JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 19, 409-430 (1967)

On Multitype Branching Processes with $\rho \leqslant 1$

A. JOFFE*

Université de Montréal, Québec

AND

F. SPITZER**

Mathematics Department, Cornell University, Ithaca, New York 14850

Submitted by Samuel Karlin

1. INTRODUCTION

For the single type branching process or Galton-Watson process several basic probabilistic phenomena have only recently been studied under natural hypotheses. For example Yaglom's theorem [1, p. 18] asserting the existence of a limiting probability measure of Z_n for processes Z_n with mean less than one, conditioned on $Z_n \neq 0$, is valid without additional hypotheses [2, 3]. Here, in treating the processes with k types ($k \geqslant 1$) our goal is a similar discussion of basic phenomena, under their weakest possible and therefore natural conditions. While all the phenomena treated here are known (except possibly the local limit result in Theorem 5), we nevertheless decided to give a self contained treatment, accessible to a reader unfamiliar with the single type theory.

Our treatment however is restricted entirely to the case $\rho \leqslant 1$, ρ being the largest (positive) eigenvalue of the expectation matrix M, ($\rho = m = E[Z_1]$ in the case $k = 1$). The reason is the existence of a complete theory by Kesten and Stigum [4-6] for the case $\rho > 1$.

For the specialist we give the following literature references concerning known results and methods used in this paper (The theorems referred to below are listed in Section 3.) In Theorem 1 only the statement of uniformity in (3.3) is new, while (3.1) and (3.2) are proved in Harris [1]. Equation (3.3) without uniformity, and under excessive moment assumptions, is due to Jiřina [7] for $\rho < 1$, and to Mullikin [8], when $\rho = 1$. Similar remarks apply to Theorem 2 which is an immediate consequence of Theorem 1. Theorem 3

* This research was supported by the N.I.H., University of Wisconsin, and by the N.R.C., Canada.

** This research was supported by the N.S.F.

was proved by Jiřina [7] under the hypothesis that second moments exist, but the independence of the limiting measure on the initial state of the process is not made clear in [7], nor was it pointed out in [1]. Theorem 4 was recently proved by Heathcote *et al.* [2] in the case of a single particle ($k = 1$). Theorem 5 seems to be a new, but very simple observation. It can be used to give a complete description of the entrance boundary (invariant measures) for processes with $\rho < 1$, satisfying the condition (3.10) of Theorem 4. Finally, in Theorem 6 both the formalism (see Eq. (4.35) in the proof) and its corollaries (3.14) and (3.15) are due to Mullikin [8], who developed them under stronger hypotheses but in a far more general context. To obtain the full strength of Theorem 6 it seems necessary to use the uniformity assertion in (3.3). The result (3.13) of Theorem 6 is the obvious k-dimensional analogue of Theorem 1 in [9]. In fact, straightforward imitation of the methods in [9], with careful use of (3.3) will yield the extension of several other results in [9] from $k = 1$ to an arbitrary number of types.

2. Definitions and Notation

We call X the set of all k-tuples $i = (i_1, i_2, \ldots, i_k)$ whose elements are non-negative integers i_ν. The zero element is 0, and e_α will denote the basis vector i with $i_\alpha = 1$, all other $i_\nu = 0$.

The k-dimensional cube of points $s = (s_1, s_2, \ldots, s_\nu)$, $0 \leqslant s_\nu \leqslant 1$, is denoted C. It has zero element 0, and unit element $\mathbf{1} = (1, 1, \ldots, 1)$. The obvious partial order on C is $s \leqslant t$, when $s_\nu \leqslant t_\nu$ for all ν and $s < t$ when $s_\nu < t_\nu$ for all ν. A mapping f of C into C or of C into the reals will be called monotone (nondecreasing) if $s \leqslant t$ implies $f(s) \leqslant f(t)$, and $f(t) \nearrow \mathbf{1}$ means that f is monotone and tends to $\mathbf{1}$ as $t \to \mathbf{1}$ in C. Similarly if $F(t)$ is a k by k matrix (quadratic form) for each $t \in C$, then $F(t) \searrow 0$ will mean that $t \leqslant s$ implies $F(t) - F(s)$ has non-negative elements (coefficients) and $F(t)$ tends to the 0 matrix (quadratic form) as $t \to \mathbf{1}$ in C. Finally we shall use the notation s^i, when $i \in X$, $s \in C$, to denote the product $\prod_{\nu=1}^{k} (s_\nu)^{i_\nu}$.

For each integer α, $1 \leqslant \alpha \leqslant k$, we assume given a probability measure p_α on X, and in terms of these given measures one proceeds to define the branching process Z_n, as a Markov chain with state space X. Its transition probabilities are given by

$$P(0, 0) = 1, \qquad P(e_\alpha, i) = p_\alpha(i), \qquad i \in X,$$

$$P(i, j) = p_1^{(i_1)} * p_2^{(i_2)} * \cdots * p_k^{(i_k)}(j), \qquad j \in X, \tag{2.1}$$

where $p_\alpha^{(n)}$ is the n-fold convolution of p_α with itself. Thus $P(i, j)$ is an $(i_1 + i_2 + \cdots + i_k)$ fold convolution. The n-step transition probabilities are now uniquely determined by the requirement that Z_n be a Markov chain. For each initial position $Z_0 = i \in X$, we therefore obtain a probability measure $P_i[\ ]$ on the sample space of the process. The corresponding expectation will be denoted $E_i[\ ]$. Whenever A is of positive P_i-measure, we will write $P_i[B \mid A]$ for $P_i[A \cap B]/P_i[A]$.

The so called generating function of (Z_n) is defined by

$$f(s) = f_1(s) = (f_{1,1}(s), f_{1,2}(s), \ldots, f_{1,k}(s)),$$
$$f_{1,\alpha}(s) = E_{e_\alpha}[s^{Z_1}] = \sum_{i \in X} p_\alpha(i)\, s^i, \qquad s \in C. \tag{2.2}$$

Since p_α is a probability measure, f maps C into C. It follows from (2.1) together with a little computation based on the Markov property (cf. [1], p. 36), that the n-fold composition of f,

$$f_n(s) = (f_{n,1}(s), f_{n,2}(s), \ldots, f_{n,k}(s)) = f \circ f \circ \cdots \circ f(s), \qquad n \geqslant 1,$$

satisfies

$$f_{n,\alpha}(s) = E_{e_\alpha}[s^{Z_n}], \qquad n \geqslant 1, \qquad 1 \leqslant \alpha \leqslant k, \qquad s \in C. \tag{2.3}$$

It is consistent with the above to define $f_0(s) = s$, $s \in C$.

Throughout we shall assume the finiteness of the k by k matrix M,

$$M_{\alpha\beta} = E_{e_\alpha}[Z_{1,\beta}] = \sum_{i \in X} p_\alpha(i)\, i_\beta\,, \qquad 1 \leqslant \alpha, \beta \leqslant k. \tag{2.4}$$

Furthermore it is assumed that there exists a positive integer n such that $M^n > 0$ (all elements positive). This is known [10; Appendix 2] to entail the existence of a positive number ρ (the largest eigenvalue of M) such that

$$\lim_{n\to\infty} \frac{M^n}{\rho^n} = u \otimes v \ ((u \otimes v)_{\alpha,\beta} = u_\alpha v_\beta). \tag{2.5}$$

where $u > 0$, $v > 0$ (all components positive) are the unique nonnegative eigenvectors of M, corresponding to ρ. Thus

$$Mu = \rho u, \qquad vM = \rho v, \qquad v \cdot u = 1, \qquad u \cdot 1 = 1, \tag{2.6}$$

the normalization in (2.6) being assumed for the sake of convenience.

(*Note:* A product $x \cdot y$ when x, y are in euclidean space R_k, will always denote the scalar product, and xMy the obvious bilinear form. It is not required that x and y lie in C; we may for instance have $x \in X$, and $y \in C$.)

The above assumptions concerning M suffice when $\rho < 1$; In the case $\rho = 1$, however, we assume more. First of all we rule out the degenerate case described by

$$f(t) = Mt, \qquad t \in C. \tag{2.7}$$

Since (2.7) implies that $1 = f(1) = M1$, we see that $\rho = 1$, so that (2.7) cannot occur when $\rho < 1$. In fact (2.7) describes the degenerate case when the total number of particles is independent of time. Then the study of Z_n, with $Z_0 = e_\alpha$, reduces to the study of an irreducible Markov chain with k states, whose transition matrix is M.

Finally we assume, for the sole purpose of Theorem 6, with $\rho = 1$, that all the second moments of Z_1 are finite. Equivalently

$$(q[t])_\alpha = \frac{1}{2} \sum_{\nu=1}^{k} \sum_{\mu=1}^{k} t_\mu t_\nu E_{e_\alpha}[Z_{1,\nu} Z_{1,\mu} - \delta(\mu, \nu) Z_{1,\nu}] < \infty \tag{2.8}$$

for all $t \in C$, $0 \leqslant \alpha \leqslant k$. Here δ is the Kronecker delta, and $q[\ ]$ is clearly a vector whose components are quadratic forms of order k. In terms of $q[\ ]$ and the left eigenvector v in (2.6) we shall find it useful to define the quadratic form $Q[\ ]$, by

$$Q[t] = v \cdot q[t], \qquad t \in C. \tag{2.9}$$

We shall be particularly interested in the value of Q at u (see Theorem 6). For the simple Galton-Watson process ($k = 1$) with $\rho = 1$, we have $q[t] = t^2\sigma^2/2$ where σ^2 is the variance of Z_1 when $Z_0 = 1$. Since $M = \rho = 1$, $u = v = 1$, one has $Q[u] = \sigma^2/2$.

3. Principal Results

THEOREM 1. *When $\rho < 1$ and when $\rho = 1$ and $f(t) \neq Mt$,*

$$f_n(t) \neq 1 \qquad on \qquad C - \{1\}, \qquad for \qquad n \geqslant 0. \tag{3.1}$$

$$1 - f_n(t) \to 0, \qquad as \qquad n \to \infty, \qquad uniformly\ on\ C. \tag{3.2}$$

$$\frac{1 - f_n(t)}{v \cdot [1 - f_n(t)]} \to u, \qquad as \qquad n \to \infty, \qquad uniformly\ on\ C - \{1\}. \tag{3.3}$$

THEOREM 2. *When $\rho < 1$, there is a monotone nonincreasing real function $\gamma(t)$ on C, such that*

$$\frac{v \cdot [1 - f_n(t)]}{\rho^n} \searrow \gamma(t) \geqslant 0, \qquad as \qquad n \to \infty, \qquad t \in C, \tag{3.4}$$

$$\frac{1 - f_n(t)}{\rho^n} \to \gamma(t)\, u, \qquad as \qquad n \to \infty, \qquad t \in C, \tag{3.5}$$

$$\rho^{-n} P_i[Z_n \neq 0] \to \gamma(0)\,(i \cdot u), \qquad as \qquad n \to \infty, \tag{3.6}$$

for each $i \in X$.

THEOREM 3. *When $\rho < 1$, the conditional probability measure*

$$P_i[Z_n = j \mid Z_n \neq 0] = \frac{P_i[Z_n = j]}{P_i[Z_n \neq 0]} \to \chi(j), \qquad j \in X - \{0\}, \quad \text{as} \quad n \to \infty \tag{3.7}$$

for each $i \in X - \{0\}$. Here χ is a probability measure on $X - \{0\}$, which is independent of the initial point $Z_0 = i$ in $X - \{0\}$. The mean vector $m = (m_1, m_2, \ldots m_k)$ of χ defined by

$$m_\nu = \sum_{j \in X - \{0\}} \chi(j)\, j_\nu \tag{3.8}$$

is finite if and only if the constant $\gamma = \gamma(0)$ in Theorem 2 is positive. In the latter case

$$m = \frac{1}{\gamma}\, v. \tag{3.9}$$

THEOREM 4. *If $\rho < 1$, then the constant $\gamma = \gamma(0)$ in Theorems 2 and 3 is positive if and only if*

$$E_{e_\alpha}[Z_{1,\beta} \log Z_{1,\beta}] = \sum_{i \in X - \{0\}} p_\alpha(i)\, i_\beta \log i_\beta < \infty \tag{3.10}$$

for all $1 \leqslant \alpha \leqslant k$ and all $1 \leqslant \beta \leqslant k$.

THEOREM 5. *Suppose that $\rho < 1$ and that (3.10) holds (so that $\gamma > 0$). Suppose also that $j \to \infty$ in the sense that $j \cdot u \to +\infty$, and that $n \to +\infty$, in such a way that $(j \cdot u)\, \rho^n \gamma \to A > 0$. Then*

$$\lim P_j[Z_n = i] = \chi_A(i), \qquad i \in X, \tag{3.11}$$

where χ_A is a probability measure on X. It has generating function

$$\sum_{i \in X} \chi_A(i)\, s^i = e^{-A[1 - g(s)]}, \qquad g(s) = \sum_{i \in X - \{0\}} \chi(i)\, s^i, \qquad s \in C, \tag{3.12}$$

where χ is the limiting measure in Theorem 3.

THEOREM 6. *Suppose that $\rho = 1, f(t) \neq Mt$, and that the second moments of Z_1 are finite. Then*

$$\frac{1}{n}\left\{\frac{1}{v \cdot [1 - f_n(t)]} - \frac{1}{v \cdot [1 - t]}\right\} \to Q[u] > 0, \qquad \text{as} \qquad n \to \infty, \tag{3.13}$$

uniformly for all $t \in C - \{1\}$. *In particular*

$$nP_i[Z_n \neq 0] \to \frac{i \cdot u}{Q[u]}, \qquad as \qquad n \to \infty, \tag{3.14}$$

for each fixed i *in* $X - \{0\}$. *Also, for every vector* x *in* R_k, *and each* i *in* $X - \{0\}$.

$$P_i[Z_n \leqslant nx \mid Z_n \neq 0] \to \begin{cases} 0 & \text{if} \quad \min_{1 \leqslant \alpha \leqslant k} x_\alpha < 0 \\ 1 - \exp\left\{\frac{\sqrt{k}}{Q[u]} \min_{1 \leqslant \alpha \leqslant k} \left(\frac{x_\alpha}{v_\alpha}\right)\right\}, & x \geqslant 0. \end{cases} \tag{3.15}$$

This limit measure has its support on the ray $x = cv$, $c > 0$.

4. Proofs and Discussion

As in the case $k = 1$ (the Galton-Watson process) the elementary theory depends on little more than the expansion of $f(t)$ about $t = 1$. We give this expansion to first order now, and to second order later on, in the proof of Theorem 6. For fixed i in X, let

$$\varphi(t) = t^i = \prod_1^k t_\nu^{i_\nu}, \qquad \psi(t) = \varphi(1 - t), \qquad t \in C. \tag{4.1}$$

Then

$$\psi(t) - 1 = \psi(t) - \psi(0) = \int_0^1 \frac{d}{d\xi} \psi(\xi t)\, d\xi = \int_0^1 t \cdot \operatorname{grad} \psi(\xi t)\, d\xi \tag{4.2}$$

$$= - \sum_{\beta=1}^k t_\beta i_\beta + \sum_{\beta=1}^k t_\beta \epsilon_\beta(i, t), \tag{4.2}$$

where

$$\epsilon_\beta(i, t) = i_\beta \left[1 - \int_0^1 \frac{\prod_1^k (1 - \xi t_\nu)^{i_\nu}}{1 - \xi t_\beta}\, d\xi\right], \qquad t \in C, \qquad i \in X. \tag{4.3}$$

We shall now replace t by $1 - t$ in (4.2), multiply (4.2) by $p_\alpha(i)$, and then sum i over X. To simplify the notation we introduce first the matrix $E(t)$, $t \in C$, defined by

$$E_{\alpha\beta}(t) = \sum_{i \in X} p_\alpha(i)\, \epsilon_\beta(i, 1 - t), \qquad t \in C, \qquad 1 \leqslant \alpha, \qquad \beta \leqslant k. \tag{4.4}$$

Observe that (4.3) and (4.4) imply, for s, $t \in C$,

$$0 \leqslant E(t) \leqslant M, \qquad t \leqslant s \Rightarrow E(t) \geqslant E(s),$$
$$E(t) \to 0 \quad \text{as} \quad t \to 1 \quad \text{in} \quad C. \tag{4.5}$$

Now the result of summing (4.2) with respect to the measure p_α is

$$\sum_{i \in X} (t^i - 1)\, p_\alpha(i) = f_{1,\alpha}(t) - 1 = - \sum_{\beta=1}^{k} [M_{\alpha\beta} - E_{\alpha\beta}(t)]\,(1 - t_\beta),$$

or in vector notation

$$1 - f(t) = [M - E(t)]\,(1 - t), \qquad t \in C. \tag{4.6}$$

The proof of Theorem 1 begins with the observation that it suffices to prove (3.1) when $n = 1$. (For $f_0(t) = t \neq 1$ when $t \neq 1$, and if $f_n(t) = 1$ for $n > 1$ and $t \neq 1$ we may choose $m \geqslant 1$ as the smallest integer such that $f_m(t) = 1$. But then $f_{m-1}(t) \neq 1$ and $f[f_{m-1}(t)] = 1$ shows that (3.1) is false when $n = 1$.) If $f(t) = 1$, $t \neq 1$, then by the monotonicity of f, we may choose $t' \in C$, with some component $t'_\beta < 1$ and all other components one, so that also $f(t') = 1$. But this implies (see 2.2) that $p_\alpha(i) = 0$ for all α and all i with $i_\beta > 0$. Hence (see 2.4) $M_{\alpha\beta} = 0$ for all α which contradicts the hypothesis that $M^n > 0$ for some n. Thus (3.1) holds.

For the proof of (3.2) observe that

$$0 \leqslant 1 - f_n(t) \leqslant 1 - f_n(0), \qquad f_n(0) \nearrow q \in C, \qquad \text{as} \qquad n \to \infty. \tag{4.7}$$

Setting $t = f_n(0)$ in (4.6) and using the monotonicity of E in (4.5),

$$1 - q = [M - E(q)]\,(1 - q). \tag{4.8}$$

Forming the scalar product with the left eigenvector v

$$v \cdot (1 - q) = \rho v \cdot (1 - q) - v \cdot E(q)\,(1 - q).$$

Since $\rho \leqslant 1$, $v > 0$ and $E(q) \geqslant 0$, it is clear that $E(q)\,(1 - q) = 0$. But then (4.8) reduces to $1 - q = M(1 - q)$, which shows that $q = 1$ in the case when $\rho < 1$. In the case when $\rho = 1$, the following reasoning will prove that $q = 1$. Suppose it is not so. Since $1 - q = M(1 - q)$, $1 - q$ must be a positive multiple of the right eigenvector u. Hence $E(1 - cu)\,u = 0$ for some $c > 0$. Since $u > 0$, also $E(1 - cu) = 0$, and $1 - cu < 1$. In other words $E(t) = 0$ for some $t < 1$ in C. Inspection of (4.3) and (4.4) shows that this implies that $E(t) \equiv 0$ on C. This in turn gives $1 - f(t) = M(1 - t)$ or $f(t) - Mt = 1 - M1$. Setting $t = 0$ yields $1 - M1 = f(0) \geqslant 0$. Multiplying this inequality by v gives $0 = v \cdot \mathbf{1} - v \cdot \mathbf{1} = v \cdot f(0) \geqslant 0$ so that $f(0) = 0$. Hence $\mathbf{1} = M\mathbf{1}$, and $f(t) = Mt$, which means that we are in the degenerate case which was ruled out.

We have shown that $q = \mathbf{1}$, and the inequality in (4.7) assures uniform convergence of $\mathbf{1} - f_n(t)$ to $\mathbf{1} - q = 0$. Therefore the proof of (3.2) is complete.

REMARK. Statement (3.2) implies that the map $f: C \to C$ has no fixed point, other than $\mathbf{1}$. Probabilistically the content of (3.2) is that

$$(\mathbf{1} - f_n(0))_\alpha = P_{e_\alpha}[Z_n \neq 0] \to 0, \qquad n \to \infty,$$

which of course implies

$$P_i[Z_n \neq 0] \to 0, \qquad n \to \infty, \tag{4.9}$$

for each $i \in X$, since $P_i[Z_n = 0]$ is a finite product of terms which tend to one.

To complete the proof of Theorem 1 we require information concerning infinite products of positive matrices, which is worth presenting in some detail. We take a non-negative matrix P, such that $P^n > 0$ for some $n \geqslant 1$, and suppose the largest eigenvalue of P to be one (this is no restriction as far as part (i) of Lemma 1 is concerned). The eigenvectors $u > 0$, $v > 0$ can be chosen to satisfy

$$Pu = u, \qquad vP = v, \qquad v \cdot u = 1, \qquad u \cdot \mathbf{1} = 1.$$

We suppose given a sequence of matrices A_k, $0 \leqslant A_k \leqslant P$, define

$$B_n = (P - A_n)(P - A_{n-1}) \cdots (P - A_1) = \prod_1^n (P - A_k), \qquad n \geqslant 1,$$

and finally consider a vector x, satisfying

$$x \geqslant 0, \qquad B_n x \neq 0 \qquad \text{for all} \qquad n \geqslant 1.$$

LEMMA 1. *Under the above conditions concerning P, u, v, A_n, B_n, and x we have*

(i) *if*

$$\lim_{n\to\infty} A_n = 0, \qquad \textit{then} \qquad \lim_{n\to\infty} \frac{B_n x}{v B_n x} = u;$$

(ii) $\lim_{n\to\infty} B_n x$ *exists always. It is non-zero if and only if the sum* $\sum_1^\infty A_k$ *converges.*

PROOF. Let $\lim P^n = R = u \otimes v$. Thus $R > 0$ and we can find a null sequence $\delta_n \geqslant 0$ such that

$$(1 - \delta_n) R \leqslant P^n \leqslant (1 + \delta_n) R, \qquad n \geqslant 1.$$

Using the hypothesis of part (i) that $A_n \to 0$, we may also choose a null sequence $\alpha_n \geqslant 0$ such that

$$0 \leqslant A_n \leqslant \alpha_n R, \qquad n \geqslant 1.$$

Note further that since $PR = RP = R$ we have for arbitrary nonnegative real numbers $\beta_1, \beta_2, \ldots, B_n$,

$$\prod_{k=1}^{n} (P - \beta_k R) = P^n - \left\{1 - \prod_{k=1}^{n} (1 - \beta_k)\right\} R \geqslant P^n - \sum_{1}^{n} \beta_k R.$$

Combining this identity with the two preceding inequalities, one obtains

$$\begin{aligned}\left\{1 - \delta_m - \sum_{k=n-m+1}^{n} \alpha_k\right\} R &\leqslant (P - \alpha_n R)(P - \alpha_{n-1} R) \cdots (P - \alpha_{n-m+1} R) \\ &\leqslant (P - A_n)(P - A_{n-1}) \cdots (P - A_{n-m+1}) \leqslant P^m \\ &\leqslant 1 + \delta_m R, \end{aligned} \tag{4.10}$$

whenever $n \geqslant m \geqslant 1$. Now we take $w = B_{n-m} x$ and apply (4.10) to

$$B_n x = (P - A_n) \cdots (P - A_{n-m+1}) w,$$

and also to $vB_n x$. Then (4.10) gives

$$\frac{1 - \sum_{n-m+1}^{n} \alpha_k - \delta_m}{1 + \delta_m} \frac{Rw}{vRw} \leqslant \frac{B_n x}{vB_n x} \leqslant \frac{1 + \delta_m}{1 - \sum_{n-m+1}^{n} \alpha_n - \delta_m} \frac{Rw}{vRw}.$$

Observe, however, that $Rw/vRw = u$, in view of the normalization $v \cdot u = 1$, regardless of the point w. Hence if $\| \ \|$ is the supremum norm,

$$\left\| \frac{B_n x}{vB_n x} - u \right\| \leqslant \frac{2\delta_m + \sum_{n-m+1}^{n} \alpha_k}{1 - \sum_{n-m+1}^{n} \alpha_k - \delta_m}. \tag{4.11}$$

By letting first n tend to infinity, and then m, we see that (i) is true.

REMARK. In the proof of the last part of Theorem 1 we shall use the full strength of (4.11). The bound in (4.11) is independent of x, and it will also remain valid if the sequence A_k were replaced by any sequence A_k', satisfying $0 \leqslant A_k' \leqslant A_k \leqslant \alpha_k R$.

Now we shall prove part (ii) of the lemma, at first under the additional hypothesis that $\lim A_n = 0$. We may then use (i), in the form

$$B_n x = (vB_n x)(u + \delta_n(x)), \tag{4.12}$$

where $\delta_n(x) \to 0$ as $n \to \infty$. Let $\Delta_n = vB_n x$, observing that

$$\Delta_{n+1} = v(P - A_{n+1})B_n x = \Delta_n - vA_{n+1}B_n x = \Delta_n\{1 - vA_{n+1}(u + \delta_n(x))\}.$$

Therefore

$$\Delta_n = \prod_{k=2}^{n} \{1 - vA_k(u + \delta_{k-1}(x))\}\,\Delta_1, \qquad n \geqslant 2. \tag{4.13}$$

Since $u > 0$, (4.13) shows that Δ_n decreases (for large enough n) to a limit $L \geqslant 0$. Further $L > 0$ if and only if

$$\sum_{k=2}^{\infty} vA_k(u + \delta_{k-1}(x)) < \infty.$$

As $u > 0$, $v > 0$ and $\delta_k \to 0$, this can occur if and only if $\sum A_k < \infty$. But now (4.12) shows that $B_n x$ always has the limit Lu, which is non-zero if and only if $\sum A_k < \infty$.

Suppose finally that $0 \leqslant A_k \leqslant P$, and that $\lim A_k$ either fails to exist, or is nonzero. Then of course $\sum A_k = \infty$, and we have to prove that $B_n x \to 0$. We may choose a sequence of matrices A_k' such that $0 \leqslant A_k' \leqslant A_k$ and in addition $A_k' \to 0$, $\sum A_k' = \infty$. If $B_n' = \prod_1^n (P - A_k')$, then $A_k' \leqslant A_k$ gives $B_n' x \geqslant B_n x \geqslant 0$. But since (ii) has been proved for the sequence A_k' we have $\lim B_n' x = 0$ and hence $\lim B_n x = 0$.

To complete the proof of Theorem 1 we write, using (4.6)

$$1 - f_n(t) = (M - E_n(t))(M - E_{n-1}(t)) \cdots (M - E_1(t))(1 - t),$$

where $E_n(t) = E[f_{n-1}(t)]$. Now divide by ρ^n, call $M/\rho = P$, and $E_n(t)/\rho = A_n(t)$. Finally let $B_n(t) = (P - A_n)(P - A_{n-1}) \cdots (P - A_1)$. Then

$$\frac{1 - f_n(t)}{v \cdot [1 - f(t)]} = \frac{B_n(t)(1 - t)}{vB_n(t)(1 - t)}.$$

Then the hypotheses of Lemma 1 are all satisfied. (Note that one must use

(3.1) to verify that $x = 1 - t$ has the property that $B_n(t)\, x \neq 0$ for all n, when $t \neq 1$.) Thus part (i) of Lemma 1 gives

$$\lim_{n\to\infty} \frac{1 - f_n(t)}{v \cdot [1 - f_n(t)]} = u, \qquad t \in C - \{1\}.$$

But in fact the convergence is uniform, in view of the remark following the estimate (4.11). If we use this estimate when $t = 0$, then it gives the uniform convergence in (3.3), since (4.11) is uniform in x and since we know from (4.5) that $0 \leqslant A_n(t) \leqslant A_n(0)$, for all $n \geqslant 1$ and $t \in C - \{1\}$. This completes the proof of Theorem 1.

Theorem 2 is almost immediate. If

$$\Delta_n(t) = \frac{v \cdot [1 - f_n(t)]}{\rho^n}, \qquad t \in C,$$

then by (2.6) and (4.6)

$$\Delta_{n+1}(t) = \frac{v \cdot [1 - f_n(t)]}{\rho^n} - vE[f_n(t)]\, \frac{1 - f_n(t)}{\rho^{n+1}} \leqslant \Delta_n(t). \tag{4.14}$$

Hence $\Delta_n(t)$ decreases to a limit $\gamma(t)$ which is monotone and nonnegative on C, since each Δ_n has these properties. That proves (3.4), and (3.5) is obtained by combining (3.3) and (3.4).

The probabilistic statement (3.6) follows from (3.5) with $t = 0$. For $i = 0$, (3.6) is trivial, and for $i \neq 0$

$$P_i[Z_n \neq 0] = 1 - P_i[Z_n = 0] = 1 - \prod_{\nu=1}^{k} [f_{n,\nu}(0)]^{i_\nu}$$

$$= 1 - \left[1 - \rho^n \frac{1 - f_n(0)}{\rho^n}\right]^i. \tag{4.15}$$

To obtain (3.6) from (3.5) it therefore suffices to prove the following.

Lemma 2. *Let $x(n)$ be a sequence in C with* $\lim x(n) = x$, *and suppose that* $\epsilon_n > 0$ *with* $\lim \epsilon_n = 0$. *Then*

$$\lim_{n\to\infty} \frac{1}{\epsilon_n} \{1 - [1 - \epsilon_n x(n)]^i\} = i \cdot x.$$

The proof is immediate from (4.1) and (4.2), and the application of the lemma with $\epsilon_n = \rho^n$, $x(n) = \rho^{-n}(1 - f_n(0))$, and $x = \gamma(0)\, u$, completes the proof of Theorem 2.

The proof of Theorem 3 will be conducted by the method of generating functions. For a fixed $i \in X - \{0\}$, the generating functions of the probability measures in (3.7) are

$$g_n(i, s) = \frac{\sum_{j \in X} P_i[Z_n = j]\, s^j - P_i[Z_n = 0]}{1 - P_i[Z_n = 0]}$$

$$= 1 - \frac{1 - [f_n(s)]^i}{1 - [f_n(0)]^i}, \qquad s \in C. \tag{4.16}$$

Suppose now that we are able to show that there exists a real function g on C, such that

$$\lim_{n \to \infty} g_n(e_\alpha, s) = g(s), \qquad 1 \leqslant \alpha \leqslant k, \tag{4.17}$$

for each basis vector e_α of X. Then call $f_n(s) = \mathbf{1} - a(n)$, $f_n(0) = \mathbf{1} - b(n)$, observe that $a(n) \to 0$ and $b(n) \to 0$ by (3.2) and that $(a(n))_\alpha/(b(n))_\alpha$ tends to $1 - g(s)$. For any such pair of sequences in C it is easy to see that

$$\lim_{n \to \infty} \frac{1 - [\mathbf{1} - a(n)]^i}{1 - [\mathbf{1} - b(n)]^i} = \lim \frac{i \cdot a(n)}{i \cdot b(n)} = 1 - g(s).$$

Thus (4.17) will imply that the generating functions $g_n(i, s)$ in (4.16) all have the same limit, independent of i.

To prove (4.17) we require another fact concerning the cube C.

LEMMA 3. *Take a and b in C, with $a \geqslant b$, $b < \mathbf{1}$, and let*

$$r = \min_{1 \leqslant \nu \leqslant k} \frac{1 - a_\nu}{1 - b_\nu}.$$

Then, for all $j \in X - \{0\}$, $(1 - a^j)/(1 - b^j) \geqslant r$.

PROOF. It clearly suffices to prove the lemma when $\mathbf{1} - a = r(\mathbf{1} - b)$, or $a = \mathbf{1} - r(\mathbf{1} - b)$. Then the lemma reduces to the inequality

$$1 - r \geqslant [(1 - r)\, \mathbf{1} + rb]^j - rb^j.$$

In terms of the function $\varphi(t) = t^j$,

$$\varphi[(1 - r)\, \mathbf{1} + rb] \leqslant (1 - r)\, \varphi(\mathbf{1}) + r\varphi(b), \qquad b \in C,$$

which is obvious since $0 \leqslant r \leqslant 1$ and φ is concave.

This lemma will now be used to construct a sequence $e(n)$, $n \geqslant 0$, of basis vectors in X, with the property that the sequence

$$g_n(e(n), s) = E_{e(n)}[s^{Z_n} \mid Z_n \neq 0],$$

is monotone nonincreasing in n, and therefore tends to a limit. The choice of $e(n)$ from among the set $\{e_1, e_2, \dots e_k\}$ will depend on $s \in C$, but that is immaterial since we shall hold s fixed. The Markov property gives, for any pair x, y in $X - \{0\}$, and $n \geqslant 1$

$$g_{n+1}(y, s) - g_n(x, s) = \sum_{j \in X - \{0\}} P_y[Z_1 = j \mid Z_{n+1} \neq 0] \{g_n(j, s) - g_n(x, s)\} \tag{4.18}$$

Hence $g_{n+1}(y, s) - g_n(x, s) \leqslant 0$ provided x is chosen in such a way that $g_n(j, s) \leqslant g_n(x, s)$ for all $j \in X - \{0\}$. According to the definition of g_n in (4.16) this will be true if

$$\frac{1 - [f_n(s)]^j}{1 - [f_n(0)]^j} \geqslant \frac{1 - [f_n(s)]^x}{1 - [f_n(0)]^x}, \qquad j \in X - \{0\},$$

and by Lemma 3 that can be achieved by choosing $x = e(n) = e_\alpha$, where $\nu = \alpha$ minimizes the ratio $[1 - f_{n,\nu}(s)]/[1 - f_{n,\nu}(0)]$. Now we are free to choose $e(n + 1)$, and inductively one obtains a sequence $e(n)$ chosen from the basis vectors of X, such that

$$\lim_{n \to \infty} g_n(e(n), s) = g(s). \tag{4.19}$$

To prove (4.17) it suffices to show that the limits

$$\lim_{n \to \infty} \frac{x \cdot [1 - f_n(s)]}{x \cdot [1 - f_n(0)]} = 1 - g(s), \qquad s \in C, \tag{4.20}$$

exist for an arbitrary $x \in C - \{0\}$. (We recover (4.17) by taking a unit vector for x.)

To obtain (4.20) from (4.19) decompose

$$\frac{x \cdot [1 - f_n(s)]}{x \cdot [1 - f_n(0)]} = \frac{x \cdot [1 - f_n(s)]}{v \cdot [1 - f_n(s)]} \frac{v \cdot [1 - f_n(0)]}{x \cdot [1 - f_n(0)]} \frac{v \cdot [1 - f_n(s)]}{e(n) \cdot [1 - f_n(s)]}$$

$$\times \frac{e(n) \cdot [1 - f_n(0)]}{v \cdot [1 - f_n(0)]} \frac{e(n) \cdot [1 - f_n(s)]}{e(n) \cdot [1 - f_n(0)]} = A_n B_n C_n D_n E_n .$$

In view of (3.3) we have $A_n \to x \cdot u$ and $B_n \to (x \cdot u)^{-1}$ so that $A_n B_n$ tends to one as $n \to \infty$. The sequences C_n and D_n need not converge separately, but since $e(n)$ can only assume k distinct values one may still conclude from (3.3) that the products $C_n D_n$ tend to one. Finally E_n converges to $1 - g(s)$ by (4.19). Therefore we have now also proved (4.17), which was shown to imply that all the generating functions in (4.16) converge to $g(s)$, independent of $i \in X - \{0\}$.

At this point the continuity theorem for generating functions (a simple compactness argument) will complete the proof of (3.7) in Theorem 3, if one can show that $g(1) = 1$, and that g is continuous at the point 1 of C. That $g(1) = 1$ is obvious from its definition as the limit in (4.19). From (4.20), (4.6), and (4.5)

$$\begin{aligned} 1 - g[f(t)] &= \lim_{n\to\infty} \frac{v \cdot [1 - f_{n+1}(t)]}{v \cdot [1 - f_n(0)]} \\ &= [1 - g(t)] \lim_{n\to\infty} \frac{v \cdot [1 - f_{n+1}(0)]}{v \cdot [1 - f_n(0)]} = \rho[1 - g(t)]. \end{aligned} \tag{4.21}$$

By iteration

$$g[f_n(0)] = 1 - \rho^n, \qquad n \geqslant 1. \tag{4.22}$$

Now g is monotone on C in view of (4.20) and if $t(n) \to 1$ in C we can use (3.2) and (3.3) to extract a subsequence $k(n)$ such that $k(n) \to \infty$, $f_{k(n)}(0) \leqslant t(n)$, so that by (4.22)

$$1 - \rho^{k(n)} = g[f_{k(n)}(0)] \leqslant g[t(n)] \leqslant 1.$$

Thus g is continuous at 1 and (3.7) holds.

The continuity theorem implies that $g(s) = \sum \chi(j)\, s^j$. Therefore the mean vector m, defined in (3.8) is the gradient of g evaluated at 1. If m is finite we therefore have $[1 - g(x)]/m \cdot (1 - x) \to 1$ as $x \to 1$ in C. Since $v > 0$ it follows that

$$\lim_{n\to\infty} \frac{1 - g[f_n(0)]}{v \cdot [1 - f_n(0)]} < \infty. \tag{4.23}$$

Conversely, if (4.23) holds, then for arbitrary α, $1 \leqslant \alpha \leqslant k$

$$\frac{\partial}{\partial t_\alpha} g(1) = \lim_{\epsilon\to 0} \frac{1 - g(1 - \epsilon e_\alpha)}{\epsilon} \leqslant C \lim_{n\to\infty} \frac{1 - g[f_n(0)]}{v \cdot [1 - f_n(0)]} < \infty$$

for some $C > 0$, since for each n, α, $1 - \epsilon e_\alpha \geqslant f_n(0)$ for sufficiently small ϵ. Since g is concave it follows that $m = \operatorname{grad} g(1)$ is finite if and only if (4.23) holds. But in view of (4.22) and (3.4), (4.23) is equivalent to

$$\lim_{n\to\infty} \frac{\rho^n}{v \cdot [1 - f_n(0)]} = \frac{1}{\gamma(0)} < \infty, \tag{4.24}$$

which proves that m exists if and only if $\gamma > 0$.

Suppose finally that m exists. Differentiation of (4.21) gives

$$\sum_{\alpha=1}^{k} \frac{\partial f_{1,\alpha}(t)}{\partial t_\beta} \frac{\partial g}{\partial t_\alpha} [f(t)] = \rho \frac{\partial g}{\partial t_\beta}(t), \qquad 1 \leqslant \alpha, \qquad \beta \leqslant k,$$

and setting $t = 1$, this reduces to $mM = \rho m$. Hence m is a positive multiple of the left eigenvector v. Finally (4.23) and (4.24) permit the evaluation $m = \gamma^{-1}v$, since

$$1 = \lim_{x\to 1} \frac{1 - g(x)}{\operatorname{grad} g(1) \cdot (1 - x)} = \lim_{n\to\infty} \frac{\rho^n}{m \cdot [1 - f_n(0)]}.$$

The proof of Theorem 3 is therefore complete.

We proceed to Theorem 4. According to Theorem 2

$$\rho^{-n}v \cdot [1 - f_n(0)] \searrow \gamma(0) = \gamma \geqslant 0,$$

and using the notation $P = \rho^{-1}M$, $E_k = E[f_{k-1}(0)]$, $A_k = \rho^{-1}E_k$, this becomes by (4.6)

$$v(P - A_n)(P - A_{n-1}) \cdots (P - A_1)\, 1 \searrow \gamma \geqslant 0.$$

Therefore part (ii) of Lemma 1 asserts that $\gamma > 0$ if and only if $\sum A_k < \infty$, so that

$$\gamma > 0 \Rightarrow \sum_1^\infty E_n < \infty, \qquad E_n = E[f_{n-1}(0)].$$

We can be a little more explicit, however. When $\gamma > 0$, then $\rho^{-n}[1 - f_n(0)]$ converges to $\gamma u > 0$, and hence there exists a scalar θ, $0 < \theta < 1$, such that $1 - f_{n-1}(0) \geqslant \rho^n\theta 1$ for sufficiently large n, or $f_{n-1}(0) \leqslant (1 - \rho^n\theta)\, 1$. In view of the monotonicity of $E(t)$ (see (4.5)) this gives $E[f_{n-1}(0)] \geqslant E[(1 - \rho^n\theta)\, 1]$ for sufficiently large n, so that

$$\gamma > 0 \Rightarrow \sum_1^\infty E[(1 - \rho^n\theta)\, 1] < \infty \qquad \text{for some} \qquad 0 < \theta < 1. \tag{4.25}$$

If on the other hand $\gamma = 0$, then $\rho^{-n}[1 - f_n(0)] \to 0$, and by a similar reasoning as above

$$\gamma = 0 \Rightarrow \sum_1^\infty E[(1 - \rho^n\theta)\, 1] = \infty \qquad \text{for some} \qquad 0 < \theta < 1. \tag{4.26}$$

The next step is to evaluate explicitly $E(\theta 1)$ for $0 < \theta < 1$, using the definition of E in (4.3) and (4.4). For fixed $i \in X - \{0\}$, let

$$|\, i \,| = i_1 + i_2 + \cdots + i_k.$$

Then if $t = \theta 1$,

$$i_\beta \left[1 - \int_0^1 \frac{\prod_{\nu=1}^{k} (1 - \xi t_\nu)^{i_\nu}}{1 - \xi t_\beta} \, d\xi\right] = i_\beta \left[1 - \int_0^1 (1 - \xi\theta)^{|i|-1} \, d\xi\right]$$

$$= i_\beta \left[1 - \frac{1 - (1 - \theta)^{|i|}}{|i| \, \theta}\right]$$

$$= \frac{i_\beta}{|i|} \sum_{\nu=0}^{|i|-1} [1 - (1 - \theta)^\nu]$$

Consulting (4.3) and (4.4) one has

$$\sum_{\beta=1}^{k} E_{\alpha\beta}[(1 - \rho^n\theta) \, 1] = \sum_{i \in X - \{0\}} p_\alpha(i) \sum_{\nu=0}^{|i|-1} [1 - (1 - \rho^n\theta)^\nu]. \qquad (4.27)$$

We also require the following estimate.

LEMMA 4. *If θ is a constant, $0 < \theta < 1$, and $0 < \rho < 1$, one can find positive constants $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, depending only on θ and ρ, such that*

$$\alpha_1 m \log m + \alpha_2 m \leqslant \sum_{n=1}^{\infty} \sum_{p=1}^{m-1} [1 - (1 - \theta\rho^n)^p] \leqslant \alpha_3 m \log m + \alpha_4 m \qquad (4.28)$$

for all integers $m \geqslant 1$.

PROOF. We decompose the sum on n into two sums, from 1 to $[B \log p]$ and from $[B \log p] + 1$ to ∞, with a judicious choice of $B > 0$. Between 1 and $[B \log p]$ we have, since $B > 0$, $\log \rho < 0$

$$\alpha_5 \leqslant 1 - \exp\left[- \theta p^{1 - B \log \rho}\right] \leqslant 1 - (1 - \theta\rho^n)^p \leqslant 1 = \alpha_6$$

and between $[B \log p] + 1$ and ∞,

$$\alpha_7 p\rho^n \leqslant 1 - (1 - \theta\rho^n)^p \leqslant \alpha_8 p\rho^n,$$

where α_5, α_6, α_7, and α_8 are positive constants depending on A, ρ, and B. Hence the double sum S_m in (4.28) satisfies

$$\sum_{p=1}^{m-1} \left\{\alpha_5 B \log p + \alpha_7 p \sum_{n > B \log p} \rho^n\right\}$$

$$\leqslant S_m \leqslant \sum_{p=1}^{m-1} \left\{\alpha_6 B \log p + \alpha_8 p \sum_{n > B \log p} \rho^n\right\}.$$

It follows that

$$S_m \leqslant \alpha_6 Bm \log m + \alpha_9 \sum_{p=1}^{m-1} p^{1+B\log\rho},$$

and if B is chosen sufficiently large (so that $1 + B \log \rho \leqslant 0$) we have $S_m \leqslant \alpha_3 m \log m + \alpha_4 m$. The lower estimate is handled in exactly the same way.

To prove the first half of Theorem 4, suppose that $\gamma > 0$. Then, combining (4.25), (4.27), and the first half of (4.28), we have

$$\sum_{i \in X} p_\alpha(i) \,[\alpha_1 \mid i \mid \log \mid i \mid + \alpha_2 \mid i \mid] < \infty$$

for a certain pair of positive constants α_1, α_2. This implies that $E_{e_\alpha}[Z_1 \cdot 1 \log (Z_1 \cdot 1)] < \infty$ for each α, and hence also the expectations of the individual components in (3.10). To prove the other half of Theorem 4, suppose that $\gamma = 0$. Then again, (4.26) (4.27) and the second half of (4.28) yield

$$\sum_{i \in X} p_\alpha(i) \,[\alpha_3 \mid i \mid \log \mid i \mid + \alpha_4 \mid i \mid] = \infty, \qquad 1 \leqslant \alpha \leqslant k$$

and it follows that for each α, at least one of the expectations $E_{e_\alpha}[Z_{1,\beta} \log Z_{1,\beta}]$ must be infinite.

Note. It is remarkable that condition (3.10) plays a crucial role also in the theory of branching processes with k types and $\rho > 1$. It was shown by Kesten and Stigum [4] that Z_n/ρ^n then always converges with probability one to a random variable W, and that $W \neq 0$ if and only if condition (3.10) holds.

Theorem 5 is a simple but probabilistically significant corollary of Theorems 2, 3, and 4. It states that when $\gamma > 0$, then the number among the original j particles whose descendants are not extinct at time n, has approximately a Poisson distribution if $j \to \infty$ and $n \to \infty$ in the way described in Theorem 5. This explains why the limiting distribution obtained is compound Poisson, with respect to the measure χ, which governs the number of particles conditioned by the fact that no extinction occurred.

For the formal proof we write

$$1 - f_{n,\alpha}(0) = \rho^n \gamma u_\alpha [1 + o(1)], \qquad 1 \leqslant \alpha \leqslant k, \qquad n \to \infty, \tag{4.29}$$

and by Theorem 3

$$\frac{f_{n,\alpha}(s) - f_{n,\alpha}(0)}{1 - f_{n,\alpha}(0)} = g(s)\,[1 + o(1)], \qquad 1 \leqslant \alpha \leqslant k, \qquad n \to \infty, \tag{4.30}$$

for each $s \in C$. We may represent $f_{n,\alpha}$ in the form

$$f_{n,\alpha}(s) = 1 - [1 - f_{n,\alpha}(0)] \left\{1 - \frac{f_{n,\alpha}(s) - f_{n,\alpha}(0)}{1 - f_{n,\alpha}(0)}\right\},$$

which in view of (4.29) and (4.30) yields

$$f_n(s) = 1 - \rho^n \gamma[1 - g(s)]\, u + \rho^n o(1), \qquad n \to \infty, \qquad s \in C. \tag{4.31}$$

LEMMA 5. *If $x(n) \in C$ and $j(n) \in X$, so that $x(n) \to 0$ and $j(n) \cdot x(n) \to \delta \geqslant 0$ as $n \to \infty$, then*

$$\lim_{n\to\infty} [1 - x(n)]^{j(n)} = e^{-\delta}.$$

This lemma (whose proof is evident) permits the desired conclusion. Just let $j = j(n)$, $f_n(s) = 1 - x(n)$, and $\delta = A[1 - g(s)]$. Then (4.31) combined with the hypotheses of Theorem 5 shows that $x(n) \to 0$ and $j(n) \cdot x(n) \to \delta$. Hence

$$\lim\, [f_n(s)]^j = \lim \sum_{i\in X} P_j[Z_n = i]\, s^i = e^{-A[1-g(s)]}, \qquad s \in C. \tag{4.32}$$

The continuity of g on C, established in the proof of Theorem 3, together with the fact that $g(1) = 1$, permits the conclusion that $P_j[Z_n = i]$ converges to a probability measure χ_A on X, whose generating function is given by (4.32). That completes the proof of Theorem 5.

From now on we consider only processes with $\rho = 1, f(t) \neq Mt$, such that the vector of quadratic forms $q[\ ]$ in (2.8) is finite. Theorem 6 will be seen to depend on a second order Taylor expansion analogous to the first order expansion in (4.1) through (4.6). Take $\varphi(t)$, $\psi(t)$ as defined in (4.1) and let

$$\psi_\alpha(t) = \frac{\partial}{\partial t_\alpha} \psi(t), \qquad \psi_{\alpha\beta}(t) = \frac{\partial^2}{\partial t_\alpha\, \partial t_\beta} \psi(t).$$

Then

$$\begin{aligned}
\psi(t) - \psi(0) = \psi(t) - 1 &= \left[\frac{d}{d\xi}\varphi(\xi t)\right]_{\xi=0} + \int_0^1 (1-\xi)\frac{d^2}{d\xi^2}\varphi(\xi t)\, d\xi \\
&= -\sum_{\nu=1}^{k} t_\nu i_\nu + \int_0^1 (1-\xi)\frac{d}{d\xi}\sum_{\nu=1}^{k} t_\nu \psi_\nu(\xi t)\, d\xi \\
&= -\sum_{\nu=1}^{k} t_\nu i_\nu + \int_0^1 (1-\xi)\sum_{\nu=1}^{k}\sum_{\mu=1}^{k} t_\nu t_\mu \psi_{\nu\mu}(\xi t)\, d\xi \\
&= -\sum_{\nu=k}^{k} t_\nu i_\nu + \sum_{\nu=1}^{k}\sum_{\mu=1}^{k} t_\nu t_\mu [i_\nu i_\mu - \delta(\nu,\mu)\, i_\nu] \\
&\qquad \times \int_0^1 \frac{\prod_{r=1}^{k}(1-\xi t_r)^{i_r}}{(1-\xi t_\nu)(1-\xi t_\mu)}(1-\xi)\, d\xi.
\end{aligned} \tag{4.33}$$

We now replace t by $\mathbf{1} - t$, and just as in the derivation of (4.3) through (4.6) one sums with respect to the probability measure p_α. The error terms will now be represented by a quadratic form valued vector $e_s[t]$ (for each $s \in C$, $e_s[\;]$ is a vector whose components are quadratic forms). We define

$$(e_s[t])_\alpha = \sum_{i \in X} p_\alpha(i) \sum_{\nu=1}^{k} \sum_{\mu=1}^{k} t_\nu t_\mu [i_\nu i_\mu - \delta(\nu, \mu)\, i_\nu]$$

$$\times \left\{ \frac{1}{2} - \int_0^1 \frac{\prod_{r=1}^{k} [1 - \xi(1 - s_r)]^{i_r}}{[1 - \xi(1 - s_\nu)]\,[1 - \xi(1 - s_\mu)]} (1 - \xi)\, d\xi \right\}. \qquad (4.34)$$

Then (4.33), (2.8), and (4.34) give

$$\mathbf{1} - f(t) = M(\mathbf{1} - t) - q[\mathbf{1} - t] + e_t[\mathbf{1} - t], \qquad t \in C. \qquad (4.35)$$

(This representation is due to Mullikin [8], who obtained and used it in the more general setting of processes with infinitely many types.) Inpesction of (4.34) shows that

$$0 \leqslant e_s[\;] \leqslant q[\;], \qquad t \leqslant s \Rightarrow e_t \geqslant e_s, \qquad s, t \in C,$$
$$\lim_{s \to \mathbf{1}} e_s = 0. \qquad (4.36)$$

(As indicated in the introduction, $e_t \geqslant e_s$ means that the coefficients of the corresponding quadratic forms satisfy these inequalities.)

We shall actually use a simplified version of (4.35) and (4.36) obtained by taking the inner product with the left eigenvector v. If Q and E_s are quadratic forms defined by

$$Q[t] = v \cdot q[t], \qquad E_s[t] = v \cdot e_s[t], \qquad (4.37)$$

then (4.35) implies

$$v \cdot f(t) - v \cdot t = Q[\mathbf{1} - t] - E_t[\mathbf{1} - t], \qquad t \in C, \qquad (4.38)$$

and (4.36) becomes

$$0 \leqslant E_s \leqslant Q, \qquad t \leqslant s \Rightarrow E_t \geqslant E_s,$$
$$E_s \searrow 0 \quad \text{as} \quad s \nearrow \mathbf{1}. \qquad (4.39)$$

Now set (imitating the method of [9])

$$\alpha(s) = Q\left[\frac{\mathbf{1} - s}{v \cdot (\mathbf{1} - s)}\right], \qquad \epsilon(s) = E_s\left[\frac{\mathbf{1} - s}{v \cdot (\mathbf{1} - s)}\right],$$

$$\delta(s) = \frac{1}{v \cdot (\mathbf{1} - s)} + \alpha(s) - \frac{1}{v \cdot [\mathbf{1} - f(s)]}, \qquad s \in C - \{\mathbf{1}\}. \qquad (4.40)$$

Simple algebraic manipulation of (4.38) and (4.40) gives

$$\delta(s) = \frac{\epsilon(s) - \alpha(s)\, v \cdot (1-s)\,[\alpha(s) - \epsilon(s)]}{1 - v \cdot (1-s)\,[\alpha(s) - \epsilon(s)]}, \qquad s \in C - \{1\}. \tag{4.41}$$

Further we know from (4.39) and (4.40) that $\epsilon(s) \leqslant \alpha(s)$, so that

$$-\alpha^2(s)\, v \cdot (1-s) \leqslant \epsilon(s) - \alpha(s)\, v \cdot (1-s)\,[\alpha(s) - \epsilon(s)] \leqslant \delta(s) \leqslant \epsilon(s), \qquad s \in C. \tag{4.42}$$

Setting $s = f_k(t)$ in the expression for $\delta(s)$ in (4.40) and summing k from 0 to $n-1$, gives (note that $f_0(t) = t$)

$$\frac{1}{n}\sum_{k=0}^{n-1} \delta[f_k(t)] = \frac{1}{n}\left\{\frac{1}{v \cdot (1-t)} - \frac{1}{v \cdot [1 - f_n(t)]} + \sum_{k=0}^{n-1} \alpha[f_k(t)]\right\}, \qquad t \in C - \{1\}. \tag{4.43}$$

Using (4.42) we estimate

$$-v \cdot [1 - f_k(t)]\, Q\left[\frac{1 - f_k(t)}{v \cdot (1 - f_k(t))}\right] \leqslant \delta[f_k(t)] \leqslant E_{f_k(0)}\left[\frac{1 - f_k(t)}{v \cdot (1 - f_k(t))}\right], \qquad t \in C - \{1\},$$

As $k \to \infty$, $1 - f_k(t) \to 0$ by Theorem 1, and $[1 - f_k(t)]/v \cdot [1 - f_k(t)] \to u$, uniformly on $C - \{1\}$ by Theorem 1. Thus

$$Q\left[\frac{1 - f_k(t)}{v \cdot (1 - f_k(t))}\right] = \alpha[f_k(t)] \to Q[u], \qquad \text{as} \qquad k \to \infty, \tag{4.44}$$

uniformly on $C - \{1\}$, and in view of (4.39)

$$E_{f_k(0)}\left[\frac{1 - f_k(t)}{v \cdot (1 - f_k(t))}\right] \to 0, \qquad \text{as} \qquad k \to \infty, \tag{4.45}$$

uniformly on $C - \{1\}$. This implies that $\delta[f_k(t)] \to 0$ uniformly, and so does the Césaro average on the left in (4.43). By (4.44) and (4.45) applied to (4.43) we therefore have

$$\lim_{n\to\infty} \frac{1}{n}\left\{\frac{1}{v \cdot (1-t)} - \frac{1}{v \cdot [1 - f_n(t)]}\right\} = -Q[u],$$

uniformly on $C - \{1\}$, which is the principal assertion (3.13) of Theorem 6.

For the proof of (3.14), observe that (3.13) with $t = 0$ states that $nv \cdot [1 - f_n(0)] \to Q[u]^{-1}$, as $n \to \infty$. (Note that this limit is finite since

$Q[u] > 0$. This follows from the assumption that $f(t) \neq Mt$ which implies that at least one component of $q[u]$ in (2.8) must be positive. Otherwise the total number of particles would be a constant.) Applying (3.3) of Theorem 1, we see that

$$\lim_{n\to\infty} nP_{e_\alpha}[Z_n \neq 0] = \lim_{n\to\infty} n[1 - f_{n,\alpha}(0)] = \frac{u_\alpha}{Q[u]}, \qquad 1 \leqslant \alpha \leqslant k. \tag{4.46}$$

This is not quite (3.14), but it was shown in (4.15) and Lemma 2 how to conclude from (4.46) that

$$\lim_{n\to\infty} nP_i[Z_n \neq 0] = \frac{i \cdot u}{Q[u]}, \qquad i \in X.$$

Finally we sketch the proof of (3.15). (See [8] for the details.) As in the proof of Theorem 4, one easily shows that the limit in (3.15) is independent of i, provided that it exists. If $i = e_\alpha$,

$$t(n) = (e^{-(\lambda_1/n)}, e^{-(\lambda_2/n)}, \ldots, e^{-(\lambda_k/n)}), \qquad \lambda \in R_k.$$

Then the Laplace-Stieltje's transform of the probability measure of $n^{-1}Z_n$, under the condition that $Z_0 = i = e_\alpha$, and $Z_n \neq 0$, is given by

$$\frac{x \cdot [f_n(t(n)) - f_n(0)]}{x \cdot [1 - f_n(0)]} = \varphi_n(x, \lambda) \tag{4.47}$$

with $x = e_\alpha$. Next a simple computation, based on (3.13), (3.14), and (3.3) shows that for each $x > 0$, $x \neq 0$,

$$\lim_{n\to\infty} \varphi_n(x, \lambda) = \varphi(\lambda) = \frac{1}{1 + (v \cdot \lambda) Q[u]}. \tag{4.48}$$

But $\varphi(\lambda)$ is the Laplace Stieltje's transform of the limiting probability measure on the right in (3.15). The continuity theorem therefore completes the proof of (3.15).

References

1. T. E. Harris. "The Theory of Branching Processes." Springer-Verlag, Berlin 1963.
2. C. R. Heathcote, E. Seneta, and D. Vere-Jones. A refinement of two theorems in the theory of branching processes; to appear in *Teor. Verojatnost. Primenen.*
3. A. Joffe. On the Galton-Watson branching process with mean less than one. *Ann. Math. Stat.* **38** (1967), 264-266.

4. H. Kesten and B. P. Stigum. A limit theorem for multi-dimensional Galton-Watson processes. *Ann. Math. Stat.* **37** (1966), 1211-1223.
5. H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Stat.* **37** (1966), 1463-1481.
6. H. Kesten and B. P. Stigum. Limit theorems for decomposable multidimensional Galton-Watson processes. *J. Math. Anal. Appl.* **17** (1967), 309-338.
7. M. Jiřina. The asymptotic behavior of branching stochastic processes. *Czech. Math. J.* **7** (1957), 130-153 (Russian).
8. T. W. Mullikin. Limiting distributions for critical multitype branching processes with discrete time. *Trans. Am. Math. Soc.* **106** (1963), 469-494.
9. H. Kesten, P. Ney, and F. Spitzer. The Galton-Watson process with mean one and finite variance. To appear in *Teor. Verojatnost. Primenen.*
10. S. Karlin. "A First Course in Stochastic Processes." Academic Press, New York, 1966.