THE ESTIMATION OF A FUNDAMENTAL INTERACTION PARAMETER IN AN EMIGRATION-IMMIGRATION PROCESS

By Harold Ruben

Columbia University[1] *and Sheffield University*

0. Synopsis. A convenient method for estimating the infinitesimal transition parameters of a continuous time, multivariate, Markovian emigration-immigration process $\{\mathbf{n}(t)\}$ is described when these can be expressed as known functions of a single fundamental interaction (or migration) parameter θ. The estimator for θ is constructed from the observed generalized mean square consecutive fluctuation in a realised sequence of $\{\mathbf{n}(t)\}$ at a finite number, k, of points in time: $t = 0, \tau, 2\tau, \cdots, (k-1)\tau$. Formulae are derived for the large-sample variance (k large) of the estimator, and its relative efficiency in such samples is investigated.

1. Introduction and summary. Consider the multivariate emigration-immigration or Poisson-Markov process $\{\mathbf{n}(t)\}$ in continuous time with state space the set of all vectors in Euclidean m-space having non-negative integral components. Some properties of this process were derived in a previous paper (Ruben, 1962) with special emphasis on the joint distribution of $\mathbf{n}(t_1), \mathbf{n}(t_2), \cdots, \mathbf{n}(t_k)$ from the point of view of generating functions and on the moments and product moments of the process. We recall in particular that the process is strictly stationary and Markovian and is completely described in terms of $\boldsymbol{\nu}$ and $\boldsymbol{\Lambda}$, where $\boldsymbol{\nu}$ is the process mean, $\boldsymbol{\nu} = E\mathbf{n}(t)$, and $\boldsymbol{\Lambda}$ is an interaction-rate matrix defined in terms of infinitesimal transition parameters $\lambda_{rs}, \lambda_r^*$ $(r, s = 1, 2, \cdots, m; r \neq s)$. Further, the covariance matrix of the process is $\mathbf{NP}(t)$ for $t \geqq 0$, where $\mathbf{N}$ is the diagonal matrix with diagonal elements $\nu_1, \cdots, \nu_m$ (the components of $\boldsymbol{\nu}$) and[2]

$$\mathbf{P}(t) = \boldsymbol{\Theta}^{-1}\mathbf{K}(t)\boldsymbol{\Theta}. \tag{1.1}$$

Here $\boldsymbol{\Theta}$ is the matrix of row eigenvectors of $\boldsymbol{\Lambda}$ ($\boldsymbol{\Theta}$ reduces $\boldsymbol{\Lambda}$ to diagonal canonical form) and $\mathbf{K}$ is the diagonal matrix with diagonal elements $\exp(-\kappa_1 t), \cdots, \exp(-\kappa_m t)$, κ_r denoting the (real positive) eigenvalues of $\boldsymbol{\Lambda}$.

The purpose of the present paper is to derive a convenient method of estimation for the infinitesimal transition parameters, and therefore for the transition probabilities,[3] of $\{\mathbf{n}(t)\}$ when these can be related in a known manner to a single unknown fluctuation parameter θ, that is, when the given stochastic process is a member of a singly-infinite class of vector Poisson-Markov processes.

Received March 21, 1961; revised July 16, 1962.

[1] Research sponsored in part under Contract AF 41(657)-214, Teachers College, Columbia University, 1959–60.

[2] Note incidentally that (1.1) implies $\mathbf{P}(t) = \exp(-\boldsymbol{\Lambda}t)$ for $t \geqq 0$.

[3] Any estimate of θ is, of course of predictive value. Prediction of $\mathbf{n}(t+\tau)$ from $\mathbf{n}(t)$ is here facilitated by the property of linear regression of $\mathbf{n}(t+\tau)$ on $\mathbf{n}(t)$ (Ruben, 1962).

The process $\{\mathbf{n}(t)\}$ reduces formally to a birth and death process[4] when $m = 1$. The estimation of the infinitesimal transition parameters in such a process has been considered by D. G. Kendall (1952), Anscombe (1953), Moran (1951, 1953), Darwin (1959), Rothschild (1953) and Bartlett (1955, pp. 246–252; 1962, pp. 77–79 and pp. 84–85). More generally, there is an extensive literature on the estimation of parameters of Markov chains and processes with a finite state space, and the reader is referred to Billingsley (1961b) for a large number of references on this topic. (See also Billingsley, 1961a, for problems of statistical inference in time continuous Markov processes with an infinite state space.) We may cite the papers by Zahl (1955), Meier (1955) and Albert (1962) as typical examples. In Zahl's paper, the transition probabilities of stationary Markov processes in continuous time with finite state space are estimated (as in Markov chains) from the observed transition frequencies in a fixed number of independent realisations of the process over a given span of time, observations being taken at periodic intervals of time. Again, if observations are taken in continuous time, then further information on the parameters may become available from the actual time instants at which transitions occur (Albert, 1962, and Meier's extension, 1955, of Zahl's results). Now, as pointed out previously (Ruben, 1962), the Poisson-Markov process may in specific applications be generated by the independent, temporally homogeneous Markovian "motion" of an infinite ensemble of systems (particles, individuals, etc.) with respect to a finite set of states (denoted by $\{E_1, E_2, \cdots, E_m, E^*\}$ in the latter paper), and accordingly the observed transition frequencies, or more comprehensively the actual sample paths over a given interval of time, would here too provide estimates of the transition probabilities. This, however, assumes distinguishability of the systems, and if this property is lacking, or if it is not possible or expedient to trace the history of individual systems, then clearly the occupancy numbers $n_r(t)$ ($n_r(t)$, the rth component of $\mathbf{n}(t)$, is the number of systems in E_r at time t, for $r = 1, 2, \cdots, m$) provide the only basis for estimating the transition probabilities.

Consider, then, the fluctuation of the occupancy numbers from the point of view of estimation. For *Markov* processes in general it does not appear unreasonable to expect that relatively efficient estimates of the transition probabilities might frequently be obtained from *consecutive* fluctuations of realised values of the process at discrete points in time. In particular, for one-parameter Markov processes, including (stationary) one-parameter vector Poisson-Markov processes, the parameter in question may be estimated from a *single* estimation

[4] The phrase "birth and death process" is here (as in Ruben, 1962) used elliptically and purely formally in the general sense apparently adopted by Feller (1957, p. 407) for any process in continuous time t ($t \geqq 0$) with state space the set of all non-negative integers in which the only possible one-step transitions are those to neighbouring states *and the zero state is not necessarily absorbing*. This is to be contrasted with (for example) Bartlett's more precise and generic terminology (1955, p. 78) in which the Poisson-Markov process with $m = 1$ is still referred to as an emigration-immigration process.

equation involving the consecutive fluctuations. For this purpose, we construct a quantity D^2 as an appropriate measure of the difference between the two (non-independent) observations $\mathbf{n}(t)$ and $\mathbf{n}(t + \tau)$ (equ. (2.2)), D^2 being a positive definite quadratic form in the components of $\mathbf{n}(t + \tau) - \mathbf{n}(t)$ for $\tau > 0$. (Compare Mahalanobis's D^2, the sample analogue of the population metric Δ^2, used in multivariate normal statistical analysis for *independent* samples, as discussed for example, by Mahalanobis *et al*, 1936.) Since $ED^2 = 1$, the proposed estimation equation for θ is $M(D^2) = 1$, where $M(D^2)$ denotes the mean of a set of observed consecutive D^2, while the expectation vector $\boldsymbol{\nu}$, if unknown, is estimated from the mean of the observed $\mathbf{n}(t)$. In brief, a generalised mean square successive difference is used to provide an estimator $\tilde{\theta}$, henceforth referred to as the mean square consecutive fluctuation estimator (MSCFE), for θ. The large sample variance of $\tilde{\theta}$ is derived in Section 3, and an interesting consequence is the existence of a critical value[5] of τ at which the MSCFE attains maximum precision. This may be regarded as the design aspect of the problem and implies that observations should be made neither too frequently nor too infrequently. The critical value of $\tau \to 0$ as the $\nu_r \to \infty$.

We have previously conjectured that the proposed mode of estimation may be quite efficient. How efficient is it in fact? A partial answer to this difficult question is provided by Sections 4 and 5. These sections enable the relative efficiency of the MSCFE to be determined for the rather special *infinite symmetric linear model*, specifically, when (i) the ν_r are equal and large ($\nu_r \to \infty$), (ii) $\mathbf{\Lambda}$ (and therefore $\mathbf{P}$) is symmetric and (iii) the λ_{rs} and λ_r^* are proportional to θ. It appears in fact that under these conditions the large sample relative efficiency approaches 1 as $\tau \to 0$. (More precisely, we need also $\nu\tau \gg 1$, ν being the common value of the ν_r.) In its turn, this suggests that the general large sample MSCFE has uniformly high efficiency (high efficiency over an entire range of θ) when the ν_r are not small, provided also τ is not small (that is, provided observations are not made too infrequently), this conclusion holding at any rate over the effectively linear portions of λ_{rs} and λ_r^* (regarded here as analytic functions of θ). Further investigation may well strengthen this tentative conclusion.

Notation.

(i) For convenience of printing, the arithmetic mean of a finite set of numbers $a_1, a_2, \cdots$ will be denoted by $M(a)$ (rather than by the more usual $\bar{a}$). The number of elements used to construct the mean will be understood from the context.

(ii) The range of i and j will be 1 to k, unless otherwise specified, and the range of p, q, r, s will be 1 to m. (Recall that k is the number of observations and m the dimensionality of each observation.)

(iii) The arguments of matrices, and of elements of matrices, which are functions of time will be suppressed, wherever convenient, and will then be under-

[5] It is obvious on general grounds that such a critical value must exist for any reasonable method of estimation, since when $\tau = 0$ and when $\tau = \infty$ no information is available about θ.

stood to be τ. Thus $\mathbf{P}$, $\mathbf{R}$, P_{pq}, R^{pq} denote $\mathbf{P}(\tau)$, $\mathbf{R}(\tau)$, $P_{pq}(\tau)$ and $R^{pq}(\tau)$, respectively.

(iv) All vectors in this paper are to be regarded as *row* vectors.

(v) For a given matrix with elements $Q_{\alpha\beta}$, the sum of the elements in the αth row and βth column will be denoted by $Q_{\alpha\cdot}$ and $Q_{\cdot\beta}$, respectively, and the sum of all the elements of the matrix by $Q_{\cdot\cdot}$.

2. Derivation of the estimation equations. Since the covariance matrix of the process is $\mathbf{NP} \equiv \mathbf{NP}(\tau)$, $\mathbf{P}$ being given by (1.1), while the dispersion matrix of $\mathbf{n}(t)$ is $\mathbf{N}$, i.e.,

$$E[(\mathbf{n}(t) - \boldsymbol{\nu})'(\mathbf{n}(t+\tau) - \boldsymbol{\nu})] = \mathbf{NP}, \quad E[(\mathbf{n}(t) - \boldsymbol{\nu})'(\mathbf{n}(t) - \boldsymbol{\nu})] = \mathbf{N},$$

it follows that the dispersion matrix of $\mathbf{d} = \mathbf{n}(\tau) - \mathbf{n}(0)$ is

$$\mathbf{R}(\tau) \equiv \mathbf{R} = \mathbf{N}(\mathbf{I} - \mathbf{P}) + (\mathbf{N}(\mathbf{I} - \mathbf{P}))', \tag{2.1}$$

where $\mathbf{I}$ denotes the unit $m \times m$ matrix. Define

$$D^2 = m^{-1}(\mathbf{d}\mathbf{R}^{-1}\mathbf{d}'). \tag{2.2}$$

Then

$$ED^2 = 1. \tag{2.3}$$

Formula (2.3) suggests the use of

$$M(D^2) = 1 \tag{2.4}$$

as an estimation equation for θ, where

$$M(D^2) = \sum_{\alpha=1}^{k-1} D_\alpha^2/(k-1),$$

$$D_\alpha^2 = m^{-1}(\mathbf{d}_\alpha \mathbf{R}^{-1}\mathbf{d}_\alpha'), \quad \mathbf{d}_\alpha = \mathbf{n}_{\alpha+1} - \mathbf{n}_\alpha \qquad (\alpha = 1, \cdots, k-1),$$

and the $\mathbf{n}_i \equiv \mathbf{n}(t + (i-1)\tau) = (n_{i1}, \cdots, n_{im})$, for $i = 1, \cdots, k$, are k consecutive observed values of $\mathbf{n}(\cdot)$ at times t, $t + \tau$, $\cdots$, $t + (k-1)\tau$. More explicitly, (2.4) is equivalent to

$$m^{-1} \sum_{p,q} R^{pq} M(d_p d_q) = 1, \tag{2.4'}$$

where $((R^{pq})) = \mathbf{R}^{-1} (p, q = 1, \cdots, m)$ and

$$M(d_p d_q) = (k-1)^{-1} \sum_{\alpha=1}^{k-1} (n_{\alpha+1,p} - n_{\alpha p})(n_{\alpha+1,q} - n_{\alpha q}).$$

Finally, if the quantities ν_r entering in the specification of $\mathbf{R}$ through (2.1) are unknown, then unbiassed estimates of these are given by

$$\tilde{\boldsymbol{\nu}} = k^{-1} \sum_i \mathbf{n}_i. \tag{2.5}$$

It should be noted that (2.4) (or (2.4′)) is a transcendental equation in θ

(the R^{pq} are functions of θ) and cannot therefore be solved explicitly for this parameter. However, numerical solution by trial and error with the aid of interpolation is quite practicable. (See also Section 4 for the form of (2.4) and (2.4′) in a special case.)

3. Large-sample variance of the MSCFE. Denote the sampling errors of $\tilde{\theta}$ and $\tilde{\nu}_r$ by $\Delta\tilde{\theta}$ and $\Delta\tilde{\nu}_r$, and, similarly, denote the sampling deviation of $M(d_p\, d_q)$ from its expected value R_{pq} by $\Delta M(d_p\, d_q)$. Then neglecting terms which are $o(k^{-\frac{1}{2}})$ in probability,[6] we obtain from (2.4′)

$$\Big(\sum_{p,q} R_{pq}\partial R^{pq}/\partial\theta\Big)\Delta\tilde{\theta} + \sum_r \sum_{p,q} (R_{pq}\partial R^{pq}/\partial\nu_r)\Delta\tilde{\nu}_r + \sum_{p,q} R^{pq}\Delta M(d_p\, d_q) = 0.$$

It is convenient to simplify the first and second terms with the aid of the identity $\sum_{p,q} R_{pq}R^{pq} = m$. This gives

$$\sum_{p,q} R_{pq}\partial R^{pq}/\partial\nu_r = -\sum_{p,q} R^{pq}\partial R_{pq}/\partial\nu_r$$

and

$$\sum_{p,q} R_{pq}\partial R^{pq}/\partial\theta = -\sum_{p,q} R^{pq}\partial R_{pq}/\partial\theta.$$

Use of the last two relations then yields (to the stated degree of approximation)

$$\Delta\tilde{\theta} = -\frac{\sum_r \sum_{p,q} (R^{pq}\partial R_{pq}/\partial\nu_r)\Delta\tilde{\nu}_r - \sum_{p,q} R^{pq}\Delta M(d_p\, d_q)}{\sum_{p,q} R^{pq}\partial R_{pq}/\partial\theta},$$

whence, to order k^{-1},

$$\text{Var}\,\tilde{\theta} \sim \frac{\Big\{\sum_{r,s} h_r\, h_s \operatorname{Cov}(\tilde{\nu}_r, \tilde{\nu}_s) - 2\sum_{p,q,r} h_r \operatorname{Cov}[\tilde{\nu}_r, M(d_p\, d_q)] + \sum_{p,q,r,s} R^{pq}R^{rs}\operatorname{Cov}[M(d_p\, d_q), M(d_r\, d_s)]\Big\}}{\Big(\sum_{p,q} R^{pq}\partial R_{pq}/\partial\theta\Big)^2} \tag{3.1}$$

where $h_r = \sum_{p,q} R^{pq}\, \partial R_{p,q}/\partial\nu_r$. To evaluate the h_r, observe from (2.1) that

$$\frac{\partial R_{pq}}{\partial\nu_r} = \begin{cases} = (\mathbf{I} - \mathbf{P})_{rq} & (p = r,\ q \neq r), \\ = (\mathbf{I} - \mathbf{P})_{rp} & (p \neq r,\ q = r), \\ = 2(\mathbf{I} - \mathbf{P})_{rr} & (p = r,\ q = r), \\ = 0 & (\text{all other values of } p, q), \end{cases}$$

[6] Rigourization of the familiar differential method used here to obtain the large-sample standard error of $\tilde{\theta}$ is straightforward but tedious. The rigourization can be carried out along the lines of Cramér (1946, pp. 353–356) or, more elegantly, of Hoeffding and Robbins (1948). Note also that a similar argument shows the large-sample bias of $\tilde{\theta}$ to be $O(k^{-1})$.

so that

$$
\begin{aligned}
h_r &= \sum_{q \neq r} R^{rq}(\mathbf{I} - \mathbf{P})_{rq} + \sum_{p \neq r} R^{pr}(\mathbf{I} - \mathbf{P})_{rp} + R^{rr} \cdot 2(\mathbf{I} - \mathbf{P})_{rr} \\
&= \sum_{q} R^{rq}(\mathbf{I} - \mathbf{P})_{rq} + \sum_{p} R^{pr}(\mathbf{I} - \mathbf{P})_{rp} = 2[(\mathbf{I} - \mathbf{P})\mathbf{R}^{-1}]_{rr} . \qquad (3.2)
\end{aligned}
$$

The first two terms in the numerator of (3.1) represent that portion of sampling error (of $\tilde{\theta}$) due to the sampling variability of the $\tilde{\nu}_r$; in particular, if the ν_r are known, (3.1) reduces to

$$
\text{Var } \tilde{\theta} \sim \frac{\sum_{p,q,r,s} R^{pq} R^{rs} \operatorname{Cov}[M(d_p\, d_q), M(d_r\, d_s)]}{\left(\sum_{p,q} R^{pq} \partial R_{pq} / \partial \theta\right)^2} . \qquad (3.1')
$$

The last term in the numerator of (3.1) decomposes (as is shown below) into two parts, one of which would appear, to the exclusion of the other part, if the process were Gaussian and another which corrects for the finiteness of the ν_r. (Recall from Ruben, 1962, that $\{\mathbf{n}(t)\}$ is in the limit a Gaussian process.[7]) The first two terms in the numerator of (3.1) are similar correcting terms. These correcting terms induce a hump in the plot of the variance of $\tilde{\theta}$, both in (3.1) and (3.1′), against τ, giving a critical value of τ at which maximum precision is attained.

We now proceed to evaluate the three covariances appearing in the numerator of (3.1). These clearly involve the product moments of order two, three and four, respectively, of the process, formulae for which are available (Ruben, 1962). It will appear subsequently that the three terms in the numerator may in certain situations be evaluated explicitly (i.e. algebraic formulae can be derived for the *sums* of the covariances defining the three terms).

The following two points should be noted in using subsequent formulae. (a) As is usual in large-sample variance formulae, parameters are replaced by estimates, i.e. ν_r and θ are replaced by $\tilde{\nu}_r$ and $\tilde{\theta}$ in (3.1) and (3.1′). (Correspondingly, $\mathbf{P}$ and $\mathbf{R}$ are replaced, with the same degree of approximation, by $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{R}}$.) (b) Any element of $\phi(\mathbf{P})$, where $\phi(\cdot)$ is a non-singular rational function with scalar co-

[7] The precise sense in which the limit property holds is as follows. Define the process $\{\mathbf{y}(t)\}$ by $\mathbf{y}(t) = (\mathbf{n}(t) - \boldsymbol{\nu})\mathbf{N}^{-\frac{1}{2}}$, and let $\nu_r \to \infty$ such that $\lim_{\nu_r, \nu_s \to \infty} (\nu_r/\nu_s)^{\frac{1}{2}} = C_{rs} < \infty$. Further, for given $t_1 < t_2 < \cdots < t_k$, let $\mathbf{Q}$ denote a matrix which may be partitioned into k^2 submatrices, each of order $m \times m$, such that the (i, j)th submatrix has $C_{rs}P_{rs}(|t_j - t_i|)$ or $C_{sr}P_{sr}(|t_j - t_i|)$ for its (r, s)th element according as $j \geqq i$ or $j < i$. Then the joint distribution of $\mathbf{y}(t_1), \mathbf{y}(t_2), \cdots, \mathbf{y}(t_k)$ tends, as $\nu_r \to \infty$, to a km-dimensional normal distribution with expectation vector zero and dispersion (or correlation) matrix $\mathbf{Q}$.

Note in particular that for equally spaced observations, $t_i = t + (i - 1)\tau$, the limiting dispersion (or correlation) matrix of the km standardized random variables $y_{ir} = (n_{ir} - \nu_r)/\nu_r^{\frac{1}{2}}$ is as given above with $\mathbf{P}(|t_j - t_i|) = \mathbf{P}^{|j-i|}(\tau)$, and the dispersion matrix for *finite* ν_r has $\mathbf{N}^{\frac{1}{2}}\mathbf{P}^{|j-i|}(\tau)\mathbf{N}^{-\frac{1}{2}}$ or $\mathbf{N}^{-\frac{1}{2}}(\mathbf{P}^{|j-i|}(\tau))'\mathbf{N}^{-\frac{1}{2}}$ for its (i, j)th submatrix according as $j \geqq i$ or $j < i$. (I regret that an irritating, though rather obvious, transcription error inadvertently crept into the last paragraph of my previous paper (1962) where the limiting dispersion matrix of the y_{ir} is incorrectly specified.)

efficients[8] is obtained from the corresponding element of $\mathbf{P}$ on replacing $\exp(-\kappa_r\tau)$ by $\phi(\exp(-\kappa_r\tau))$ (Ruben, 1962).

(i) *Evaluation of* $\mathrm{Cov}(\tilde{\nu}_r, \tilde{\nu}_s)$.

$$\begin{aligned}\mathrm{Cov}(\tilde{\nu}_r, \tilde{\nu}_s) &= k^{-2}\sum_{i,j}\mathrm{Cov}(n_{ir}, n_{js})\\ &= k^{-2}\Big\{\sum_{j\geqq i}\nu_r P_{rs}((j-i)\tau) + \sum_{j<i}\nu_s P_{sr}((i-j)\tau)\Big\}\\ &= k^{-2}\Big\{k\nu_r\,\delta_{rs} + \sum_{\alpha=1}^{k-1}(k-\alpha)(\mathbf{NP}^\alpha)_{rs} + \sum_{\alpha=1}^{k-1}(k-\alpha)(\mathbf{NP}^\alpha)_{sr}\Big\}\end{aligned}$$

($\delta_{rs} \equiv P_{rs}(0)$ denotes, as usual, the Kronecker delta function). Defining

$$\begin{aligned}G_k(\mathbf{P}) &= \sum_{\alpha=1}^{k-1}(k-\alpha)\mathbf{P}^\alpha\\ &= (k-1)\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1} - \mathbf{P}^2(\mathbf{I}-\mathbf{P})^{k-1}(\mathbf{I}-\mathbf{P})^{-2},\end{aligned}\tag{3.3}$$

we have the exact formula $\mathrm{Cov}(\tilde{\nu}_r, \tilde{\nu}_s) = k^{-2}\{k\nu_r\delta_{rs} + [\mathbf{N}G_k(\mathbf{P})]_{rs} + [\mathbf{N}G_k(\mathbf{P})]_{sr}\}$. Since

$$G_k(\mathbf{P}) \sim (k-1)\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}\tag{3.4}$$

to order k, the large-sample approximation for the required covariance is given by

$$\mathrm{Cov}(\tilde{\nu}_r, \tilde{\nu}_s) \sim k^{-1}\{\nu_r\delta_{rs} + \nu_r[\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{rs} + \nu_s[\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{sr}\}.\tag{3.5}$$

(ii) *Evaluation of* $\mathrm{Cov}[\tilde{\nu}_r, M(d_p d_q)]$. Since $\tilde{\nu}_r$ is unbiassed,

$$\mathrm{Cov}[\tilde{\nu}_r, M(d_p d_q)] = E[(\tilde{\nu}_r - \nu_r)M(d_p d_q)],$$

and therefore

$$\begin{aligned}&k(k-1)\,\mathrm{Cov}[\tilde{\nu}_r, M(d_p\,d_q)]\\ &= E\Big[\sum_{i=1}^{k}\sum_{j=1}^{k-1}(n_{ir}-\nu_r)\{(n_{jp}-\nu_p)-(n_{j+1,p}-\nu_p)\}\{(n_{jq}-\nu_q)-(n_{j+1,q}-\nu_q)\}\Big]\\ &= E\Big[\sum_{i=1}^{k}\sum_{j=1}^{k-1}\{(n_{ir}-\nu_r)(n_{jp}-\nu_p)(n_{jq}-\nu_q) - (n_{ir}-\nu_r)(n_{jp}-\nu_p)(n_{j+1,q}-\nu_q)\\ &\quad - (n_{ir}-\nu_r)(n_{jq}-\nu_q)(n_{j+1,p}-\nu_p) + (n_{ir}-\nu_r)(n_{j+1,p}-\nu_p)(n_{j+1,q}-\nu_q)\}\Big].\end{aligned}$$

On splitting up the summation with respect to i into two parts, namely, $i \leqq j$ and $i > j$, and using a previous formula for third order product moments (Ruben,

[8] Such a function is well-defined, for if $\phi(\mathbf{P}) = \phi_1(\mathbf{P})(\phi_2(\mathbf{P}))^{-1}$ where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are polynomials with scalar coefficients, $\phi_2(\mathbf{P})$ being non-singular, then also $\phi(\mathbf{P}) = (\phi_2(\mathbf{P}))^{-1}\phi_1(\mathbf{P})$. (See Ruben, 1962.)

1962, Equ. (3.20)), we obtain

$$k(k-1)\operatorname{Cov}[\tilde{\nu}_r, M(d_p\,d_q)] = \sum_{i=1}^{j}\sum_{j=1}^{k-1}\nu_r\{P_{rp}((j-i)\tau)\delta_{pq}$$
$$- P_{rp}((j-i)\tau)P_{pq}(\tau) - P_{rq}((j-i)\tau)P_{qp}(\tau)$$
$$+ P_{rp}((j-i+1)\tau)\delta_{pq}\} + \sum_{i=j+1}^{k-1}\sum_{j=1}^{k-1}\{\nu_p\delta_{pq}P_{qr}((j-i)\tau)$$
$$- \nu_p P_{pq}(\tau)P_{qr}((i-j-1)\tau)$$
$$- \nu_q P_{qp}(\tau)P_{pr}((i-j-1)\tau) + \nu_p\delta_{pq}P_{qr}((i-j-1)\tau)\}.$$

Multiply both members of the last equation by R^{pq} and sum with respect to p and q. The first term then contributes $\nu_r \sum_p R^{pp}[\mathbf{P}^{-1}G_k(\mathbf{P})]_{rp}$, the second and third terms each contribute $-\nu_r \sum_{p,q} R^{pq}[\mathbf{P}^{-1}G_k(\mathbf{P})]_{rp}P_{pq}$, while the fourth term contributes $\nu_r \sum_p R^{pp}[G_k(\mathbf{P})]_{rp}$; similarly, the fifth term contributes $\sum_p \nu_p R^{pp}[G_k(\mathbf{P})]_{pr}$, the sixth and seventh terms each contribute

$$-\sum_{p,q} \nu_q R^{pq}P_{qp}[\mathbf{P}^{-1}G_k(\mathbf{P})]_{pr},$$

while the eighth term contributes $\sum_p \nu_p R^{pp}[\mathbf{P}^{-1}G_k(\mathbf{P})]_{pr}$. Collection of terms followed by some reduction yields

$$k(k-1)\sum_{p,q} R^{pq}\operatorname{Cov}[\tilde{\nu}_r, M(d_p\,d_q)]$$
$$(3.6)\qquad = \sum_p R^{pp}\{[\mathbf{N}(\mathbf{I}+\mathbf{P})\mathbf{P}^{-1}G_k(\mathbf{P})]_{rp} + [\mathbf{N}(\mathbf{I}+\mathbf{P})\mathbf{P}^{-1}G_k(\mathbf{P})]_{pr}\}$$
$$- 2\sum_p\{(\mathbf{P}\mathbf{R}^{-1})_{pp}[\mathbf{N}\mathbf{P}^{-1}G_k(\mathbf{P})]_{rp} + (\mathbf{R}^{-1}\mathbf{N}\mathbf{P})_{pp}[\mathbf{P}^{-1}G_k(\mathbf{P})]_{pr}\}.$$

This formula is exact. For k large, (3.6) becomes, on using (3.4),

$$\sum_{p,q} R^{pq}\operatorname{Cov}[\tilde{\nu}_r, M(d_p\,d_q)]$$
$$(3.7)\qquad = k^{-1}\sum_p R^{pp}\{[\mathbf{N}(\mathbf{I}+\mathbf{P})(\mathbf{I}-\mathbf{P})^{-1}]_{rp} + [\mathbf{N}(\mathbf{I}+\mathbf{P})(\mathbf{I}-\mathbf{P})^{-1}]_{pr}\}$$
$$- 2k^{-1}\sum_p\{\nu_r(\mathbf{P}\mathbf{R}^{-1})_{pp}(\mathbf{I}-\mathbf{P})^{rp} + (\mathbf{R}^{-1}\mathbf{N}\mathbf{P})_{pp}(\mathbf{I}-\mathbf{P})^{pr}\}.$$

(iii) *Evaluation of* Cov $[M(d_p\,d_q), M(d_r\,d_s)]$. First (after some straightforward though rather lengthy and tedious algebra), on using the previous formula for the fourth order product moments of the process (Ruben, 1962, Equ. (3.21)),

$$E[(n_{ip}-n_{i+1,p})(n_{iq}-n_{i+1,q})(n_{jr}-n_{j+1,r})(n_{js}-n_{j+1,s})]$$
$$(3.8)\qquad = \begin{cases} R_{pq}R_{rs} + X^{(i,j)}_{p,q;r,s} + Y^{(i,j)}_{p,q;r,s} & (j > i),\\ R_{pq}R_{rs} + X^{(j,i)}_{r,s;p,q} + Y^{(j,i)}_{r,s;p,q} & (j < i),\\ R_{pr}R_{qs} + R_{ps}R_{qr} + R_{pq}R_{rs} + Z_{p,q;r,s} & (j = i),\end{cases}$$

where

$$X^{(\alpha,\beta)}_{p,q;r,s} = U^{(\alpha,\beta)}_{pr} U^{(\alpha,\beta)}_{qs} + U^{(\alpha,\beta)}_{ps} U^{(\alpha,\beta)}_{qr} \qquad (\alpha < \beta)$$

with

$$U^{(\alpha,\beta)} = \mathbf{N}\mathbf{P}^{\beta-\alpha-1}(\mathbf{I} - \mathbf{P})^2 \qquad (\alpha < \beta),$$

$$\begin{aligned} Y^{(\alpha,\beta)}_{p,q;r,s} = {} & \nu_p(\mathbf{I} - \mathbf{P})_{rs}\{\delta_{pq}(\mathbf{P}^{\beta-\alpha})_{qr} - P_{pq}(\mathbf{P}^{\beta-\alpha-1})_{qr}\} \\ & + \nu_p\delta_{pq}\{\delta_{sr}(\mathbf{P}^{\beta-\alpha+1})_{qs} - P_{sr}(\mathbf{P}^{\beta-\alpha})_{qs}\} \\ & - \nu_p P_{pq}\{\delta_{sr}(\mathbf{P}^{\beta-\alpha})_{qs} - P_{sr}(\mathbf{P}^{\beta-\alpha-1})_{qs}\} \\ & + \nu_q(\mathbf{I} - \mathbf{P})_{qp}\{(\mathbf{I} - \mathbf{P})_{rs}(\mathbf{P}^{\beta-\alpha-1})_{pr} \\ & + \delta_{sr}(\mathbf{P}^{\beta-\alpha})_{ps} - P_{sr}(\mathbf{P}^{\beta-\alpha-1})_{ps}\} \qquad (\alpha < \beta), \end{aligned}$$

and

$$\begin{aligned} Z_{p,q;r,s} = {} & \nu_p\delta_{pq}\delta_{qr}\delta_{qs} - \nu_p\delta_{pq}\delta_{qr}P_{rs} - \nu_p\delta_{pq}\delta_{qs}P_{sr} + \nu_p\delta_{pq}\delta_{rs}P_{qr} \\ & - \nu_p\delta_{pr}\delta_{rs}P_{sq} + \nu_p\delta_{pr}\delta_{sq}P_{rs} + \nu_p\delta_{ps}\delta_{qr}P_{sq} - \nu_p\delta_{qs}\delta_{qr}P_{ps} \\ & - \nu_q\delta_{qr}\delta_{rs}P_{sp} + \nu_q\delta_{qr}\delta_{sp}P_{rs} + \nu_p\delta_{qs}\delta_{pr}P_{sp} - \nu_q\delta_{sp}\delta_{pr}P_{qs} \\ & + \nu_r\delta_{rs}\delta_{pq}P_{sp} - \nu_r\delta_{sp}\delta_{pq}P_{rs} - \nu_s\delta_{rp}\delta_{pq}P_{sr} + \nu_s\delta_{sr}\delta_{rp}\delta_{pq}\,. \end{aligned}$$

Therefore,

$$\begin{aligned} E[M(d_p\, d_q)M(d_r\, d_s)] = {} & (k-1)^{-2}\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} E[(n_{ip} - n_{i+1,p})(n_{iq} - n_{i+1,q}) \\ & (n_{jr} - n_{j+1,r})(n_{js} - n_{j+1,s})] \\ = {} & (k-1)^{-2}\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}(R_{pq}R_{rs} + X^{(i,j)}_{p,q;r,s} + Y^{(i,j)}_{p,q;r,s}) \\ & + (k-1)^{-2}\sum_{i=2}^{k-1}\sum_{j=1}^{i-1}(R_{pq}R_{rs} + X^{(j,i)}_{r,s;p,q} + Y^{(j,i)}_{r,s;p,q}) \\ & + (k-1)^{-1}(R_{pr}R_{qs} + R_{ps}R_{qr} + R_{pq}R_{rs} + Z_{p,q;r,s}), \end{aligned}$$

or

$$\begin{aligned} \operatorname{Cov}[M(d_p\, d_q), M(d_r\, d_s)] = {} & (k-1)^{-2}\Bigg\{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k-1}(X^{(i,j)}_{p,q;r,s} + Y^{(i,j)}_{p,q;r,s}) \\ & + \sum_{i=2}^{k-1}\sum_{j=1}^{i-1}(X^{(j,i)}_{r,s;p,q} + Y^{(j,i)}_{r,s;p,q})\Bigg\} + (k-1)^{-1}\{R_{pr}\,R_{qs} + R_{ps}\,R_{qr} + Z_{p,q;r,s}\}. \end{aligned} \tag{3.9}$$

Formula (3.9) is the required covariance formula. On noting further that $\sum_{p,q,r,s} R^{pq}R^{rs}R_{pr}\,R_{qs} = m = \sum_{p,q,r,s} R^{pq}R^{rs}R_{ps}\,R_{qr}$, the numerator in (3.1′), and the third term in the numerator of (3.1), may be expressed in the form

$$
\begin{aligned}
\sum_{p,q,r,s} & R^{pq}R^{rs} \operatorname{Cov}[M(d_p d_q), M(d_r d_s)] \\
&= (k-1)^{-2}\{ \sum_{p,q,r,s} \sum_{j>i} R^{pq}R^{rs}(X^{(i,j)}_{p,q;r,s} + Y^{(i,j)}_{p,q;r,s}) \\
&\quad + \sum_{p,q,r,s} \sum_{j<i} R^{pq}R^{rs}(X^{(j,i)}_{r,s;p,q} + Y^{(j,i)}_{r,s;p,q})\} \\
&\quad + (k-1)^{-1} \sum_{p,q,r,s} R^{pq}R^{rs}Z_{p,q;r,s} + (k-1)^{-1}\cdot(2m).
\end{aligned} \tag{3.10}
$$

We remark that in (3.10) the Y and Z terms, which are linear in the ν_r, may be regarded as correction factors for the finiteness of the ν_r or for the non-Gaussian character of $\{\mathbf{n}(t)\}$, the remaining terms, which are quadratic in the ν_r, then giving the purely Gaussian components.

The series in (3.10) cannot in general be further reduced algebraically. Such a reduction is however possible in special cases (see Section 4).

4. Some special cases.

(A) *Known ratios of the* ν_r. Assume that the ν_r may be expressed in the form $\nu_r = g_r\phi$, where the g_r are known and ϕ is to be estimated. In relation to the states $E_1, \cdots, E_m$ of the introductory section, this implies that the relative "sizes" of these states are assumed known (at least approximately). Clearly, ϕ may be regarded as a kind of density parameter.

An unbiassed estimator for ϕ is

$$\tilde{\phi} = \sum_{i,r} n_{ir}/k \sum_r g_r, \tag{4.1}$$

and the corresponding unbiassed estimators for the ν_r are

$$\tilde{\nu}_r' = g_r\tilde{\phi}. \tag{4.2}$$

The estimation equation (2.4) for θ is used as before with $\tilde{\nu}_r$ replaced by $\tilde{\nu}_r'$.

The large-sample variance formula (3.1) is no longer valid, but proceeding analogously as in Section 3 the corresponding formula is readily obtained as

$$\operatorname{Var}\tilde{\theta} \sim \frac{\begin{aligned}\{m^2 \operatorname{Var}\tilde{\phi}/\phi^2 - 2m \operatorname{Cov}[\tilde{\phi}, \textstyle\sum_{p,q} R^{pq}M(d_p d_q)]/\phi \\ + \textstyle\sum_{p,q,r,s} R^{pq}R^{rs}\operatorname{Cov}[M(d_p d_q), M(d_r d_s)]\}\end{aligned}}{(\sum_{p,q} R^{pq}\partial R_{pq}/\partial\theta)^2}. \tag{4.3}$$

The third term in the numerator of (4.3) as well as the denominator are identical with the corresponding terms of (3.1). To evaluate the new terms (i.e. the first and second terms in the numerator of (4.3)), recall that the dispersion matrix (cf. footnote 7) of the km variables n_{ir} may be partitioned into k^2 submatrices, each of size $m \times m$, such that the (i, j)th submatrix is $\mathbf{NP}^{|j-i|}$ or $(\mathbf{P}^{|j-i|})'\mathbf{N}$, according as $j \geqq i$ or $j < i$ (the covariance between n_{ir} and n_{js} being given by the element in the rth row and sth column of the (i, j)th submatrix). Hence (4.1) gives on summation of the elements in the dispersion matrix,

$$\begin{aligned} \operatorname{Var} \tilde{\phi} &= (k \sum_r g_r)^{-2} \left[k\mathbf{N} + 2 \sum_{\alpha=1}^{k-1} (k - \alpha)\mathbf{N}\mathbf{P}^{\alpha} \right]_{\cdot\cdot} \\ &= (k \sum_r g_r)^{-2} [k\mathbf{N} + 2\mathbf{N}G_k(\mathbf{P})]_{\cdot\cdot}, \end{aligned} \tag{4.4}$$

where $G_k(\mathbf{P})$ is evaluated in (3.3). Equation (4.4) is exact. For large k, (4.4) reduces with the aid of (3.4) to

$$\operatorname{Var} \tilde{\phi} \sim k^{-1} (\sum_r g_r)^{-2} [\mathbf{N}(\mathbf{I} + \mathbf{P})(\mathbf{I} - \mathbf{P})^{-1}]_{\cdot\cdot}\,. \tag{4.5}$$

The second term in the numerator of (4.3) follows directy from the corresponding term in (3.1). For

$$\operatorname{Cov} [\tilde{\phi}, \sum_{p,q} R^{pq} M(d_p\, d_q)] = (\sum_r g_r)^{-1} \sum_r \operatorname{Cov} [\tilde{\nu}_r, \sum_{p,q} R^{pq} M(d_p\, d_q)],$$

where the $\tilde{\nu}_r$ are the unrestricted estimators given by (2.5). On referring to (3.6), we then obtain the exact formula

$$\begin{aligned} & k(k - 1)(\sum_r g_r) \cdot \operatorname{Cov} [\tilde{\phi}, \sum_{p,q} R^{pq} M(d_p\, d_q)] \\ &= \sum_p R^{pp} \{[\mathbf{N}(\mathbf{I} + \mathbf{P})\mathbf{P}^{-1} G_k(\mathbf{P})]_{\cdot p} + [\mathbf{N}(\mathbf{I} + \mathbf{P})\mathbf{P}^{-1} G_k(\mathbf{P})]_{p\cdot}\} \\ &\quad - 2 \sum_p \{(\mathbf{P}\mathbf{R}^{-1})_{pp} [\mathbf{N}\mathbf{P}^{-1} G_k(\mathbf{P})]_{\cdot p} + (\mathbf{R}^{-1}\mathbf{N}\mathbf{P})_{pp} [\mathbf{P}^{-1} G_k(\mathbf{P})]_{p\cdot}\} \end{aligned} \tag{4.6}$$

for all k, while for k large,

$$\begin{aligned} & k(\sum_r g_r) \cdot \operatorname{Cov} [\tilde{\phi}, \sum_{p,q} R^{pq} M(d_p\, d_q)] \\ &\sim \sum_p R^{pp} \{[\mathbf{N}(\mathbf{I} + \mathbf{P})(\mathbf{I} - \mathbf{P})^{-1}]_{\cdot p} + [\mathbf{N}(\mathbf{I} + \mathbf{P})(\mathbf{I} - \mathbf{P})^{-1}]_{p\cdot}\} \\ &\quad - 2 \sum_p \{(\mathbf{P}\mathbf{R}^{-1})_{pp} \sum_q \nu_q (\mathbf{I} - \mathbf{P})^{qp} + (\sum_q \nu_q P_{qp} R^{pq})(\mathbf{I} - \mathbf{P})^{p\cdot}\}. \end{aligned} \tag{4.7}$$

(B) *The symmetric model.* Consider the case where $\boldsymbol{\Lambda}$ is symmetric and $E\mathbf{n}(t)$ has equal components: (i) $\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}$, (ii) $\mathbf{N} = \nu\mathbf{I}$. Thus, if $\{\mathbf{n}(t)\}$ is generated by independent "migration" of elements between $m + 1$ states $E_1, \cdots, E_m, E^*$ (see introductory section), then (i) implies that there is no preference of motion as regards any two states E_r, $E_s(\lambda_{rs} = \lambda_{sr})$ and (ii) states that the "sizes" of $E_1, \cdots, E_m$ are equal (ν may either be known or has to be estimated). Model (B) is a specialisation of model (A) with $g_r = 1$, i.e. $\phi = \nu$, and the further striction that $\mathbf{P}$ is symmetric, so that the matrix $\mathbf{R}$ of Section 2 reduces to

$$\mathbf{R} = 2\nu(\mathbf{I} - \mathbf{P}). \tag{4.8}$$

The estimation equation (2.4′) for θ simplifies to

$$(2m\tilde{\nu})^{-1} \sum_{p,q} (\mathbf{I} - \mathbf{P})^{pq} M(d_p\, d_q) = 1 \tag{4.9}$$

where

$$\tilde{\nu} = (km)^{-1} \sum_{i,r} n_{ir} . \tag{4.10}$$

The variance formula (4.3) is still applicable (with $\phi = \nu$), and formulae (4.4), (4.5), (4.6) and (4.7) for the first two terms in the numerator of (4.3) reduce to

$$\operatorname{Var} \tilde{\nu} = (km)^{-2}\nu[k\mathbf{I} + 2G_k(\mathbf{P})]_{..} \sim k^{-1}m^{-2}\nu[\mathbf{I} + \mathbf{P})(\mathbf{I} - \mathbf{P})^{-1}]_{..} \tag{4.11}$$

and

$$k(k - 1)m \cdot \operatorname{Cov}[\tilde{\nu}, \sum_{p,q} R^{pq}M(d_p\, d_q)]$$

$$\begin{aligned} &= \sum_p (\mathbf{I} - \mathbf{P})^{pp}[(\mathbf{I} + \mathbf{P})\mathbf{P}^{-1}G_k(\mathbf{P})]_{p\cdot} - 2 \sum_p [\mathbf{P}(\mathbf{I} - \mathbf{P})^{-1}]_{pp}[\mathbf{P}^{-1}G_k(\mathbf{P})]_{p\cdot} \\ &\sim (k - 1)\{ \sum_p (\mathbf{I} - \mathbf{P})^{pp}[(\mathbf{I} + \mathbf{P})(\mathbf{I} - \mathbf{P})^{-1}]_{p\cdot} \\ &\qquad - 2\sum_p [\mathbf{P}(\mathbf{I} - \mathbf{P})^{-1}]_{pp}(\mathbf{I} - \mathbf{P})^{p\cdot}\}. \end{aligned} \tag{4.12}$$

The third term in the numerator of (4.3), evaluated in (3.10), may also be reduced to more explicit form (as noted at the end of Section 3). In fact, since

$$\begin{aligned} \sum_{p,q,r,s} R^{pq}R^{rs}U_{pr}^{(i,j)}U_{qs}^{(i,j)} &= \tfrac{1}{4} \sum_p [\mathbf{P}^{2(j-i-1)}(\mathbf{I} - \mathbf{P})^2]_{pp} \\ &= \sum_{p,q,r,s} R^{pq}R^{rs}U_{ps}^{(i,j)}U_{qr}^{(i,j)}, \end{aligned}$$

so that

$$\sum_{p,q,r,s} R^{pq}R^{rs}X_{p,q;r,s}^{(i,j)} = \tfrac{1}{2} \sum_p [\mathbf{P}^{2(j-i-1)}(\mathbf{I} - \mathbf{P})^2]_{pp} = \sum_{p,q,r,s} R^{pq}R^{rs}X_{r,s;p,q}^{(j,i)},$$

we obtain

$$\begin{aligned} \sum_{p,q,r,s} \sum_{j>i} R^{pq}R^{rs}X_{p,q;r,s}^{(i,j)} &= \tfrac{1}{2} \sum_r [\mathbf{P}^{-2}G_{k-1}(\mathbf{P}^2)(\mathbf{I} - \mathbf{P})^2]_{pp} \\ &= \sum_{p,q,r,s} \sum_{j<i} R^{pq}R^{rs}X_{r,s;p,q}^{(j,i)} \end{aligned}$$

(this formula being exact), and to order k^{-1},

$$\begin{aligned} \sum_{p,q,r,s} \sum_{j>i} R^{pq}R^{rs}X_{p,q;r,s}^{(i,j)} &\sim \tfrac{1}{2}(k - 1) \sum_p [(\mathbf{I} + \mathbf{P})^{-1}(\mathbf{I} - \mathbf{P})]_{pp} \\ &= \sum_{p,q,r,s} \sum_{j<i} R^{pq}R^{rs}X_{r,s;p,q}^{(j,i)} . \end{aligned}$$

Thus

$$\begin{aligned} (k - 1)^{-2} \{ &\sum_{p,q,r,s} \sum_{j>i} R^{pq}R^{rs}X_{p,q;r,s}^{(i,j)} + \sum_{p,q,r,s} \sum_{j<i} R^{pq}R^{rs}X_{r,s;p,q}^{(j,i)}\} \\ &+ (k - 1)^{-1}(2m) \sim (k - 1)^{-1} \sum_r \{(1 - e^{-\kappa_r\tau})/(1 + e^{-\kappa_r\tau})\} \\ &\qquad + (k - 1)^{-1}(2m) = (k - 1)^{-1} \sum_r \{(3 + e^{-\kappa_r\tau})/(1 + e^{-\kappa_r\tau})\}, \end{aligned}$$

since the trace of the matrix $\psi(\mathbf{P}(\tau))$ is $\sum_r \psi(e^{-\kappa_r \tau})$ for any rational function $\psi(\cdot)$ with scalar coefficients. Formula (3.10) now becomes (to order k^{-1})

$$\begin{aligned}
&\sum_{p,q,r,s} R^{pq}R^{rs}\,\mathrm{Cov}\,[M(d_p\,d_q),\,M(d_r\,d_s)] \\
&\qquad\sim (k-1)^{-1}\sum_r\{(3+e^{-\kappa_r\tau})/(1+e^{-\kappa_r\tau})\} \\
&\qquad + (k-1)^{-2}\ \{\sum_{p,q,r,s}\sum_{j>i} R^{pq}R^{rs}Y^{(i,j)}_{p,q;r,s} \\
&\qquad + \sum_{p,q,r,s}\sum_{j<i} R^{pq}R^{rs}Y^{(j,i)}_{r,s;p,q}\} \\
&\qquad + (k-1)^{-1}\sum_{p,q,r,s} R^{pq}R^{rs}Z_{p,q;r,s}\,.
\end{aligned} \tag{4.13}$$

From the previous definition of Y and Z, we find that to order k^{-1} (again omitting the algebraic details) the last three terms on the right of (4.13) are given as follows:[9]

$$\begin{aligned}
&\sum_{p,q,r,s}\sum_{j>i} R^{pq}R^{rs}Y^{(i,j)}_{p,q;r,s} \\
&= (4\nu)^{-1}(k-1)\sum_{p,q}(\mathbf{I}-\mathbf{P})^{pq}\{1-2[\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{pp} \\
&+ [\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{pp}[\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{qq}\} \\
&+ (2\nu)^{-1}(k-1)\sum_{p,q}(\mathbf{I}-\mathbf{P})^{pp}\{1-[\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{qq}\}[\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{pq} \\
&+ (4\nu)^{-1}(k-1)\sum_{p,q}(\mathbf{I}-\mathbf{P})^{pp}(\mathbf{I}-\mathbf{P})^{qq}[\mathbf{P}^2(\mathbf{I}-\mathbf{P})^{-1}]_{pq} \\
&= \sum_{p,q,r,s}\sum_{j<i} R^{pq}R^{rs}Y^{(j,i)}_{r,s;p,q}
\end{aligned} \tag{4.13.1}$$

and

$$\begin{aligned}
&\sum_{p,q,r,s} R^{pq}R^{rs}Z_{p,q,r,s} \\
&= (4\nu)^{-1}\{2\sum_p[(\mathbf{I}-\mathbf{P})^{pp}]^2 - 8\sum_p(\mathbf{I}-\mathbf{P})^{pp}[\mathbf{P}(\mathbf{I}-\mathbf{P})^{-1}]_{pp} \\
&\qquad + 2\sum_{p,q}(\mathbf{I}-\mathbf{P})^{pp}(\mathbf{I}-\mathbf{P})^{qq}P_{pq} + 4\sum_{p,q}[(\mathbf{I}-\mathbf{P})^{pq}]^2P_{pq}\}.
\end{aligned} \tag{4.13.2}$$

(4.13) (taken in conjunction with (4.13.1) and (4.13.2)) is the desired formula for the third term in the numerator of (4.3). We note that the dominant part ($\nu \to \infty$) of this term is the expression involving exponentials.

Finally, a simple representation of the denominator in (4.3) is now possible.

[9] Y and Z have been previously evaluated as the sum of nine and sixteen terms, respectively. Consequently, the left-hand members of (4.13.1) and (4.13.2) are likewise expressible as the sum of nine and sixteen terms, respectively. The separate terms in these two sets of terms are, however, not distinct, and collection of like terms, after suitable simplification, produces the expressions on the right of (4.13.1) and (4.13.2).

Define $a_r = \partial\kappa_r/\partial\theta$, and denote the diagonal matrix with diagonal elements $a_1, \cdots, a_m$ by **A**. Since (recall the definitions of **P** and **R** in (1.1) and (4.8))

$$\partial\mathbf{R}/\partial\theta = (2\nu\tau)\cdot\mathbf{\Theta}^{-1}\mathbf{AK\Theta},$$

we have $\mathbf{R}^{-1}\partial\mathbf{R}/\partial\theta = \tau\cdot\mathbf{\Theta}^{-1}\mathbf{AK}(\mathbf{I} - \mathbf{K})^{-1}\mathbf{\Theta}$, and consequently

$$(4.14)\qquad \begin{aligned}(\sum_{p,q} R^{pq}\partial R_{pq}/\partial\theta)^2 &= \tau^2 tr^2\{\mathbf{\Theta}^{-1}\mathbf{AK}(\mathbf{I} - \mathbf{K})^{-1}\mathbf{\Theta}\}\\ &= \tau^2\left(\sum_r \frac{a_r e^{-\kappa_r\tau}}{1 - e^{-\kappa_r\tau}}\right)^2.\end{aligned}$$

This completes the evaluation of the four terms in formula (4.3) for the large-sample variance of the MSCFE in the symmetric model, these terms being given by (4.11), (4.12), (4.13) (taken in conjunction with (4.13.1) and (4.13.2)) and (4.14). Observe in particular that the dominant part ($\nu \to \infty$) of the large-sample variance is given by

$$(4.15)\qquad \operatorname{Var}\tilde{\theta} \sim (k - 1)^{-1}\tau^{-2}\cdot\sum_r\left(\frac{3 + e^{-\kappa_r\tau}}{1 + e^{-\kappa_r\tau}}\right)\Big/\left(\sum_r \frac{a_r e^{-\kappa_r\tau}}{1 - e^{-\kappa_r\tau}}\right)^2,$$

and that the right-hand member of (4.15) is an increasing function in τ which tends to

$$(4.16)\qquad \frac{2m}{k - 1}\frac{1}{(\sum_r \partial\log\kappa_r/\partial\theta)^2}$$

as $\tau \to 0$ (the corresponding coefficient of variation being

$$\{2m/(k - 1)\}^{\frac{1}{2}}/\sum_r(\theta\partial\log\kappa_r/\partial\theta))$$

and to ∞ as $\tau \to \infty$. However, when the correcting non-Gaussian components of variance are taken into account the variance function is seen to tend to ∞ both as $\tau \to \infty$ and as $\tau \to 0$, and to have a minimum at some $\tau \neq 0$. For *finite* ν, then, there is a critical (non-zero) value of τ at which maximum precision is attained. A similar conclusion holds for the general Poisson-Markov model.

We remark that (4.16) is the limiting value, as $\tau \to 0$, of the expression in (4.15), itself the limiting value, as $\nu \to \infty$, of the large-sample variance (i.e., (4.16) is the intercept of the particular variance curve $\nu = \infty$, plotted against τ, on the vertical axis). It is in fact readily established that the large-sample variance in (4.3) tends to the expression (4.16) as $\nu \to \infty$ and $\tau \to 0$ such that $\nu\tau \to \infty$.[10]

(C) *The symmetric linear*[11] *model.* (Known ratios of the transition parameters.)

[10] A further interesting limiting result of a rather different kind may be noted here. We find from (4.3) (and the subsequent expressions for the four terms on the right of (4.3)) that, for continuous observations over a fixed span of time T, $\operatorname{Var}\tilde{\theta}/\theta^2$ is proportional to $1/T$. This follows on letting $\tau \to 0$ and $k \to \infty$ such that $(k - 1)\tau = T$.

[11] The *general* Poisson-Markov process is linear in the sense that $\mathbf{n}(t + \tau) - \boldsymbol{\nu} = (\mathbf{n}(t) - \boldsymbol{\nu})\mathbf{P}(\tau) + \boldsymbol{\varepsilon}(t)$, where $E[(\boldsymbol{\varepsilon}(t_1))'\boldsymbol{\varepsilon}(t_2)] = \mathbf{0}$ for $t_1 \neq t_2$. Linearity in the text refers specifically to the *transition* parameters.

This is the symmetric model of (B) with the further condition that the transition parameters λ_{rs} and λ_r^* are *linear* in θ: more precisely, these parameters are multiples of θ. Thus the symmetric linear model has the properties (i) $\mathbf{\Lambda}' = \mathbf{\Lambda}$ (ii) $\mathbf{N} = \nu\mathbf{I}$ and (iii) $\lambda_{rs} = c_{rs}\theta$, $\lambda_r^* = c_r^*\theta$, where the c_{rs} and c_r^* are assumed to be known for our present purposes. The quantity θ represents an intrinsic migration or interaction parameter, while the c_{rs} may clearly be regarded as measures of the separation between the states $E_1, \cdots, E_m$. Note further that (iii) may be a reasonable approximation if the transition parameters are regular in the neighbourhood of $\theta = 0$.

It may perhaps appear that this model is too restrictive to cover any practical applications. Our justification for considering it is two-fold. First, the model allows as to gain some insight into the efficiency of the MSCFE under more general conditions, inasmuch as a lower bound to the variance of any unbiassed estimator of θ may be determined when (i), (ii) and (iii) hold. This is done in Section 5. Next, practical applications of the model, even if only approximate, *can* be found. One such application occurs in the study of the migration of particles and organisms (Ruben and Rothschild, 1953; Patil, 1955, 1957) between disjoint regions $E_1, \cdots, E_m$ and E^*, where $E_1, \cdots, E_m$ are finite and of equal size while E^*, the complement to $\bigcup_1^m E_r$, is infinite, $n_r(t)$ denotes the number of particles in E_r at time t and ν is the expected value of $n_r(t)$, estimation of ν being equivalent to estimation of the (constant) number density of the particles. On a kinematic hypothesis derived from general gas kinetic theory, it may be shown that if $\{\mathbf{n}(t)\}$ is approximated by a Markovian emigration-immigration process (the adequacy of such an approximation has been examined from a heuristic standpoint by Ruben and Rothschild, 1953, and more rigorously by Patil, 1955, 1957), then the infinitesimal transition parameters are simple functions of the regions. Here θ is the mean speed of the particles. We remark further that in this application (ii) (i.e. regions of equal size) automatically implies (i).

Since (C) is a special case of (B), the variance formulae in (B) hold here too with the simplification that $a_r = \kappa_r/\theta$. In particular, (4.16) reduces to

$$\frac{2\theta^2}{(k-1)m} \tag{4.17}$$

(the corresponding coefficient of variation being $2/\{(k-1)m\}^{\frac{1}{2}}$).

As a special example of (C), consider the case where the λ_{rs} are all equal and where the λ_r^* are also equal. This will illustrate the use of the estimation Equation (2.4′) in more general situations. In relation to the particle migration problem discussed previously, equality of the λ_{rs} and λ_r^* will obtain under certain conditions of symmetry with regard to the regions $E_1, E_2, \cdots, E_m$ (e.g., for laminar motion and $m = 3$, when E_1, E_2, E_3 are formed by division of a circle into three equiangular sectors).

For equal λ_{rs} and λ_r^*, $\mathbf{\Lambda}$ has all its diagonal elements equal to $-\lambda$ and all its off-diagonal elements equal to $(m-1)\lambda + \lambda^*$, where λ denotes the common value of the λ_{rs} and λ^* the common value of the λ_r^*. Here

$$P_{rs}(\tau) = \begin{cases} m^{-1}[e^{-\lambda^*\tau} - e^{-(\lambda^*+m\lambda)\tau}] & (r \neq s), \\ m^{-1}e^{-\lambda^*\tau} + (1 - m^{-1})e^{-(\lambda^*+m\lambda)\tau} & (r = s), \end{cases}$$

and therefore for $l = 0, 1, \cdots$,

$$(\mathbf{P}^l(\mathbf{I} - \mathbf{P})^{-1})_{rs} = \begin{cases} \dfrac{1}{m}\left[\dfrac{e^{-l\lambda^*\tau}}{1 - e^{-\lambda^*\tau}} - \dfrac{e^{-l(\lambda^*+m\lambda)\tau}}{1 - e^{-(\lambda^*+m\lambda)\tau}}\right] & (r \neq s) \\ \dfrac{1}{m}\dfrac{e^{-l\lambda^*\tau}}{1 - e^{-\lambda^*\tau}} + \left(1 - \dfrac{1}{m}\right)\dfrac{e^{-l(\lambda^*+m\lambda)\tau}}{1 - e^{-(\lambda^*+m\lambda)\tau}} & (r = s). \end{cases}$$

Note that $l = 0$ in the last formula gives $R^{rs} = (2\nu)^{-1}(\mathbf{I} - \mathbf{P})^{rs}$ which are needed for the use of (2.4′). The latter equation reduces to

$$\{m^{-1}v_1 + (1 - m^{-1})v_2\}M(d_1^2 + d_2^2) + 2m^{-1}(v_1 - v_2)M(d_1d_2) = 2m\tilde{\nu},$$

where $v_1 = (1 - e^{-\lambda^*\tau})^{-1}$, $v_2 = (1 - e^{-(\lambda^*+m\lambda)\tau})^{-1}$, and $\tilde{\nu}$ is given by (4.10). Also,

$$[\mathbf{P}^l(\mathbf{I} - \mathbf{P})^{-1}]_{..} = me^{-l\lambda^*\tau}/(1 - e^{-\lambda^*\tau}) \qquad (l = 0, 1, \cdots).$$

On applying these results in the precision formulae, we find after some reduction

$$\text{Var } \tilde{\theta} \sim (k - 1)^{-1}\theta^2(M + \nu^{-1}C), \tag{4.18}$$

where

$$M = \frac{(3 + x_1)(1 + x_1)^{-1} + (m - 1)(3 + x_2)(1 + x_2)^{-1}}{[x_1(1 - x_1)^{-1} \log x_1 + (m - 1)x_2(1 - x_2)^{-1} \log x_2]^2},$$

$$C = \frac{\{m(1 + x_1)(1 - x_1)^{-1} - 2mF(1 - x_1)^{-2}(1 - x_2)^{-1} + \frac{1}{4}(D + E)(1 - x_1)^{-2}(1 - x_2)^{-2}\}}{[x_1(1 - x_1)^{-1} \log x_1 + (m - 1)x_2(1 - x_2)^{-1} \log x_2]^2},$$

and D, E, F, x_1 and x_2 are defined by

$$\begin{aligned} D \equiv D_m(x_1, x_2) = {} & 2m[1 - (1 - m^{-1})x_1 - m^{-1}x_2]^2 \cdot \\ & [1 + (1 + 2m^{-1})x_1 + 2(1 - m^{-1})x_2] + 4m^{-1}(1 - m^{-1})(x_1 - x_2)^3 \\ & -8m[1 - (1 - m^{-1})x_1 - m^{-1}x_2][m^{-1}x_1 + (1 - m^{-1})x_2 - x_1x_2], \end{aligned}$$

$$\begin{aligned} E \equiv E_m(x_1, x_2) = {} & 2m(1 - x_1)^{-1} \cdot \\ & [1 - m^{-1}x_1 - (1 - m^{-1})x_1^2 - (2 - m^{-1})x_2 + (2 - m^{-1})x_1x_2]^2, \end{aligned}$$

$$F \equiv F_m(x_1, x_2) = 1 - m^{-1}x_1 - (1 - m^{-1})x_1^2 - (2 - m^{-1})x_2 + (2 - m^{-1})x_1x_2,$$

$$x_1 = e^{-\lambda^*\tau} \equiv e^{-a_1\theta\tau}, \; x_2 = e^{-(\lambda^*+m\lambda)\tau} \equiv e^{-a_2\theta\tau}.$$

The matrix $\mathbf{A}$ has a_1 as its leading element and the $m - 1$ remaining diagonal elements are each a_2.

The term $[\theta^2/(k - 1)]M$ in (4.18) represents in general the dominant part of the error of estimation, while $[\theta^2/\{(k - 1)\nu\}]C$ represents a correction factor.

5. The efficiency of the MSCFE. We now discuss the large-sample relative efficiency of the MSCFE for model (C) in the limiting case $\nu \to \infty$.

The elements of the variance-covariance matrix of a finite set of unbiassed estimators are bounded below by the corresponding elements of the inverse of the information matrix. In particular, then, the variance of any unbiassed estimator of θ is bounded below by $J^{\theta\theta}$, where $\mathbf{J}$ is the 2×2 information matrix for θ and ν based on a set $\{\mathbf{n}_i\}$ of k vector observations. Furthermore, since the process $\{\mathbf{n}(t)\}$ tends to a Gaussian process (in the sense of footnote 7) as $\nu \to \infty$, and since for a normal Markov sequence of vectors the maximum likelihood estimators (MLE's) may be shown to be governed by the classic asymptotic theory of such estimation based on *independent* observations (cf. Bartlett, 1955, pp. 246–247; Mann and Wald, 1943; Billingsley, 1961a), the above lower bounds are actually attained asymptotically and the large-sample efficiency of the MSCFE is legitimately measured in terms of the ratio of the large-sample variance of the MLE[12] of θ to that of the MSCFE when ν is sufficiently large. The likelihood surface provides indeed a whole family of curves, relating information on θ to τ, for varying values of ν (each value of ν gives a member of the family). The latter curves have each a single minimal value at a certain critical value of τ which tends to 0 as $\nu \to \infty$. To enable the efficiency of generalized mean-square fluctuation to be gauged for reasonably large ν we examine the limiting member in the family ($\nu \to \infty$) together with the associated maximal precision at $\tau = 0$.

The limiting likelihood ($\nu \to \infty$) of ν and θ based on the set of observations $\{\mathbf{n}_i\}$ is

$$(5.1) \quad L = (2\pi\nu)^{-\frac{1}{2}km}|\mathbf{W}_k|^{-\frac{1}{2}}\exp[-\sum_{i,j,p,q} W_{k;p,q}^{(i,j)}\,(n_{ip} - \nu)(n_{jq} - \nu)/(2\nu)],$$

where $\mathbf{W}_k$ is a matrix which is partitionable into k^2 submatrices, each of size $m \times m$, such that the (i, j)th submatrix is $\mathbf{P}^{|j-i|}$ ($\mathbf{W}_k$ is the correlation matrix of the n_{ip}, the correlation between n_{ip} and n_{jq} being given by the element in the pth row and qth column of the (i, j)th submatrix), and $W_{k;p,q}^{(i,j)}$ denotes the element in the pth row and qth column of the (i, j)th submatrix of $\mathbf{W}_k^{-1}$. Elementary manipulations on $\mathbf{W}_k$ (of which the first step is the subtraction of the second column of submatrices from the first) readily gives $\mathbf{W}_k = |\mathbf{I} - \mathbf{P}^2||\mathbf{W}_{k-1}|$, and since $|\mathbf{W}_1| = |\mathbf{I}| = 1$,

$$(5.2) \qquad |\mathbf{W}_k| = |\mathbf{I} - \mathbf{P}^2|^{k-1} = [\prod_1^m (1 - e^{-2\kappa_r\tau})]^{k-1} \qquad (k = 1, 2, \cdots),$$

on using a previous result (Ruben, 1962; Equ. (2.19)) for the determinant of a non-singular rational function of $\mathbf{P}$ with scalar coefficients. Next, to determine the inverse of $\mathbf{W}_k$, observe that when $\mathbf{P}$ is scalar, $\mathbf{P} \equiv x$, the inverse is given by

[12] It should be remarked that maximum likelihood estimation of θ is intractable.

a Jacobi matrix (continuant) weighted by an appropriate factor, in the form

$$\mathbf{W}_k^{-1} = \begin{bmatrix} 1 & -x & 0 & \cdots & 0 & 0 \\ -x & 1+x^2 & -x & \cdots & 0 & 0 \\ 0 & -x & 1+x^2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & -x & 1 \end{bmatrix} \times (1-x^2)^{k-1} \qquad (m = 1).$$

This suggests that the inverse of $\mathbf{W}_k$ when $m > 1$ is a generalized Jacobi matrix weighted by an appropriate generalized diagonal matrix, in the form

$$\mathbf{W}_k^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{P} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ -\mathbf{P} & \mathbf{I}+\mathbf{P}^2 & -\mathbf{P} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{P} & \mathbf{I}+\mathbf{P}^2 & \cdots & \mathbf{0} & \mathbf{0} \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{P} & \mathbf{I} \end{bmatrix}$$

$$\cdot \begin{bmatrix} (\mathbf{I}-\mathbf{P}^2)^{-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}-\mathbf{P}^2)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{I}-\mathbf{P}^2)^{-1} & \cdots & \mathbf{0} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & (\mathbf{I}-\mathbf{P}^2)^{-1} \end{bmatrix},$$

i.e., the submatrices, $\mathbf{W}_k^{(i,j)}$, of $\mathbf{W}_k^{-1}$ are given by

$$\begin{aligned} \mathbf{W}_k^{(1,1)} &= (\mathbf{I}-\mathbf{P}^2)^{-1} = \mathbf{W}_k^{(k,k)} \\ \mathbf{W}_k^{(\alpha,\alpha)} &= (\mathbf{I}+\mathbf{P}^2)(\mathbf{I}-\mathbf{P}^2)^{-1} && (\alpha = 2, 3, \cdots, k-1), \\ \mathbf{W}_k^{(\alpha,\alpha+1)} &= -\mathbf{P}(\mathbf{I}-\mathbf{P}^2)^{-1} = \mathbf{W}_k^{(\alpha+1,\alpha)} && (\alpha = 1, 2, \cdots, k-1), \\ \mathbf{W}_k^{(\alpha,\beta)} &= \mathbf{0} && (|\alpha-\beta| > 1). \end{aligned} \tag{5.3}$$

The validity of (5.3) may now be verified by multiplication of $\mathbf{W}_k$ and $\mathbf{W}_k^{-1}$ in partitioned form. Finally, to obtain the first two derivatives with respect to θ of the various submatrices of $\mathbf{W}_k^{-1}$ needed for the information matrix, note that for any nonsingular rational function $\psi(\mathbf{P})$ of $\mathbf{P}$,

$$(\partial/\partial\theta)\psi(\mathbf{P}) = (\partial/\partial\theta)\{\boldsymbol{\Theta}^{-1}\psi(\mathbf{K})\boldsymbol{\Theta}\} = -\tau\boldsymbol{\Theta}^{-1}\mathbf{A}\mathbf{K}\psi'(\mathbf{K})\boldsymbol{\Theta}. \tag{5.4}$$

We then find from (5.1) and (5.2)

$$
(5.5)\quad
\begin{aligned}
-\frac{\partial^2}{\partial \nu^2}\log L &= -\frac{km}{2\nu^2}+\frac{1}{\nu^3}\sum_{i,j,p,q} W_{k;p,q}^{(i,j)} n_{ip}n_{jq},\\
-\frac{\partial^2}{\partial \nu\partial\theta}\log L &= \frac{1}{2}\sum_{i,j,p,q}\frac{\partial}{\partial\theta}W_{k;p,q}^{(i,j)}\left(1-\frac{n_{ip}n_{jq}}{\nu^2}\right),\\
-\frac{\partial^2}{\partial \theta^2}\log L &= -2(k-1)\tau^2\sum_r \frac{a_r^2 e^{-2\kappa_r\tau}}{(1-e^{-2\kappa_r\tau})^2}\\
&\qquad +\frac{1}{2\nu}\sum_{i,j}\sum_{p,q}\frac{\partial^2}{\partial\theta^2}W_{k;p,q}^{(i,j)}(n_{ip}-\nu)(n_{jq}-\nu),
\end{aligned}
$$

whence by (5.4)

$$
(5.6)\quad
\begin{aligned}
&\partial \mathbf{W}_k^{(1,1)}/\partial\theta\\
&\quad = -2\tau\cdot\boldsymbol{\Theta}^{-1}\mathbf{A}\mathbf{K}^2(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta} = \partial\mathbf{W}_k^{(k,k)}/\partial\theta,\\
&\partial \mathbf{W}_k^{(\alpha,\alpha)}/\partial\theta\\
&\quad = -4\tau\cdot\boldsymbol{\Theta}^{-1}\mathbf{A}\mathbf{K}^2(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta} \qquad (\alpha = 2, 3, \cdots, k-1),\\
&\partial \mathbf{W}_k^{(\alpha,\alpha+1)}/\partial\theta\\
&\quad = \tau\cdot\boldsymbol{\Theta}^{-1}\mathbf{A}\mathbf{K}(\mathbf{I}+\mathbf{K}^2)(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta} = \partial\mathbf{W}_k^{(\alpha+1,\alpha)}/\partial\theta\\
&\qquad\qquad (\alpha = 1, 2, \cdots, k-1),\\
&\partial \mathbf{W}_k^{(\alpha,\beta)}/\partial\theta\\
&\quad = \mathbf{0} \qquad (|\alpha-\beta|>1),
\end{aligned}
$$

and

$$
(5.7)\quad
\begin{aligned}
&\partial^2 \mathbf{W}_k^{(1,1)}/\partial\theta^2\\
&\quad = 4\tau^2\cdot\boldsymbol{\Theta}^{-1}\mathbf{A}^2\mathbf{K}^2(\mathbf{I}+\mathbf{K}^2)(\mathbf{I}-\mathbf{K}^2)^{-3}\boldsymbol{\Theta} = \partial^2\mathbf{W}_k^{(k,k)}/\partial\theta^2,\\
&\partial^2 \mathbf{W}_k^{(\alpha,\alpha)}/\partial\theta^2\\
&\quad = 8\tau^2\cdot\boldsymbol{\Theta}^{-1}\mathbf{A}^2\mathbf{K}^2(\mathbf{I}+\mathbf{K}^2)(\mathbf{I}-\mathbf{K}^2)^{-3}\boldsymbol{\Theta} \qquad (\alpha = 2, 3, \cdots, k-1),\\
&\partial^2 \mathbf{W}_k^{(\alpha,\alpha+1)}/\partial\theta^2\\
&\quad = -\tau^2\cdot\boldsymbol{\Theta}^{-1}\mathbf{A}^2(\mathbf{K}+6\mathbf{K}^3+\mathbf{K}^5)(\mathbf{I}-\mathbf{K}^2)^{-3}\boldsymbol{\Theta} = \partial^2\mathbf{W}_k^{(\alpha+1,\alpha)}/\partial\theta^2\\
&\qquad\qquad (\alpha = 1, 2, \cdots, k-1),\\
&\partial^2 \mathbf{W}_k^{(\alpha,\beta)}/\partial\theta^2\\
&\quad = \mathbf{0} \qquad (|\alpha-\beta|>1).
\end{aligned}
$$

The elements of the information matrix are now derived from (5.5) with the aid of (5.6) and (5.7). We have

$$E\left(-\frac{\partial^2 \log L}{\partial \nu^2}\right) = -\frac{km}{2\nu^2} + \frac{1}{\nu^3}\sum_{i,j,p,q} W_{k;p,q}^{(i,j)}(\nu^2 + \nu P_{pq}(|j-i|\tau)) \tag{5.8}$$

$$= \frac{km}{2\nu^2} + \frac{1}{\nu}\sum_{i,j,p,q} W_{k;p,q}^{(i,j)},$$

since

$$\sum_{i,j,p,q} W_{k;p,q}^{(i,j)} P_{pq}(|j-i|\tau) = tr(\mathbf{W}_k^{-1}\mathbf{W}_k) = km,$$

and on further simplification

$$E\left(-\frac{\partial^2 \log L}{\partial \nu^2}\right) = \frac{km}{2\nu^2} + \frac{1}{\nu}\{2[(\mathbf{I}-\mathbf{P}^2)^{-1}]..$$

$$+ (k-2)[(\mathbf{I}+\mathbf{P}^2)(\mathbf{I}-\mathbf{P}^2)^{-1}]..$$

$$- 2(k-1)[\mathbf{P}(\mathbf{I}-\mathbf{P}^2)^{-1}]..\} \tag{5.9}$$

$$= \frac{km}{2\nu^2} + \frac{1}{\nu}[(k\mathbf{I} - (k-2)\mathbf{P})(\mathbf{I}+\mathbf{P})^{-1}]..$$

$$\sim (k/\nu)[(\mathbf{I}-\mathbf{P})(\mathbf{I}+\mathbf{P})^{-1}]..,$$

after neglecting negative powers of ν higher than the first and remembering that k is considered large. Again,

$$E\left(-\frac{\partial^2 \log L}{\partial \nu \partial \theta}\right) = -\frac{1}{2\nu}\sum_{i,j,p,q}\frac{\partial}{\partial \theta} W_{k;p,q}^{(i,j)} P_{pq}(|j-i|\tau)$$

$$= -\frac{1}{2\nu}\sum_{p,q}\{[-4\tau\boldsymbol{\Theta}^{-1}\mathbf{AK}^2(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta}$$

$$- 4(k-2)\tau\boldsymbol{\Theta}^{-1}\mathbf{AK}^2(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta}]_{pq}P_{pq}(0)$$

$$+ 2(k-1)\tau[\boldsymbol{\Theta}^{-1}\mathbf{AK}(\mathbf{I}+\mathbf{K}^2)(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta}]_{pq}P_{pq}(\tau)\}$$

$$= \frac{2(k-1)\tau}{\nu} tr\{\boldsymbol{\Theta}^{-1}\mathbf{AK}^2(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta}\} \tag{5.10}$$

$$- \frac{(k-1)\tau}{\nu} tr\{\boldsymbol{\Theta}^{-1}\mathbf{AK}^2(\mathbf{I}+\mathbf{K}^2)(\mathbf{I}-\mathbf{K}^2)^{-2}\boldsymbol{\Theta}\}$$

$$= \frac{(k-1)\tau}{\nu} tr\{\boldsymbol{\Theta}^{-1}\mathbf{AK}^2(\mathbf{I}-\mathbf{K}^2)^{-1}\boldsymbol{\Theta}\}$$

$$= \frac{(k-1)\tau}{\nu}\sum_r \frac{a_r e^{-2\kappa_r\tau}}{1-e^{-2\kappa_r\tau}}.$$

Proceeding in a similar manner,

$$E\left(-\frac{\partial^2 \log L}{\partial\theta^2}\right) = -2(k-1)\tau^2 \sum_r \frac{a_r^2 e^{-2\kappa_r\tau}}{(1-e^{-2\kappa_r\tau})^2} + \frac{1}{2}\sum_{i,j,p,q} \frac{\partial^2}{\partial\theta^2} W_{k;p,q}^{(i,j)} P_{pq}(|j-i|\tau)$$

$$= (k-1)\tau^2 \sum_r \frac{a_r^2 e^{-2\kappa_r\tau}(1+e^{-2\kappa_r\tau})}{(1-e^{-2\kappa_r\tau})^2},$$

after some reduction. The large-sample variances of the MLE's of ν and θ, denoted by $\hat{\nu}$ and $\hat{\theta}$, are therefore given from (5.8), (5.9) and (5.10) by

$$\text{(5.11)} \qquad \operatorname{Var} \hat{\nu} \sim \frac{\nu}{k} \frac{1}{[(\mathbf{I}-\mathbf{P})(\mathbf{I}+\mathbf{P})^{-1}]_{..}}$$

and

$$\text{(5.12)} \qquad \operatorname{Var} \hat{\theta} \sim \frac{1}{(k-1)\tau^2}\left(\sum_r \frac{a_r^2 e^{-2\kappa_r\tau}(1+e^{-2\kappa_r\tau})}{(1-e^{-2\kappa_r\tau})^2}\right)^{-1},$$

the two estimators being uncorrelated to order $\nu^{-\frac{1}{2}}$. Now

$$\text{(5.13)} \qquad \lim_{\tau\to 0} \operatorname{Var} \hat{\theta} = \frac{2\theta^2}{(k-1)m},$$

and since this is also the limiting value of Var $\tilde{\theta}$ as $\tau \to 0$ on the variance curve of $\tilde{\theta}$ corresponding to $\nu = \infty$ (see Equ. (4.17), we conclude that the proposed method of estimation based on generalized mean-square fluctuations in a long sequence of vector observations (k large) is highly efficient in model (C) over all θ if the frequency of observations is high[13] (τ small) and ν is sufficiently large for these observations to be regarded as normally distributed. This result suggests that the method of estimation may have high efficiency in more general circumstances for high ν_r and sufficiently frequent observations, at any rate over the effectively linear portions of λ_{rs} and λ_r^*, regarded as analytic functions of θ.

REFERENCES

[1] ALBERT, ARTHUR (1962). Estimating the infinitesimal generator of a continuous time, finite state Markov process. *Ann. Math. Statist.* **33** 727–753.

[2] ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–21.

[3] BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge Univ. Press.

[4] BARTLETT, M. S. (1962). *Essays on Probability and Statistics*. Wiley, New York.

[5] BILLINGSLEY, PATRICK (1961 a). *Statistical Inference for Markov Processes*. Institute of Mathematical Statistics—Univ. of Chicago Statistical Research Monographs, Univ. of Chicago Press.

[6] BILLINGSLEY, PATRICK (1961 b). Statistical methods in Markov chains. *Ann. Math. Statist.* **32** 12–40.

[13] Note from (4.15) and (5.12) that $\lim_{\tau\to\infty} (\operatorname{Var} \hat{\theta}/\operatorname{Var} \tilde{\theta}) = \omega/(3m)$ for $\nu \to \infty$, where ω is the multiplicity of $\min_r \kappa_r$, so that the large-sample efficiency of the MSCFE as $\tau \to \infty$ (and $\nu \to \infty$) cannot exceed 1/3.

[7] CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[8] DARWIN, J. H. (1959). The behaviour of an estimator for a simple birth and death process. *Biometrika* **46** 412–419.

[9] FELLER, WILLIAM (1957). *An Introduction to Probability Theory and its Applications*, **1** (2nd ed.). Wiley, New York.

[10] HOEFFDING, WASSILY and ROBBINS, HERBERT (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15** 773–780.

[11] KENDALL, D. G. (1952). Les processus stochastiques de croissance en biologie. *Ann. Inst. Poincaré* **13**, fasc. 1 43–108.

[12] MAHALANOBIS, P. C., BOSE, R. C. and ROY, S. N. (1936). Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. *Sankhyā* **3** 1–40.

[13] MANN, H. B. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–220.

[14] MEIER, PAUL (1955). Note on estimation in a Markov process with constant transition probability. *Human Biology* **27** 121–124.

[15] MORAN, P. A. P. (1951). Estimation methods for evolutive processes. *J. Roy. Statist. Soc. Ser. B* **13** 141–146

[16] MORAN, P. A. P. (1953). The estimation of the parameters of a birth and death process. *J. Roy. Statist. Soc. Ser. B* **15** 241–245.

[17] PATIL, V. T. (1955). A study of some stochastic migration models with particular reference to certain biological phenomena. Ph.D. Thesis, Manchester University. Unpublished.

[18] PATIL, V. T. (1957). The consistency and adequacy of the Poisson-Markoff model for density fluctuations. *Biometrika* **44** 43–56.

[19] ROTHSCHILD, LORD (1953). A new method of measuring the activity of spermatozoa. *J. Experimental Biology* **30** 178–199.

[20] RUBEN, HAROLD and ROTHSCHILD, LORD (1953). Estimation of mean speeds of organisms and particles by counting. Unpublished.

[21] RUBEN, HAROLD (1962). Some aspects of the emigration-immigration process. *Ann. Math. Statist.* **33** 111–129.

[22] ZAHL, SAMUEL (1955). A Markov process for follow-up studies. *Human Biology* **27** 90–120.