Søren Asmussen Niels Keiding

Martingale Central Limit Theorems and Asymptotic Estimation Theory for Multitype Branching Processes

Preprint
March
1977
1

Institute of Mathematical Statistics
University of Copenhagen

Søren Asmussen and Niels Keiding

MARTINGALE CENTRAL LIMIT THEOREMS AND ASYMPTOTIC ESTIMATION THEORY FOR MULTITYPE BRANCHING PROCESSES

Preprint 1977 No. 1

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

March 1977

Abstract.

This paper considers the problem of estimating the growth rate ρ of a p-type Galton-Watson process $\{Z_n\}$. To this end, a general approach of possible independent interest to central limit theorems for discrete-time branching processes is developed. The idea is to adapt martingale central limit theory to martingale difference triangular arrays indexed by the set of all individuals ever alive. Iterated logarithm laws are derived by similar methods. Asymptotic distribution results and the a.s. asymptotic behaviour are derived for a maximum likelihood estimator based upon all parent-offspring combinations in a given number N of generations, and for the estimator $(|Z_1|+ \dots +|Z_N|)/(|Z_0|+ \dots +|Z_{N-1}|)$ which depends on the total generation sizes $|Z_n|$ only.

1. Introduction.

Consider a p-type Galton-Watson process $\{Z_n\} = \{(Z_n(1),\ldots,Z_n(p))\}$, see Athreya and Ney (1972, Chapter V) for background material. Let P^i, E^i etc. refer to the case $Z_0(j) = \delta_{ij}$ and assume that $m_{ij} = E^i Z_1(j)$ and $\Sigma^i_{jk} = \mathrm{Cov}^i(Z_1(j), Z_1(k))$ are all finite. We consider the positively regular and supercritical case throughout. Let v and u be the left and right eigenvectors associated with the principal eigenvalue $\rho > 1$ of the offspring mean matrix $M = (m_{ij})$, normalized such that

$$v\cdot u = \sum_{i=1}^{p} v(i)u(i) = 1, \quad |v| = v\cdot \underset{\sim}{1} = 1$$

where $\underset{\sim}{1} = (1,\ldots,1)$. The basic a.s. limit results state that $W = \lim W_n = \lim \rho^{-n} Z_n\cdot u$ exists and is finite, that $\{W > 0\} = \{Z_n \neq 0 \text{ for all } n\}$ and that $\lim \rho^{-n} Z_n = Wv$.

In the present paper we study the problem of estimating ρ and other parameters associated with M, and in conjunction herewith, we develop a general approach to central limit theorems for discrete-time branching processes. As is explained in detail in Section 2, the idea is here to adapt martingale central limit theory, see e.g. Lévy (1954), Brown (1971), Dvoretzky (1972), McLeish (1974), to martingale difference triangular arrays indexed by the set of all individuals ever alive. It seems that most central limit theorems of the literature come out as special cases of our results, in some cases even with a much simpler proof. We illustrate this by an example in Section 3 on the central limit theorem for a linear functional $Z_n\cdot a$, where a is some vector with $v\cdot a = 0$, see Kesten and Stigum (1966), Athreya (1969a,b,1971). Section 3 also provides some further background material for the sequel, other aspects of the limiting behaviour of linear functionals and closely related to this, introduction of the Jordan canonical form of M.

In Section 4 we briefly review the known results concerning the maximum likelihood estimator $\hat{m}$ of the offspring mean m in a single-type Galton-Watson process and we take the opportunity to point out an iterated logarithm result for the a.s. behaviour of $\hat{m}-m$. Sections 5 and 6 then consider similar estimation problems for the multitype case, where no systematic exposition seems to exist (see, however, a very recent report by Quine and Durham (1977) on the subcritical immigration-case). In Section 5 we study maximum likelihood estimation of M and a fortiori the principal eigenvalue ρ, based upon knowledge of all parent-offspring combinations in the N first generations. Though seldom of direct practical applicability, such results will provide baselines against which to compare simpler estimators based upon more realistic data, like the estimator

$$\tilde{\rho} = \frac{|Z_1|+ \dots +|Z_N|}{|Z_0|+ \dots +|Z_{N-1}|}$$

based upon the total population sizes $|Z_n|$, $n = 0,\dots,N$, which is the object of study of Section 6. As pointed out by Becker (1976), $\tilde{\rho}$ is clearly strongly consistent on the set $\{W > 0\}$ of nonextinction. The asymptotic behaviour of $\tilde{\rho}-\rho$ is seen to depend qualitatively on the relative sizes of ρ and λ^2, where λ is the absolute value of a certain other eigenvalue of M. Thus the form of our theorems is analogous to the limit results for linear functionals $Z_n \cdot a$, see Kesten and Stigum (1966), Athreya (1969a,b,1971), Asmussen (1977). In fact, when $\lambda^2 \geqq \rho$, we shall even be able to exhibit a linear functional $Z_{N-1} \cdot b^*$ with the same limiting behaviour as $\tilde{\rho}-\rho$, while if $\lambda^2 < \rho$, the central limit theorems for $\tilde{\rho}-\rho$ and $Z_n \cdot a$ are contained in a more general result in Section 2. The final Section 7 disproves a conjecture by Athreya and Keiding (1975) concerning the asymptotic properties of an occurrence/exposure rate estimating the Malthusian parameter of a Bellman-Harris process.

In the rest of the paper, it is assumed that $P(W=0) = 0$. This is merely a notational convenience in order to avoid making trivial exceptions on the set of extinction. The possibility of extinction may present a problem in concrete statistical applications, but the discussion of this does not seem different for the multitype rather than the single-type Galton-Watson process.

2. Central limit theorems.

Let $\mathcal{H}_{k,N}$; $N = 0,1,2,\ldots$, $k = 0,\ldots,k(N)$ (with $k(N)$ constant or $\mathcal{H}_{1,N}$-measurable) be σ-fields such that $\mathcal{H}_{1,N} \subseteq \cdots \subseteq \mathcal{H}_{k(N),N}$. $\{X_{k,N}\}$; $N = 0,1,2,\ldots$, $k = 1,\ldots,k(N)$ is a martingale difference triangular array w.r.t. $\{\mathcal{H}_{k,N}\}$ if $X_{k,N}$ is $\mathcal{H}_{k,N}$-measurable and $E(X_{k,N} \mid \mathcal{H}_{k-1,N}) = 0$. Central limit theorems for such arrays have been developed by a number of authors, see e.g. Lévy (1954), Brown (1971), Dvoretzky (1972), McLeish (1974). Our initial Theorem 2.1 is an adaptation of this theory to branching processes and is valid (with obvious modifications) for many other discrete-time branching processes than the multitype Galton-Watson process. In contrast, Theorem 2.2 is a specialization of Theorem 2.1 with some concrete applications in mind.

Let $\tau_n = |Z_0| + \ldots + |Z_n|$ and suppose in the present section that the n^{th} generation is represented as $\{k \in \mathbb{N}: \tau_{n-1} < k \leq \tau_n\}$. Here $n = n(k)$ always denotes the generation of k and we let U_k be the offspring vector produced by k and

$$\mathcal{G}_k = \sigma(U_\ell; \ell \leq k) \quad , \quad \mathcal{F}_N = \sigma(U_k; n \leq N) = \mathcal{G}_{\tau_N} .$$

Theorem 2.1. Let $\{X_{k,N}\}$ $(N = 0,1,2,\ldots, k = 0,1,2,\ldots)$ be a martingale difference triangular array w.r.t. $\{\mathcal{G}_k\}$ and define

$$s^2_{k,N} = \sum_{\ell=1}^{k} E(X^2_{\ell,N} \mid \mathcal{G}_{\ell-1}) \quad , \quad s^2_N = \sum_{\ell=1}^{\infty} E(X^2_{\ell,N} \mid \mathcal{G}_{\ell-1}) = \sup_k s^2_{k,N} .$$

Suppose

(2.1) For any N, $S_N = \sum_{k=0}^{\infty} X_{k,N}$ converges in L^2, i.e. $E\, s^2_N < \infty$,

(2.2) $s^2_N \xrightarrow{P} W$,

(2.3) $\sum_{k=0}^{\infty} E(X^2_{k,N}\, I(|X_{k,N}| > \varepsilon) \mid \mathcal{G}_{k-1}) \xrightarrow{P} 0 \quad \forall \varepsilon > 0.$

Then the limiting distribution of $W^{-1/2}\, S_N$ is standard normal.

In the proof, we need

Lemma 2.1. Let $\{X_{k,N}\}$, $N = 0,1,2,\ldots$, $k = 1,\ldots,k(N)$, be a martingale difference triangular array w.r.t. $\{\mathcal{H}_{k,N}\}$ and define

$$s^2_{k,N} = \sum_{\ell=1}^{k} E(X^2_{\ell,N} \mid \mathcal{H}_{\ell-1,N}),\quad s^2_N = s^2_{k(N),N},\quad S_{k,N} = \sum_{\ell=1}^{k} X_{\ell,N},\quad S_N = S_{k(N),N}.$$

Then $s^2_N \overset{P}{\to} 0$ implies that $S_N \overset{P}{\to} 0$ as well.

Proof. Define for $\varepsilon > 0$

$$\nu(N) = \inf\{k:\ 1 \leqq k < k(N),\ s^2_{k+1,N} > \varepsilon\},$$

$\nu(N) = k(N)$ if no such k exists. Then

(2.4) $\qquad P(\nu(N) = k(N)) \to 1,$

(2.5) $\qquad ES^2_{k(N)\wedge\nu(N),N} = Es^2_{k(N)\wedge\nu(N),N} \leqq \varepsilon$

(the first assertion of (2.5) requires an argument as given in Neveu (1972, pg. 148)). It follows from (2.4),(2.5) and Chebycheff's inequality applied to $S_{k(N)\wedge\nu(N),N}$ that

$$\limsup_{N\to\infty} P(|S_N| > \delta) \leqq \varepsilon/\delta^2$$

and since ε is arbitrary, the proof is complete.

Remark. The lemma and its proof are easily adapted to the case $k(N) = \infty$, since it is well-known, cf. Neveu (1972, pg. 148), that

$$\sum_{k=1}^{\infty} X_{k,N} \quad \text{exists on} \quad \{\sum_{k=1}^{\infty} E(X^2_{k,N} \mid \mathcal{H}_{k-1,N}) < \infty\}\ .$$

Proof of Theorem 2.1. We approximate $W^{-1/2}\, S_N$ by

$$T_N = \sum_{k=\underline{k}(N)}^{\bar{k}(N)} \tilde{X}_{k,N}$$

where $\tilde{X}_{k,N} = W_{R(N)}^{-1} X_{k,N}$ and $R(N)$, $\underline{k}(N) = \tau_{R(N)}+1$, $\bar{k}(N) = \underline{k}(N)+\Delta(N)$ are to be determined later such that $R(N)$ and $\Delta(N)$ are non-random,

$$(2.6) \qquad R(N) \uparrow \infty,$$

$$(2.7) \qquad s^2_{\underline{k}(N)-1,N} \overset{P}{\to} 0,$$

$$(2.8) \qquad s^2_N - s^2_{\bar{k}(N),N} \overset{P}{\to} 0.$$

Note that $W_{R(N)} = \rho^{-n} Z_{R(N)} \cdot a$ is $\mathcal{G}_{\underline{k}(N)}$-measurable and that thus the $\tilde{X}_{k,N}$ are again martingale differences. By Lemma 2.1 and the remark, (2.7) and (2.8) ensure that the limiting distribution of $W^{-1/2} S_N$ is that of T_N, and to show that this in fact is standard normal it suffices by Theorem 2.2 of Dvoretzky (1972) that

$$(2.9) \qquad \sum_{k=\underline{k}(N)}^{\bar{k}(N)} E(\tilde{X}^2_{k,N} \mid \mathcal{G}_{k-1}) = W_{R(N)}^{-1} (s^2_{\bar{k}(N),N} - s^2_{\underline{k}(N)-1,N}) \overset{P}{\to} 1,$$

$$(2.10) \qquad \sum_{k=\underline{k}(N)}^{\bar{k}(N)} E(\tilde{X}^2_{k,N}\, I(|\tilde{X}_{k,N}| > \varepsilon) \mid \mathcal{G}_{k-1}) \overset{P}{\to} 0 \qquad \forall \varepsilon > 0.$$

But (2.10) is an easy consequence of (2.3) and (2.9) follows from (2.2),(2.6)-(2.8). Thus we only have to specify $R(N)$, $\Delta(N)$ such that (2.6)-(2.8) hold. Recalling that $\underline{k}(N)-1 = \tau_{R(N)}$, define for fixed $R = 0,1,2,\ldots$ and $\varepsilon > 0$

$$A_{k,N} = \sum_{k=0}^{\tau_R} E(X^2_{k,N}\, I(|X_{k,N}| \leq \varepsilon) | \mathcal{G}_{k-1}), \quad B_{k,N} = \sum_{k=0}^{\tau_R} E(X^2_{k,N}\, I(|X_{k,N}| > \varepsilon) | \mathcal{G}_{k-1}).$$

Let $N \to \infty$ with R fixed. Then by (2.3), $B_{R,N} \overset{P}{\to} 0$ and since ε is arbitrary and $A_{R,N} \leq \varepsilon^2 \tau_R$,

$$s^2_{\tau_R,N} = A_{R,N} + B_{R,N} \overset{P}{\to} 0$$

as well. Thus if $R(N) \uparrow \infty$ sufficiently slowly, (2.7) will hold. By (2.1), we can choose $\Delta(N)$ such that

$$\sum_{k=\Delta(N)}^{\infty} EX_{k,N}^2 \to 0$$

and the proof is completed by observing that (2.8) follows from

$$s_N^2 - s_{k(N),N}^2 \leqq \sum_{k=\Delta(N)}^{\infty} E(X_{k,N}^2 \mid G_{k-1}) \overset{P}{\to} 0.$$

<u>Theorem 2.2.</u> <u>Let</u> $a_1, a_2, \ldots$ <u>be vectors and</u> $\gamma_1, \gamma_2, \ldots$ <u>constants,</u> $0 < \gamma \leqq \gamma_n < \infty$, <u>and define</u>

$$\sigma_n^2 = \gamma_n^{-2}\, v\cdot\mathrm{Var}^{\cdot} Z_1 \cdot a_n = \gamma_n^{-2} \sum_{i=1}^{p} v(i)\; a_n \Sigma^i a_n, \text{ 1)}$$

$$\beta_N^2 = \sum_{n=1}^{N} \rho^{-n} \gamma_n^2 \sigma_n^2, \quad \alpha_N^2 = \sum_{n=1}^{N} \rho^{-n} \gamma_n^2.$$

<u>Suppose</u> $\{a_n/\gamma_n\}$ [and thus $\{\sigma_n^2\}$] <u>is relatively compact and that</u>

(2.11) $$\gamma_N^2 = o(\rho^N \alpha_N^2)$$

(2.12) $$\liminf_{N\to\infty} \beta_N/\alpha_N > 0$$

<u>Then the limiting distribution of</u>

$$(W\rho^N \beta_N^2)^{-1/2} \sum_{n=0}^{N-1} \{Z_{n+1} - Z_n M\}\cdot a_{N-n}$$

<u>is standard normal</u>.

<u>Proof</u>. We let $X_{k,N} = 0$, $n \geqq N$,

$$X_{k,N} = \rho^{-N/2} \beta_N^{-1} \{U_k - E(U_k \mid F_n)\}\cdot a_{N-n} \qquad n < N.$$

Then

$$S_N = \rho^{-N/2} \beta_N^{-1} \sum_{n=0}^{N-1} \{Z_{n+1} - Z_n M\}\cdot a_{N-n},$$

$$s_N^2 = \rho^{-N} \beta_N^{-2} \sum_{n=0}^{N-1} Z_n \cdot\mathrm{Var}^{\cdot} Z_1 \cdot a_{N-n}.$$

1) With a slight abuse of notation, vectors to the left of M, Σ^i etc. are always assumed to be written as row vectors and those to the right as column vectors.

When $n \geqq n_0$, we have $\rho^{-n} Z_n \leqq W^*_{n_0} v$ with $W^*_{n_0} \to W$ as $n_0 \to \infty$ and it follows that for each n_0

$$\limsup_{N\to\infty} s_N^2 = \limsup_{N\to\infty} \rho^{-N} \beta_N^{-2} \sum_{n=n_0}^{N-1} Z_n \cdot \mathrm{Var}^{\cdot} Z_1 \cdot a_{N-n} \leqq$$

$$W^*_{n_0} \limsup_{N\to\infty} \beta_N^{-2} \sum_{n=n_0}^{N-1} \rho^{n-N} \gamma_{N-n}^2 \sigma_{N-n}^2 = W^*_{n_0} .$$

Here one has to use (2.11),(2.12) for the first and last equality. As $n_0 \to \infty$, $\limsup s_N^2 \leqq W$ follows. A similar argument gives $\liminf s_N^2 \geqq W$ and (2.2). Also (2.1) is obvious, and we only have to verify (2.3). Suppose without loss of generality that all components of a_n/γ_n are bounded by one for all n. Then

$$Y_{k,N} = |\{U_k - E(U_k|\mathcal{F}_n)\} \cdot a_{N-n}/\gamma_{N-n}| \leqq |U_k| + E(|U_k||\mathcal{F}_n)$$

and if F is the distribution function defined by

$$1 - F(y) = \max_{i=1,\dots,p} P^i(|Z_1| + E^i|Z_1| > y),$$

it follows that $P(Y_{k,N} > y|\mathcal{F}_n) \leqq 1 - F(y)$ and that therefore $E(g(Y_{k,N})|\mathcal{F}_n) \leqq \int_0^\infty g(y)dF(y)$ for any non-decreasing g. Letting $g(y) = y^2 I(y > \varepsilon\rho^{N/2}\beta_N/\gamma_n)$, we see that the quantity to be inspected in (2.3) can be bounded by

$$(\rho^N \beta_N^2)^{-1} \sum_{n=0}^{N-1} \gamma_n^2 |Z_{N-n}| \int_0^\infty g(y)dF(y) \leqq$$

$$\sup_{n=0,1,2,\dots} \rho^{-n}|Z_n| \int_0^\infty g^2 I(y > \varepsilon\rho^{N/2}\beta_N/\gamma) dF(y)\ \beta_N^{-2} \sum_{n=0}^{N-1} \rho^{-n}\gamma_n^2$$

which is finite and tends to zero because of (2.12) and because the second moment of F obviously exists.

3. The Jordan canonical form of M and the limiting behaviour of linear functionals $Z_N \cdot a$.

In the present section, we first briefly review some well-known facts on the Jordan canonical form of M. The reader can skip this at the cost of specializing Section 6 to $p=2$, where explicit expressions for the relevant parameters are given in Example 6.1. As is well known, writing M on the Jordan canonical form amounts to operating with (possibly complex) vectors $u_{\nu,j}$ $(\nu=1,\dots,\nu_0;\ j=1,\dots,j_0(\nu))$ and eigenvalues ρ_ν, such that

$$(3.1)\quad Mu_{\nu,1}=\rho_\nu u_{\nu,1},\quad Mu_{\nu,j}=u_{\nu,j-1}+\rho_\nu u_{\nu,j}\qquad j=2,\dots,j_0(\nu)$$

and that each (complex) vector a can be uniquely expanded in the $u_{\nu,j}$,

$$a=\sum_{\nu=1}^{\nu_0}\sum_{j=1}^{j_0(\nu)}(v_{\nu,j}\cdot a)u_{\nu,j}.$$

Obviously, for some ν, $\rho_\nu = \rho$, $j_0(\nu) = 1$, $u_{\nu,1} = u$, $v_{\nu,1} = v$. Whenever ρ_ν is complex, then for some μ ρ_μ must be the complex conjugate of ρ_ν, $\rho_\mu = \bar{\rho}_\nu$, and in that case we can assume $u_{\mu,j} = \bar{u}_{\nu,j}$ $j = 1,\dots,j_0(\nu) = j_0(\mu)$. Then also $v_{\mu,j} = \bar{v}_{\nu,j}$. Define

$$\lambda = \lambda(a) = \max\{|\rho_\nu|:\ v_{\nu,j}\cdot a \neq 0 \text{ for some } j\}$$

$$\gamma = \gamma(a) = \max\{j:\ v_{\nu,j}\cdot a \neq 0 \text{ for some } \nu \text{ with } |\rho_\nu| = \lambda(a)\}$$

Then, since

$$M^n u_{\nu,j} = \sum_{i=1}^{j}\rho_\nu^{n-j+i}\binom{n}{j-1}u_{\nu,i},$$

$M^n a$ is of magnitude $\lambda^n n^{\gamma-1}$ in the sense that $M^n a = 0$ for $\lambda = 0$ and $n \geq \gamma$ and that otherwise $\{M^n a/\lambda^n n^{\gamma-1}\}$ is relatively compact and bounded away from zero.

Let I denote the $p \times p$ identity matrix. In Section 6 we shall need

Lemma 3.1. *Suppose a is real with $\lambda(a) > 1$. Then we can write $a = (M-I)b + c$, with b,c real and $\lambda(b) = \lambda(a)$, $\gamma(b) = \gamma(a)$, $\lambda(c) \leq 1$.*

Proof. Define

$$c = \sum_{\nu,j:|\rho_\nu|\leq 1} (v_{\nu,j} \cdot a) u_{\nu,j}, \quad d = \sum_{\nu,j:|\rho_\nu|>1} (v_{\nu,j} \cdot a) u_{\nu,j} .$$

Then c,d are real, $a = d + c$ and $\lambda(c) \leq 1$. Let V be the manifold spanned by the $u_{\nu,j}$ with $|\rho_\nu| > 1$. The eigenvectors of the restriction of $M - I$ to V are the $\rho_\nu - 1$ with $|\rho_\nu| > 1$. It follows that $M - I$ is one-one on V and since $d \in V$, $d = (M-I)b$ for some (necessarily real) $b \in V$. That $\lambda(b) = \lambda(a)$, $\gamma(b) = \gamma(a)$ is obvious by using (3.1) to compute $v_{\nu,j} \cdot a$ in terms of the $v_{\nu,i} \cdot b$, $v_{\nu,i} \cdot c$.

Let b be real. Then $\rho^{-n} Z_n \cdot b \to Wv \cdot b$ and if $v \cdot b \neq 0$, this settles the question on the limiting behaviour of $Z_n \cdot b$. If $v \cdot b = 0$ (i.e. $\lambda < \rho$), we quote the following results from Kesten and Stigum (1966), Athreya (1969a,b,1971), Asmussen (1977), which form a trichotomy depending on whether $\lambda^2 > \rho$, $\lambda^2 = \rho$ or $\lambda^2 < \rho$. If $\lambda^2 > \rho$, one can exhibit a sequence $\{H_N\}$ with $\lim\sup|H_N| < \infty$ and, except for special structures of the Σ^i, $\lim\inf|H_N| \neq 0$, such that

$$\lim_{N\to\infty} \lambda^{-N} N^{-(\gamma-1)} Z_N \cdot b - H_N = 0 \text{ a.s.}$$

If $\lambda^2 \leq \rho$ it is easily seen that $\sigma^2 = \sigma^2(b)$ given by

$$(3.2)\quad \sigma^2 = \lim_{N\to\infty} \frac{v \cdot \mathrm{Var}^* Z_N \cdot b}{\rho^N N^{2\gamma-1}} \text{ if } \lambda^2 = \rho, \quad \sigma^2 = \lim_{N\to\infty} \frac{v \cdot \mathrm{Var}^* Z_N \cdot b}{\rho^N} \text{ if } \lambda^2 < \rho$$

exists and that, again with the exception of special Σ^i where $\sigma^2 = 0$, $0 < \sigma^2 < \infty$. Furthermore the limiting distribution of $(\sigma^2 W \rho^N N^{2\gamma-1})^{-1/2} Z_N \cdot b$ if $\lambda^2 = \rho$, $(\sigma^2 W \rho^N)^{-1/2} Z_N \cdot b$ if $\lambda^2 < \rho$, is standard normal and in the two cases, respectively, a.s.

$$\lim_{N\to\infty} {\sup \atop \inf} \frac{Z_N \cdot b}{(2\sigma^2 W\rho^N N^{2\gamma-1} \log\log N)^{1/2}} = \pm 1, \quad \lim_{N\to\infty} {\sup \atop \inf} \frac{Z_N \cdot b}{(2\sigma^2 W \rho^N \log N)^{1/2}} = \pm 1.$$

To illustrate our methods, we shall give a proof based upon Theorem 2.2 of the central limit theorem for the case $\lambda^2 = \rho$, where the proofs of the literature are particular cumbersome. Write $a_n = M^{n-1}b$,

(3.3) $$Z_N = Z_0 M^N + \sum_{n=0}^{N-1} \{Z_{n+1} - Z_n M\} M^{N-n-1},$$

(3.4) $$Z_N \cdot b = Z_0 M^N b + \sum_{n=0}^{N-1} \{Z_{n+1} - Z_n M\} \cdot a_{N-n}.$$

Let $\gamma_n = \rho^{n/2} n^{\gamma-1}$. Then in the notation of Theorem 2.2,

$$\beta_N^2 = \rho^{-N} v \cdot \mathrm{Var} \cdot Z_N \cdot a, \quad \frac{\alpha_N^2}{N^{2\gamma-1}} \to \frac{1}{2\gamma-1}, \quad \frac{\beta_N^2}{N^{2\gamma-1}} \to \sigma^2,$$

with $\sigma^2 = \sigma^2(b)$ as in (3.2). The conditions of Theorem 2.2 are obvious ((2.12) follows from $\sigma^2 > 0$) and since $M^N a = o(\rho^{N/2}\beta_N)$, it follows that indeed $(W\rho^N \beta_N^2)^{-1/2} Z_N \cdot b$ or equivalently $(\sigma^2 W\rho^N N^{2\gamma-1})^{-1/2} Z_N \cdot b$ is asymptotically standard normal.

The central limit theorem when $\lambda^2 < \rho$ comes out with equal ease from Theorem 2.2. The laws of the iterated logarithm for $Z_n \cdot b$ were discussed by Asmussen (1977). Whereas the proof is rather involved when $\lambda^2 = \rho$, the result for $\lambda^2 < \rho$ is an application of the following lemma, which will be used a number of times in the sequel (see also Heyde and Leslie (1971)).

Lemma 3.2. Let $Y = Y(Z_0, \ldots, Z_r)$ be some functional depending on the r first generations only and with $E^i Y = 0$, $i = 1, \ldots, p$, $0 < \sigma^2 = v \cdot \mathrm{Var} \cdot Y < \infty$. Then, if $Y_{k,N}$ is the functional corresponding to Y evaluated in the line of descent initiated by the k^{th} individual alive at time N,

$$\lim_{N\to\infty} {\sup \atop \inf} \sum_{k=1}^{Z_N} Y_{k,N} / (2\sigma^2 W\rho^N \log N)^{1/2} = \pm 1.$$

4. Estimation of the offspring mean in a simple Galton-Watson process.

Let $Z_0,\dots,Z_N$ be the first N+1 generation sizes of a single-type Galton-Watson process and consider the problem of estimating the offspring mean m in the "nonparametric" statistical model given by $Z_0 = z_0$ and all offspring distributions with $0 < m < \infty$. Harris (1948) derived the estimator

$$\hat{m} = (Z_1 + \dots + Z_N)/(Z_0 + \dots + Z_{N-1})$$

as the maximum likelihood estimator of m based on recording the individual offspring size for each individual in each generation, not only $Z_0,\dots,Z_N$. The first complete proof of the fact that $\hat{m}$ is also the maximum likelihood estimator of m based on $Z_0,\dots,Z_N$ only was provided by Feigin (1976). Keiding (1975) remarked on the derivation of $\hat{m}$ in more restricted models specified by parametric classes of offspring distributions.

Asymptotic properties of $\hat{m}$ as $z_0 \to \infty$ are easy exercises in standard i.i.d. asymptotic theory for maximum likelihood estimators. This is due to the branching property which implies that a process $Z_0, Z_1, \dots$ with $Z_0=z_0$ may be interpreted as the sum of z_0 i.i.d. processes with the same offspring distribution but only one ancestor. The most complete and careful treatment was given by Yanev (1975) and despite its unquestionable practical importance we shall not consider this theory further in the present paper, but assume $Z_0 = 1$.

To study asymptotic properties of $\hat{m}$ as $N \to \infty$, that is, based on one long realization of the process, it is necessary to single out the supercritical case $m > 1$. If $m \leq 1$, $Z_N \to 0$ and there is no hope for consistency or asymptotic distribution results.

Theorem 4.1. Assume $1 < m < \infty$, $P(Z_N \to \infty) = 1$. Then

(a) as $N \to \infty$, $\hat{m} \to m$ a.s.

If furthermore $0 < \sigma^2 < \infty$, where σ^2 is the offspring variance, then

(b) the limiting distribution as $N \to \infty$ of $[W(1+m+\ldots+m^{N-1})]^{1/2}(\hat{m}-m)$ is normal $(0,\sigma^2)$.

(c) $$\lim_{N\to\infty} \begin{matrix}\sup\\ \inf\end{matrix} \left[\frac{W(1+\ldots+m^{N-1})}{2\sigma^2 \log N}\right]^{1/2} (\hat{m}-m) = \pm 1.$$

Remark. The usefulness of statement (b) will most likely be to obtain asymptotic confidence intervals. For this purpose, rephrase (b) to assert that $\sigma^{-1}(Z_0 + \ldots + Z_{N-1})^{1/2}(\hat{m}-m)$ is asymptotically standard normal. If nothing is known about σ^2, use may be made of the estimators studied by Heyde (1975) and Dion (1975). Similar remarks apply in the following.

Proof. The strong consistency (a) was noted in its final form by Heyde (1970) and the asymptotic normality (b) was shown by Dion (1974), cf. also Jagers (1973,1975). Alternatively, (b) is a special case of Theorem 2.2 with $p = a_r = \gamma_r = 1$. The iterated logarithm law (c) may be obtained by a rather direct application of Lemma 3.2. Choosing first $Y = Z_1 - m$ yields $\Sigma_1^{Z_n} Y_{k,n} = Z_{n+1} - m Z_n$ and it follows that

(4.1) $$\sup_n |Z_{n+1} - m Z_n| / (m^n \log n)^{1/2} < \infty \text{ a.s.},$$

and therefore, letting $U_N = \Sigma_{n=0}^{N-1}\{Z_{n+1} - m Z_n\}$, that

(4.2) $$U_N \leqq \sum_{n=0}^{N-1} \frac{|Z_{n+1} - m Z_n|}{(m^n \log n)^{1/2}} (m^n \log n)^{1/2} = O([m^N \log N]^{1/2}).$$

Next, let $Y = U_r$,

$$c_n = \mathrm{Var}(U_n | Z_0 = 1) = \sigma^2(1 + \ldots + m^{n-1}), \quad A_n = \frac{U_n}{(2c_n W \log n)^{1/2}} .$$

Since $U_{n+r} = U_n + \Sigma_{k=1}^{Z_n} Y_{k,n}$, it follows that

$$\limsup_{n\to\infty} A_{n+r} \leq \limsup_{n\to\infty} \frac{U_n}{(2c_{n+r} W \log (n+r))^{1/2}} + \limsup_{n\to\infty} \frac{\Sigma_1^{Z_n} Y_{k,n}}{(2c_{n+r} W \log (n+r))^{1/2}} .$$

By (4.2), the first term is $O(m^{-r/2})$ as $r \to \infty$, and by Lemma 3.2, the second is

$$\limsup_{n\to\infty} \frac{\Sigma_1^{Z_n} Y_{k,n}}{(2c_r W m^n \log n)^{1/2}} \left(\frac{c_r m^n}{c_{n+r}}\right) = 1 \cdot (1 - m^{-r})^{1/2} .$$

Similarly

$$\limsup_{n\to\infty} A_{n+r} \geq -O(m^{-r/2}) + (1 - m^{-r})^{1/2}$$

so that $\limsup A_n = 1$, which is equivalent to the lim sup-part of (c). The lim inf-part is similar.

5. Estimation of the asymptotic growth rate ρ of multitype branching processes: Maximum likelihood estimation based on all parent-offspring combinations.

Let the offspring distributions of a multitype Galton-Watson process be given by

$$p_i(z) = P^i(Z_1 = z), \quad z = (z(1),\dots,z(p)) \in \{0,1,\dots\}^p.$$

Assume that the offspring vector $U^n_{k,i}$ produced by the k^{th} individual of type i alive at time n is observable for $n = 0,1,\dots,N-1$. In the statistical model specified by all possible offspring distributions it is then obvious by similar considerations as for the single-type Galton-Watson process that the maximum likelihood estimator is given by

$$\hat{p}_i(z) = \left(\sum_{n=0}^{N-1} \sum_{k=1}^{Z_n(i)} I(U^n_{k,i} = z) \right) \Big/ \sum_{n=0}^{N-1} Z_n(i)$$

whenever the denominator is positive. In this case, since

$$m_{ij} = \sum_{z(1)=1}^{\infty} \cdots \sum_{z(p)=1}^{\infty} z(j) p_i(z)$$

it follows that the maximum likelihood estimator of m_{ij} based upon the $U^n_{k,i}$ is

$$\hat{m}_{ij} = \sum_{z(1)=1}^{\infty} \cdots \sum_{z(p)=1}^{\infty} z(j)\hat{p}_i(z) = \sum_{n=0}^{N-1} Z^i_{n+1}(j) \Big/ \sum_{n=0}^{N-1} Z_n(i),$$

where $Z^i_{n+1}(j) = \Sigma_{k=1}^{Z_n} U^n_{k,i}(j)$ is the number of individuals of type j in the n+1'st generation whose parents were of type i. We conjecture that $\hat{M} = (\hat{m}_{ij})$ is also the maximum likelihood estimator of M based upon the $Z^i_n(j)$ only.

Theorem 5.1. (a) *As* $N \to \infty$, $\hat{M} \to M$ a.s.

(b) *The limiting distribution of the matrix*

$$\left([W(1+\rho+\dots+\rho^{N-1})v(i)]^{1/2} (\hat{m}_{ij} - m_{ij}) \right)$$

is that of $(Y_i(j))$, *where* $Y_1,\dots,Y_p$ *are independent and the distribution of* Y_i *is p-dimensional normal* $(0,\Sigma^i)$.

(c) *Let* $\alpha_1,\dots,\alpha_p$ *be vectors such that*

$$V = \sum_{i=1}^{p} v(i)^{-1} \sum_{j,k=1}^{p} \alpha_i(j)\alpha_i(k)\, \Sigma^i_{jk} > 0 .$$

Then a.s.

$$\lim_{N\to\infty} \begin{matrix}\sup\\ \inf\end{matrix} \left[\frac{W(1+\dots+\rho^{N-1})}{2V\log N}\right]^{1/2} \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i(j)(\hat{m}_{ij} - m_{ij}) = \pm 1.$$

Remark. When $\Sigma^i_{jj} > 0$, we get an iterated logarithm law for $\hat{m}_{ij} - m_{ij}$ by taking $\alpha_i(j) = \delta_{ij}$. If $\Sigma^i_{jj} = 0$, then obviously $\hat{m}_{ij} = m_{ij}$ a.s.

Proof. Similar methods as for the single-type case studied in Section 4 apply. We outline here a martingale proof of part (b). By the Cramér-Wold device we must show that

$$(5.1) \qquad [W(1+\dots+\rho^{N-1})]^{1/2} \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i(j)(\hat{m}_{ij} - m_{ij})$$

tends in probability to zero for $V = 0$ and is asymptotically normal $(0,V)$ for $V > 0$, for any given set $\alpha_1,\dots,\alpha_p$ of p-vectors and V defined as in part (c). Clearly, (5.1) behaves like $[W(1+\dots+\rho^{N-1})]^{1/2} S_N$, where

$$S_N = \sum_{n=0}^{N-1} \sum_{i=1}^{p} v(i)^{-1} \sum_{j=1}^{p} \alpha_i(j)\{Z^i_{n+1}(j) - m_{ij} Z_n(i)\}$$

$$= \sum_{n=0}^{N-1} \sum_{i=1}^{p} \{Z^i_{n+1}\cdot a^i - E(Z^i_{n+1}\cdot a^i | \mathcal{F}_n)\}, \quad a^i(j) = v(i)^{-1}\alpha_i(j).$$

The case $V=0$ comes out easily by noting that

$$\sum_{n=0}^{N-1} \mathrm{Var}\left(\sum_{i=1}^{p} Z^i_{n+1}\cdot a^i \middle| \mathcal{F}_n\right) = \sum_{n=0}^{N-1} \sum_{i=1}^{p} Z_n(i)\; a^i \Sigma^i a^i = \sum_{n=0}^{N-1} \sum_{i=1}^{p} v(i)^{-2}\; Z_n(i)\; \alpha_i \Sigma^i \alpha_i$$

$$\mathrm{Var}^j S_N = E^j \sum_{n=0}^{N-1} \sum_{i=1}^{p} v(i)^{-2}\; Z_n(i)\; \alpha_i \Sigma^i \alpha_i \simeq u(j)(1+\ldots+\rho^{N-1})V,$$

while for $V>0$ a straightforward modification of Theorem 2.2 and its proof applies.

The maximum likelihood estimator of a functional of the mean matrix M, given the detailed information here considered, is given as the same functional of $\hat{M}$. Asymptotic properties will be immediate by standard transformaiton techniques. We illustrate the procedure by considering the maximum-likelihood estimator $\hat{\rho}$ of the growth rate ρ. Of course, $\hat{\rho} = \rho(\hat{M})$, where $\rho(A)$ is the principal eigenvalue of A. Note that $\rho(A)$ is well-defined whenever A is positively regular and that thus $\rho(\hat{M})$ is so for $\hat{M}$ sufficiently close to M. For small p, $\rho(A)$ can be expressed explicitly as function of the a_{ij}. E.g. for $p=2$, $\rho(A) = (a_{11}+a_{22}+D^{\frac{1}{2}})/2$, where $D=(a_{11}-a_{22})^2+4a_{12}a_{21}$ is the discriminant of the characteristic polynomial of A. For larger p, $\rho(A)$ is implicitly determined as one of the solutions of the equation $|A-\rho I|=0$. In particular, ρ is a smooth function of the a_{ij}.

Corollary 5.1. (a) As $N\to\infty$, $\hat{\rho}\to\rho$ a.s.

(b) The asymptotic distribution of $[W(1+\ldots+\rho^{N-1})]^{\frac{1}{2}}(\hat{\rho}-\rho)$ is normal $(0,V)$, with

$$V = \sum_{i=1}^{p} v(i)^{-1} \sum_{j,k=1}^{p} \frac{\partial\rho}{\partial m_{ij}} \frac{\partial\rho}{\partial m_{ik}} \Sigma^i_{jk} = v\cdot \mathrm{Var}^{\boldsymbol{\cdot}} Z_1 \cdot u \, .$$

(c) If $V>0$, then a.s.

$$\lim_{N\to\infty} \begin{matrix}\sup\\ \inf\end{matrix} \left[\frac{W(1+\ldots+\rho^{N-1})}{2V\;\log N}\right]^{1/2} (\hat{\rho}-\rho) = \pm 1.$$

Proof. Part (a) is obvious and part (b), with the first expression for V, is standard from the theorem on differentiable transformation of asymptotically normal variables. In a similar manner, part (c) follows from the expansion

$$\hat{\rho} - \rho = \sum_{i,j=1}^{p} \alpha_i(j)(\hat{m}_{ij} - m_{ij}) + O(\max_{i,j}(\hat{m}_{ij} - m_{ij})^2)$$

with $\alpha_i(j) = \partial\rho/\partial m_{ij}$, cf. part (c) of Theorem 4.1 and the following remark, and it only remains to verify the last expression for V, i.e. that $\partial\rho/\partial m_{k\ell} = v(k)u(\ell)$. To this end, differentiating the equation $Mu = \rho u$ with respect to $m_{k\ell}$ yields

$$\underset{\sim}{u}_{k\ell} + M\Delta = \frac{\partial\rho}{\partial m_{k\ell}} u + \rho\Delta$$

where $\underset{\sim}{u}_{k\ell}(i) = 0$ when $i \neq k$, $\underset{\sim}{u}_{k\ell}(k) = u(\ell)$ and $\Delta(i) = \partial u_i/\partial m_{k\ell}$. The conclusion follows by taking the inner product with v and using $vM = \rho v$, $v \cdot u = 1$.

Remark. From Theorem 2.2, it follows easily that the asymptotic distribution of $\hat{\rho}$ is the same as that of the estimator $\Sigma_1^N Z_n \cdot u / \Sigma_0^{N-1} Z_n \cdot u$, which seems natural when u is known and the observations are $(Z_n(1), \ldots, Z_n(p);\ 0 \leq n \leq N)$.

6. Estimators of ρ based on the total generation sizes.

Becker (1976) suggested the estimator

$$\tilde{\rho} = \frac{|Z_1| + \ldots + |Z_N|}{|Z_0| + \ldots + |Z_{N-1}|}$$

and pointed out that $\tilde{\rho}$ is strongly consistent. When $p = 1$, $\tilde{\rho}$ clearly reduces to the maximum likelihood estimator and the limiting behaviour of $\tilde{\rho} - \rho$ is given by the results of Section 4. In the present section we study the corresponding problem for $p > 1$, where the results are much more complex. We formulate the results for general p and remark in Example 6.1 on some of the explicit expressions for $p = 2$.

Let $a = \underset{\sim}{1} - u$. Our analysis is based on the identity

$$(6.1)\quad \tilde{\rho} - \rho = \sum_{n=0}^{N-1} \{Z_{n+1}\cdot\underset{\sim}{1} - \rho Z_n\cdot\underset{\sim}{1}\}/(|Z_0|+\ldots+|Z_{N-1}|) = (S_N + T_N)/(|Z_0|+\ldots+|Z_{N-1}|)$$

where

$$S_N = \sum_{n=0}^{N-1} \{Z_{n+1} - Z_n M\}\cdot\underset{\sim}{1},\quad T_N = \sum_{n=0}^{N-1} Z_n\cdot d,\quad d = (M - \rho I)\underset{\sim}{1} = (M - \rho I)a.$$

Here S_N is a martingale and it follows by Theorem 2.2 that S_N, normalized by $\rho^{-N/2}$, converges in distribution. Furthermore, it is easily checked that $\lambda(d) = \lambda(a) = \lambda$ in the notation of Section 3. Therefore the limit results for linear functionals suggest that S_N and T_N are of the same order of magnitude when $\lambda^2 < \rho$, while T_N dominates S_N if $\lambda^2 \geqq \rho$. The first of these assertions is made precise in

<u>Theorem 6.1.</u> <u>Suppose</u> $\lambda^2 < \rho$.

(a) <u>As</u> $N \to \infty$, $[W(1 + \ldots + \rho^{N-1})]^{\frac{1}{2}}(\tilde{\rho} - \rho)$ <u>is asymptotically normal</u> $(0,V)$, <u>with</u> V <u>specified in (6.3) below</u>.

(b) $$\lim_{N\to\infty} \begin{matrix}\sup\\ \inf\end{matrix} \left[\frac{W(1+\ldots+\rho^{N-1})}{2\,V\log N}\right]^{1/2} (\tilde{\rho}-\rho) = \pm 1 \text{ a.s.}$$

Proof. Inserting (3.3) yields

$$T_N = \sum_{n=0}^{N-1} Z_0 \cdot M^n d + \sum_{n=0}^{N-2} \{Z_{n+1} - Z_n M\} \cdot \sum_{k=0}^{N-n-1} M^k d. \tag{6.2}$$

When $\lambda^2 < \rho$, the first term is $o(\rho^{N/2})$ and it is easily checked that Theorem 2.2 with $\gamma_n = \mu^n$ where $\lambda < \mu < \rho^{\frac{1}{2}}$, $a_1 = \underset{\sim}{1}$, $a_n = \underset{\sim}{1} + \Sigma_0^{n-1} M^k d$, $n > 1$, applies to show that $(W\rho^N \beta_N^2)^{-\frac{1}{2}}(S_N + T_N)$ tends to the standard normal distribution. Here

$$\beta_N^2 = \sum_{n=1}^{N} \rho^{-n}\, v \cdot \mathrm{Var}^{\bullet} Z_1 \cdot a_n$$

and (a) now follows with

$$V = \lim_{N\to\infty} \frac{\rho^N \beta_N^2}{1+\ldots+\rho^{N-1}} = (\rho - 1) \sum_{n=1}^{\infty} \rho^{-n}\, v \cdot \mathrm{Var}^{\bullet} Z_1 \cdot a_n. \tag{6.3}$$

Part (b) is proved exactly as in Section 4, with

$$U_N = \sum_{n=0}^{N-1} \{Z_{n+1} - m\, Z_n\} \cdot a_{N-n}.$$

When $\lambda^2 \geqq \rho$, we rewrite (6.2) somewhat. Choose b,c as in Lemma 3.1. Then

$$\sum_{k=0}^{N-n-1} M^k d = (M-\rho I) \sum_{k=0}^{N-n-1} \{M^k (M-I) b + M^k c\}$$

$$= (M-\rho I) M^{N-n} b + (M-\rho I)\{\sum_{k=0}^{N-n-1} M^k c - b\} = M^{N-n} b^* + e_{N-n}$$

with $b^* = (M-\rho I)b$. It follows that

$$T_N = \sum_{n=0}^{N-1} Z_0 \cdot M^n d + Z_{N-1} \cdot b^* - Z_0 \cdot M^{N-1} b^* + \sum_{n=0}^{N-2} \{Z_{n+1} - Z_n M\} \cdot e_{N-n}. \tag{6.4}$$

The last term of (6.4), normalized by $\rho^{-N/2}$, is easily seen to converge in distribution, and can also be proved to be $O([\rho^N \log N]^{1/2})$ a.s.

It follows that the dominant term of $S_N + T_N$ is $Z_{N-1} \cdot b^*$ (since, as in the proof of Lemma 3.1, $\lambda(b^*) = \lambda(b) = \lambda(a)$), and from the results cited in Section 3, we have at once

Theorem 6.2. Suppose $\lambda^2 \geq \rho$. Then there exist vectors b^*, c^* with $(M - \rho I)a = (M - I)b^* + c^*$, $\lambda(b^*) = \lambda$, $\gamma(b^*) = \gamma(a) = \gamma$, $\lambda(c^*) \leq 1$. If $\lambda^2 = \rho$, define $V = (1 - \rho^{-1})\sigma^2(b^*)$ [cf. Section 3]. Then if $V > 0$,

(a) The limiting distribution of

$$N^{\frac{1}{2}-\gamma}[W(1 + \ldots + \rho^{N-1})]^{\frac{1}{2}}(\tilde{\rho} - \rho) \quad \text{or} \quad \frac{Z_{N-1} \cdot b^*}{[WN^{2\gamma-1}(1 + \ldots + \rho^{N-1})]^{\frac{1}{2}}}$$

is normal $(0, V)$.

(b)
$$\lim_{N\to\infty} \begin{matrix}\sup\\ \inf\end{matrix} \left[\frac{W(1 + \ldots + \rho^{N-1})}{2\,V N^{2\gamma-1} \log\log N}\right]^{1/2} (\tilde{\rho} - \rho) = \pm 1 \text{ a.s.}$$

Theorem 6.3. Suppose $\lambda^2 > \rho$. Then there exist random variables $\{H_N^*\}$ such that a.s.

$$\lim_{N\to\infty} \left\{\frac{W(1 + \ldots + \rho^{N-1})}{\lambda^{N-1}(N-1)^{\gamma-1}}(\tilde{\rho} - \rho) - H_{N-1}^*\right\} = \lim_{N\to\infty}\left\{\frac{Z_N \cdot b^*}{\lambda^N N^{\gamma-1}} - H_N^*\right\} = 0.$$

Furthermore, $\lim\sup|H_N^*| < \infty$ and, except for special Σ^i, $\lim\inf|H_N^*| \neq 0$ a.s.

Example 6.1. Suppose $p = 2$. Then M has a real eigenvalue $\rho_1 \neq \rho$ and necessarily $\lambda = |\rho_1|$, $\gamma = 1$, $Ma = \rho_1 a$, $d = (\rho_1 - \rho)a$. If $\lambda^2 < \rho$, then $a_1 = \underset{\sim}{1} = u + a$,

$$a_n = \underset{\sim}{1} + (\rho_1 - \rho)\sum_{k=0}^{n-1}\rho_1^k a = \begin{cases} u + [n + 1 - n\rho]a & \rho_1 = 1 \\ u + \left[\dfrac{(\rho_1 - \rho)(\rho_1^n - 1)}{\rho_1 - 1} + 1\right]a & \rho_1 \neq 1 \end{cases}$$

If $\lambda^2 \geq \rho$, then $b = (\rho_1 - 1)^{-1}a$, $c = 0$, $b^* = (\rho_1 - \rho)(\rho_1 - 1)^{-1}a$. If $\lambda^2 = \rho$, it follows after some calculations that

$$V = (1-\rho^{-1})\left(\frac{\rho_1-\rho}{\rho_1-1}\right)^2 \sigma^2(a) = (1-\rho^{-1})\left(\frac{\rho_1-\rho}{\rho_1-1}\right)^2 \rho_1^{-2} v\cdot\mathrm{Var}^{\bullet} Z_1\cdot a$$

If $\lambda^2 > \rho$, it is well-known that $\{W_n^*\} = \{\rho_1^{-n} Z_n \cdot a\}$ is a L^2-bounded martingale. The variance of the limit W^* can be verified to be

$$\mathrm{Var}^i W^* = u(i)(\rho_1^2 - \rho)^{-1} v\cdot\mathrm{Var}^{\bullet} Z_1\cdot a$$

and Theorem 6.3 reduces to

$$\lim_{N\to\infty} \frac{W(1+\ldots+\rho^{N-1})}{\rho_1^{N-1}}(\tilde{\rho}-\rho) = \lim_{N\to\infty} \frac{Z_N\cdot b^*}{\rho_1^N} = \frac{\rho_1-\rho}{\rho_1-1} W^* .$$

Example 6.2. If the P^i-distribution of $|Z_1|$ is independent of i, then $|Z_0|, |Z_1|, \ldots$ is a single-type Galton-Watson process (this will be so, for example, if the process is a single-type Galton-Watson process with the particles moving according to a Markov chain with p states). Then $u = \underset{\sim}{1}$, $\rho = \Sigma_j m_{ij}$ for all i and the offspring variance $\sigma^2 = \mathrm{Var}^i|Z_1| = \mathrm{Var}^i Z_1\cdot u$ is independent of i. Since we here have $a = 0$, we get $d = 0$ and therefore $\lambda^2 = 0 < \rho$ so that Theorem 6.1 applies. Since $a_n = \underset{\sim}{1}$ for all n, (6.3) reduces to $V = \sigma^2$.

It is interesting to notice that in this case $\tilde{\rho}$ is the maximum likelihood estimator of ρ based on observation of $|Z_0|, \ldots, |Z_N|$, cf. Section 4, so that the asymptotic properties of $\tilde{\rho}$ are in this case already obvious from Theorem 4.1.

Finally, we notice regarding the estimator $\hat{\rho}$ based on the more detailed information and studied in Section 5 that the asymptotic distribution of $\{W(1+\ldots+\rho^{N-1})\}^{\frac{1}{2}}(\hat{\rho}-\rho)$ is normal $(0,\sigma^2)$. In this situation, $\tilde{\rho}$ thus has full asymptotic efficiency.

Example 6.3. Let $p=2$ and assume $m_{11}=m_{12}=\mu$, $m_{21}=m_{22}=\beta$, all variances equal to σ^2, all covariances 0. Then $\rho=\mu+\beta$, $\rho_1=0$, $v_1=v_2=\frac{1}{2}$, $u_1=2\mu/(\mu+\beta)$, $u_2=2\beta/(\mu+\beta)$, $a_0=\underset{\sim}{1}$, $a_n=\underset{\sim}{1}-\rho a=(1+\mu-\beta, 1-\mu+\beta)'$, $n>1$. Hence

$$v\cdot\mathrm{Var}(Z_1\cdot a_0) = 2\sigma^2,$$

$$v\cdot\mathrm{Var}(Z_1\cdot a_j) = 2\sigma^2[1+(\mu-\beta)^2], \quad n = 1,\ldots,N-1$$

and applying Theorem 6.1, $[W(1+\rho+\ldots+\rho^{N-1})]^{1/2}(\tilde{\rho}-\rho)$ is asymptotically normal $(0,V)$, with

$$V=\lim_{n\to\infty} 2\sigma^2 \frac{(\mu+\beta)^{n-1}+(1+[\mu-\beta]^2)\sum_{k=1}^{n-1}(\mu+\beta)^{n-1-k}}{1+\ldots+(\mu+\beta)^{n-1}} = 2\sigma^2\left[1+\frac{(\mu-\beta)^2}{\mu+\beta}\right].$$

This may be compared with the asymptotic variance from the estimator $\hat{\rho}$ based on more detailed information and studied in Section 5. From Corollary 5.1, it follows that $[W(1+\rho+\ldots+\rho^{N-1})]^{1/2}(\hat{\rho}-\rho)$ is asymptotically normal with zero mean and variance $(u_1^2+u_2^2)\sigma^2 = 4\sigma^2(\mu^2+\beta^2)/(\mu+\beta)^2$. Notice that, since $\mu+\beta=\rho>1$,

$$2\sigma^2\left[1+\frac{(\mu-\beta)^2}{\mu+\beta}\right] > 2\sigma^2\left[\frac{(\mu+\beta)^2+(\mu-\beta)^2}{(\mu+\beta)^2}\right] = 4\sigma^2\,\frac{\mu^2+\beta^2}{(\mu+\beta)^2}\,.$$

Remark. Results analogous to Theorems 6.1,6.2,6.3 follow by slight modifications of the arguments for the estimator $|Z_N|/|Z_{N-1}|$, which for $p=1$ is the so-called Lotka-Nagaev estimator.

7. A remark on the asymptotic distribution of an occurrence/exposure rate in Bellman-Harris processes.

In their study of estimation theory for continuous-time branching processes Athreya and Keiding (1975) proposed the occurrence/exposure rate $\tilde{\alpha} = (X_T - X_0)/\int_0^T X_t dt$ as estimator of the Malthusian parameter α of a Bellman-Harris process $\{X_t\}$. Assume $P\{X_t = 0\} = 0$ and $E(X_t \log X_t) < \infty$ for all t. Then $\tilde{\alpha}$ is strongly consistent as $T \to \infty$. If the life-length distribution is exponential, then $\{X_t\}$ is a Markov branching process and $\tilde{\alpha}$ is the maximum likelihood estimator. Based upon an analogy with this case, and on a hope on fast convergence of the relative age distribution to the stable age distribution, Athreya and Keiding conjectured that as $T \to \infty$,

$$(7.1) \qquad \left(\int_0^T X_t dt\right)^{1/2} (\tilde{\alpha} - \alpha)$$

is in the general Bellman-Harris case asymptotically normal. We disprove below this conjecture by studying the particular case of the p-phase birth process introduced by Kendall (1948) where the life-length distribution is gamma with form parameter p. If p is sufficiently large (greater than 57) the above mentioned convergence is too slow. This result indicates that a similar trichotomy as discussed above applies for the convergence rate of functionals of the age distribution of Bellman-Harris processes. A difficulty in a detailed study of these problems is that little is known in general about the spectral properties of the relevant mean operator.

The p-phase birth process may be described by first defining a continuous time Markov branching process (Z_t) with p types $j = 0,\ldots,p-1$ and infinitesimal generator A of the mean semigroup $\{M_t\} = \{e^{At}\}$ given by

$$A = \beta \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & & 0 & 0 \\ \vdots & & & \vdots & & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \\ 2 & 0 & 0 & 0 & & 0 & -1 \end{pmatrix}$$

Then $|Z_t|$ is a p-phase birth process, that is a Bellman-Harris process with offspring distribution degenerate at 2 and gamma (p,β^{-1}) life-length distribution. Estimation of the intensity β or, equivalently, the Malthusian parameter $\alpha = a_0\beta$ (a_0 to be computed below), was considered by Hoel and Crump (1974) and further discussed by Athreya and Keiding (1975).

We first recapitulate some facts on $\{M_t\}$. They are in part contained in Kendall (1948), but we give a slightly different derivation. Assume without loss of generality that $\beta = 1$. It is easily seen that the characteristic equation $|A - aI| = 0$ for A has the form $(1+a)^p = 2$, which has p different complex solutions

$$a_\nu = 2^{1/p} e^{i2\pi\nu/p} - 1, \quad \nu = 0,1,\dots,p-1.$$

If v_ν, u_ν are the corresponding left and right eigenvectors normalized such that $v_\nu \cdot u_\mu = \delta_{\nu\mu}$, $|v| = v\cdot u = 1$ where $v = v_0$, $u = u_0$, then for $j = 0,\dots,p-1$

$$v_\nu(j) = \frac{(1+a_\nu)^{-j}}{2(1-2^{-1/p})} = \frac{2^{-j/p} e^{-i2\pi\nu j/p}}{2(1-2^{-1/p})}$$

$$u_\nu(j) = \frac{2}{p}(1-2^{-1/p})(1+a_\nu)^j = \frac{2}{p}(1-2^{-1/p})2^{j/p} e^{i2\pi\nu j/p},$$

(7.2) $$I = \sum_{\nu=0}^{p-1} u_\nu \otimes v_\nu \ , \qquad M_t = \sum_{\nu=0}^{p-1} e^{a_\nu t} u_\nu \otimes v_\nu,$$

where $u_\nu \otimes v_\nu = (u_\nu(i)v_\nu(j))$. Note that $a_{p-\nu} = \bar{a}_\nu, v_{p-\nu} = \bar{v}_\nu$, $u_{p-\nu} = \bar{u}_\nu$.

In particular,

$$|Z_t| = Z_t \cdot \underset{\sim}{1} = \sum_{\nu=0}^{p-1} (v_\nu \cdot \underset{\sim}{1}) Z_t \cdot u_\nu$$

$$= Z_t \cdot u + \sum_{\nu=1}^{[(p-1)/2]} 2\mathrm{Re}[(v_\nu \cdot \underset{\sim}{1}) Z_t \cdot u_\nu] + (v_{p/2} \cdot \underset{\sim}{1}) Z_t \cdot u_{p/2},$$

the last term being interpreted as zero for p uneven. By standard facts on Markov branching processes, $W = \lim e^{-a_0 t} Z_t \cdot u$ exists and $P(0 < W < \infty) = 1$, (more generally, $\lim e^{-a_0 t} Z_t = Wv$), while for $\nu \geq 1$ the behaviour of the ν^{th} term depends on the relative size of a_0 and $2\mathrm{Re}\, a_\nu$. It is easily seen that for large p

$$a_0 = 2^{1/p} - 1 \cong \frac{\log 2}{p} \quad , \quad \mathrm{Re}\, a_1 = 2^{1/p} \cos 2\pi/p - 1 \cong a_0 - \frac{2\pi^2}{p^2}$$

and it is clearly possible to choose p such that $\gamma = 2\,\mathrm{Re}\, a_1 > a_0$. It is then well-known that $\{W^*(t)\} = \{\omega e^{-a_1 t} (v_1 \cdot \underset{\sim}{1}) Z_t \cdot u_1\}$, $\omega = a_1^{-1} - a_0^{-1}$, is an L^2-bounded complex martingale and letting $W^* = W_1^* + iW_2^*$ denote its limit, we have with reference to Asmussen (1977, Section 4, slightly extended), that

$$|Z_t| = e^{a_0 t} W + 2\mathrm{Re}[\omega^{-1} e^{a_1 t} W^*] + o(e^{\gamma t}) \text{ a.s.}$$

and thus that

$$\int_0^T |Z_t| dt - a_0^{-1} |Z_T| = 2\mathrm{Re}[e^{a_1 T} W^*] + o(e^{\gamma T})$$

$$= 2e^{\gamma T} \{\cos 2\pi T/p\, W_1^* - \sin 2\pi T/p\, W_2^*\} + o(e^{\gamma T}).$$

To see that this is of order of magnitude $e^{\gamma T}$ (which clearly contradicts the convergence in distribution of (7.1)), we must check that W_1^* and W_2^* are not both degenerate at zero. To see this, it suffices to note that $E^0 W^* = W^*(0) = \omega (v_1 \cdot \underset{\sim}{1}) u_1(0) \neq 0$.

Kendall remarked that $\mathrm{Re}\, a_1 > 0$ when $p > 28$. Correspondingly, $2\mathrm{Re}\, a_1 > a_0$ when $p > 57$. Otherwise, $2\,\mathrm{Re}\, a_1 < a_0$ and, as was to be expected from

Section 6, indeed (7.1) is asymptotically normal. We sketch the argument to show how the methods of Section 2 can be used also in central limit problems in continuous time which does not reduce in a trivial way to the consideration of discrete skeletons (as is the case e.g. for linear functionals $Z_N \cdot a$). The continuity properties of (7.1) ensure that it suffices to show the central limit theorem along any discrete skeleton $\{N\delta\}, \delta > 0$. Let $Y = \int_0^\delta |Z_t| dt$, $\kappa = (e^{a_0 \delta} - 1)/a_0$. Then from (7.2),

$$E^i Y = \int_0^\delta M_t \underset{\sim}{1}(i) dt = \kappa u(i) + b(i)$$

where $M_t b = O(e^{a_1 t})$. The study of (7.1) (with $T = N\delta$) is equivalent to considering

$$a_0^{-1} |Z_{N\delta}| - \int_0^{N\delta} |Z_t| dt = a_0^{-1} |Z_{N\delta}| - \sum_{n=0}^{N-1} \{\kappa Z_{n\delta} \cdot u + Z_{n\delta} \cdot b + \sum_{k=1}^{Z_{n\delta}} (Y^{n\delta,k} - E(Y^{n\delta,k} | F_{n\delta})\}$$

where $Y^{n\delta,k}$ is the functional corresponding to Y evaluated in the line of descent initiated by the k^{th} individual alive at time $n\delta$. Writing $Z_{N\delta}$, $Z_{n\delta}$ in a similar manner as in (3.3) leads as in Section 6 to an expansion in martingale increments and the methods of Section 2 apply. We shall not give the details or compute the variance.

Acknowledgement. Torgny Lindvall read the manuscript and offered helpful suggestions.

References.

Asmussen, S. (1977) Almost sure behavior of linear functionals of supercritical branching processes. Trans.Amer.Math.Soc. (to appear)

Athreya, K.B. (1969a) Limit theorems for multitype continuous time Markov branching processes.I. The case of an eigenvector linear functional. Z.Wahrscheinlichkeitstheorie verw.Geb. 12,320-332.

Athreya, K.B. (1969b) Limit theorems for multitype continuous time Markov branching processes.II. The case of an arbitrary linear functional. Z.Wahrscheinlichkeitstheorie verw.Geb. 13,204-214.

Athreya, K.B. (1971) Some refinements in the theory of supercritical multitype Markov branching processes. Z.Wahrscheinlichkeitstheorie verw.Geb. 20,47-57.

Athreya, K.B. and Keiding, N.(1975) Estimation theory for continuous-time branching processes. Sankhyā Ser. A (to appear).

Athreya, K.B. and Ney, P.E.(1972) Branching processes. Springer, Berlin.

Becker, N.(1976) Estimation for a Galton-Watson process with application to epidemics. Manuscript, Cornell University.

Brown, B.M.(1971) Martingale central limit theorems. Ann.Math.Statist. 42,59-66.

Dion, J.-P.(1974) Estimation of the mean and the initial probabilities of a branching process. J.Appl.Prob. 11,687-694.

Dion, J.-P.(1975) Estimation of the variance of a branching process. Ann.Statist. 3,1183-1187.

Dvoretzky, A.(1972) Asymptotic normality for sums of dependent random variables. Proc.Sixth Berk.Symp.Math.Statist.Prob. 2, 513-535.

Feigin, P.(1976) A note on maximum likelihood estimation for simple branching processes. Manuscript, Technion-Israel Institute of Technology, Haifa.

Harris, T.E.(1948) Branching processes. Ann.Math.Statist. 19,474-494.

Heyde, C.C.(1970) Extension of a result of Seneta for the supercritical Galton-Watson process. Ann.Math.Statist. 41,739-742.

Heyde, C.C.(1974) On estimating the variance of the offspring distribution in a simple branching process. Adv.Appl.Prob. 6, 421-433.

Heyde, C.C. and Leslie, J.R.(1971) Improved classical limit analogues for Galton-Watson processes with or without immigration. Bull.Austral.Math.Soc. 5,145-155.

Hoel, D.G. and Crump, K.S.(1974) Estimating the generation-time distribution of an age-dependent branching process. Biometrics 30,125-135.

Jagers, P.(1973) A limit thorem for sums of random numbers of i.i.d. random variables. In: Jagers, P. and Råde, L.(ed.) Mathematics and Statistics. Essays in honour of Harald Bergström. Göteborg, pp.33-39.

Jagers, P.(1975) Branching processes with biological applications. Wiley, New York.

Keiding, N.(1975) Estimation theory for branching processes. Bull. Int.Statist.Inst. 46 (4),12-19.

Kendall, D.G.(1948) On the role of variable generation time in the development of a stochastic birth process. Biometrika 35, 316-330.

Kesten, H. and Stigum, B.P.(1966) Additional limit theorems for indecomposable multidimensional Galton-Watson processes. Ann.Math. Statist. 37,1463-1481.

Lévy, P.(1954) Théorie de l'addition des variables aléatoires. Deuxième édition. Gauthier-Villars, Paris.

McLeish, D.L.(1974) Dependent central limit theorems and invariance principles. Ann.Probab. 2,620-628.

Neveu, J.(1972) Martingales à temps discret . Masson, Paris.

Quine, M.P. and Durham, P.(1977) Estimation for multitype branching processes. Manuscript, University of Sydney.

Yanev, N.M.(1975) On the statistics of branching processes. Theor. Prob.Appl. 20,612-622.

PREPRINTS 1976

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Becker, Niels: Estimation for an Epidemic Model.

No. 2 Kaplan, Norman: A Generalization of a Result of Erdös and Renyi and a Related Problem.

No. 3 Lauritzen, Steffen & Remmer, Ole: An Investigation of the Asymptotic Properties of Geodetic Least Squares Estimates in Certain Types of Network.

No. 4 Asmussen, Søren: Almost Sure Behavior of Linear Functionals of Supercritical Branching Processes.

No. 5 Hald, Anders & Møller, Uffe: On the SPRT of the Mean of a Poisson Process.

No. 6 Lauritzen, Steffen: Projective Statistical Fields.

No. 7 Keiding, Niels: Inference and Tests for Fit in the Birth-Death-Immigration Process.

No. 8 Møller, Uffe: OC and ASN og the SPRT for the Poisson Process.

No. 9 Keiding, Niels: Population Growth and Branching Processes in Random Environments.

No. 10 Andersen, Per K., Andersen, Søren & Lauritzen, Steffen: The Average Noise from a Poisson Stream of Vehicles.

No. 11 Hald, Anders & Møller, Uffe: Multiple Sampling Plans of Given Strength for the Poisson and Binomial Distributions.

No. 12 Johansen, Søren: Two Notes on Conditioning and Ancillarity.

No. 13 Asmussen, Søren: Some Martingale Methods in the Limit Theory of Supercritical Branching Processes.

No. 14 Johansen, Søren: Homomorphisms and general exponential families.

No. 15 Asmussen, Søren & Hering, Heinrich: Some Modified Branching Diffusion Models.

No. 16 Hering, Heinrich: Minimal Moment Conditions in the Limet theory for General Markov Branching Processes.

PREPRINTS 1977

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Asmussen, Søren & Keiding, Niels: Martingale Central Limit Theorems and Asymptotic Estimation Theory for Multitype Branching Processes.