ESTIMATING THE INFINITESIMAL GENERATOR OF A CONTINUOUS TIME, FINITE STATE MARKOV PROCESS[1]

By Arthur Albert[2]

Columbia University

0. Summary. Let $\{Z(t), t > 0\}$ be a separable, continuous time Markov Process with stationary transition probabilities $P_{ij}(t), i, j = 1, 2, \cdots, M$. Under suitable regularity conditions, the matrix of transition probabilities, $P(t)$, can be expressed in the form $P(t) = \exp tQ$, where Q is an $M \times M$ matrix and is called the "infinitesimal generator" for the process.

In this paper, a density on the space of sample functions over $[0, t)$ is contructed. This density depends upon Q. If Q is unknown, the maximum likelihood estimate $\hat{Q}(k, t) = \|\hat{q}_{ij}(k, t)\|$, based upon k independent realizations of the process over $[0, t)$ can be derived.

If each state has positive probability of being occupied during $[0, t)$ and if the number of independent observations, k, grows larger (t held fixed), then $\hat{q}_{ij}$ is strongly consistent and the joint distribution of the set $\{k^{\frac{1}{2}}(\hat{q}_{ij} - q_{ij})\}_{i \neq j}$ (suitably normalized), is asymptotically normal with zero mean and covariance equal to the identity matrix.

If k is held fixed (at one, say) and if t grows large, then $\hat{q}_{ij}$ is again strongly consistent and the joint distribution of the set $\{t^{\frac{1}{2}}(\hat{q}_{ij} - q_{ij})\}_{i \neq j}$ (suitably normalized), is asymptotically normal with zero mean and covariance equal to the identity matrix, provided that the process $\{Z(t), t > 0\}$ is positively regular.

The asymptotic variances of the $\hat{q}_{ij}$ are computed in both cases.

1. Mathematical formulation. The probabilistic behavior of a finite state, continuous time Markov process $\{Z(t), t \geqq 0\}$ is determined by a knowledge of the *transition probability* matrix $P(t, s)$, whose entries are

$$P_{ij}(t, s) = \Pr [Z(s) = j \mid Z(t) = i]$$

(where $t \leqq s$ and i and j range over the possible states of the process). The process is said to have *stationary transition probabilities* if $P(t, s)$ depends only on the difference between s and t:

$$P(t, s) = P(s - t).$$

Received October 2, 1960; revised June 5, 1961.

[1] This work was initiated while the author was a National Science Postdoctoral Fellow at Stockholm's Institut för Försäkringsmatematik och Matematisk Statistik in 1959. It first appeared in August, 1960 as Technical Report No. 60 at the Applied Mathematics and Statistics Laboratories, Stanford University, where it was prepared under ONR Contract No. Nonr-225(52) (NR 342-022). The present revision was prepared while the author was at Columbia University.

[2] Present address: Communications Biophysics Laboratory, Room 20B-221, Massachusetts Institute of Technology, Cambridge 39, Mass.

If certain regularity conditions on the behavior of the (stationary) transition probabilities are met, the so-called *forward* and *backward systems* of differential equations can be derived:

$$P'(t) = QP(t) = P(t)Q; \qquad P(0) = I,$$

where Q(called the *infinitesimal generator* for the process) is a square matrix whose value is constant in time.

The (unique bounded) solution to the forward and backward equations is

$$P(t) = \exp tQ = \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!}.$$

In this paper, we will discuss the problem of estimating Q when it is not known. In particular, a maximum likelihood estimate for Q will be derived and its large-sample properties investigated. The results obtained here are the continuous time analogues of the results stated by Anderson and Goodman in [1], which describe the asymptotic behavior of the maximum likelihood estimate of the transition probability matrix of a (discrete time) Markov chain.

The question of defining a maximum likelihood estimate (hereafter denoted by m.l.e.) is not an altogether trivial one. Customarily, an m.l.e. is defined relative to a density (which depends upon the parameter to be estimated) over the sample space of the experiment. In the case at hand, the sample space will be a function space (though happily a simple one), and so, one of the first major tasks to be attended to is the construction of a density over the set of realizations for the process. Once this is accomplished, we will see that the m.l.e. is quite simply expressed as a function of the observations, and the remainder of the paper will concern itself with questions of consistency and asymptotic distribution theory.

2. General properties of finite-state Markov processes. Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\{Z(t, \cdot); t \geqq 0\}$ be a *separable* Markov process on that space, which takes its values in a finite set (which for convenience will be taken to be the set $\{1, 2, \cdots, M\}$). (Occasionally, the second argument of $Z(\cdot, \cdot)$ will be suppressed as a notational convenience.) The value of $Z(\cdot, \cdot)$ at time t is a random variable which will be called the *state of the process* at time t.

It is assumed that the transition probability function is *homogeneous in time* (stationary):

For any $s, t \geqq 0$,

$$P[Z(t + s) = j \mid Z(s) = i] = P_{ij}(t)$$

depends only upon the time difference t. It is further assumed that

$$\lim_{h\to 0}[(1 - P_{ii}(h))/h](= q(i)) \quad \text{exists for all } i,$$

and

$$\lim_{h\to 0}[(P_{ij}(h))/h](= q(i, j)) \quad \text{exists for all } i \text{ and } j, \quad j \neq i.$$

The following theorem can be established on the basis of these assumptions, and will serve as the basis for constructing a density on the space of sample functions. For a proof, the reader is referred to Chapter VI of [6].

THEOREM 2.1.

(a) *Let*

$$q(i, j) = \begin{cases} -q(i) = -\sum_{j \neq i} q(i, j) & \text{if } i = j, \\ q(i, j) & \text{if } i \neq j, \end{cases}$$

and let Q *be the* $(M \times M)$ *matrix whose* (i, j)th *entry is* $q(i, j)$. *The matrix of transition probability functions is given by* $P(t) = \exp tQ$.

(b) $$P[Z(t) = i, \quad t_0 \leqq t \leqq t_0 + \alpha \mid Z(t_0) = i] = \exp - q(i)\alpha$$

for all non-negative t_0 *and* α.

(c) *If* $Z(t_0) = i$ *and* $q(i) > 0$, *there is, with probability one, a sample function discontinuity for some* $t > t_0$, *and in fact, a first discontinuity which is a jump. If* $0 < \alpha \leqq \infty$, *the conditional probability that the first discontinuity in* $[t_0, t_0 + \alpha)$ *is a jump to* j, *given that* $Z(t_0) = i$ *and that there is a discontinuity in* $[t_0, t_0 + \alpha)$, *is* $q(i, j)/q(i)$.

(d) *Almost all sample functions are step functions with a finite number of jumps in any finite time interval.*

3. The space of sample functions. Suppose observations are made on the process $\{Z(t); 0 \leqq t < T\}$ (where T is finite). By virtue of Theorem 2.1, a sample function can be specified by a knowledge of the number of jumps made in $[0, T)$, the (ordered) lengths of time between jumps, and the succession of values taken on by the process in $[0, T)$.

To be more precise, suppose Ω' is that subset of the underlying probability space for which sample functions of the process $\{Z(t, \cdot), t \geqq 0\}$ are step functions with a finite number of jumps in any finite interval. By Theorem 2.1, this set has probability one. Now, let us define the following random variables (r.v.'s):

$$\tau_0(\omega) = 0, \qquad \tau_i(\omega) = \begin{cases} \text{The time at which the } i\text{th jump occurs if } \omega \in \Omega' \\ + \infty \text{ otherwise,} \end{cases}$$

$$T_i(\omega) = \begin{cases} \tau_{i+1}(\omega) - \tau_i(\omega) & \text{if } \tau_i(\omega) < \infty \\ 0 & \text{otherwise,} \end{cases}$$

$$N(T, \omega) = \text{The largest integer, } n, \text{ for which } \tau_n(\omega) < T,$$

$$Z_i(\omega) = Z(\tau_i(\omega), \omega) \qquad i = 0, 1, 2, \cdots.$$

(With probabilities one, $T_i(\cdot)$ is the time spent in the ith state, $N(T, \cdot)$ is the number of jumps made by the process in $[0, T)$, and $Z_i(\cdot)$ is the state of the process immediately after the ith jump.)

In fact, with probability one, a sample function of $\{Z(t, \cdot), 0 \leqq t < T\}$ can be represented as an ordered sequence:

$$\{Z(t, \omega), 0 \leqq t < T\} = ((Z_0(\omega), T_0(\omega)), \cdots, (Z_{N(T,\omega)-1}(\omega), T_{N(T,\omega)-1}(\omega)), Z_{N(T,\omega)}).$$

By this we mean: If

$$\{Z(t, \omega), 0 \leqq t < T\} = ((z_0, t_0), \cdots, (z_{n-1}t_{n-1}), z_n),$$

then the path function starts at z_0 at time zero, remains in z_0 for t_0 units of time, makes a jump to z_1, remains in z_1 for t_1 units of time, $\cdots$, jumps to z_{n-1}, remains there for t_{n-1} units of time and then makes the final jump to z_n, and remains there at least until time T. (Note that n jumps, in all, have been made.)

We can write down the probability distribution on the space of sample functions quite easily now:

THEOREM 3.1. *Let*

$$q'(i, j) = \begin{cases} 0 & \text{if } i = j \\ q(i, j) & \text{if } i \neq j. \end{cases}$$

Then

$$\begin{aligned} \Pr [N(T) = n \;\&\; Z_0 = z_0 \;\&\; T_0 \leqq \alpha_0 \;\&\; \cdots \;\&\; Z_{n-1} = z_{n-1} \;\&\; T_{n-1} \\ \leqq \alpha_{n-1} \;\&\; Z_n = z_n] = \Pr [Z(0) = z_0] \exp - q(z_n)T \\ \cdot \int_{S_n} \prod_{j=0}^{n-1} dt_j \, q'(z_j, z_{j+1}) \exp - [q(z_j) - q(z_n)] \, t_j \end{aligned}$$

where

$$S_n = \left\{(t_0, t_1, \cdots, t_{n-1}): \sum_{j=0}^{n-1} t_j < T \;\&\; 0 \leqq t_j \leqq \alpha_j\right\} \qquad \text{if } n > 0.$$

$$\Pr [N(T) = 0 \;\&\; Z_0 = z_0] = \Pr [Z(0) = z_0] \exp - q(z_0)T.$$

PROOF. The second assertion follows directly from Theorem 2.1(b). If $n > 0$, then

$$\begin{aligned} \Pr [N(T) = n \;\&\; Z_0 = z_0 \;\&\; T_0 \leqq \alpha_0 \;\&\; \cdots \;\&\; Z_{n-1} = z_{n-1} \;\&\; T_{n-1} \leqq \alpha_{n-1} \;\&\; Z_n = z_n] \\ = \Pr\left\{\left[Z_0 = z_0 \;\&\; \cdots \;\&\; Z_n = z_n \;\&\; \sum_{j=0}^{n} T_j \geqq T\right] \cap \tilde{S}_n\right\}, \end{aligned}$$

where

$$\tilde{S}_n = [(T_0, T_1, \cdots, T_{n-1}) \varepsilon S_n].$$

By Theorem 2.1(b) and (c)

$$\Pr [Z_n = z_n \mid Z_0 = z_0, \cdots, Z_{n-1} = z_{n-1}, T_0, \cdots, T_{n-1}] = \frac{q'(z_{n-1}, z_n)}{q(z_{n-1})}$$

and

$$\Pr [T_n \leqq \alpha_n \mid Z_0 = z_0, \cdots, Z_n = z_n, T_0, \cdots, T_{n-1}] = 1 - e^{-q(z_n)\alpha_n}.$$

By induction,

$$\Pr[Z_0 = z_0, \cdots, Z_n = z_n, T_0 \leqq \alpha_0, \cdots, T_n \leqq \alpha_n] =$$
$$\Pr[Z(0) = z_0]\left\{\prod_{j=0}^{n-1} \frac{q'(z_j, z_{j+1})}{q(z_j)} [1 - e^{-q(z_j)\alpha_j}]\right\} [1 - e^{-q(z_n)\alpha_n}].$$

It follows that

$$\Pr\left\{\left[Z_0 = z_0, \cdots, Z_n = z_n \,\&\, \sum_{j=0}^{n} T_j \geqq T\right] \cap \tilde{S}_n\right\} =$$
$$\Pr[Z(0) = z_0] \int_{S_n} \sum_{j=0}^{n-1} dt_j \left\{\left\{\int_{T-\sum_{j=0}^{n-1} t_j}^{\infty} q(z_n)\, e^{-q(z_n)t_n}\, dt_n\right\} \cdot \prod_{j=0}^{n-1} q'(z_j, z_{j+1})\, e^{-q(z_j)t_j}\right.\, .$$

The conclusion is obtained by performing the inner integration.

It is now (conceptually) quite simple to construct the desired density:

Every sample function which makes n jumps in $[0, T)$ can be represented as a point in

$$\mathcal{W}_n = \left[\prod_{j=1}^{n} (\mathcal{W}_0 \otimes R^1)\right] \otimes \mathcal{W}_0,$$

where $\mathcal{W}_0 = \{1, 2, \cdots, M\}$ (the state space of the process) and R^1 is the real line.

Let l be the Lebesgue measure on R^1 and let C be the counting measure on $\mathcal{W}_0$:

$$C(\{z\}) = 1 \quad \text{if} \quad z \in \mathcal{W}_0.$$

Let $\sigma^{(n)}$ be the (sigma-finite) product measure on $\mathcal{W}_n$, defined by the relation:

$$\sigma^{(n)} = \left[\prod_{j=1}^{n} (C \times l)\right] \times C.$$

By Theorem 2.1(d), almost every sample function of the process $\{Z(t), 0 \leqq t < T\}$ can be represented as a point in

$$\mathcal{W} = \bigcup_{n=0}^{\infty} \mathcal{W}_n.$$

For each set $W \subseteq \mathcal{W}$ for which $W \cap \mathcal{W}_n$ is $\sigma^{(n)}$ measurable define $\sigma^*(W) = \sum_{n=0}^{\infty} \sigma^{(n)}(W \cap \mathcal{W}_n)$. ($\sigma^*$ is defined on the Borel-field $\mathcal{B}^*$, which is the smallest Borel-field containing all sets $W \subseteq \mathcal{W}$ whose projection on $\prod_{j=1}^{n} \otimes R^1$ is a Borel set for each n.) Let σ be a measure on the space of *all* sample functions, defined for all subsets B whose intersection with $\mathcal{W}$ is in $\mathcal{B}^*$:

$$\sigma(B) = \sigma^*(B \cap \mathcal{W}).$$

(σ is a sigma-finite measure.)

The desired density function can now be exhibited. (It is a density with respect to σ.)

THEOREM 3.2. *If B is a subset of the space of all sample functions over $[0, T)$ which is measurable with respect to σ, then*

$$\Pr[B] = \int_B f_Q(v)\, d\sigma(v),$$

where

$$f_Q(v) = \begin{cases} \Pr[Z(0) = z_0] e^{-q(z_0)T} & \text{if } v = (z_0); \\ \Pr[Z_0 = z_0] e^{-q(z_n)T} \prod_{j=0}^{n-1} q'(z_j, z_{j+1}) \exp - [q(z_j) - q(z_n)] t_j & \\ \quad \text{if } v = ((z_0, t_0), \cdots, (z_{n-1}, t_{n-1}), z_n) & \\ \quad \text{with } n > 0,\ z_j \varepsilon \mathcal{W}_0,\ t_j \geqq 0\ (j = 0, 1, \cdots, n-1), & \\ \quad \text{and } \sum_{j=0}^{n-1} t_j < T; & \\ 0 & \text{otherwise.} \end{cases}$$

PROOF. By Theorem 3.1, the conditional distribution of $((Z_0, T_0) \cdots (Z_{n-1}, T_{n-1}), Z_n)$ given that $N(T) = n$ is, for each n, absolutely continuous with respect to $\sigma^{(n)}$. If the derivative exists,

$$g_n(z_0, z_1, \cdots, z_n, t_0, t_1, \cdots, t_{n-1}) = \frac{\partial^n}{\partial t_0 \cdots \partial t_{n-1}} \cdot \frac{\Pr[N(T) = n \,\&\, Z_0 = z_0 \,\&\, \cdots \,\&\, Z_n = z_n \,\&\, T_0 \leqq t_0 \,\&\, \cdots \,\&\, T_{n-1} \leqq t_{n-1}]}{\Pr[N(T) = n]}$$

is the (joint) conditional density of $((Z_0, T_0), \cdots, (Z_{n-1}, T_{n-1}), Z_n)$ with respect to $\sigma^{(n)}$, given that $N(T) = n$ (for $n > 0$). In this case,

$$\Pr[B \mid N(T) = n] = \int_{B \cap \mathcal{W}_n} g_n\, d\sigma^{(n)}$$

for all sets B whose intersection with $\mathcal{W}_n$ is $\sigma^{(n)}$ measurable, so that if B is a subset of the space of all possible sample functions, and $B \cap \mathcal{W}_n$ is $\sigma^{(n)}$ measurable for each n, then

$$\Pr[B] = \sum_{n=0}^{\infty} \int_{B \cap \mathcal{W}_n} \Pr[N(T) = n]\, g_n\, d\sigma^{(n)}.$$

A routine computation (in connection with Theorem 3.1) shows that the derivative in question does indeed exist and that $g_n \Pr[N(T) = n]$ corresponds to f_Q on $\mathcal{W}_n$. Thus,

$$\Pr[B] = \sum_{n=0}^{\infty} \int_{B \cap \mathcal{W}_n} f_Q\, d\sigma^{(n)} = \int_{B \cap \mathcal{W}} f_Q\, d\sigma^* = \int_B f_Q\, d\sigma.$$

Although we shall not discuss the hypothesis testing problem here, the following result is quite important in that context and is included here since it follows so readily from the results established thus far.

THEOREM 3.3. *Let* $Q_0 = \|q_0(i, j)\|$ *and* $Q_1 = \|q_1(i, j)\|$ $i, j = 1, 2, \cdots, M$, *be two values of the parameter* Q.

(a) *If* $q_0(i, j)$ *vanishes whenever* $q_1(i, j)$ *does, then the probability measure on* $(\Omega, \mathfrak{A}, P)$ *under* Q_0 *is absolutely continuous with respect to the one under* Q_1.

(b) *If for some* i *and* $j (j \neq i)$, $q_1(i,j) = 0$ *and* $q_0(i,j) > 0$ *and* $P_{Q_0}[Z(t_0) = i] > 0$ *for some* $t_0 \geqq 0$, *then* P_{Q_0} *is not absolutely continuous with respect to* P_{Q_1}.

PROOF.

(a) Under stated conditions, f_{Q_0} vanishes with f_{Q_1}.

(b) Let B be the set of sample functions whose value at time t_0 is i and whose next (distinct) value is j. By Theorem 2.1(c), $P_{Q_1}[B] = 0$ and $P_{Q_0}[B] > 0$.

It can be shown that $P_{ij}(t)$ never vanishes for $t > 0$ unless it vanishes identically (see [6], Ch. VI). It follows then, that Pr $[Z(t) = i]$ never vanishes for $t > 0$ unless it too vanishes identically. If the parameter space is restricted to those values of Q for which every state i has positive probability of being occupied eventually, then a necessary and sufficient condition that P_{Q_0} be absolutely continuous with respect to P_{Q_1} is that $q_0(i, j)$ vanish whenever $q_1(i, j)$ does.

4. Sufficient statistics and maximum likelihood estimates. Suppose k independent realizations $v_1, v_2, \ldots, v_k$ of $\{Z(t), 0 \leqq t < T\}$ are observed. The likelihood function, $\mathfrak{L}_Q^{(k)}$, has been traditionally defined by the equation

$$\mathfrak{L}_Q^{(k)} = \prod_{j=1}^{k} f_Q(v_j).$$

If we let $N_T^{(k)}(i, j)$ = the total number of transitions from state i to state j observed during the k trials and $A_T^{(k)}(i)$ = the total length of time that state i is occupied during the k trials, we may write

$$\log \mathfrak{L}_Q^{(k)} = C_k + \sum_i \sum_{j \neq i} N_T^{(k)}(i, j) \log q(i, j) - \sum_i A_T^{(k)}(i) q(i),$$

where C_k is finite with probability one and does not depend upon Q.

The Halmos-Savage factorization theorem can be applied to the last expression, and by inspection we see that the set

$$\{N_T^{(k)}(i, j), A_T^{(k)}(i)\}_{j \neq i}$$

is a sufficient statistic for Q. The maximum likelihood estimates (m.l.e) for $q(i, j)$ (i.e., those values of $q(i, j)$ which maximize $\log \mathfrak{L}_Q^{(k)}$) are seen to be

$$\hat{q}_T^{(k)}(i, j) = N_T^{(k)}(i, j)/A_T^{(k)}(i) \quad \text{if} \quad i \neq j \ \& \ A_T^{(k)}(i) > 0.$$

If $A_T^{(k)} = 0$, the m.l.e. does not exist and so we adopt the convention that

$$\hat{q}_T^{(k)}(i, j) = 0 \quad \text{if} \quad i \neq j \quad \text{and} \quad A_T^{(k)}(i) = 0.$$

5. Moments. In the sections to follow, we will investigate the large sample properties of the set $\{\hat{q}_T^{(k)}(i, j)\}_{j \neq i}$ both as T approaches infinity while k is fixed, and as k approaches infinity with T fixed. A knowledge of certain moments of $N_T^{(k)}(i, j)$ and $A_T^{(k)}(i)$ proves to be of use, and the following lemma reduces the determination of these moments to a fairly routine computational task:

For notational convenience, we will let $N_T(i, j) = N_T^{(1)}(i, j)$ and $A_T(i) = A_T^{(1)}(i)$ for all i and j. $N(T)$ is as defined in Section 3.

LEMMA 5.1.

(a) *There are constants* α *and* β *such that*

$$\Pr[N(h) \geqq n] \leqq \beta \int_0^h \frac{t^{n-1} e^{\alpha t}}{1-t} dt$$

for all h between zero and one. Consequently,

$$\sum_{n=2}^{\infty} \Pr[N(h) \geqq n] \leqq \beta \int_0^h \frac{te^{\alpha t}\, dt}{(1-t)^2} = o(h), \qquad \text{as } h \to 0.$$

(b) $\Pr[N_h(i, j) = 1] = p(i)q(i, j)h + o(h)$ *as* $h \to 0$ *and* $EN_h(i, j) = p(i)q(i, j)h + o(h)$ *as* $h \to 0$. (*Here*, $p(i) = \Pr[Z(0) = i]$.)

PROOF.

(a) By Theorem 3.1, there are constants K, α, and γ, such that

$$\Pr[N(h) = n] \leqq K\gamma^n \int_{0 \leqq \sum_{j=1}^n t_j \leqq h} \cdots \int \prod_{j=1}^n e^{\alpha t_j}\, dt_j$$

$$= K' \frac{(\gamma')^{n-1}}{(n-1)!} \int_0^h t^{n-1} e^{\alpha t}\, dt,$$

$$\leqq \beta \int_0^h t^{n-1} e^{\alpha t}\, dt$$

where $\gamma' > 0$ and $\beta = K'e^{\gamma'}$.

(b) By Theorem 2.1,

$$\Pr[N_h(i, j) = 1 \mid N(h) = 1 \;\&\; Z(0) = i] = q(i, j)/q(i).$$

By Theorem 3.1,

$$\Pr[N(h) = 1 \mid Z(0) = i] = q(i)h + o(h).$$

Whence,

$$\Pr[N_h(i, j) = 1] = p(i)q(i, j)h + o(h)$$

Since

$$EN_h(i, j) = \sum_{n=1}^{\infty} [\Pr N_h(i, j) \geqq n]$$

$$= \Pr[N_h(i, j) = 1] + \Pr[N_h(i, j) \geqq 2] + \sum_{n=2}^{\infty} \Pr[N_h(i, j) \geqq n].$$

From part (a)

$$\Pr[N_h(i,j) \geqq 2] \leqq \Pr[N(h) \geqq 2] = o(h)$$

and

$$\sum_{n=2}^{\infty} \Pr[N_h(i,j) \geqq n] \leqq \sum_{n=2}^{\infty} \Pr[N(h) \geqq n] = o(h).$$

THEOREM 5.1.

(a)
$$EN_T(i,j) = q(i,j) \int_0^T [\Pr Z(t) = i]\, dt.$$

(b)
$$EA_T(i) = \int_0^T \Pr[Z(t) = i]\, dt.$$

(c)
$$EN_T(i,j)N_T(r,s) = q(i,j)q(r,s) \int_0^T \int_0^x \{P_{si}(x-t)\Pr[Z(t) = r] + P_{jr}(x-t)\Pr[Z(t) = i]\}\, dt\, dx + \delta(i,j;r,s)q(i,j) \int_0^T \Pr[Z(t) = i]\, dt,$$

where

$$\delta(i,j;r,s) = \begin{cases} 1 & \text{if } r = i \text{ and } j = s \\ 0 & \text{otherwise.} \end{cases}$$

(d)
$$EA_T(i)A_T(r) = \int_0^T \int_0^x \{P_{ri}(x-t)\Pr[Z(t) = r] + P_{ir}(x-t)\Pr[Z(t) = i]\}\, dt\, dx.$$

(e)
$$EN_T(r,s)A_T(i) = q(r,s) \int_0^T \int_0^x \{P_{ir}(x-t)\Pr[Z(t) = i] + P_{si}(x-t)\Pr[Z(t) = r]\}\, dt\, dx.$$

(f)
$$E\{[N_T(r,s) - q(r,s)A_T(r)][N_T(i,j) - q(i,j)A_T(i)]\} = \delta(i,j;r,s)q(i,j) \int_0^T \Pr[Z(t) = i]\, dt.$$

Part (f) of the theorem is rather surprising: Since the r.v.'s $\{A_T(i)\}_{i=1}^M$ are constrained by the relation $\sum_{i=1}^M A_T(i) = T$, it is unexpected to find that

$$[N_T(i,j) - q(i,j)A_T(i)] \quad \text{and} \quad [N_T(i',j') - q(i',j')A_T(i')]$$

are uncorrelated, (even if $i = i'$) when $j \neq j'$.

PROOF. We will prove (a), (b) and (f) here and relegate the proofs of (c), (d) and (e) to Appendix I.

(a) Divide $[0, T)$ into $n + 1$ equal parts of length $h = T/n + 1$, and let

$X_k(i, j)$ be the number of transitions from i to j during the (time) interval $k \cdot h$, $(k + 1)h)$, $k = 0, 1, 2, \cdots, n$. Then

$$EN_T(i, j) = \sum_{k=0}^{n} EX_k(i, j) = \sum_{k=0}^{n} \Pr[Z((k + 1)h) = j \;\&\; Z(k \cdot h) = i] + o(h)$$

$$= \sum_{k=0}^{n} \Pr[Z(k \cdot h) = i] q(i, j)h + o(h) \to q(i, j) \int_0^T \Pr[Z(t) = i]\, dt,$$

as $n \to \infty$.

(b) Let
$$Y_i(t) = \begin{cases} 1 & \text{if } Z(t) = i, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$A_T(i) = \int_0^T Y_i(t)\, dt, \qquad \text{so that}$$

$$EA_T(i) = \int_0^T EY_i(t)\, dt = \int_0^T \Pr[Z(t) = i]\, dt.$$

(c)
$$E\{[N_T(i, j) - q(i, j)A_T(i)][N_T(r, s) - q(r, s)A_T(r)]\}$$
$$= EN_T(i, j)N_T(r, s) + q(r, s)q(i, j)EA_T(i)A_T(r)$$
$$- (q(i, j)EN_T(r, s)A_T(i) + q(r, s)EN_T(i, j)A_T(r)),$$

and the result follows from comparison with (c), (d) and (e).

6. Large sample properties of the M.L.E. The term "large sample" can be interpreted in two ways in relation to the problem at hand: Many independent realizations of the process $\{Z(t), 0 \leqq t < T\}$ could be observed. (This corresponds to an investigation of the behavior of $\hat{q}_T^{(k)}(i, j)$ as $k \to \infty$.) On the other hand, a single realization of the process can be observed over a long period of time. (This corresponds to an investigation of $\hat{q}_T^{(1)}(i, j)$ as $T \to \infty$.) We will obtain results pertaining to the consistency and asymptotic normality of these estimates in both cases.

First, we will study the (easier) problem of holding T fixed and letting k grow large. The reader will recall that if $i \neq j$,

$$\hat{q}_T^{(k)}(i, j) = \begin{cases} N_T^{(k)}(i, j)/A_T^{(k)}(i) & \text{if } A_T^{(k)}(i) > 0 \\ 0 & \text{if } A_T^{(k)}(i) = 0. \end{cases}$$

(For convenience's sake, we will sometimes suppress the "T" in this part of the discussion.)

If $\Pr[A_T(i) = 0] = 1$, then $\hat{q}^{(k)}(i, j) = 0$ with probability one for every k, even if $q(i, j) > 0$. In this case, the estimate is a bad one. This situation only occurs when we try to estimate parameters associated with transitions *out of* a state that can never be reached (for $\Pr[A_T(i) = 0] = 1$ iff $\Pr[Z(t) = i] = 0$

for every t). This case must be excluded from consideration. If $\Pr[Z(t) = i] > 0$, then $EA_T(i) > 0$, so that by Theorem 5.2,

$$\frac{EN_T(i, j)}{EA_T(i)} = q(i\ j).$$

Since $N_T^{(k)}(i, j)$ and $A_T^{(k)}(i)$ are sums of independent, identically distributed variables,

$$\lim_{k\to\infty} \hat{q}^{(k)}(i, j) = \lim_{k\to\infty} \frac{N_T^{(k)}(i, j)/k}{A_T^{(k)}(i)/k} = q(i, j)$$

with probability one.

By applying a theorem of Cramér (c.f., [4], Theorem 2, or [5], page 254), the set of r.v.'s

$$\{k^{\frac{1}{2}}(\hat{q}^{(k)}(i, j) - q(i, j))\}_{i\neq j}$$

has the same asymptotic distribution as the set

$$\left\{\frac{1}{EA_T(i)}\left(\frac{N_T^{(k)}(i, j) - q(i, j)A_T^{(k)}(i)}{k^{\frac{1}{2}}}\right)\right\}_{i\neq j}$$

and by the multivariate central limit theorem, the last is asymptotically normal, with mean zero and covariances

$$\begin{aligned} C(i, j; k, l) &= \frac{1}{EA_T(i)EA_T(k)} E[N_T(i, j) - q(i, j)A_T(i)][N_T(k, l) - q(k, l)A_T(k)] \\ &= \delta(i, j; k, l)q(i, j) \Big/ \int_0^T \Pr[Z(t) = i]\, dt. \end{aligned}$$

This completes the proof of

THEOREM 6.1.

(a) *If there is positive probability of the* ith *state being occupied at some time* $t \geqq 0$, *then*

$$\lim_{k\to\infty} \hat{q}^{(k)}(i, j) = q(i, j)$$

with probability one.

(b) *If every state has positive probability of being occupied, then the set of* r.v.'s

$$\{k^{\frac{1}{2}}(\hat{q}^{(k)}(i, j) - q(i, j))\}_{j\neq i}$$

are asymptotically normal and independent with zero mean and variances

$$q(i, j) \Big/ \int_0^T \Pr[Z(t) = i]\, dt.$$

(Notice that $\hat{q}^{(k)}(i, j) = 0$ with probability one for every k if $q(i, j) = 0$, since, by Theorem 5.1, $N_T(i, j)$ has zero mean in this case.)

Now let us turn our attention to the more interesting problems of consistency and asymptotic normality when k is held fixed ($k = 1$) and T grows large.

First we will derive a series of results which lead to a demonstration of the (joint) asymptotic normality of the set of r.v.'s

$$\left\{\frac{N_T(i,j) - q(i,j)A_T(i)}{T^{\frac{1}{2}}}\right\}_{j\neq i}.$$

From this, we can conclude that

$$(1/T)[N_T(i,j) - q(i,j)A_T(i)]$$

converges to zero in probability as $T \to \infty$.

Then, we will show that $(1/T)N_T(i,j)$ and $(1/T)A_T(i)$ converge with probability one as $T \to \infty$, and, in fact, converge to constants with probability one. If $\lim_{T\to\infty} (1/T)A_T(i)$ is positive, Cramér's theorem can be used to show that the set of r.v.'s

$$\{T^{\frac{1}{2}}(\hat{q}_T(i,j) - q(i,j))\}_{j\neq i}$$

have a joint distribution which is asymptotically normal.
Furthermore, $\hat{q}_T(i,j)$ is consistent in the strong sense, since

$$[N_T(i, j)/A_T(i)] - q(i,j)$$

converges to zero with probability one as $T \to \infty$.

DEFINITION.

(a) If $\boldsymbol{\alpha} = \{\alpha_{ij}\}_{i\neq j}$ and $\boldsymbol{\beta} = \{\beta_{ij}\}_{i\neq j}$ are two sequences indexed by double subscripts which run over the integers $1, 2, \cdots, M$, we define

$$\langle \boldsymbol{\alpha}, \boldsymbol{\beta}\rangle = \sum_i \sum_{j\neq i} \alpha_{ij}\beta_{ij}.$$

(b) If

$$\boldsymbol{\delta} = \{\delta_i\}_{i=1}^M \quad \text{and} \quad \boldsymbol{\varepsilon} = \{\epsilon_i\}_{i=1}^M$$

we define

$$(\boldsymbol{\delta}, \boldsymbol{\varepsilon}) = \sum_{i=1}^M \delta_i\epsilon_i$$

(c) For any matrix A, let A^* be the transpose of A.

THEOREM 6.2. *Let* $\boldsymbol{\xi}_T = \{\xi_T(i,j)\}_{i\neq j}$, *where*

$$\xi_T(i,j) = T^{\frac{1}{2}}[N_T(i,j) - q(i,j)A_T(i)].$$

Let $\boldsymbol{\omega} = \{\omega(i,j)\}_{i\neq j}$, *and* $\varphi(\boldsymbol{\omega}; T) = E \exp - \langle \boldsymbol{\xi}_T, \boldsymbol{\omega}\rangle$ *be the joint moment generating function of the set* $\{\xi_T(i,j)\}_{i\neq j}$. Then

$$\varphi(\boldsymbol{\omega}; t) = (e^{tR^*(\boldsymbol{\omega},t)}\mathbf{I}, \mathbf{n}),$$

where $R(\boldsymbol{\omega}, t)$ *is an* $M \times M$ *matrix whose* (i, j)th *entry is*

$$r_{ij}(\boldsymbol{\omega}, t) = \begin{cases} -q(i) + \sum_{k \neq i} t^{-\frac{1}{2}} q(i, k)\omega(i, k) & \text{if } i = j, \\ q(i, j) \exp - \dfrac{\omega(i, j)}{t^{\frac{1}{2}}} & \text{if } i \neq j, \end{cases}$$

and $\mathfrak{n}$ *and* $\mathbf{I}$ *are* M*-vectors:*

$$\mathbf{I} = \begin{pmatrix} \Pr [Z(0) = 1] \\ \Pr [Z(0) = 2] \\ \vdots \\ \Pr [Z(0) = M] \end{pmatrix}, \qquad \mathfrak{n} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

PROOF.

Let $g_k(\boldsymbol{\omega}, t) = E(\exp - \langle \boldsymbol{\xi}_t, \boldsymbol{\omega} \rangle \mid Z(t) = k)$ be the conditional moment generating function of $\boldsymbol{\xi}_t$ given that $Z(t) = k$. Since

$$A_t(k) = t - \sum_{i \neq k} A_t(i),$$

$$(6.2.1) \quad g_k(\boldsymbol{\omega}, t) = e^{\alpha_k t^{\frac{1}{2}}} \left\{ \int_{\mathfrak{N}} \int_{\mathfrak{A}} \exp - t^{-\frac{1}{2}}[\langle \mathbf{n}, \boldsymbol{\omega} \rangle + \sum_{i \neq k} (\alpha_k - \alpha_i) a_i] \, d\, G_k \right.$$
$$\left. + \int_{\mathfrak{N}} \int_{\mathfrak{B}} \exp - t^{-\frac{1}{2}}[\langle \mathbf{n}, \boldsymbol{\omega} \rangle + \sum_{i \neq k} (\alpha_k - \alpha_i) a_i] \, d\, G_k \right\}$$

where $G_k(\mathbf{n}, \mathbf{a})$ is the joint conditional distribution function of the $\{N_t(i, j)\}_{i \neq j}$ and $\{A_t(i)\}$ given that $Z(t) = k$, $\mathfrak{A}$ is the positive orthant in $M - 1$ dimensional Euclidean space, $\mathfrak{A} \cup \mathfrak{B}$ is the non-negative orthant in $M - 1$ Euclidean space, $\mathfrak{N}$ is the set of all vectors with $M^2 - M$ non-negative, integer components and

$$\alpha_i = \sum_{j \neq i} q(i, j)\omega(i, j) \qquad i = 1, 2, \cdots, M.$$

Clearly,

$$(6.2.2) \qquad \varphi(\boldsymbol{\omega}, t) = \sum_{k=1}^{M} \tilde{g}_i(\boldsymbol{\omega}, t)$$

where $\tilde{g}_i(\boldsymbol{\omega}, t) = g_i(\boldsymbol{\omega}, t) \Pr [Z(t) = i]$.

The inner integral in the first term of the expression for $g_k(\boldsymbol{\omega}, t)$ can be integrated by parts (see Appendix II) and we find that

$$(6.2.3) \qquad \tilde{g}_k(\boldsymbol{\omega}, t) = f_k(\mathbf{u}^0, \mathbf{v}^0, t)$$

where $f_k(\mathbf{u}, \mathbf{v}; t)$ is the Laplace transform of

$$F_k(\mathbf{n}, \mathbf{a}; t) = \Pr \{(\bigcap_i \bigcap_{j \neq i} [N_t(i, j) = n_{ij}]) \cap (\bigcap_{i \neq k} [A_t(i) \leqq a_i] \cap [Z(t) = k])\}:$$

$$f_k(\mathbf{u}, \mathbf{v}; t) = \sum_{\mathbf{n} \in \mathfrak{N}} \int_{\mathfrak{A}} \prod_{i=1}^{M} da_i \, F_k(\mathbf{n}, \mathbf{a}; t) \exp - [\langle \mathbf{n}, \mathbf{u} \rangle + (\mathbf{a}, \mathbf{v})],$$

and

$$(6.2.4)\qquad \begin{aligned} u_{ij}^0 &= \omega(i, j)/t^{\frac{1}{2}} && i, j = 1, 2, \cdots, M(i \neq j), \\ v_i^0 &= (\alpha_k - \alpha_i)/t^{\frac{1}{2}} && \text{if } i = 1, 2, \cdots, k-1, k+1, \cdots, M, \end{aligned}$$

and

$$v_k^0 = \left[e^{\alpha_k t^{\frac{1}{2}}} \prod_{i \neq k} \left(\frac{\alpha_k - \alpha_i}{t^{\frac{1}{2}}} \right) \right]^{-1}.$$

In Appendix III it is shown that the functions F_k, $k = 1, 2, \cdots, M$, satisfy a system of first order linear differential-difference equations:

$$\frac{\partial F_k}{\partial t} = -q(k)F_k + \sum_{\nu \neq k} q(\nu, k)\Delta_{\nu k} F_\nu,$$

$$\begin{aligned} F_k &= \Pr[Z(0) = k] && \text{if } t = 0 \text{ and } n_{ij} = 0 \text{ for all } i, j, \\ F_k &= 0 && \text{if } n_{ij} < 0 \text{ for some } i \text{ and } j \\ F_k &= 0 && \text{if } t = 0 \text{ and } n_{ij} > 0 \text{ for some } i \text{ and } j. \end{aligned}$$

(Here, $\Delta_{\nu k}$ is the first order difference operator:

$$\Delta_{\nu k} F_k(\mathbf{n}, \mathbf{a}; t) = F_k(n_{12}, n_{13}, \cdots, n_{\nu k} - 1, \cdots, n_{M,M-1}, \mathbf{a}; t).$$

If we take Laplace transforms (t fixed) and adopt a matrix notation, we find

$$\frac{\partial \mathbf{f}(\mathbf{u}, \mathbf{v}; t)}{\partial t} = W^*(\mathbf{u})\mathbf{f}(\mathbf{u}, \mathbf{v}; t),$$

$$\mathbf{f}(\mathbf{u}, \mathbf{v}; 0) = \left(\prod_{i=1}^M v_i \right)^{-1} \mathbf{I}$$

where

$$\mathbf{f}(\mathbf{u}, \mathbf{v}; t) = \begin{pmatrix} f_1(\mathbf{u}, \mathbf{v}; t) \\ f_2(\mathbf{u}, \mathbf{v}; t) \\ \vdots \\ f_M(\mathbf{u}, \mathbf{v}; t) \end{pmatrix} \qquad \mathbf{I} = \begin{pmatrix} \Pr[Z(0) = 1] \\ \vdots \\ \Pr[Z(0) = M] \end{pmatrix}$$

and $W(\mathbf{u})$ is an $M \times M$ matrix whose (i, j)th entry is

$$w_{i,j}(\mathbf{u}) = \begin{cases} -q(i) & \text{if } i = j \\ q(i, j)e^{-u_{ij}} & \text{if } i \neq j. \end{cases}$$

The only bounded solution to this system of equations is

$$(6.2.5)\qquad \mathbf{f}(\mathbf{u}\ \mathbf{v}; t) = \left(\prod_{j=1}^M v_j \right)^{-1} e^{tW^*(\mathbf{u})}\,\mathbf{I},$$

where

$$e^{tW^*} = \sum_{n=0}^{\infty} \frac{(tW^*)^n}{n!}.$$

If we let $\boldsymbol{\varepsilon}^{(k)}$ be the M-vector whose components are all zero except for the kth which is unity,

$$f_k(\mathbf{u}, \mathbf{v}; t) = \left(\prod_{i=1}^{M} v_i\right)^{-1} (e^{tW^*(\mathbf{u})}\mathbf{I}, \boldsymbol{\varepsilon}^{(k)}).$$

Referring to equations 6.2.3 and 6.2.4, we find that

$$\tilde{g}_k(\omega, t) = (e^{tW^*(\omega/t^{\frac{1}{2}})}\mathbf{I}, e^{\alpha_k t^{\frac{1}{2}}}\boldsymbol{\varepsilon}^{(k)})$$

so that

$$\varphi(\omega, t) = (e^{tW^*(\omega/t^{\frac{1}{2}})}\mathbf{I}, e^{tA(\alpha/t^{\frac{1}{2}})}\mathbf{n})$$

where $\mathbf{n}$ is the M-vector whose entries are all unity, and

$$A(\boldsymbol{\alpha}) = \begin{pmatrix} \alpha_1, & 0, & \cdots & 0 \\ 0 & \alpha_2, & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots & \alpha_M \end{pmatrix}.$$

Since $A = A^*$ commutes with W^*, we find

$$\varphi(\omega, t) = (\exp\{t[W^*(\omega/t^{\frac{1}{2}}) + A^*(\boldsymbol{\alpha}/t^{\frac{1}{2}})]\}\mathbf{I}, \mathbf{n}).$$

Since $R^*(\omega, t) = W^*(\omega/t^{\frac{1}{2}}) + A^*(\boldsymbol{\alpha}/t^{\frac{1}{2}})$, the conclusion follows.

The asymptotic behavior of $\varphi(\omega, t)$ is determined by the next lemma:

DEFINITION.

(a) For any matrix A, adj A is the transposed matrix of cofactors of A.

(b) If $B(t)$ and A are $M \times M$ matrices, we say that $B(t) = A + o(1)$ as $t \to t_0$ if each entry of $B(t)$ approaches the corresponding entry of A as $t \to t_0$.

LEMMA 6.3. *Let $R(t)$ be an $M \times M$ matrix such that $R(t) = Q + o(1)$ as $t \to \infty$. Let $r(t) = \det R(t)$. If zero is a simple eigenvalue of Q and $\gamma = \lim_{t\to\infty} t\, r(t)$ exists and is finite, then*

$$\lim_{t\to\infty} e^{tR(t)} = (1/\rho)e^{\gamma/\rho} \operatorname{adj} Q,$$

where ρ is the product of the non-zero eigenvalue of Q.

PROOF. Let $\mu_1, \mu_2, \cdots, \mu_M$ be the (not necessarily distinct) eigenvalues of Q arranged in lexicographical order: Re $\mu_i \leqq$ Re μ_{i+1}, and Im $\mu_i \leqq$ Im μ_{i+1} if Re $\mu_i =$ Re μ_{i+1}. It is known that zero is always an eigenvalue of Q(the row sums of Q are all zero), and that the non-zero eigenvalues of Q have negative real parts. (*c.f.*, [2], page 52.) Hence, $\mu_M = 0$. Let $\mu_1(t), \cdots, \mu_M(t)$ be the eigenvalues of $R(t)$ similarly arranged. Since,

$$R(t) = Q + o(1) \quad \text{as} \quad t \to \infty,$$

$$\mu_k(t) = \mu_k + o(1) \quad \text{as} \quad t \to \infty, \quad k = 1, 2, \cdots, M.$$

(In particular, $\mu_M(t) = o(1)$ as $t \to \infty$.)

If A is any matrix possessing eigenvalues $\nu_1, \cdots, \nu_r$, with multiplicities $m_1, \cdots, m_r$, then

$$e^A = \sum_{i=1}^{r} \frac{1}{(m_i - 1)!} \left[\frac{d^{m_i-1}}{d\nu^{m_i-1}} \frac{e^{\nu} \operatorname{adj}(\nu I - A)}{\prod_{j \neq i} (\nu - \nu_j)^{m_j}} \right]_{\nu=\nu_i}.$$

(Sylvester's theorem, c.f., [2], page 32.)

If $m_r = 1$,

$$e^A = \frac{e^{\nu_r} \operatorname{adj}(\nu_r I - A)}{\prod_{j \neq r} (\nu_r - \nu_j)^{m_j}} + \sum_{i \neq r} e^{\nu_i} B_i(\nu_i),$$

where B_i is a matrix whose entries are rational functions of ν_i.

Since $\mu_1(t), \cdots, \mu_{M-1}(t)$ all have negative real parts for all t sufficiently large,

$$\begin{aligned} e^{tR(t)} &= \frac{e^{t\mu_M(t)}}{\prod_{j \neq M} (\mu_M(t) - \mu_j(t))} [\operatorname{adj}(\mu_M(t)I - R(t))] + o(1) \quad \text{as} \quad t \to \infty, \\ &= \frac{e^{t\mu_M(t)}}{\rho + o(1)} [\operatorname{adj} R(t) + o(1)] + o(1) \\ &= e^{t\mu_M(t)} \left[\frac{\operatorname{adj} Q + o(1)}{\rho + o(1)} \right] \quad \text{as} \quad t \to \infty. \end{aligned}$$

Since $\det R(t) = \prod_{j=1}^{M} \mu_j(t)$,

$$\mu_M(t) = \frac{r(t)}{\rho + o(1)} \quad \text{as} \quad t \to \infty.$$

$$\therefore \lim_{t \to \infty} e^{tR(t)} = \lim_{t \to \infty} (1/\rho) e^{tr(t)/\rho} \operatorname{adj} Q,$$

if the right hand limit exists.

The application of Lemma 6.3 to Theorem 6.2 is facilitated by

LEMMA 6.4. *Let B be a square matrix whose row sums all vanish. Let $B^{(i,j)}$ be the (i, j)th cofactor of B. Then for every i and j, $B^{(i,j)} = B^{(i,i)}$.*

PROOF. Let $D(i, j)$ be the determinant of the matrix obtained from B by replacing the ith entry of the ith row of B by 1, the jth entry of the ith row by -1, and all other elements of the ith row by zero. Expand $D(i, j)$ by cofactors of the ith row and we find that $D(i, j) = B^{(i,i)} - B^{(i,j)}$. Since the row sums of the modified matrix are all zero, $D(i, j) = 0$.

THEOREM 6.5. *If zero is a simple eigenvalue of* Q,

$$\lim_{t\to\infty} \varphi(\omega, t) = \exp \tfrac{1}{2} \sum_{i} \sum_{j\neq i} \frac{q(i,j)Q^{(i,i)}}{\rho} \omega^2(i,j).$$

PROOF. Let $r(t) = \det R(\omega, t)$, where $R(\omega, t)$ is as defined in Theorem 6.2. Clearly,

$$R(\omega, t) = Q + o(1) \quad \text{as} \quad t \to \infty,$$

so if $\gamma = \lim_{t\to\infty} t\, r(t)$ exists, Lemma 6.3 applies:

$$\lim_{t\to\infty} e^{tR^*(\omega,t)} = (1/\rho)e^{\gamma/\rho} \operatorname{adj} Q^*,$$

so that

$$\lim_{t\to\infty} \varphi(\omega, t) = (1/\rho)e^{\gamma/\rho} (\operatorname{adj} Q^*\mathbf{I}, \mathfrak{n}).$$

Let $q^*(i,j)$ be the (i,j)th entry of adj Q^*. Then $q^*(i,j) = Q^{(i,j)}$ and by Lemma 6.4, $Q^{(i,j)} = Q^{(i,i)}$. Whence, since $\mathbf{I}$ is a probability measure,

$$(\operatorname{adj} Q^*\mathbf{I}, \mathfrak{n}) = \sum_{i=1}^{M} Q^{(i,i)}.$$

Let $\Phi(\mu) = \det(\mu I - Q) = \prod_{j=1}^{M} (\mu - \mu_j)$. Since $\mu_M = 0$ is a simple eigenvalue of Q, $\Phi(\mu) = \mu \prod_{j\neq M} (\mu - \mu_j)$ and by the rule for differentiating determinants,

$$\frac{d}{d\mu} \Phi(\mu)\bigg|_{\mu=0} = (-1)^{M-1}\rho = (-1)^{M-1} \sum_{i=1}^{M} Q^{(i,i)}$$

(c.f., [3]). Hence, if $\lim_{t\to\infty} tr(t) = \gamma$ exists and is finite, $\lim_{t\to\infty} \varphi(\omega, t) = e^{\gamma/\rho}$.

It remains to evaluate γ: If either exists,

$$\lim_{t\to\infty} t\, r(t) = \lim_{y\to 0} (1/y) r(1/y).$$

Since the right-hand side is indeterminate ($\lim_{y\to 0} r(1/y) = \det Q = 0$), we resort to L'Hospital's rule: If $\lim_{y\to 0} (d/dy) r(1/y)$ exists and is finite, then

$$\gamma = \lim_{t\to\infty} t\, r(t) = \lim_{y\to 0} (d/dy) r(1/y).$$

By the rule for differentiating determinants (c.f., [3] again), $(d/dy)r(1/y)$ exists for every positive y. In fact,

$$\frac{d}{dy} r\left(\frac{1}{y}\right) = \sum_{i} \sum_{j} R^{(i,j)} \left(\frac{1}{y}\right) \frac{\partial}{\partial y} r_{ij}\left(\omega, \frac{1}{y}\right),$$

where $R^{(i,j)}(1/y)$ is the (i,j)th cofactor of $R(\omega; 1/y)$. Explicitly:

$$\frac{\partial}{\partial y} r_{ij}\left(\omega, \frac{1}{y}\right) = \begin{cases} \dfrac{1}{2y^{\frac{1}{2}}} \displaystyle\sum_{k\neq i} q(i,k)\omega(i,k) & \text{if} \quad i = j \\[2ex] -\dfrac{1}{2y^{\frac{1}{2}}} \omega(i,j) q(i,j) e^{-\omega(i,j)y^{\frac{1}{2}}} & \text{if} \quad i \neq j. \end{cases}$$

Since $R^{(i,j)}(1/y) \to Q^{(i,j)} = Q^{(i,i)}$ as $y \to 0$, and since

$$e^{-\omega(i,j)y^{\frac{1}{2}}} = 1 - \omega(i,j)y^{\frac{1}{2}} + O(y) \qquad \text{as} \qquad y \to 0,$$

we can write

$$\frac{\partial}{\partial y} r\left(\frac{1}{y}\right) = \frac{1}{2y^{\frac{1}{2}}} \sum_{i=1}^{M} \left[\sum_{k \neq i} q(i,k)\omega(i,k)\left(R^{(i,i)}\left(\frac{1}{y}\right) - R^{(i,k)}\left(\frac{1}{y}\right)\right)\right]$$
$$+ \frac{1}{2} \sum_{i=1}^{M} \sum_{k \neq i} q(i,k) Q^{(i,i)} \omega^2(i,k) + o(1) \qquad \text{as} \qquad y \to 0.$$

It is shown in Appendix IV that

$$\lim_{y \to 0} \frac{1}{2y^{\frac{1}{2}}} \sum_{i=1}^{M} \left[\sum_{k \neq i} q(i,k)\omega(i,k)\left(R^{(i,i)}\left(\frac{1}{y}\right) - R^{(i,k)}\left(\frac{1}{y}\right)\right)\right] = 0.$$

We have, as an immediate

COROLLARY 6.6. *For each i and j, $j \neq i$, $[N_T(i,j) - q(i,j)A_T(i)]/T \to 0$ in probability as $T \to \infty$.*

PROOF. Each $\xi_T(i, j)$ converges in law to a normal distribution. Hence $(1/T^{\frac{1}{2}})\xi_T(i,j)$ converges in law (and hence in probability) to zero.

Actually, we need a stronger result than this in order to carry out the proposed agenda. An investigation of the almost-sure convergence of $(1/T)A_T(i)$ and $(1/T)N_T(i, j)$ must be made, and, as one might expect, such an investigation utilizes such concepts as metric transitivity and stationarity. A brief digression along these lines is now appropriate and the results of this discussion will point the way to our main theorem.

We say that the process $\{Z(t), t \geqq 0\}$ is *strictly stationary*, if P $[Z(t) = i]$ does not depend upon t for any i. Let

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{pmatrix},$$

where $\lambda_i = \Pr[Z(0) = i]$. Since

$$\Pr[Z(t) = j] = \sum_i P_{ij}(t)\lambda_i,$$

it seems that $\mathbf{\Lambda}$ must possess some mysterious power if the left-hand side is not to depend upon t. Specifically, $\mathbf{\Lambda}$ must make the matrix equation, $P^*(t)\mathbf{\Lambda} = \mathbf{\Lambda}$, work. $\mathbf{\Lambda}$ can be characterized in the following way:

LEMMA 6.7.

(a) *The process $\{Z(t), t \geqq 0\}$ is strictly stationary if and only if the initial distribution $\mathbf{\Lambda}$ satisfies the equation $Q^*\mathbf{\Lambda} = 0$.*

(b) *If zero is a simple eigenvalue of Q, there is exactly one such* $\mathbf{\Lambda}$*:*

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{pmatrix},$$

with $\lambda_i = Q^{(i,i)}/\rho$.

(c) *If zero is a simple eigenvalue of Q, then* $\lim_{t\to\infty} \Pr[Z(t) = i] = \lambda_i$, $(i = 1, 2, \cdots, M)$, *independent of the initial distribution of the process.*

PROOF.

(a) The process is strictly stationary if and only if $P^*(t)\mathbf{\Lambda} = \mathbf{\Lambda}$. Since

$$P'(t) = QP(t) = P(t)Q; \qquad P(0) = I,$$

the conclusion follows.

(b) Under the hypothesis of the lemma, the null space of Q is one dimensional. The vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix},$$

where $x_i = Q^{(i,i)}$, is always a solution of $Q^*\mathbf{x} = 0$, since $\det Q = 0$. All solutions of $Q^*\mathbf{\Lambda} = 0$ are therefore multiples of $\mathbf{x}$.

Let $P = \lim_{t\to\infty} P(t)$. It is well known that the matrix limit exists when the limit is taken term-by-term, and that the column vectors $\mathbf{p}^{(1)} \cdots \mathbf{p}^{(M)}$ of P are stationary distributions for the $Z(t)$ process, and hence are eigenvectors of Q^*. Whence, $\mathbf{p}^{(j)} = c\mathbf{x}$ for $j = 1, 2, \cdots, M$. Since $\sum_{i=1}^{M} Q^{(i,i)} = \rho$ (c.f., Theorem 6.5), $c = 1/\rho$.

(c) For any other initial distribution, $\mathbf{\Lambda}'$,

$$\lim_{t\to\infty} P^*(t)\mathbf{\Lambda}' = P^*\mathbf{\Lambda}' = \mathbf{\Lambda}.$$

LEMMA 6.8. *The process $\{Z(t), t \geqq 0\}$ is metrically transitive if and only if there is exactly one initial distribution $\mathbf{\Lambda}$ which satisfies* $Q^*\mathbf{\Lambda} = 0$.

PROOF. It is well known that $\{Z(t), t \geqq 0\}$ is metrically transitive if and only if there is exactly one $\mathbf{\Lambda}$ satisfying $P^*(t)\mathbf{\Lambda} = \mathbf{\Lambda}$ (c.f., [6], page 238, in conjunction with page 511). The last results permit us to state:

THEOREM 6.9. *If zero is a simple eigenvalue of Q, and if* $Q^{(k,k)} > 0$ *for* $k = 1, 2, \cdots, M$,

$$\Pr[\lim_{T\to\infty} (1/T)A_T(i) = \lambda_i] = 1$$

and

$$\Pr[\lim_{T\to\infty} (1/T)N_T(i,j) = q(i,j)\lambda_i] = 1.$$

(Here, $\lambda_i = Q^{(i,i)}/\rho$.)

PROOF. If zero is a simple eigenvalue of Q, the process $\{Z(t), t \geqq 0\}$ is metrically transitive. So then, are the processes $\{Y_i(t), t \geqq 0\}$, where,

$$Y_i(t) = \begin{cases} 1 & \text{if } Z(t) = i \\ 0 & \text{otherwise}, \quad i = 1, 2, \cdots, M. \end{cases}$$

If the initial distribution on $\{Z(t), t \geqq 0\}$ is

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{pmatrix}, \qquad (\lambda_i = Q^{(i,i)}/\rho > 0), \qquad i = 1, 2, \cdots, M,$$

then, by the ergodic theorem,

$$\lim_{T\to\infty} \frac{A_T(i)}{T} = \lim_{T\to\infty} \frac{1}{T} \int_0^T Y_i(t)\, dt = EY_i(0) = \lambda_i$$

with probability one.

Since $\lambda_k > 0$ for each k,

$$\Pr\,[\lim_{T\to\infty} (1/T) A_T(i) = \lambda_i \mid Z(0) = k] = 1 \quad \text{for each } k.$$

Hence, no matter what the initial distribution,

$$\Pr\,[\lim_{T\to\infty} (1/T) A_T(i) = \lambda_i] = 1.$$

Let $h > 0$ be given and let $t_n = nh$. Let $X_k(i, j)$ be the number of transitions from i to j made in the interval $[(k - 1)h, k \cdot h)$, $k = 1, 2, \cdots$. If the $Z(t)$ process is strictly stationary, so then is the process

$$\{X_k(i, j), k = 1, 2, \cdots\}.$$

Since

$$N_{t_n}(i, j) = \sum_{k=1}^{n} X_k(i, j),$$

the ergodic theorem tells us that

$$y(h, i, j) = \lim_{n\to\infty} \frac{1}{t_n} N_{t_n}(i, j) = \frac{1}{h} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k(i, j)$$

exists with probability one for every positive h.

By Theorem 5.2, $EN_T(i, j) = q(i, j)\lambda_i T$ and

$$\operatorname{Var} N_T(i, j) = 2q^2(i, j)\lambda_i \int_0^T \int_0^x P_{ji}(x - t)\, dt\, dx + q(i, j)\lambda_i T - q^2(i, j)\lambda_i^2 T^2$$

if the process has the initial distribution $\mathbf{\Lambda}$.

Since $\lim_{T\to\infty} (1/T^2) \operatorname{Var} N_T(i, j) = q^2(i, j)\lambda_i \lim_{T\to\infty} (P_{ji}(T) - \lambda_i)$, and since (by Lemma 6.7) $\lim_{T\to\infty} P_{ji}(T) = Q^{(ii)}/\rho$, Tchebycheff's inequality reveals that $[N_T(i, j)/T] \to q(i, j)\lambda_i$ in probability as $T \to \infty$. Hence $[N_{t_n}(i, j)/t_n] \to q(i, j)\lambda_i$ in probability, so that $y(h, i, j) = q(i, j)\lambda_i$ identically in h.

If $\mathbf{\Lambda}$ is the initial distribution, then, $\lim_{T\to\infty} [N_T(i, j)/T] = q(i, j)\lambda_i$ with probability one. By proceeding as in the first part of this theorem, it follows that

$$\Pr\{[N_T(i, j)/T] \to q(i, j)\lambda_i\} = 1$$

independent of the initial distribution.

The conditions that zero be a simple eigenvalue of Q and that $Q^{(i,i)}$ be positive for every i are sometimes abbreviated by the phrase: "$\{Z(t), t \geqq 0\}$ is positively regular."

The main result can be stated.

THEOREM 6.10.

(a) *If* $\{Z(t), t \geqq 0\}$ *is positively regular, then*

$$\lim_{T\to\infty} \hat{q}_T(i, j) = q(i, j)$$

with probability one, and

(b) *The joint distribution of the set of* r.v.'s

$$\{T^{\frac{1}{2}}[\hat{q}_T(i, j) - q(i, j)]\}_{i,j=1, i\neq j}^{M}$$

is asymptotically normal and independent with zero mean and variances $q(i, j)\rho/Q^{(i,i)}$, *where* ρ *is the product of the non-zero-eigenvalues of* Q *and* $Q^{(i,i)}$ *is the* (i, i)th *cofactor of* Q.

PROOF.

(a) From Theorem 6.9,

$$\lim_{T\to\infty} N_T(i, j)/A_T(i) = q(i, j)$$

with probability one. Since $A_T(i)/T$ tends to a positive limit,

$$\lim_{T\to\infty} \hat{q}_T(i, j) = \lim_{T\to\infty} N_T(i, j)/A_T(i)$$

with probability one.

(b) By Cramér's theorem ([5], page 254), the (joint) asymptotic distribution of the set $\{T^{\frac{1}{2}}[\hat{q}_T(i, j) - q(i, j)]\}_{j\neq i}$ is the same as that of the set

$$\left\{T^{\frac{1}{2}}\left[\frac{N_T(i, j) - q(i, j)A_T(i)}{A_T(i)}\right]\right\}_{j\neq i}$$

and this, in term, has the same limiting distribution as

$$\left\{\frac{\xi_T(i, j)}{\lambda_i}\right\}_{j\neq i}.$$

From Theorem 6.5, the joint moment generating function of this set tends to

$$\exp \frac{1}{2} \sum_i \sum_{j\neq i} \frac{q(i, j)Q^{(i,i)}}{\lambda_i^2 \rho} \omega^2(i, j) \qquad \text{as} \qquad T \to \infty.$$

But, $\lambda_i = Q^{(i,i)}/\rho$.

As an afterthought, we point out that $\hat{q}_T(i, j) = q(i, j)$ with probability one if $q(i, j) = 0$. For then, $EN_T(i, j) = 0$ (by Theorem 5.2) so that $\Pr[N_T(i, j) = 0] = 1$.

7. Asymptotic efficiency. If a random variable, X, has a density, $g(x, \alpha)$, (where α ranges over an interval of the real line), then under certain regularity conditions (c.f., [5], pp. 478–481), the Cramér-Rao inequality states that for any estimate α^* of α, which is based upon k independent observations of the r.v. X,

$$E(\alpha^* - \alpha)^2 \geqq \left(1 + \frac{db(\alpha)}{d\alpha}\right)^2 \left[kE\left(\frac{\partial \log g}{\partial \alpha}\right)^2\right]^{-1},$$

where $b(\alpha) = E(\alpha^* - \alpha)$ is the bias of the estimate α^*.

If α^* is an unbiased estimate of α, $(b(\alpha) \equiv 0)$, the inequality states that the variance of α^* must be at least as large as

$$\left[kE\left(\frac{\partial \log g}{\partial \alpha}\right)^2\right]^{-1}.$$

If this lower bound is achieved, the estimate is called *efficient*.

More generally, if the asymptotic distribution of $k^{\frac{1}{2}}(\alpha^* - \alpha)$ is normal with zero mean and variance c^2 (as $k \to \infty$), α^* is said to be *asymptotically efficient* if

$$c^2 = \left[E\left(\frac{\partial \log g}{\partial \alpha}\right)^2\right]^{-1}$$

(c.f., [5], p. 489).

In the case at hand, if we wish to estimate a particular $q(i, j)$ on the basis of k independent observations of the process $\{Z(t),\ 0 \leqq t < T\}$, we take $\hat{q}^{(k)}(i, j) = N_T^{(k)}(i, j)/A_T^{(k)}(i)$ as our estimate.

Under the conditions of Theorem 6.1, $k^{\frac{1}{2}}(\hat{q}^{(k)}(i, j) - q(i, j))$ is asymptotically normally distributed with zero mean and variance

$$c_T^2 = q(i, j) \Big/ \int_0^T \Pr[Z(t) = i]\, dt.$$

A routine computation in connection with Theorem 3.2 shows that

$$\frac{\partial \log f_Q}{\partial q(i, j)} = \frac{1}{q(i, j)} [N_T(i, j) - q(i, j)A_T(i)]$$

so that by Theorem 5.2

$$\left[E\left(\frac{\partial \log f_Q}{\partial q(i, j)}\right)^2\right]^{-1} = q\,(i, j) \Big/ \int_0^T \Pr[Z(t) = i]\, dt = c_T^2 .$$

Thus, under the conditions of Theorem 6.1, $q^{(k)}(i, j)$ is asymptotically efficient as the number of independent observations grows infinite.

It seems reasonable to extend the notion of asymptotic efficiency to the case where the estimate for $q(i, j)$ is based upon one very long realization of the process $Z(t)$, instead of many independent realizations of the process over a finite time interval.

The expression $E(\partial \log g/\partial\alpha)^2$ is sometimes called "Fisher's Information Number" and the Cramér-Rao lower bound clearly exhibits the relevance of this quantity. It should be observed, that the information is multiplied by k when k independent observations are taken. Conversely, if k independent observations yield an information I, the information contained in a single observation is I/k.

If one observation consists of a single realization of $Z(t)$ over the time interval $[0, T)$, it therefore seems natural to define

$$\frac{1}{T} E\left(\frac{\partial \log f_Q}{\partial q(i, j)}\right)^2$$

as the amount of information obtained per unit time of observation. We will call an estimate α_T^* of $q(i, j)$ *asymptotically efficient* as $T \to \infty$ if $T^{\frac{1}{2}}(\alpha_T^* - q(i, j))$ is asymptotically normally distributed with zero mean and variance

$$c^2 = \lim_{T\to\infty}\left[\frac{1}{T} E\left(\frac{\partial \log f_Q}{\partial q(i, j)}\right)^2\right]^{-1}.$$

We already know that under the conditions of Theorem 6.10, $T^{\frac{1}{2}}(\hat{q}_T(i, j) - q(i, j))$ is asymptotically normal with zero mean and variance $c^2 = q(i, j)\rho/Q^{(i,i)}$. Since

$$\frac{1}{T} E\left(\frac{\partial \log f_Q}{\partial q(i, j)}\right)^2 = \int_0^T \Pr [Z(t) = i]\, dt/Tq(i, j),$$

and since, by Lemma 6.7,

$$\int_0^T \Pr [Z(t) = i]\, dt = Q^{(i,i)}T/\rho + o(T) \qquad \text{as} \qquad T \to \infty,$$

we see that

$$\lim_{T\to\infty}\left[\frac{1}{T} E\left(\frac{\partial \log f_Q}{\partial q(i, j)}\right)^2\right]^{-1} = c^2.$$

$\hat{q}_T(i, j)$ is therefore asymptotically efficient as $T \to \infty$.

8. Acknowledgments. I wish to thank Prof. Ulf Grenander for suggesting this investigation and for offering many fruitful ideas whenever things threatened to come to a grinding halt. I also wish to thank the National Science Foundation for their financial support during the initial stages of this investigation.

APPENDIX I

Proof of Theorem 5.2 (c, d, e)

Throughout this discussion we will let

$$Y_i(t) = \begin{cases} 1 & \text{if } Z(t) = 1 \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, 2, \cdots, M,$$

$$X_k(i, j) = \text{the number of jumps from } i \text{ to } j \text{ in } [(k - 1)h, kh),\ k = 1, 2, \cdots,$$

and

$$\pi_i(t) = \Pr[Z(t) = i], \qquad i = 1, 2, \cdots, M.$$

(c) Divide $[0, T)$ into n parts of length $T/n = h$.

$$\begin{aligned}
EN_T(i,j)N_T(r,s) &= \sum_{k=1}^{n}\sum_{m=1}^{n} EX_k(i,j)X_m(r,s) \\
&= \sum_{k=1}^{n}\sum_{m=1}^{n} \Pr[X_k(i,j) = 1 \& X_m(r,s) = 1] \\
&\quad + o(1) \qquad \text{as} \qquad h \to 0, \\
&= \sum_{k=3}^{n}\sum_{m=1}^{k-2} \Pr[X_k(i,j) = 1 \& X_m(r,s) = 1] \\
&\quad + \sum_{k=2}^{n} \Pr[X_k(i,j) = 1 \& X_{k-1}(r,s) = 1] \\
&\quad + \sum_{k=1}^{n} \Pr[X_k(i,j) = 1 \& X_k(r,s) = 1] \\
&\quad + \sum_{m=3}^{n}\sum_{k=1}^{m-2} \Pr[X_k(i,j) = 1 \& X_m(r,s) = 1] \\
&\quad + \sum_{m=2}^{n} \Pr[X_{m-1}(i,j) = 1 \& X_m(r,s) = 1] \\
&\quad + o(1) \qquad \text{as} \qquad h \to 0, \\
&= \sum_{k=3}^{n}\sum_{m=1}^{k-2} \pi_r(mh)q(r,s)hP_{si}(k\cdot h - (m+1)h)q(i,j)h \\
&\quad + \delta(i,j;r,s)\sum_{k=0}^{n-1} \pi_i(kh)q(i,j)h \\
&\quad + \sum_{m=3}^{n}\sum_{k=1}^{m-2} \pi_i(kh)q(i,j)hP_{jr}(m\cdot h - (k+1)h)q(r,s)h \\
&\quad + o(1) \qquad \text{as} \qquad h \to 0.
\end{aligned}$$

As $h \to 0$, this expression tends to

$$q(i,j)q(r,s)\int_0^T\int_0^x \{\pi_r(t)P_{si}(x-t) + \pi_i(t)P_{jr}(x-t)\}\,dt\,dx + \delta(i,j;r,s)q(i,j)\int_0^T \pi_i(t)\,dt.$$

(d)

$$\begin{aligned}
EA_T(i)A_T(r) &= \int_0^T\int_0^T EY_i(x)Y_r(t)\,dt\,dx \\
&= \int_0^T\int_0^T \Pr[Z(x) = i \& Z(t) = r]\,dt\,dx \\
&= \int_0^T\int_0^x P_{ri}(x-t)\pi_r(t)\,dt\,dx \\
&\quad + \int_0^T\int_0^t P_{ir}(t-x)\pi_i(x)\,dx\,dt.
\end{aligned}$$

(e) Divide the interval $[0, T)$ into n equal parts of length $h = T/n$. Then, $N_T(i, j) = \sum_{k=1}^{n} X_k(i, j)$, so that,

$$EA_T(r)N_T(i,j) = \sum_{k=1}^{n} \int_0^T E[Y_r(t)X_k(i,j)]\,dt.$$

By Lemma 5.1,

$$E[Y_r(t)X_k(i,j)] = \Pr[Z(t) = r \& X_k(i,j) = 1] + o(h)$$

as $h \to 0$, so that

$$\begin{aligned} EA_T(r)N_T(i,j) &= \sum_{k=1}^{n} \Big\{ \int_0^{kh} \Pr[Z(t) = r \& Z(kh) = i \& Z((k+1)h) = j]\,dt \\ &\quad + \int_{kh}^{(k+1)h} \Pr[Z(kh) = i \& Z(t) = r \& Z((k+1)h) = j]\,dt \\ &\quad + \int_{(k+1)h}^{T} \Pr[Z(kh) = i \& Z((k+1)h) = j \& Z(t) = r]\,dt \Big\} \\ &\quad + o(1) \qquad \text{as} \qquad h \to 0, \\ &= \sum_{k=0}^{n} \left[\int_0^{kh} \pi_r(t) P_{ri}(kh - t) q(i,j)\,dt \right] h \\ &\quad + \sum_{k=0}^{n} \left[\int_{(k+1)h}^{T} \pi_i(k \cdot h) q(i,j) P_{jr}(t - (k+1)h)\,dt \right] h \\ &\quad + o(1) \qquad \text{as} \qquad h \to 0, \\ &\to q(i,j) \int_0^T \int_0^x \Big\{ \pi_r(t) P_{ri}(x - t)\,dt\,dx + \pi_i(x) P_{jr}(t - x)\,dt\,dx \Big\} \\ &= q(i,j) \int_0^T \int_0^x \{ \pi_r(t) P_{ri}(x - t) + \pi_i(t) P_{jr}(x - t) \}\,dt\,dx \end{aligned}$$

as $n \to \infty$. (We have used Lemma 5.1 repeatedly.)

APPENDIX II

Justification of the Integration-by-Parts in Equation 6.2.1 *of Theorem* 6.2

It will suffice to show that the set function J_k, where

$$\text{(A2.1)} \qquad \begin{aligned} J_k(E) = \Pr\{[\mathbf{N}_T = \mathbf{n}] \cap [(A_T(1), \cdots, A_T(k-1), \\ A_T(k+1), \cdots, A_T(M)) \in E] \mid Z(T) = k\} \end{aligned}$$

is, for fixed k and fixed $\mathbf{n} \in \mathfrak{N}$, absolutely continuous with respect to $M - 1$ dimensional Lebesgue measure over $\mathfrak{A}$, where $\mathbf{N}_T = (N_T(1, 2), N_T(1, 3), \cdots, N_T(M, M - 1))$. It will then follow that the mass function $G_k(\mathbf{n}, \cdot, T)$ has no singular component with respect to Lebesgue measure over $\mathfrak{A}$ and this, in turn, justifies the integration-by-parts step.

First, we notice that $J_k(\cdot)$ can be expressed as a countable sum over a certain subset $\prod_{j=1}^{n} \otimes \mathcal{W}_0$ of set functions of the form

$$J_k^*(E; n, z_0, \cdots, z_{n-1}) = \Pr[N(T) = n, Z_0 = z_0, \cdots, Z_{n-1} = z_{n-1}, Z_n = k, (A_T(1), \cdots, A_T(k-1), A_T(k+1), \cdots, A_T(M)) \varepsilon E]/\Pr[Z(T) = k]$$

where $n = \sum_{i,j} n_{ij}$. Therefore it suffices to show that set functions of the form J_k^* (with $n, z_0, \cdots, z_{n-1}$ fixed) are absolutely continuous over $\mathfrak{A}$.

Let $I(j)$ be the set of indices ν for which $z_\nu = j$. Then, $A_T(j) = \sum_{\nu \varepsilon I(j)} T_\nu$, (where, as in Section 3, T_ν is the time spent in Z_ν) and

$$J_k^*(E; n, z_0, \cdots, z_{n-1}) = \frac{1}{\Pr[Z(T) = k]} \Pr[N(T) = n, Z_0 = z_0, \cdots, Z_n = k, (\sum_{\nu \varepsilon I(1)} T_\nu, \cdots, \sum_{\nu \varepsilon I(k-1)} T_\nu, \sum_{\nu \varepsilon I(k+1)} T_\nu, \cdots, \sum_{\nu \varepsilon I_M} T_\nu) \varepsilon E].$$

If $I(j)$ is empty for some $j \neq k$, then $J_k^*(E; n, z_0, \cdots, z_{n-1}) = 0$ since E is assumed to be a subset of the positive orthant $\mathfrak{A}$. Otherwise, by Theorem 3.1, $J_k^*(E; n, z_0, \cdots, z_{n-1})$ can be expressed as an $M - 1$ dimensional Lebesgue integral over E.

APPENDIX III

$$\frac{\partial}{\partial t} F_k = -q(k)F_k + \sum_{\nu \neq k} q(\nu, k) \Delta_{\nu k} F_\nu .$$

PROOF.

$$F_k(n_{12}, \cdots, n_{M,M-1}, a_1, \cdots, a_{k-1}, a_{k+1}, \cdots, a_M; t + h) = \sum_{m=0}^{1} \Pr\{(\bigcap_i \bigcap_{j \neq i} [N_{t+h}(i, j) = n_{ij}] \cap (\bigcap_{i \neq k} [A_{t+h}(i) \leqq a_i]) \cap [Z(t + h) = k] \cap B_m\}$$

$+ O(h^2)$ as $h \to 0$, (where B_m is the event: "[m changes in $[t, t + h)$]." (c.f., Lemma 5.1)); $= F_k(n_{12}, n_{\nu k} - 1, \cdots, n_{M,M-1}, a_1, \cdots, a_{k-1}, a_{k+1}, \cdots, a_M; t)$ $(1 - q(k)h) + \sum_{\nu \neq k} F_\nu(n_{12}, \cdots, n_{\nu k} - 1, \cdots, n_{M,M-1}, a_1 \cdots, a_{\nu-1}, a_{\nu+1}, \cdots$ $a_M; t) q(\nu, k)h + o(h)$ as $h \to 0$. Thus,

$$\lim_{h \to 0} (1/h)[F_k(\mathbf{n}, \mathbf{a}; t + h) - F_k(\mathbf{n}, \mathbf{a}; t)] = -q(k)F_k + \sum_{\nu \neq k} q(\nu, k)\Delta_{\nu k} F_\nu .$$

The initial conditions are obvious: If $n_{ij} > 0$ for some i and j,

$$F_k(n_{12}, \cdots, n_{M,M-1}, \mathbf{a}; 0) = 0.$$

Otherwise,

$$F_k(0, \cdots, 0, \mathbf{a}; 0) = \Pr[Z(0) = k].$$

APPENDIX IV

Proof of the last step in Theorem 6.5

$$\lim_{y \to 0} \frac{1}{2y^{\frac{1}{2}}} \sum_{i=1}^{M} \left[\sum_{k \neq i} q(i, k)\omega(i, k) \left(R^{(i,i)}\left(\frac{1}{y}\right) - R^{(i,k)}\left(\frac{1}{y}\right)\right)\right] = \lim_{x \to 0} \frac{1}{2} \frac{d}{dx} \sum_{i=1}^{M} \sum_{k \neq i} q(i, k)\omega(i, k) \left[R^{(i,i)}\left(\frac{1}{x^2}\right) - R^{(i,k)}\left(\frac{1}{x^2}\right)\right]$$

Let the matrix $R(\omega, t)$ be defined as in Theorem 6.2, and let $S^{(i)}(x) = \|s_{n,m}(x)\|$ be the matrix obtained from $R(\omega, 1/x^2)$ by replacing the ith row of R by the row whose ith element is $\sum_{k\neq i} q(i,k)\omega(i,k)$ and whose mth element is $-q(i, m)\omega(i, m)$ if $m \neq i$.

If we expand $s^{(i)}(x) = \det S^{(i)}(x)$ by cofactors of its ith row, we find

$$s^{(i)}(x) = \sum_{k\neq i} q(i, k)\omega(i, k)[R^{(i,i)}(\omega, 1/x^2) - R^{(i,k)}(\omega, 1/x^2)].$$

It suffices to show that for each i, $\lim_{x\to 0} (d/dx)s^{(i)}(x) = 0$. But

$$\frac{d}{dx} s^{(i)}(x) = \sum_n \sum_m S^{(n,m)}(x) \frac{d}{dx} s_{n,m}(x)$$

where $S^{(n,m)}(x)$ is the (n, m)th cofactor of $S^{(i)}(x)$.

We use the explicit form of $s_{n,m}(x)$ along with the fact that

$$S^{(n,m)}(x) = S^{(n,m)}(0) + o(1) \text{ as } x \to 0$$

to deduce that

$$\frac{d}{dx} s^{(i)}(x) = \sum_{n\neq i} \sum_{m\neq n} q(n, m)\omega(n, m) [S^{(n,n)}(0) - S^{(n,m)}(0)] + o(1).$$

Since the row runs of $S^{(i)}(0)$ are all zero, Lemma 6.4 applies:

$$S^{(n,n)}(0) = S^{(n,m)}(0).$$

Whence

$$(d/dx)s^{(i)}(x) = o(1) \quad \text{as} \quad x \to 0.$$

REFERENCES

[1] Anderson, T. W. and Goodman, Leo (1957). Statistical inference about Markov chains. *Ann. Math. Statist.* **28** 89–110.

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

[3] Birkhoff, Garrett and MacLane, Saunders (1953). *A Survey of Modern Algebra.* Macmillan, New York.

[4] Chernoff, Herman (1956). Large sample theory: Parametric case. *Ann. Math. Statist.* **27** 1–22.

[5] Cramér, Harald (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[6] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.