_l)\}, \\ \mathbf{M}_l' &= \{N_1(t_l), \ldots, N_{k-1}(t_l)\} \end{aligned} \tag{3.1}$$

for $l = 0, 1, \ldots, m$. In addition, let $Y_{ij}(l)$ be the number of individuals who occupy State i at Time t_{l-1} and State j at Time t_l. Note that only values of $N_j(t_l)$, and not of $Y_{ij}(l)$, are observed.

The joint distribution of $\mathbf{N}_1, \ldots, \mathbf{N}_m$, given $\mathbf{N}_0$, is easily derived. In particular, given $N_j(t_{l-1})$, $\{Y_{j1}(l), \ldots, Y_{jk}(l)\}$ has a k-class multinomial distribution with parameters $\{N_j(t_{l-1}); P_{j1}(u_l), \ldots, P_{jk}(u_l)\}$, where the matrix $P(u)$ is defined as in (2.2). Since

$$N_j(t_l) = Y_{1j}(l) + \cdots + Y_{kj}(l), \quad j = 1, \ldots, k, \quad l = 1, \ldots, m, \tag{3.2}$$

the conditional distribution of $\mathbf{N}_l$, given $\mathbf{N}_{l-1}$, is a convolution of multinomials. Because the $\mathbf{N}_l$ terms have the Markov property, the joint distribution of $\mathbf{N}_1, \ldots, \mathbf{N}_m$, given $\mathbf{N}_0$, is built up as the product of these conditional distributions thus: $\mathrm{pr}(\mathbf{N}_1 \mid \mathbf{N}_0)\mathrm{pr}(\mathbf{N}_2 \mid \mathbf{N}_1) \ldots \mathrm{pr}(\mathbf{N}_m \mid \mathbf{N}_{m-1})$.

Likelihood construction requires the probability function of the $\mathbf{N}_l$, and this is computationally intractable. However, conditional means and variances (and generating functions) of $\mathbf{N}_l$, given $\mathbf{N}_{l-1}$, are easily obtained, and these can be used to develop least squares or approximate maximum likelihood estimation procedures. From (3.2), it is apparent that

$$\mathrm{E}\{N_j(t_l) \mid \mathbf{N}_{l-1}\} = \sum_{i=1}^{k} N_i(t_{l-1})P_{ij}(u_l), \quad j = 1, \ldots, k,$$

and similar calculations give the conditional variances and covariances. In matrix notation, the results are

$$\mathrm{E}(\mathbf{N}_l \mid \mathbf{N}_{l-1}) = \mathbf{P}'(u_l)\mathbf{N}_{l-1}$$

and

$$\begin{aligned}\text{cov}(\mathbf{N}_l \mid \mathbf{N}_{l-1}) &= \boldsymbol{\Sigma}_l \\ &= \text{diag}\{\mathbf{P}'(u_l)\mathbf{N}_{l-1}\} - \mathbf{P}'(u_l)\text{diag}(\mathbf{N}_{l-1})\mathbf{P}(u_l),\end{aligned}$$

where, for a vector $\mathbf{x}' = (x_1, \ldots, x_s)$, $\text{diag}(\mathbf{x}) = \text{diag}(x_1, x_2, \ldots, x_s)$. Further, the conditional distribution of $\mathbf{N}_l$, given $\mathbf{N}_{l-1}$, is approximately (singular) k-variate normal. In terms of an equivalent nonsingular distribution, given $\mathbf{N}_{l-1}$, $\mathbf{M}_l$ is approximately $(k-1)$-variate normal with

$$E(\mathbf{M}_l \mid \mathbf{N}_{l-1}) = \mathbf{P}_1'(u_l)\mathbf{N}_{l-1} \tag{3.3}$$

and

$$\begin{aligned}\text{cov}(\mathbf{M}_l \mid \mathbf{N}_{l-1}) &= \boldsymbol{\Sigma}_{1l} \\ &= \text{diag}\{\mathbf{P}_1'(u_l)\mathbf{N}_{l-1}\} - \mathbf{P}_1'(u_l)\text{diag}(\mathbf{N}_{l-1})\mathbf{P}_1(u_l),\end{aligned} \tag{3.4}$$

where $\mathbf{P}_1(u_l)$ is the $k \times (k-1)$ matrix obtained by deleting the last column of $\mathbf{P}(u_l)$, and $\boldsymbol{\Sigma}_{1l}$ is the $(k-1) \times (k-1)$ principal submatrix of $\boldsymbol{\Sigma}_l$.

3.2 *Conditional Least Squares Estimation and Approximate Maximum Likelihood Estimation*

We shall now present several methods of estimation, making use of (3.3) and (3.4). First, conditional least squares estimates of the parameter $\boldsymbol{\theta}$ in $\{q_{ij}(\boldsymbol{\theta})\}$ may be obtained by minimizing

$$\mathbf{S}_1 = \sum_{l=1}^{m} \{\mathbf{M}_l - \mathbf{P}_1'(u_l)\mathbf{N}_{l-1}\}'\{\mathbf{M}_l - \mathbf{P}_1'(u_l)\mathbf{N}_{l-1}\}. \tag{3.5}$$

Some improvement in the estimates may be obtained by considering instead a weighted least squares criterion,

$$\mathbf{S}_2 = \sum_{l=1}^{m} \{\mathbf{M}_l - \mathbf{P}_1'(u_l)\mathbf{N}_{l-1}\}\, \boldsymbol{\Sigma}_{1l}^{-1}\, \{\mathbf{M}_l - \mathbf{P}_1'(u_l)\mathbf{N}_{l-1}\}. \tag{3.6}$$

A third possibility arises through consideration of the approximate log likelihood of $\boldsymbol{\theta}$, obtained from the normal approximation to the distribution of $\mathbf{M}_l$, given $\mathbf{N}_{l-1}$. Since $\mathbf{N}_1, \ldots, \mathbf{N}_m$ is a Markov process, the approximate negative log likelihood is proportional to

$$\mathbf{S}_3 = \sum_{l=1}^{m} \log \det(\boldsymbol{\Sigma}_{1l}) + \mathbf{S}_2, \tag{3.7}$$

which can be minimized with respect to $\boldsymbol{\theta}$.

Direct minimization of any of these equations could be accomplished by utilizing a search or other minimization procedure that requires only functional evaluations. For specified $\boldsymbol{\theta}$, $\mathbf{P}(u_l)$ can virtually always be computed from (2.3). The restriction that $\mathbf{Q} = \{q_{ij}(\boldsymbol{\theta})\}$ has distinct eigenvalues in (2.3) is of little practical import since for most models $\mathbf{Q}$ will have distinct eigenvalues at almost all $\boldsymbol{\theta}$ values.

Quasi-Newton procedures tend, however, to be computationally more efficient and automatically provide variance estimates. A Gauss–Newton procedure for obtaining ordinary or weighted conditional least squares estimates can be implemented by utilizing (2.4) for $\partial\mathbf{P}(t)/\partial\theta_j$. Let $\boldsymbol{\theta}_0$ be a trial value of $\boldsymbol{\theta}$. Consider $\mathbf{S}_2$ with the $\boldsymbol{\Sigma}_{1l}$ treated as fixed ($\mathbf{S}_1$ is a special case of this), and expand the other terms of (3.6) to first order about $\boldsymbol{\theta}_0$. This leads to the term $\mathbf{M}_l - \mathbf{P}_1'(u_l)\mathbf{N}_{l-1}$ in (3.6) being replaced with $\mathbf{Z}_l - C_l(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, where

$$\mathbf{Z}_l = \mathbf{M}_l - \mathbf{P}_1'(u_l)\mathbf{N}_{l-1}\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{3.8}$$

and $C_l(\boldsymbol{\theta})$ is the $(k-1) \times r$ matrix whose jth column contains the vector

$(\partial/\partial\theta_j)\mathbf{P}_1'(u_{l-1})\,\mathbf{N}_{l-1}$. The resulting approximation to $\mathbf{S}_2$ is minimized at

$$\boldsymbol{\theta}_u = \boldsymbol{\theta}_0 + \left\{\sum_{l=1}^{m} C_l'(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_{1l}^{-1} C_l(\boldsymbol{\theta}_0)\right\}^{-1}\left\{\sum_{l=1}^{m} C_l'(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_{1l}^{-1} \mathbf{Z}_l\right\}, \tag{3.9}$$

where $\boldsymbol{\Sigma}_{1l}$ is evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. This provides an updated estimate of $\boldsymbol{\theta}$, and we now repeat the procedure with $\boldsymbol{\theta}_u$ replacing $\boldsymbol{\theta}_0$. Upon convergence the algorithm provides an estimate, $\tilde{\boldsymbol{\theta}}_2$, of $\boldsymbol{\theta}$. An asymptotic covariance matrix estimate for $\tilde{\boldsymbol{\theta}}_2$ is

$$\left\{\sum_{l=1}^{m} C_l'(\tilde{\boldsymbol{\theta}}_2)\boldsymbol{\Sigma}_{1l}^{-1}(\tilde{\boldsymbol{\theta}}_2) C_l(\tilde{\boldsymbol{\theta}}_2)\right\}^{-1}, \tag{3.10}$$

which is a byproduct of the calculations. The ordinary conditional least squares estimate, $\tilde{\boldsymbol{\theta}}_1$, that minimizes $\mathbf{S}_1$ is obtained from this algorithm by replacing $\boldsymbol{\Sigma}_{1l}$ with the $(k-1)\times(k-1)$ identity matrix in (3.9). An asymptotic covariance matrix estimate for $\tilde{\boldsymbol{\theta}}_1$ is

$$\left\{\sum_{l=1}^{m} C_l'(\tilde{\boldsymbol{\theta}}_1) C_l(\tilde{\boldsymbol{\theta}}_1)\right\}^{-1}\left\{\sum_{l=1}^{m} C_l'(\tilde{\boldsymbol{\theta}}_1)\boldsymbol{\Sigma}_{1l}(\tilde{\boldsymbol{\theta}}_1) C_l(\tilde{\boldsymbol{\theta}}_1)\right\}\left\{\sum_{l=1}^{m} C_l'(\tilde{\boldsymbol{\theta}}_1) C_l(\tilde{\boldsymbol{\theta}}_1)\right\}^{-1}. \tag{3.11}$$

The computation of $\tilde{\boldsymbol{\theta}}_1$ or $\tilde{\boldsymbol{\theta}}_2$ by the Gauss–Newton procedure embodied in (3.9) is straightforward and easily programmed. Provided that $\mathbf{Q}$ has distinct eigenvalues at almost all $\boldsymbol{\theta}$ values, we can proceed as in the following way. Given the trial value $\boldsymbol{\theta}_0$, obtain the eigenvalues $d_1, \ldots, d_k$ and the diagonalizing eigenvector matrix $\mathbf{A}$ for $\mathbf{Q}(\boldsymbol{\theta}_0)$. Matrices $\mathbf{P}(u_l)$, and thus $\mathbf{P}_1(u_l)$, can now be easily computed via (2.3). In addition, for each θ_h, $h = 1, \ldots, r$, compute $\partial\mathbf{P}(t)/\partial\theta_h$ via (2.4), and obtain the matrices $C_l(\boldsymbol{\theta}_0)$ required by (3.9). This yields an updated estimate of $\boldsymbol{\theta}$, and the process is repeated until convergence. Note that with this approach it is nowhere necessary to have explicit algebraic expressions for $\mathbf{P}(t)$, or its derivatives, in terms of $\theta_1, \ldots, \theta_r$. This is important, since with most models it is not feasible to develop or work with such expressions. In addition, this approach allows for the development of general computer programs to handle data from an arbitrary model.

It should be mentioned that although the above algorithm yields the estimate, $\tilde{\boldsymbol{\theta}}_1$, that minimizes $\mathbf{S}_1$ when $\boldsymbol{\Sigma}_{1l} = \mathbf{I}$ is used in (3.9), the estimate $\tilde{\boldsymbol{\theta}}_2$ obtained from the general version of (3.9) does not minimize $\mathbf{S}_2$. In situations where the number of individuals in the system is large, reasonable parameter estimates can also be obtained by minimizing $\mathbf{S}_2$, though it is preferable to minimize the approximate negative log likelihood $\mathbf{S}_3$ given by (3.7). We shall denote the estimate obtained by minimizing $\mathbf{S}_3$ as $\tilde{\boldsymbol{\theta}}_3$, and that obtained by minimizing $\mathbf{S}_2$ as $\tilde{\boldsymbol{\theta}}_4$. These can be found by quasi-Newton procedures or, more simply, by an optimization procedure that does not require the calculation of exact first derivatives.

The determination of a good starting value, $\boldsymbol{\theta}_0$, is, in general, difficult. A suitable choice can usually be obtained by calculation of $\mathbf{S}_2$ over a grid of $\boldsymbol{\theta}$ values. In some special cases, other methods may be available. See, for example, the discussion in §4.

Before presenting an example, we note that least squares estimation has been investigated for continuous-time Markov models with aggregate data by several authors (for example, Matis and Hartley, 1971; Kodell and Matis, 1976), and also for discrete Markov chain models (Lee *et al.*, 1970). Interestingly, the simple structure embodied in (3.3), (3.4), and (3.5) or (3.6) has not been used in the previous studies. Instead, the joint distribution of $(\mathbf{N}_1, \ldots, \mathbf{N}_m)$ has been studied without utilizing the conditional (Markov) structure of the $\mathbf{N}_l$. This has led to enormous computational problems as, for example, in the paper by Kodell and Matis (1976) where a Gauss–Newton procedure is suggested that requires inversion of matrices of dimension $m(k-1)\times m(k-1)$. The earlier procedures, in addition, do not lend themselves to the development of computer packages to handle arbitrary models. It should also be noted that Kodell and Matis (1976) discussed somewhat different estimates than those given here. Their ordinary least squares estimate does not minimize $\mathbf{S}_1$ as does $\tilde{\boldsymbol{\theta}}_1$, but rather minimizes $\mathbf{S}_1$ with $\mathbf{P}_1'(u_l)\mathbf{N}_{l-1}$ replaced by the unconditional mean of $\mathbf{M}_l$. They did not consider the approximate maximum likelihood estimate $\tilde{\boldsymbol{\theta}}_3$, nor the Gauss–Newton estimate $\tilde{\boldsymbol{\theta}}_2$. Finally, their weighted least squares estimate is based on a quadratic form that uses the unconditional mean and covariance structure of $\mathbf{M}_1, \ldots, \mathbf{M}_m$.

3.3 *An Example*

We consider what is sometimes referred to as a two-compartment open model. The model has three states, with transition intensity matrix of the form

$$\mathbf{Q} = \begin{bmatrix} -\lambda_1 - \mu_1 & \lambda_1 & \mu_1 \\ \lambda_2 & -\lambda_2 - \mu_2 & \mu_2 \\ 0 & 0 & 0 \end{bmatrix}.$$

The third state is absorbing. In compartmental-model work, States 1 and 2 represent the two compartments of a system, and State 3 represents the exterior of the system. This model is much discussed in biomedical work (see, for example, Kodell and Matis, 1976) and in other areas such as sociology (see, for example, Tuma *et al.*, 1979).

For this model, explicit forms for $\mathbf{P}(t)$ are easily obtained. The eigenvalues of $\mathbf{Q}$ are $d_1 = 0$ and

$$d_2, d_3 = -\tfrac{1}{2}(\theta_1 + \theta_2) \pm \tfrac{1}{2}\{(\theta_1 - \theta_2)^2 + 4\lambda_1\lambda_2\}^{\frac{1}{2}},$$

where $\theta_1 = \lambda_1 + \mu_1$ and $\theta_2 = \lambda_2 + \mu_2$. Thus, from (2.3),

$$\mathbf{P}(t) = \frac{1}{d_2 - d_1}\begin{bmatrix} (\theta_1 + d_2)\exp(d_1 t) - (\theta_1 + d_1)\exp(d_2 t) & -\lambda_1\{exp(d_1 t) - \exp(d_2 t)\} & p_{13}(t) \\ -\lambda_2\{\exp(d_1 t) - \exp(d_2 t)\} & -(\theta_1 + d_1)\exp(d_1 t) + (\theta_1 + d_2)\exp(d_2 t) & p_{23}(t) \\ 0 & 0 & 1 \end{bmatrix}, \quad (3.12)$$

where $p_{13}(t)$ and $p_{23}(t)$ are obtained by using the fact that $\mathbf{P}(t)$ is stochastic. Estimation procedures that require only calculation of $\mathbf{P}(t)$ can thus be implemented directly by means of (3.12) or, more simply, (2.3). The Gauss–Newton procedures, on the other hand, also require first derivatives of the $P_{ij}(t)$ with respect to θ_1, θ_2, λ_1 and λ_2. Even for this simple model, these are most conveniently calculated via (2.4) since explicit differentiation of the $P_{ij}(t)$ in (3.12) is tedious.

As an illustration, we consider some data given by Kodell and Matis (1976) and reproduced here in Table 1. In Table 2 four sets of estimates are shown: (i) ordinary least squares estimates, $\tilde{\boldsymbol{\theta}}_1$, obtained by minimizing (3.5); (ii) estimates, $\tilde{\boldsymbol{\theta}}_2$, obtained by the Gauss–Newton iteration procedure based on (3.9); (iii) the approximate MLE, $\tilde{\boldsymbol{\theta}}_3$, obtained by minimizing (3.7); (iv) the estimate, $\tilde{\boldsymbol{\theta}}_4$, which minimizes $\mathbf{S}_2$ of (3.6). Asymptotic variance estimates for $\tilde{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_2$ can be obtained from (3.11) and (3.10), respectively. For example, estimated standard deviations for $\tilde{\lambda}_1$, $\tilde{\mu}_1$, $\tilde{\lambda}_2$ and $\tilde{\mu}_2$ from $\tilde{\boldsymbol{\theta}}_2$ are .047, .042, .104 and .085, respectively. When the

Table 1
Aggregate data from a two-compartment model

t_i	$N_1(t_i)$	$N_2(t_i)$	t_i	$N_1(t_i)$	$N_2(t_i)$	t_i	$N_1(t_i)$	$N_2(t_i)$
0	1000	0	1.75	217	156	3.50	58	76
0.25	772	103	2.00	183	152	3.75	45	70
0.50	606	169	2.25	159	126	4.00	42	58
0.75	477	191	2.50	142	107	4.25	36	49
1.00	386	198	2.75	124	98	4.50	35	39
1.25	317	181	3.00	106	79	4.75	26	27
1.50	278	162	3.25	78	80	5.00	21	24

Table 2
Parameter estimates obtained by four methods

Method		$\tilde{\lambda}_1$	$\tilde{\mu}_1$	$\tilde{\lambda}_2$	$\tilde{\mu}_2$
(i)	$\tilde{\theta}_1$	.568	.506	.498	.683
(ii)	$\tilde{\theta}_2$	.555	.492	.409	.731
(iii)	$\tilde{\theta}_3$	.561	.495	.425	.730
(iv)	$\tilde{\theta}_4$	.589	.501	.485	.732

number of individuals under observation is large, as here, (3.10) also provides asymptotic variance estimates for $\tilde{\boldsymbol{\theta}}_3$ and $\tilde{\boldsymbol{\theta}}_4$.

There is generally good agreement among the four estimation procedures. We evaluated $\tilde{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_4$ by both the Gauss–Newton procedure and by general purpose optimization programs not requiring derivatives. The same estimates were obtained by both methods. The weighted unconditional least squares estimates found by Kodell and Matis (1976) are close to $\tilde{\boldsymbol{\theta}}_4$, and the estimated covariance matrices agree closely. The estimates of these authors, however, are much more difficult to obtain.

4. Equally Spaced Observations

When the observations are equally spaced ($u_1 = u_2 = \cdots = u_m = u$), the conditional means $E(\mathbf{M}_l \mid \mathbf{N}_{l-1})$ can all be expressed as linear combinations of the $P_{ij}(u)$. If γ is a vector of parameters corresponding to the nonzero elements of $\mathbf{P}_1(u)$ in some order, then there exist matrices $\mathbf{B}_{l-1}$ with entries from $\mathbf{N}_{l-1}$, such that $\mathbf{P}_1'(u_l)\mathbf{N}_{l-1} = \mathbf{B}_{l-1}\gamma$. The conditional weighted least squares criterion (3.6) can now be rewritten as

$$\mathbf{S}_2 = \sum_{l-1}^{m} (\mathbf{M}_l - \mathbf{B}_{l-1}\gamma)'\boldsymbol{\Sigma}_{1l}^{-1}(\mathbf{M}_l - \mathbf{B}_{l-1}\gamma). \tag{4.1}$$

This raises the possibility of using linear rather than nonlinear least squares.

If the $\boldsymbol{\Sigma}_{1l}$ in (4.1) are known, then the vector γ which minimizes $\mathbf{S}_2$ is

$$\tilde{\gamma} = \left(\sum_{l=1}^{m} \mathbf{B}_{l-1}'\boldsymbol{\Sigma}_{1l}^{-1}\mathbf{B}_{l-1}\right)^{-1}\left(\sum_{l=1}^{m} \mathbf{B}_{l-1}'\boldsymbol{\Sigma}_{1l}^{-1}\mathbf{M}_l\right). \tag{4.2}$$

If $\boldsymbol{\Sigma}_{1l} = \mathbf{I}$, then $\tilde{\gamma}_1$ from (4.2) is the ordinary least squares estimate. In the general case where the $\boldsymbol{\Sigma}_{1l}$ depend on γ, an iterative scheme yields an estimate, $\tilde{\gamma}_2$, of γ that is analogous to $\tilde{\boldsymbol{\theta}}_2$ in §3. The algorithm begins with an initial estimate γ_0, evaluates $\boldsymbol{\Sigma}_{1l}$ at γ_0 and then calculates $\tilde{\gamma}$ from (4.2). The process is then repeated with $\tilde{\gamma}$ replacing γ_0 until convergence is reached. Estimates $\tilde{\gamma}_3$ and $\tilde{\gamma}_4$, which are analogous to $\tilde{\boldsymbol{\theta}}_3$ and $\tilde{\boldsymbol{\theta}}_4$, could also be defined, though for these the advantage of linear least squares calculation is lost.

An estimate, $\tilde{\gamma}$, of γ produced in this way gives an estimate, $\tilde{\mathbf{P}}(u)$, of $\mathbf{P}(u)$. It should be noted, however, that unrestricted use of (4.2) sometimes yields a $\tilde{\mathbf{P}}(u)$ with negative entries. In such cases a restricted least squares procedure has to be used to find the $\tilde{\gamma}$ that gives a stochastic matrix $\mathbf{P}(u)$. For this procedure to produce an estimate of $\boldsymbol{\theta}$, there must also be an intensity matrix $\tilde{\mathbf{Q}} = \mathbf{Q}(\tilde{\boldsymbol{\theta}})$ such that

$$\tilde{\mathbf{P}}(u) = \exp(\tilde{\mathbf{Q}}u). \tag{4.3}$$

A necessary condition for this is that the dimensions of γ and $\boldsymbol{\theta}$ are the same. Even when this holds, however, there are unfortunately no useable necessary and sufficient conditions known that will ensure that (4.3) possesses a solution. This problem is often referred to as the embeddability problem for finite continuous-time Markov chains; Singer and Spilerman (1976) have reviewed this area.

For a k-state process, a unique $\tilde{\mathbf{Q}}$ with real entries satisfying (4.3) always exists when $\tilde{\mathbf{P}}(u)$

has real distinct nonnegative eigenvalues $d_1^*, \ldots, d_k^*$: this matrix is

$$\tilde{\mathbf{Q}} = \left(\frac{1}{u}\right)\mathbf{A}_*\text{diag}(\log d_1^*, \ldots, \log d_k^*)\mathbf{A}_*^{-1}, \tag{4.4}$$

where $\mathbf{A}_*$ is a $k \times k$ matrix whose jth column is the right eigenvector of $\tilde{\mathbf{P}}(u)$ corresponding to d_j^*. There is, however, no guarantee that $\tilde{\mathbf{Q}}$ will be an admissible intensity matrix, although in the case $k = 2$, the condition $d_1^* + d_2^* > 1$ is necessary and sufficient for $\tilde{\mathbf{Q}}$ to be admissible.

In practice a good procedure is to determine $\tilde{\mathbf{P}}(u)$ and its eigenvalues. If these are real distinct and nonnegative, calculate $\tilde{\mathbf{Q}}$ via (4.4). If $\tilde{\mathbf{Q}}$ is admissible, it is the desired estimate of $\mathbf{Q}$. Otherwise, the nonlinear least squares approach of §3 can be used. It should be noted that even if $\mathbf{P}(u)$ does not yield the estimate of $\mathbf{Q}$, it may be useful in determining an initial value, $\boldsymbol{\theta}_0$, for use in the nonlinear least squares algorithms.

In the example in §3.3, the estimates $\tilde{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_2$ can be obtained by finding $\mathbf{P}(u)$ and using (4.4) to obtain $\tilde{\mathbf{Q}}$. Details are omitted in the interest of brevity.

5. Immigration

Our discussion thus far has concerned closed systems in which the total population size, N, remains fixed over time. We now consider the situation in which immigration into the system is allowed. For convenience we suppose that all immigration is into State 1. The situation where immigrants enter the system via two or more states is handled by a straightforward extension of this case. We also assume that, once in the system, immigrants act independently of other individuals in the system and move through the system according to the same probability laws as other individuals already there.

Using the notation of §3.1, we let $Y_{ij}(l)$ denote the number of individuals who occupy State i at Time t_{l-1} and State j at Time t_l, where $i, j = 1, \ldots, k$ and $l = 1, \ldots, m$. In addition, we define $R_j(l)$ to be the number of individuals immigrating into the system during (t_{l-1}, t_l) and who are in State j at Time t_l. If $N_j(t_l)$ is the number of individuals in State j at Time t_l, then for $j = 1, \ldots, k$,

$$N_j(t_l) = \sum_{i=1}^{k} Y_{ij}(l) + R_j(l). \tag{5.1}$$

As before, we consider the conditional distribution of $\mathbf{N}_l = \{N_1(t_l), \ldots, N_k(t_l)\}'$, given $\mathbf{N}_{l-1}$. Let the mean and covariance of $\mathbf{R}_l = \{R_1(l), \ldots, R_k(l)\}'$ be denoted by $\boldsymbol{\mu}_l^{(R)}$ and $\boldsymbol{\Sigma}_l^{(R)}$, respectively. Then from (3.3) and (3.4) and the conditional independence of the $R_j(l)$ and $Y_{ij}(l)$ terms it follows that

$$\text{E}(\mathbf{N}_l \mid \mathbf{N}_{l-1}) = \mathbf{P}'(u_l)\mathbf{N}_{l-1} + \boldsymbol{\mu}_l^{(R)} \tag{5.2}$$

and

$$\text{cov}(\mathbf{N}_l \mid \mathbf{N}_{l-1}) = \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_l^{(R)}.$$

The distribution of $\mathbf{R}_l$, and therefore its moments, will depend upon the process generating the immigration. If this process depends upon parameters not already in the model, these additional parameters can be estimated, along with those specifying the Markov process, by employing least squares or approximate maximum likelihood procedures based on (5.2). In many situations, however, computation of $\boldsymbol{\mu}_l^{(R)}$ and $\boldsymbol{\Sigma}_l^{(R)}$ may not be particularly easy and approximations may have to be considered.

We shall illustrate these remarks by considering two important immigration processes.

Model 1. At a known time $\tau_l \in (t_{l-1}, t_l)$, a known number, r_l, of individuals enter the system.

In this case, $\mathbf{R}_l$ has a multinomial distribution with

$$\boldsymbol{\mu}_l^{(R)} = \mathbf{P}'(t_l - \tau_l)\,\mathbf{i}_l$$

and

$$\boldsymbol{\Sigma}_l^{(R)} = \text{diag}(\boldsymbol{\mu}_l^{(R)}) - \mathbf{P}'(t_l - \tau_l)\text{ diag }(\mathbf{i}_l)\mathbf{P}(t_l - \tau_l),$$

where $\mathbf{i}_l = (r_l, 0, \ldots, 0)'$. In the special case in which $\tau_l = t_{l-1}$, the conditional mean and variance of $\mathbf{N}_{l-1}$ in (5.2) reduce to those given in (3.3) and (3.4), with $\mathbf{N}_{l-1}$ in the right-hand sides of the latter expressions replaced by $\mathbf{N}_{l-1}+\mathbf{i}_l$. Under this model, no additional parameters are introduced.

Model 2. Immigration into State 1 occurs according to a Poisson process with intensity $\alpha(t)$. In this case it is easily shown (see, for example, McLean, 1976) that the $R_j(l)$ are independent Poisson random variables with means and variances given by

$$\int_{t_{l-1}}^{t_l} \alpha(\tau)P_{1j}(t_l - \tau)\, d\tau. \tag{5.3}$$

Consequently, $\boldsymbol{\mu}_l$ and $\boldsymbol{\Sigma}_l^{(R)}$ are, in principal, easily obtained, though computation of (5.3) may be difficult.

6. An Additional Example

Table 3 gives data on the prevalence of the yellow lace birch bug in each of five developmental instars and on recruitment to the adult stage, for a single (small) tree in northern Ontario. The data were collected at intervals of three to 10 days during the summer of 1980. The adult of this species is a flying insect and leaves the tree; the moult of the final instar, however, provides evidence of its emergence. A key feature of the situation is that mortality occurs in the various stages.

Table 3
Prevalence and recruitment data for the yellow lace birch bug

Day (t_l)	$u_l = t_l - t_{l-1}$	State 1	2	3	4	5	Adult (to date)
July 2		31					
July 6	4	200					
July 10	4	411	58				
July 15	5	435	320	97	1		
July 18	3	496	294	250	48		
July 21	3	514	316	299	214	6	
July 24	3	492	339	328	332	79	
July 30	6	509	390	353	325	326	4
August 2	3	478	374	356	369	476	83
August 5	3	359	382	344	404	549	202
August 9	4	270	261	339	446	617	460
August 15	6	142	186	209	400	666	745
August 18	3	103	159	198	329	669	900
August 21	3	63	73	183	237	616	1095
August 25	4	28	40	66	196	451	1394
August 29	4	11	26	41	105	340	1581
September 8	10	0	1	6	26	97	1826

Let $\mathbf{N}_l = \{N_1(t_l), \ldots, N_6(t_l)\}'$ be the vector of aggregate totals for the instar and adult states on Day t_l. On a microscopic level, a semi-Markov model for the developmental process of an individual insect seems most reasonable; if, however, the age distribution in the various instars is fairly stable over time, the process $\mathbf{N}_l$ should have macroscopic properties approximating those of a Markov process. We therefore examined the utility of a homogeneous Markov model in studying the system.

Consider a homogeneous Markov model for States $1, \ldots, 7$, where States $1, \ldots, 5$ represent the five nymphal instars, 6 is the adult state, and 7 corresponds to death. Let μ_j represent the mortality rate from State j, $j = 1, \ldots, 6$, and let λ_j represent the intensity of transitions from State j to State $j+1$, $j = 1, 2, \ldots, 5$. Since the data contain no information on adult mortality, the adult state is treated as absorbing, with $\mu_6 = 0$. In the analysis considered here we also assume a constant mortality rate $\mu = \mu_i$, $i = 1, \ldots, 5$, over the five nymphal instars. In the full model with λ_i and μ_i left free, the estimates of λ_i and μ_i are highly correlated, which indicates that these separate parameters are nearly nonidentifiable; this same problem is seen in other estimation methods based on moments in a semi-Markov model.

The Markov model fitted thus has the 7×7 transition matrix $\mathbf{Q} = (q_{ij})$ with entries $q_{j,j+1} = \lambda_j$ and $q_{j7} = \mu_j$, $j = 1, \ldots, 5$, and $q_{jj} = -\lambda_j - \mu_j$ and $q_{ij} = 0$ otherwise. For estimation, we consider the distribution of $N_2(t_l), \ldots, N_6(t_l)$, given $N_1(t_{l-1}), \ldots, N_6(t_{l-1})$. In this situation there is immigration present since new individuals are constantly recruited into the first instar (State 1). We handle this by assuming that all immigration into State 1 during (t_{l-1}, t_l) occurs at t_l. There is a slight problem in that insects may actually enter State 1 and transfer out to States 2 or 7, all during (t_{l-1}, t_l), but we are assuming that this cannot occur. This may create a slight underestimation of λ_1.

The methods of §3 can be applied, with minor modification. Let $\mathbf{P}(t) = \exp(\mathbf{Q}t)$ as before and let $\mathbf{P}_1(t)$ be the 6×5 submatrix composed of Rows 1 through 6 and Columns 2 through 6 of $\mathbf{P}(t)$. The weighted least squares criterion is then given by (3.6) with $\mathbf{\Sigma}_{1l}$ given by (3.4).

A Gauss–Newton algorithm that utilizes the computational formulae for $\mathbf{P}(t)$ and its derivatives was applied and Table 4 presents the estimates obtained. The final column gives the estimated mean live sojourn time, $(\hat{\mu} + \hat{\lambda}_j)^{-1}$, in the j'th instar, $j = 1, \ldots, 5$. These estimates turn out to be in quite good agreement with estimates obtained by an analysis based on a semi-Markov model.

An examination of residuals $\mathbf{M}_l - \tilde{\mathbf{P}}_1'(\mu_l)\mathbf{N}_{l-1}$ reveals some lack of fit of the Markov model to these data. This may be due, in part, to the homogeneity assumption; sojourn times in instars are known to depend on temperature, and this analysis makes no attempt to account for this. In addition, the adequacy of the Markov model depends on stable age distributions in the instars, but there may not be stability outside the central part of the data.

7. Discussion

It is possible to extend the methods of this paper to certain types of time-dependent Markov processes. Essentially, immediate extension is possible to any nonhomogeneous model that

Table 4
Results of Markov analysis of the data in Table 3 using a Gauss–Newton algorithm

Instar	λ_i	μ	Mean sojourn time (days)
1	.288	.016	3.29
2	.200	.016	4.63
3	.213	.016	4.37
4	.164	.016	5.54
5	.091	.016	9.36

facilitates computation of $\mathbf{P}(s, t)$ in convenient form. Such models include those for which an operational time exists, so that the process can be made homogeneous by a monotone transformation on the time scale. In addition, processes in which $\mathbf{Q}(t)$ is constant over intervals (t_0, t_1), (t_1, t_2), ... can be handled easily.

We have not carried out an extensive study of the asymptotic properties of the estimators proposed here. It should be noted, however, that two different limits are conceivably of interest, depending upon the application. On the one hand, a closed system may be observed at a fixed number, m, of time points $t_1, t_2, \ldots, t_m$; then the limit arises as the number of individuals $N \rightarrow \infty$. On the other hand, if the system is ergodic, we may follow a fixed population of N individuals over a large number of time points and consider the limit as $m \rightarrow \infty$.

The case $N \rightarrow \infty$ is more appropriate for the examples of §§3 and 6, and simpler from a theoretical standpoint. It is easily shown that, as $N \rightarrow \infty$, $\mathbf{S}_2$ is asymptotically chi square and estimator $\tilde{\boldsymbol{\theta}}_4$ obtained by minimizing $\mathbf{S}_2$ is a minimum chi square estimator. Properties of such estimators have been considered by Chiang (1956), Wijsman (1959) and others. With only very mild conditions, their results show that $\tilde{\boldsymbol{\theta}}_2$ and $\tilde{\boldsymbol{\theta}}_4$ are consistent and asymptotically normal.

When a small number of individuals are observed at many points in time, the second limit (i.e. $m \rightarrow \infty$) may be more appropriate. The results of Klimko and Nelson (1978) and Nelson (1980) are relevant in this case. These workers have discussed the consistency of conditional ordinary and weighted least squares in a univariate setting. Extensions of their work would be necessary in a thorough study of the estimators proposed in this paper.

Problems related to the estimability of, and information about, various parameters in these models also deserve further study, as do other finite-sample properties of proposed estimators. For example, if observations are made on a system which is fluctuating about a steady equilibrium state, then precise estimation of those parameters characterizing the equilibrium should be possible, but much less information will be available about parameters describing other aspects of the model. The lengths of the intervals between observation times may also substantially influence precision of estimation.

Finally, we note that, in certain instances, $\mathbf{N}_0$ may be known to have the multinomial distribution associated with the equilibrium probabilities. If $\mathbf{M}_0 = \{N_1(t_0), \ldots, N_{k-1}(t_0)\}$ were known to have expectation $\boldsymbol{\eta}(\boldsymbol{\theta})$ and covariance matrix $\boldsymbol{\Sigma}_1(\boldsymbol{\theta})$, an additional term,

$$\{\mathbf{N}_0 - \boldsymbol{\eta}(\boldsymbol{\theta})\}' \boldsymbol{\Sigma}_1(\boldsymbol{\theta})^{-1}\{\mathbf{N}_0 - \boldsymbol{\eta}\ \boldsymbol{\theta}\ )\},$$

could be incorporated into $\mathbf{S}_2$ and more efficient estimators determined.

Acknowledgements

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada. We are grateful to the referees of an earlier draft of the paper for helpful comments and queries that led to an improved exposition. We also wish to thank Dr Steve Smith and Ms Marjorie Henderson of the Department of Biology, University of Waterloo, for the data used in §6.

Résumé

Cet article considère des situations dans lesquelles des individus se dèplacent dans un ensemble fini d'états selon un processus de Markov à temps continu. Seules les données aggrégées sont disponibles: elles consistent en nombre d'individus dans chaque état, en des temps d'observation spécifiés. Nous développons des procédures d'estimation de moindres carrés conditionnels et de maximum de vraisemblance approché, pour des modèles homogènes en fonction du temps, et nous étendons les méthodes pour prendre en compte l'entrée de nouveaux individus dans le système pendant l'obsevation. On présente des estimateurs pour les covariances asymptotiques, et quelques problèmes qui feront l'objet d'études ultérieures.

References

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes.* Cambridge: Cambridge University Press.

Bartholomew, D. J. (1973). *Stochastic Models for Social Processes*, 2nd ed. London: Wiley.

Basawa, I. V. and Rao, L. S. P. (1980). *Statistical Inference for Stochastic Processes.* New York: Academic Press.

Berlin, B., Brodsky, J. and Clifford, P. (1979). Testing disease dependence in survival experiments with serial sacrifice. *Journal of the American Statistical Association* **74,** 5–14.

Carter, M. W., Matrone, G. and Mendenhall, W. (1964). Estimation of the life span of red blood cells. *Journal of General Physiology* **47,** 851–858.

Chiang, C. L. (1956). On regular best asymptotically normal estimates. *Annals of Mathematical Statistics* **27,** 336–351.

Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics.* New York: Wiley.

Chiang, C. L. (1980). *An Introduction to Stochastic Processes and their Applications.* New York: Krieger.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes.* London: Methuen.

Dempster, J. D. (1961). The analysis of data obtained by regular sampling of an insect population. *Journal of Animal Ecology* **30,** 429–432.

Faddy, M. J. (1976). A note on the general time-dependent stochastic compartmental model. *Biometrics* **32,** 443–448.

Fix, E. and Neyman, J. (1951). A simple stochastic model of recovery, relapse, death, and loss of patients. *Human Biology* **23,** 205–241.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Annals of Mathematical Statistics* **6,** 629–642.

Kodell, R. L. and Matis, J. H. (1976). Estimating the rate constants in a two-compartmental stochastic model. *Biometrics* **32,** 377–400.

Lee, T. C., Judge, G. C. and Zellner, A. (1970). *Estimating the Parameters of the Markov Model from Aggregate Time Series Data.* Amsterdam: North-Holland.

Manly, B. F. J. (1974). Estimation of stage specific survival rates and other parameters for insect populations developing through several stages. *Oecologia* **15,** 277–285.

Matis, J. H. and Hartley, H. O. (1971). Stochastic compartment analysis: model and least squares estimation from time series data. *Biometrics* **27,** 77–102.

McClean, S. I. (1976). A continuous time population model with Poisson recruitment. *Journal of Applied Probability* **13,** 348–354.

Nelson, P. I. (1980). A note on the strong consistency of least squares estimators in regression models with martingale difference errors. *Annals of Statistics* **7,** 1057–1064.

Read, K. L. Q. and Ashford, J. R. (1968). A system of models for the life cycle of a biological organism. *Biometrika* **55,** 211–221.

Shah, B. K. (1976). Data analysis problems in the area of pharmacokinetics research. *Biometrics* **32,** 145–157.

Singer, B. and Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology* **82,** 1–54.

Temkin, N. R. (1978). An analysis for transient states with application for tumor shrinkage. *Biometrics* **34,** 571–580.

Tuma, N. B., Hannan, M. T. and Groeneveld, L. P. (1979). Dynamic analysis of event histories. *American Journal of Sociology* **84,** 820–854.

Wijsman, R. A. (1959). On the theory of BAN estimates. *Annals of Mathematical Statistics* **30,** 185–191. Correction: (1959) *Annals of Mathematical Statistics* **30,** 1268–1270.

Received December 1981; *revised June* 1982

APPENDIX

Suppose that $\mathbf{Q}$ has distinct eigenvalues $d_1, \ldots, d_k$ for all $\boldsymbol{\theta}$ in some open set. The matrix obtained by differentiating each entry in $\mathbf{P}(t)$ with respect to θ_h is, from (2.2) and the fact that $\mathbf{Q} = \mathbf{ADA}^{-1}$,

$$\frac{\partial \mathbf{P}(t)}{\partial \theta_h} = \sum_{s=1}^{\infty} \frac{\partial}{\partial \theta_h}\left(\frac{\mathbf{Q}^s t^s}{s!}\right)$$

$$= \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} \mathbf{Q}^l \frac{\partial \mathbf{Q}}{\partial \theta_h} \mathbf{Q}^{s-1-l} \frac{t^s}{s!}$$

$$= \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} \mathbf{AD}^l \mathbf{G}_h \mathbf{D}^{s-1-l} \mathbf{A}^{-1} \frac{t^s}{s!},$$

where $\mathbf{G}_h = \mathbf{A}^{-1} \dfrac{\partial \mathbf{Q}}{\partial \theta_h} \mathbf{A}$.

Continuing, we find that

$$\frac{\partial \mathbf{P}(t)}{\partial \theta_h} = \mathbf{A}\left(\sum_{s=1}^{\infty} \sum_{l=0}^{s-1} \mathbf{D}^l \mathbf{G}_h \mathbf{D}^{s-1-l} \frac{t^s}{s!}\right)\mathbf{A}^{-1}$$

$$= \mathbf{AV}_h\mathbf{A}^{-1},$$

where $\mathbf{V}_h$ is a $k \times k$ matrix with (i, j)th element

$$g_{ij}^{(h)} \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} d_i^{s-1-l} d_j^l \frac{t^s}{s!} = g_{ij}^{(h)} \frac{\{\exp(d_i t) - \exp(d_j t)\}}{d_i - d_j}, \quad i \neq j,$$

$$g_{ii}^{(h)} \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} d_i^{s-1} \frac{t^s}{s!} = g_{ii}^{(h)} t \exp(d_i t), \quad i = j,$$

where $g_{ij}^{(h)}$ is the (i, j)th element in $\mathbf{G}_h$. This establishes (2.4).