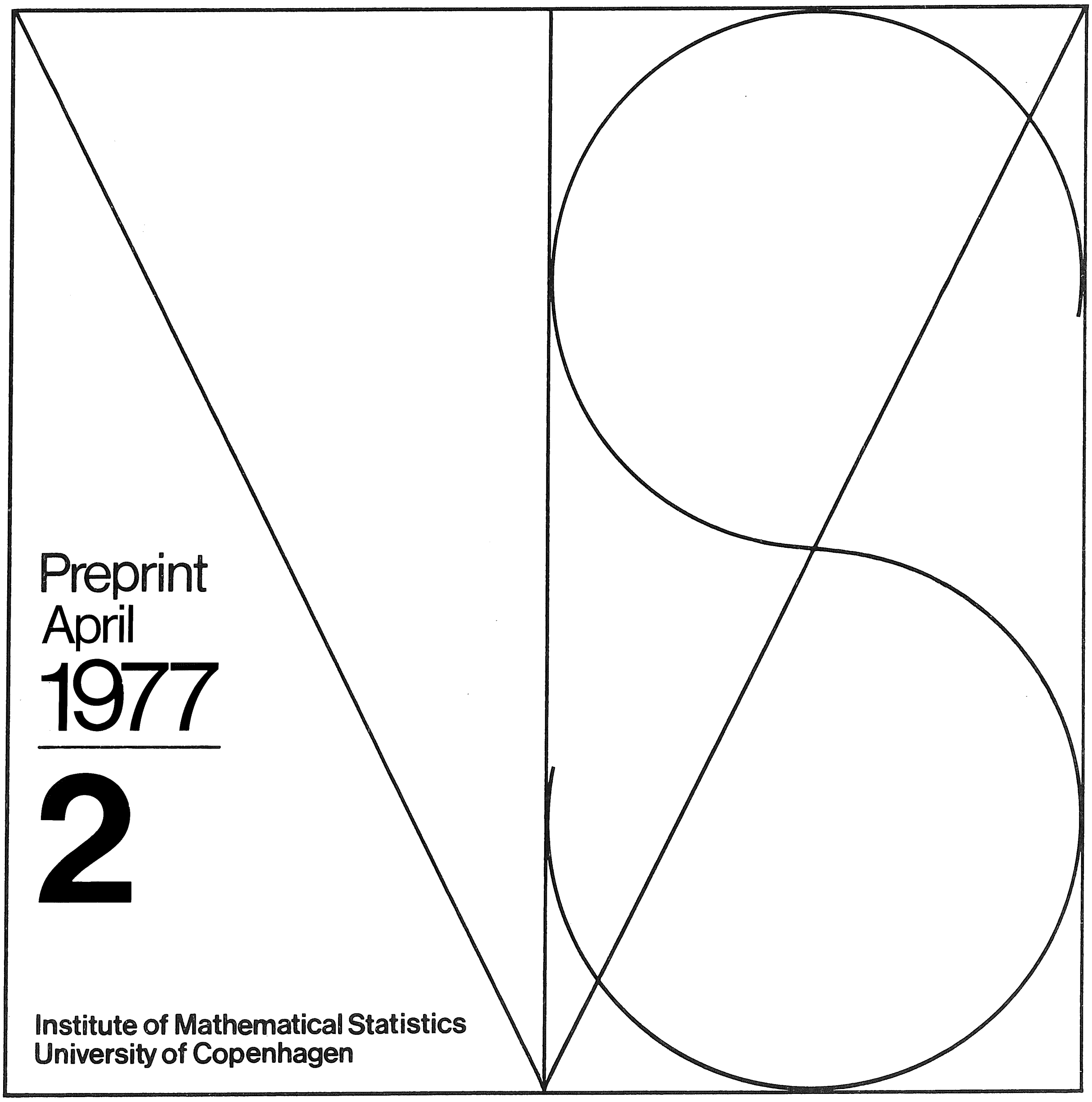
Martin Jacobsen

Stochastic Processes with Stationary Increments in Time and Space



Preprint
April
1977

2

Institute of Mathematical Statistics
University of Copenhagen

Jean-Pierre Dion and Niels Keiding

STATISTICAL INFERENCE IN BRANCHING PROCESSES*

Preprint 1977 No. 10

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

November 1977

STATISTICAL INFERENCE IN BRANCHING PROCESSES

Jean-Pierre Dion

Département de Mathématiques
Université du Québec à Montréal
Montréal, P.Q., Canada

and

Niels Keiding

Institute of Mathematical Statistics
University of Copenhagen
Copenhagen, Denmark

INTRODUCTION

This paper surveys the theoretical literature on statistical inference for branching processes, almost exclusively from the last five years. The emphasis is on estimation of the offspring mean (time-discrete case) and the Malthusian growth parameter (time-continuous case) based on a complete record of the process in a time period.

Essentially two different repetitive structures are important: increasing number of ancestors, leading to standard iid theory, and increasing number of generations with which we shall mostly be concerned. Very little thought has been given to small-sample theory, including suitable conditionality arguments for defining reasonable reference sampling distributions.

For the large sample theory for supercritical branching processes in discrete time it seems useful to view a series of generations as successive sections (of random length) of iid replications of a random variable with the offspring distribution. The continuous-time analogue is the random time transformation to a Poisson process (with has iid increments) via the inverse of the natural increasing process corresponding to the split time process, this being a trivial submartingale. The latter approach has been used by Aalen for statistical analysis of more general point processes.

The use of the natural increasing process as a measure of efficiency has been studied for both discrete and continuous time by a series of Australian authors, who have also provided the <u>offspring distribution</u> (p_k). Let z_0 be some fixed number and define the stochastic process $Z_0, Z_1, \dots$ by

$$
\begin{aligned}
Z_0 &= z_0 \, , \\
Z_1 &= Y_1 + \dots + Y_{Z_0} \, , \\
Z_2 &= Y_{Z_0+1} + \dots + Y_{Z_0+Z_1} \, , \\
&\dots \\
Z_n &= Y_{Z_0+\dots+Z_{n-2}+1} + \dots + Y_{Z_0+\dots+Z_{n-1}} \, , \\
&\dots
\end{aligned}
$$

Then (Z_n) is a Galton-Watson process with offspring distribution (p_k) and $Z_0 = z_0$.

<u>Proposition 1.1</u> Consider the completely general statistical model specified by all offspring distributions $\{(p_k) : \Sigma p_k = 1\}$ and assume that all individual offspring sizes of individuals in the n first generations have been observed. Then the maximum likelihood estimator of the (p_k) is the obvious set of relative frequencies

$$\hat{p}_k = N_k / (Z_0 + \dots + Z_{n-1})$$

where N_k is the number of times k offspring are produced, that is

$$N_k = \# \{ i \in \{1, \dots, Z_0 + \dots + Z_{n-1}\} \mid Y_i = k \} .$$

<u>Proof</u>. If $Y_1, \dots, Y_p$ had been observed, obviously

$$\hat{p}_k = \# \{ i \in \{1, \dots, p\} \mid Y_i = k \} / p .$$

In the present context, p has been replaced by the <u>stopping time</u> $Z_0 + \dots + Z_{n-1}$ which does not depend on the parameters. At the point $(y_1, \dots, y_{z_0+\dots+z_{n-1}})$ the likelihood of $(Y_1, \dots, Y_{Z_0+\dots+Z_{n-1}})$ is proportional to that of $(Y_1, \dots, Y_p)$ with $p = z_0 + \dots + z_{n-1}$ (this property is conventionally termed "the likelihood is independent of the stopping rule"). This concludes the proof. The result was first pointed out by Harris [31].

Corollary. The maximum likelihood estimator of the offspring mean $m = \Sigma k p_k$ in the general model based on the observation of all offspring sizes in the n first generations is given by

$$\hat{m} = (Z_1+\ldots+Z_n)/(Z_0+\ldots+Z_{n-1}),$$

that is, the total number of children divided by the total number of parents.

Proof. $\hat{m} = \Sigma k\hat{p}_k$.

Based on the result of the Corollary, it is an obvious conjecture that $\hat{m}$ is also the maximum likelihood estimator in the general model if only the generation sizes $Z_0,\ldots,Z_n$ are observed.

Theorem 1.2 The maximum likelihood estimator of the offspring mean m in the general model based on the observation of the generation sizes $Z_0,\ldots,Z_n$ only is given by

$$\hat{m} = (Z_1+\ldots+Z_n)/(Z_0+\ldots+Z_{n-1}).$$

Proof. Dion [21, Théorème 1.2] and Feigin [28] gave Lagrange multiplier arguments. Keiding and Lauritzen [54] used an exponential family approach which also delineates a class of statistical models in which $\hat{m}$ is the maximum likelihood estimator of the offspring mean, as shown below.

Remark. Notice that $\hat{m}$ is always finite even though the model allows for $m = \infty$.

Theorem 1.3 (Keiding and Lauritzen [54]). Let the parameter set $\Theta \subseteq R\times\Omega$, $\theta = (\xi,\omega) \in \Theta$ and the offspring probabilities be of the form

$$p_x(\theta) = a(\theta)g(x,\omega)e^{\xi x}$$

where the support $S = \{x \mid g(x,\omega) > 0\}$ is independent of θ and for each fixed ω the sections $\Xi_\omega = \{\xi \in R^k \mid (\xi,\omega) \in \Theta\}$ are open and given as

$$\Xi_\omega = \{\xi \in R^k \mid \sum_{x\in S} g(x,\omega)e^{\xi x} < \infty\}.$$

Then if the observed $\hat{m}$ belongs to the interior of the closed convex hull of S, it is the maximum likelihood estimate of m.

Corollary. If N is a stopping time not depending on the parameters, and if $Z_0,\ldots,Z_N$ is observed, then the maximum likelihood estimator of m in the models described in the previous theorems is given by

$$(Z_1+\ldots+Z_N)/(Z_0+\ldots+Z_{N-1}).$$

Examples of offspring distributions fulfilling the conditions of Theorem 1.3 are the power series distributions (Dion [21], Becker [11], Eschenbach and Winkler [26], Heyde [35], Heyde and Feigin [36], binary splitting (Jagers [46], the modified geometric distribution (Keiding [50]) and the negative binomial distribution. A nontrivial counterexample is the zeta-distributions with probabilities $p_x = (x + 1)^{-\theta}/\zeta(\theta)$, $x = 0,1,2,\ldots$, $1 < \theta < \infty$, with ζ the Riemann zeta function, for which $\hat{m}$ is not the maximum likelihood estimator (Keiding [50]).

Observation of a random number N of generation sizes has in particular been studied by Becker [11].

We conclude this section by noticing that there exist practically no small-sample distribution theory for $\hat{m}$, and the interesting question of possible conditional inference has hardly been touched upon.

1.2 Asymptotic theory for $\hat{m}$.

Essentially two different repetitive structures are important: increasing number of ancestors and increasing number of generations.

By the branching property, the distribution of $(Z_0,\ldots,Z_N)$ given $Z_0 = z$ is the same as that of a sum of z independent identically distributed replications of $(Z_0,\ldots,Z_n)$ given $Z_0 = 1$. It follows that standard large sample theory applies for large z, the most complete and careful treatment being that of Yanev [72].

Theorem 2.1 For fixed n and $z \to \infty$,

(a) if $m \to \infty$, then $\hat{m} \to m$ a.s. and $E(\hat{m}) \to m$,

(b) if the offspring variance $\sigma^2 < \infty$, then the asymptotic distribution of

$$\{\frac{z}{\sigma^2}(1+m+\ldots+m^{n-1})\}^{\frac{1}{2}}(\hat{m}-m)$$

is standard normal.

Remark. An iterated logarithm result is routine.

We next consider the second asymptotic structure, that of increasing number of generations, first with fixed number z of ancestors, conveniently assumed equal to one. Since a branching process has all states $j > 0$ transient, there is no way of appealing to the standard theory of statistical inference for Markov chains such as given by Billingsley [15]. Obviously, nontrivial results exist only for the supercritical case $m > 1$; we let A denote the set $\{Z_n \to \infty\}$ of nonextinction.

Theorem 2.2 If $1 < m < \infty$, then for fixed z and $n \to \infty$

$$\hat{m} \to m \quad \text{a.s. on } A.$$

Proof. This result was noted by Heyde [33] as a corollary to his definitive theorem on normalizing constants for the asymptotic growth of supercritical branching processes. We take here the opportunity to point out that the result is in fact an easy consequence of the strong law of large numbers in its simplest form.

Let $Y_1, Y_2, \ldots$ be the iid offspring sizes defined in Section 1. Then as $p \to \infty$, $(Y_1+\ldots+Y_p)/p \to m$ a.s. On the set $\{Z_0+\ldots+Z_{n-1} \to \infty\}$, which is a.s. the same as A, it follows that

$$\hat{m} = \frac{Y_1+\ldots+Y_{Z_0+\ldots+Z_{n-1}}}{Z_0+\ldots+Z_{n-1}} \to m$$

a.s.

Asymptotic normality is obtained in a similar manner.

Theorem 2.3 (Dion [21,22], Jagers [44]). Assume $m > 1$, $0 < \sigma^2 < \infty$ and let $P_A\{\cdot\} = P\{\cdot \mid A\}$ and $S(w) = P_A\{W \leq w\}$ where as usual $W = \lim_{n\to\infty} \text{a.s. } Z_n m^{-n}$. Then for fixed z and $n \to \infty$

(a) $P_A\{\sigma^{-1}(Z_0+\ldots+Z_{n-1})^{\frac{1}{2}}(\hat{m} - m) \leq x\} \to \Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-t^2/2}\, dt,$

(b) $P_A\{\sigma^{-1} z(1 + m+\ldots+ m^{n-1})^{\frac{1}{2}}(\hat{m} - m) \leq x\} \to \int_0^\infty \Phi(x\sqrt{w})dS(w)$

and (c), the results (a) and (b) continue to hold if P_A is replaced by $P\{\cdot|Z_n > 0\}$.

Proof. We outline the main steps and refer to the quoted papers for details. By the central limit theorem the asymptotic distribution as $p \to \infty$ of

$$\sigma^{-1}p^{-\frac{1}{2}} \sum_{i=1}^{p} (Y_i-m)$$

is standard normal. Now

$$\sigma^{-1}(Z_0+\ldots+Z_{n-1})^{\frac{1}{2}}(\hat{m} - m) = \sigma^{-1}(Z_0+\ldots+Z_{n-1})^{-\frac{1}{2}} \sum_{i=1}^{Z_0+\ldots+Z_{n-1}} (Y_i - m)$$

and at least if $p_0 = P\{Z_1 = 0\} = 0$, the result (a) follows fairly directly from a central limit theorem for a sum of a random number of independent random variables such as quoted by Billingsley [16, Theorem 17.2]. Also, (b) is a consequence of that result and (c) is elementary since the sequence of sets $\{Z_n>0\}$ decreases to A. The generalization to $p_0>0$ was carried through by Dion and Jagers in the above mentioned papers.

Notice that result (b) states that $\hat{m}$ is *not* asymptotically normal when deterministic normalizing constants are used. If the offspring distribution is geometric, then the limiting distribution of (b) is a Student-distribution with 2z degrees of freedom.

Remark. An iterated logarithm result for $\hat{m}$ was recently provided by Asmussen and Keiding [6].

Conditioning on the non-extinction set makes sense only in the supercritical case. However it would be interesting to study the asymptotic distribution of $\hat{m}$ (properly normalized) in the subcritical and critical cases, given $Z_n>0$. Results of this type were given by Pakes [63] for the estimators $\hat{p}_k$ discussed in Proposition 1.1 above.

We finally note that Yanev [72] provided asymptotic distribution theory under the assumption that both z and n become large. These results would presumably be quite important in practice. The character of the results turns out to depend on the usual criticality trichotomy.

Theorem 2.4 (Yanev [72]). Let $0<\sigma^2<\infty$. Then as $n\to\infty$ and $z\to\infty$

(a) $\hat{m}\xrightarrow{P} m$ and $E(\hat{m})\to m$.

(b) If $m<1$, then the asymptotic distribution of

$$\left(\frac{z}{\sigma^2(1-m)}\right)^{\frac{1}{2}}(\hat{m}-m)$$

is standard normal.

(c) If m=1, then if $n/z\to 0$, the asymptotic distribution of

$$\frac{(zn)^{\frac{1}{2}}}{\sigma}(\hat{m}-m)$$

is standard normal, while if $n/z^2\to\infty$, the asymptotic distribution of

$$\frac{\sigma^2}{2z^2(1-\hat{m})}$$

is a stable distribution with exponent $\frac{1}{2}$.

(d) If $m>1$ and the offspring distribution has finite fourth moment, then the asymptotic distribution of

$$\sigma^{-1}\{z(1+\ldots+m^{n-1})\}^{\frac{1}{2}}(\hat{m}-m)$$

is standard normal.

Notice that the results in Theorems 2.1 (b) and 2.3 (b) correspond to (d) in the present theorem (since W will degenerate as $z\to\infty$.)

1.3 Alternative estimators of the offspring mean.

A different estimator of the offspring mean was proposed by Lotka [58] and studied by Nagaev [62], see also Crump and Howe [19] and Dion [22]. This estimator is defined by

$$\bar{m} = \begin{cases} Z_n/Z_{n-1} & \text{if } Z_{n-1} > 0 \\ 1 & \text{otherwise .} \end{cases}$$

Then as noted by Nagaev, $E(\bar{m} \mid Z_{n-1} > 0) = m$. Asymptotic properties as $Z_0 = z \to \infty$ are standard as before. If $1 < m < \infty$, it follows directly from Heyde's [33] results that as $n \to \infty$, $\bar{m} \to m$ a.s. on A. Crump and Howe [19] showed that if $0 < \sigma^2 < \infty$, $E(\bar{m} \mid A) \to m$ as $n \to \infty$.

Asymptotic distribution results as $n \to \infty$ were derived by Nagaev [62] and Dion [22]. In the supercritical case $m > 1$ these state that

$$P_A\{\sigma^{-1} Z_{n-1}^{\frac{1}{2}} (\bar{m}-m) \le x\} \to \Phi(x),$$

$$P_A\{\sigma^{-1} m^{n/2} (\bar{m}-m) \le x\} \to \int_0^\infty \Phi(x\sqrt{w})dS(w)$$

and that these limiting distribution results continue to hold if conditioning on A is replaced by conditioning on $\{Z_{n-1} > 0\}$.

Nagaev also provided results for the critical and subcritical cases; iterated logarithm laws may be obtained as before.

Finally Heyde [35] mentioned the possibility of using $m^* = Z_n^{1/n}$ as an estimator for m. If $E(Z_1 \log Z_1) < \infty$, $Z_n m^{-n} \to W$ a.s. as $n \to \infty$, and therefore $m^* \to m$ a.s. on A. The rate of convergence is illustrated by the fact that by an application of Heyde and Leslie [37, Theorem 2], one has $n(m^*-m) \to m \log W$ a.s. on A. This "linear" rate of convergence compares badly to the "geometric" rate of $\hat{m}$ and $\bar{m}$, as pointed out by Heyde.

1.4 Estimation of the offspring variance.

In order to use results from Sections 2 and 3 to deduce confidence intervals for the mean of the offspring distribution, one has to give a consistent estimator for the variance σ^2. Such an estimator is given by

$$(4.1)\quad \overset{*}{\sigma}{}^2 = n^{-1} \sum_{k=0}^{n-1} Z_k((Z_{k+1}/Z_k - m^*)^2)$$

where m^* is either Harris' estimator $\hat{m}$ (Dion [23]) or the Lotka-Nagaev estimator $\bar{m}$ (Heyde [34]). Denote $\overset{*}{\sigma}{}^2$ by $\hat{\sigma}^2$ if $m^*=\hat{m}$ and by $\bar{\sigma}^2$ if $m^*=\bar{m}$. The properties of $\overset{*}{\sigma}{}^2$ are best studied by taking first $m^*=m$ and then replacing m by either $\hat{m}$ or $\bar{m}$.

The asymptotic behaviour of $\overset{*}{\sigma}{}^2$ has been studied in the case when $p_0 = 0$, but one expects the same asymptotic results to hold on the set of non-extinction in the general situation where $p_0 > 0$ is allowed.

Theorem 4.1 (Dion [23]) Assume $p_0 = 0$, $m > 1$, $0 < \sigma^2 < \infty$ and $E(Z_1{}^4) < \infty$. Put $\tilde{\sigma}^2 = n^{-1} \sum_{k=0}^{n-1} Z_k((Z_{k+1}/Z_k - m)^2)$. Then

(a) $(\tilde{\sigma}^2 - \hat{\sigma}^2)n^{1-\varepsilon} \overset{P}{\to} 0, \quad \forall\varepsilon>0$.

(b) $\tilde{\sigma}^2 > \hat{\sigma}^2$ a.s.

Proof. For all $\varepsilon>0$,

$$(\tilde{\sigma}^2 - \hat{\sigma}^2)n^{1-\varepsilon} = n^{-\varepsilon} \sum_{k=0}^{n-1} Z_k[(\hat{m}-m)\{2(Z_{k+1}/Z_k) - (\hat{m}+m)\}]$$

$$= (Z_0 + \ldots + Z_{n-1})(\hat{m}-m)^2 . n^{-\varepsilon} > 0 \text{ a.s.},$$

from which (b) follows. By Theorem 2.3(a) the asymptotic distribution of $\sigma^{-2}(Z_0 + \ldots + Z_{n-1})(\hat{m}-m)^2$ is χ^2 with 1 d.f., and therefore

$$|\tilde{\sigma}^2 - \hat{\sigma}^2|\, n^{1-\varepsilon} \overset{P}{\to} 0, \quad \forall\varepsilon>0.$$

Note 1. It is easy to show that $E(\tilde{\sigma}^2) = \sigma^2$. The property (b) then states that $E(\hat{\sigma}^2) < \sigma^2$.

Note 2. By straightforward but long computations with conditional expectations and variances one can show that $\mathrm{Var}(\tilde{\sigma}^2) \approx 2\sigma^4/n$. From property (a), $\hat{\sigma}^2$ is then a consistent estimator of σ^2.

Note 3. Using a central limit theorem for martingales due to Billingsley [15, p.52], the asymptotic normality of $\tilde{\sigma}^2$ (and hence of $\hat{\sigma}^2$) can be easily established under the further assumption $E(Z_1^6) < \infty$. But as is seen in the next theorem Heyde [34] has stronger results for $\bar{\sigma}^2$. These have been extended to $\hat{\sigma}^2$ by P. Feigin in an unpublished thesis.

Theorem 4.2 (Heyde [34]). Assume $p_0 = 0$, $m > 1$ and $0 < \sigma^2 < \infty$. Then

(a) $\bar{\sigma}^2 \xrightarrow{a.s.} \sigma^2$

(b) if $E(Z_1^4) < \infty$, then

$$\sqrt{n}(\bar{\sigma}^2 - \sigma^2)/(2\sigma^4)^{\frac{1}{2}} \xrightarrow{D} N(0,1) \quad \text{and}$$

(c) if $E(Z_1^{4+\delta})$ for some $\delta > 0$, then

$$\bar{\sigma}^2 = \sigma^2 + \eta(n)(4\sigma^4 n^{-1} \log \log n)^{\frac{1}{2}},$$

where $\eta(n)$ has its set of a.s. limit points confined to $[-1,+1]$ with $\limsup_{n\to\infty} \eta(n) = +1$ a.s. and $\liminf_{n\to\infty} \eta(n) = -1$ a.s.

For the proof, which is rather long and far from trivial, the reader is referred to Heyde's article. However we will give an idea of his proof which relies heavily on strong results for martingales.

Let F_n be the σ-field generated by $Z_1, \ldots, Z_n$ and put

$$U_{k+1} = (Z_{k+1} - mZ_k)^2 Z_k^{-1} - \sigma^2.$$

Using once more the underlying independent r.v. that constitute the family tree, one has that the U_K are martingale differences. Apply next a strong law of large numbers for martingales to conclude that

$$n^{-1} \sum_{k=1}^{n} U_k \xrightarrow{a.s.} 0$$

and then the property (a) follows after proving that m can be replaced by $\bar{m}$. (This last fact still requires an iterated logarithm analogue for $\bar{m}$ as given by Heyde & Leslie [37]).

The central limit result (b) is obtained from a central limit theorem for martingales (Theorem 2 of Brown [17]) applied to the martingale $\{\sum_{j=1}^{n} U_j, \mathcal{F}_n\}$ and the iterated logarithm result (c) is deduced from a corresponding iterated logarithm result for martingales due to Heyde and Scott [38, Theorem 1].

In the critical case, $m = 1$, Nagaev [62] observed that $\beta_n = (Z_n - Z_{n-1})^2 / Z_{n-1}$ could be used to estimate σ^2 since $E(\beta_n \mid Z_{n-1} > 0) = \sigma^2$ and if $E(Z_1^4) < \infty$, $\mathrm{Var}(\beta_n \mid Z_{n-1} > 0) = O(\log n/n)$.

1.5 Immigration.

Consider now $X_0 = 1, X_1, X_2, \ldots$ a Galton-Watson process with immigration whose offspring distribution has the distribution of Z_1 with $EZ_1 = m$ and $0 < \mathrm{Var}(Z_1) = \sigma^2 < \infty$. Suppose further that the immigration process has a finite mean $\lambda = \sum i b_i$.

Estimation problems in that context have been studied by Heyde & Seneta [39,40,41] and Quine [65].

In the supercritical case, $m>1$, the asymptotic properties of the Lotka-Nagaev estimator as given in Section 3 and Heyde's [34] estimator for σ^2 are both robust in the sense that the asymptotic results as $n \to \infty$ continue to apply unchanged in the case where immigration occurs. Only minor modifications of the proofs for the Galton-Watson process without immigration are necessary to establish these results.

The subcritical case ($m<1$) however presents a new situation. Put $\lambda_1 = m$ and $\lambda_2 = \lambda$. The estimators investigated are

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^{n} X_i (X_{i+1} - n^{-1} S_n)}{\sum_{i=1}^{n} (X_i - n^{-1} S_n)^2}$$

and

$$\hat{\lambda}_2 = \frac{S_n}{2n} \frac{\sum_{i=1}^{n} (X_{i+1}-X_i)}{\sum_{i=1}^{n} (X_i-n^{-1}S_n)^2}$$

where $S_n = \sum_{i=1}^{n} X_i$. Let $\mu = \lambda_2(1-\lambda_1)$, $c^2 = \mu\sigma^2 + b^2$ where b^2 is the variance of the immigration distribution. Assume further the immigration distribution is not degenerate. Then, improving on the work by Heyde & Seneta [40,41], Quine [65] proved that $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are strongly consistent and obey the central limit theorem and law of the iterated logarithm under the sole condition $c^2 < \infty$.

Theorem 5.1 (Quine [65]). If $c^2 < \infty$, then for $i = 1,2$ as $n \to \infty$

(a) $\hat{\lambda}_i \to \lambda_i$ almost surely.

(b) The asymptotic distribution of $\sqrt{n}(\hat{\lambda}_i - \lambda_i)/k_i$ is standard normal, and

(c) $$\limsup \frac{n^{\frac{1}{2}}(\hat{\lambda}_i - \lambda_i)}{k_i(2 \log\log n)^{\frac{1}{2}}} = 1 \text{ a.s.}$$

$$\liminf \frac{n^{\frac{1}{2}}(\hat{\lambda}_i - \lambda_i)}{k_i(2 \log\log n)^{\frac{1}{2}}} = -1 \quad \text{a.s.}$$

as long as $k_i < \infty$, where k_i are constants given by Quine [65, p.320, cf. the correction note].

The theorem follows from known results about Markov chains. To illustrate that, let us recall the very elegant proof of (a) as given by Quine. Note that the state space I contains a countable irreducible set I* on which $X_1, X_2, \ldots$ have their support. Clearly I* is aperiodic since

$$\kappa \equiv \inf\{i : b_i > 0\} = \inf\{i : i \in I^*\}$$

is accessible at all times $n \geqq 1$. Furthermore as $\Sigma b_j \log j < \infty$, I* forms a positive class. Let $\overline{T}_{I*}$ be the σ-field of all subsets of I*. If φ is any function from $F_{I*} \times F_{I*}$ to R, and if $\{\pi_i\}$ is the limiting distribution of $\{X_n\}$, it follows from Billingsley [15, Theorems 1.1 and 1.3] that

$$n^{-1}\Sigma\ \varphi(X_j, X_{j+1}) \rightarrow \sum_{i,j} \pi_i P(X_1=j \mid X_0=i)\varphi\,(i,j) \quad \text{a.s.}$$

as long as the limiting series converges absolutely. It can be shown that the stationary distribution $\{\pi_i\}$ has mean μ and variance $c^2/(1-\lambda_1{}^2)$.

There remains only to choose suitable φ to show that

$$n^{-1}\ S_n \rightarrow \mu \quad \text{a.s.}$$

$$n^{-1}\sum_{i=1}^{n} (X_{i+1} - X_i)^2 \rightarrow 2c^2/(1+\lambda_1) \quad \text{a.s.}$$

$$n^{-1}\sum_{i=1}^{n} X_i{}^2 \rightarrow c^2/(1-\lambda_1{}^2) + \mu^2 \quad \text{a.s.}$$

and

$$n^{-1}\sum_{i=1}^{n} (X_i - \mu)^3 \rightarrow \gamma \quad \text{a.s.}$$

The property (a) follows from these a.s. convergence results.

We conclude this section by calling attention to the interesting historical remarks and discussions of applications provided by Heyde & Seneta [40].

1.6 Multitype processes

This section contains a brief survey of some very recent results concerning the estimation of the mean matrix M, in particular the growth rate ρ in the positive regular case where ρ is given as the unique largest positive eigenvalue of M.

Motivated by a wish to be able to estimate at an early stage whether an epidemic is minor or major Becker [13] posed the problem of the behaviour of the three above mentioned estimators $\hat{m}$, $\overline{m}$ and m^* when the process is in fact a multitype Galton-Watson process

$(Z_n) = ((Z_n(1),\ldots,Z_n(p)))$ and the estimators are based on the total generation sizes $|Z_n| = Z_n(1)+\ldots+Z_n(p)$. Becker showed under mild conditions ($\rho > 1$ and "j log j") that all of these estimators will converge almost surely to ρ on the set of nonextinction as $n \to \infty$.

A more detailed analysis was provided by Asmussen and Keiding [6] who adapted martingale central limit theory to martingale difference triangular arrays indexed by the set of all individuals ever alive. Aside from $(|Z_1|+\ldots+|Z_n|)/(|Z_0|+\ldots+|Z_{n-1}|)$, which we shall henceforth denote by $\tilde{\rho}$, these authors concentrated their interest on the obvious estimator $\hat{M} = (\hat{m}_{ij})$ of the whole mean matrix based on the observation of all combinations of types of parents and offspring. Let U^{ν}_{ki} be the vector of offspring of the k'th individual of type i in generation ν, $\nu = 0,\ldots,n-1$. Then

$$\hat{m}_{ij} = \sum_{\nu=0}^{n-1} Z^{i}_{\nu+1}(j) / \sum_{\nu=0}^{n-1} Z_{\nu}(i)$$

where $Z^{i}_{\nu+1}(j) = \Sigma_{k=1}^{Z_{\nu}(i)} U^{\nu}_{ki}(j)$ is the number of individuals of type j in the $(\nu+1)$st generation whose parents were of type i and $Z_{\nu}(j) = \Sigma Z^{i}_{\nu}(j)$. As noted by Keiding and Lauritzen [54], $\hat{M}$ is the maximum likelihood estimator of M whether based on observations on the U^{ν}_{ki} or on the $Z^{i}_{\nu}(j)$ only, in the completely general model as well as in models specified by classes of offspring distributions as discussed in Theorem 1.3 above.

In the results below, we assume throughout that M is positive regular, its principal eigenvalue $\rho > 1$, and zero offspring always has probability zero, so that $W = \lim \text{a.s. } \rho^{-n} Z_n > 0$ a.s. Finally v and u are the left and right eigenvectors, $vM = \rho v$ and $Mu = \rho u$.

Theorem 6.1 As $n \to \infty$,

(a) $\hat{M} \to M$ a.s.

(b) The limiting distribution of the matrix
$([W(1+\rho+\ldots+\rho^{n-1})v(i)]^{\frac{1}{2}} (\hat{m}_{ij} - m_{ij}))$

is that of $(Y_i(j))$, where $Y_1,\ldots,Y_p$ are independent and the distribution of Y_i is p-dimensional normal with mean zero and variance matrix Σ^i given by

$$\Sigma^i_{jj} = \mathrm{Var}(Z_1(j) \mid Z_0 = e_i),$$

$$\Sigma^i_{jk} = \mathrm{Cov}(Z_1(j), Z_1(k) \mid Z_0 = e_i).$$

By the transformation invariance of maximum likelihood estimators, the maximum likelihood estimator $\hat{\rho}$ of ρ is given as the largest positive eigenvalue of $\hat{M}$, which will be well-defined at least for large n. Standard techniques then immediately yield

Corollary. As $n \to \infty$,

(a) $\hat{\rho} \to \rho$ a.s.

(b) The asymptotic distribution of $[W(1+\rho+\ldots+\rho^{n-1})]^{\frac{1}{2}}(\hat{\rho}-\rho)$ is normal $(0, v \cdot \mathrm{Var}\, Z_1 \cdot u)$.

Iterated logarithm laws were also derived by Asmussen and Keiding.

We next turn our attention to the estimator $\tilde{\rho} = (|Z_1|+\ldots+|Z_n|) \,/\, (|Z_0|+\ldots+|Z_{n-1}|)$. Becker [13] proved strong consistency. The rate of convergence turns out to depend in a qualitative way on the rate of convergence of the relative type distribution to the stable type distribution, cf. Kesten and Stigum [57] and Asmussen [5]. As it appears from the latter references, there is a trichotomy depending on the relative sizes of $|\lambda|^2$ and ρ, where λ is a certain other eigenvalue of M. In fact, if $|\lambda|^2 \geqq \rho$, no central limit result parallel to that of the Corollary above holds.

Theorem 6.2 When $|\lambda|^2 < \rho$, as $n \to \infty$, $[W(1+\rho+\ldots+\rho^{n-1})]^{\frac{1}{2}}(\tilde{\rho}-\rho)$ is asymptotically normal with mean zero and variance computed by Asmussen and Keiding [6].

We finally note that Quine and Durham [66] recently obtained multitype analogues of the results by Heyde and Seneta [40,41] and Quine [65] concerning the subcritical process with immigration and discussed in Section 5 above.

1.7 Random environments

Becker [13] continued his investigation by inquiring into the properties of $\hat{m}$, $\bar{m}$ and m^* when the process (Z_n) is in fact a branching process with (independent identically distributed) random environments. As is well known, it is here important to distinguish between the average offspring mean $\mu = E(Z_1 \mid Z_0 = 1)$ and the growth rate (criticality parameter) $\tau = \exp(E[\log Z_1 \mid Z_0, \zeta_1])$, ζ_1 denoting the environment of the first generation. To Becker, τ was of prime importance and he observed that m^* is strongly consistent for τ whereas

$$\tilde{m} = n^{-1} \sum_{j=1}^{n} (Z_j/Z_{j-1})$$

is strongly consistent for μ. An independent study by Dion and Esty [24] also pointed out these facts and went on to derive asymptotic variances and asymptotic distribution results for m^* and $\tilde{m}$.

1.8 Estimation of the age of the process

It has been suggested that it might occasionally be of interest to estimate the generation number of a Galton-Watson process, assuming that it descended from one ancestor, see Stigler [69] and Crump and Howe [19]. For a discussion of the possibility of applying this idea to rare human blood types we refer to Thompson [70,71]. It appears from Thompson's analysis that the evaluation of the age is critically dependent on the estimated mean offspring, and at least in her context it seemed difficult to estimate both at the same time.

2. ESTIMATION THEORY FOR CONTINUOUS-TIME BRANCHING PROCESSES

2.1 Maximum likelihood estimation of parameters in Markov branching processes

Consider a Markov branching process $(X_t,\ t \geqq 0)$ with split intensity $\lambda > 0$ and offspring distribution $(p_k,\ k = 0,1,2,\ldots,\ \Sigma\, p_k = 1)$. As usual we assume $p_1 = 0$ and $X_0 = x_0$ degenerate.

If this process is assumed observed continuously over a time interval $[0,t]$, and if it is known that with probability one only finitely many jumps happen in each finite interval (a sufficient condition being that the offspring mean $m = \Sigma\, kp_k < \infty$, cf. Harris [32, p.107]), then it follows by a minor modification of Albert's [4] arguments that a measure may be constructed such that the likelihood function is as given in the Theorem below.

Theorem 2.1 The likelihood function corresponding to observation of $\{X_u :\ 0 \leqq u \leqq t\}$ is

$$\lambda^{N_t}\, e^{-\lambda S_t} \prod_{k=0,2}^{\infty} p_k^{N_t(k)}$$

where $N_t(k)$ is the number of splits of size $k-1$ in $[0,t]$, $N_t = \Sigma\, N_t(k)$ is the total number of splits, and

$$S_t = \int_0^t X_u\, du$$

is the total time lived by the population (the total exposure time) in $[0,t]$.

For the statistical problem specified by $\lambda \in (0,\infty)$ and all offspring distributions with finite mean the maximum likelihood estimators are given as

$$\hat{\lambda} = N_t / S_t,\quad \hat{p}_k = N_t(k)/N_t$$

when $N_t > 0$.

Remark. The results of the Theorem continue to hold with the obvious changes if the deterministic observation interval [0,t] is replaced by [0,T], where T is a stopping time not depending on the parameters. If in particular T is chosen as a first hitting time for N_t or S_t, the distribution of $\hat{\lambda}$ will be more easily tractable, since it is no longer a ratio between random variables, cf. Moran [60,61] and Kendall [55].

Corollary. The maximum likelihood estimator of the offspring mean m and the growth rate (Malthusian parameter) $\alpha = (m-1)\lambda$ are given by

$$\hat{m} = 1 + (X_t - x_0)/N_t, \quad \hat{\alpha} = (X_t - x_0)/S_t.$$

A further look at the estimation problem reveals that as long as the split parameter λ and the parameters of the offspring distribution are variation independent, there are two separate problems: The random number N_t of i i d observations of the offspring distribution, leading to standard relative frequency estimators, and the question of estimating the split intensity λ, where the occurrence/exposure rate N_t/S_t is seen to arise in a natural way. As will be seen below, the latter question is of greatest theoretical interest and will be emphasized.

It follows from these remarks that no great new theoretical insight is to be expected from the analysis of more restricted models based on particular classes of offspring distributions. Aside from the general model discussed above we shall therefore only mention the linear birth-and-death process with $p_0+p_2 = 1$, in particular the pure birth (Yule) process $(p_2 = 1)$ and the pure death process $(p_0 = 1)$.

Very few small-sample results exist. Beyer, Keiding and Simonsen [14] gave formulas, numerical tables, and approximative results for the first three moments of the maximum likelihood estimator in the pure birth process and the pure death process. That paper also surveyed earlier literature.

As in Chapter 1, we shall therefore have to be satisfied mostly with asymptotic results.

2.2 Asymptotic results for estimators in Markov branching processes.

As for the time-discrete case, two types of asymptotic theory are relevant: large number of ancestors and large interval of observation.

Though the many-ancestors theory might often be the more important in practice, it is a fairly standard application of i i d estimation theory and we shall therefore be content with referring to the statement of the results by Keiding [49,51] and Athreya and Keiding [7].

The results for $t \to \infty$ are of greatest interest in the supercritical case $\alpha > 0$ ($\Longleftrightarrow m > 1$) since otherwise $X_t \to 0$ a.s. Let $A = \{X_t \to 0\}$.

Theorem 2.1 As $t \to \infty$,

(a) if $m < \infty$, then $\hat{\lambda} \to \lambda$ and $\hat{p}_k \to p_k$, $k = 0,2,3,\ldots$, a.s. on A.

(b) if $0 < \sigma^2 < \infty$ (where σ^2 is the offspring variance $\Sigma(k-m)^2 p_k$), then the asymptotic distribution, given A, of

$$S_t^{\frac{1}{2}}(N_t(0)/S_t - \lambda p_0,\ N_t(2)/S_t - \lambda p_2,\ N_t(3)/S_t - \lambda p_3, \ldots)$$

is that of independent normals with parameters $(0, \lambda p_k)$, $k = 0,2,3,\ldots$

Remark. Notice that $N_t(k)/S_t = \hat{\lambda}\hat{p}_k$ is the maximum likelihood estimator of λp_k.

Corollary. As $t \to \infty$,

(a) if $m < \infty$, $\hat{\alpha} = (X_t - x_0)/S_t \to \alpha$ a.s. on A,

(b) if $\sigma^2 < \infty$, the asymptotic distribution, given A, of $S_t^{\frac{1}{2}}(\hat{\alpha} - \alpha)$, is normal $(0, \lambda(\sigma^2 + (m-1)^2))$.

Remark. These statements may be transformed into statements with deterministic normalising factors in a similar way as discussed for the time-discrete case in Section 1.2 above. Also, for the purpose of suggesting asymptotic confidence intervals, conditioning on the set A may be replaced by conditioning on $\{X_t > 0\}$.

As indicated in Section 1 above, the estimators of the offspring probabilities are really just the standard relative frequencies, although with a random number of replications. To give an indication of methods of proof we may therefore restrict ourselves to the pure birth process with $p_2 = 1$.

Thus let $(X_t, t \geqq 0)$ be a pure birth process with birth intensity λ and $X_0 = x$ degenerate. We want to prove, with $\hat{\lambda} = (X_t - x)/S_t$, $S_t = \int_0^t X_u \, du$,

Theorem 2.2 As $t \to \infty$,

(a) $\hat{\lambda} \to \lambda$ a.s.,

(b) the asymptotic distribution of $(\lambda S_t)^{\frac{1}{2}}(\hat{\lambda}/\lambda - 1)$ is standard normal.

We shall indicate three different methods of proof, the two first of which are applicable for general Markov branching processes.

First proof. (Keiding [51]). Define the stochastic process V_u by $S_{V_u} = u$ (for each ω, $V.(\omega)$ is the inverse function of the continuous sample function $S.(\omega)$). Then it is well-known (Athreya and Ney [8, Theorem III.11.1], Papangelou [64], Rudemo [68] that

$$Y_u = X_{V_u} - x, \quad u \geqq 0$$

is a homogeneous Poisson process with intensity λ. This having i i d increments, it follows directly from the strong law of large numbers that $Y_u/u \to \lambda$ a.s. as $n \to \infty$. But hence

$$(X_t - x)/S_t = Y_{S_t}/S_t \to \lambda$$

a.s. on $\{S_t \to \infty\}$, that is, a.s.

Moreover, by the central limit theorem $(\lambda u)^{-\frac{1}{2}}(Y_u - \lambda u)$ is asymptotically standard normal as $u \to \infty$. If we may substitute u by S_t, this proves part (b), since

$$(\lambda S_t)^{-\frac{1}{2}}(Y_{S_t} - \lambda S_t) = (\lambda S_t)^{\frac{1}{2}}(\hat{\lambda}/\lambda - 1).$$

It follows from central limit theory for sums of a random number of i i d random variables that this substitution is permissible when as here $S_t e^{-\lambda t} \to W/\lambda$ a.s., where $P\{0 < W < \infty\} = 1$, cf. e.g. Billingsley [16, Section 17].

Second proof. (Athreya and Keiding [7]). Let $0 = T_0 < T_1 < T_2 < \dots$ be the birth times of (X_t), that is,

$$X_{T_j} = x + j, \quad X_{T_j -} = x + j - 1.$$

Let $Z_i = X_{T_{i-1}}(T_i - T_{i-1})$, $Z'_t = X_{t-}(t - T_{X_t - x})$. It is then well known, cf. Athreya and Ney [8, p.127] that

$$S_t = \sum_{i=1}^{X_t - x} Z_i + Z'_t$$

where $Z_1, Z_2, \dots$ are i i d exponential with intensity λ. Therefore $(Z_1 + \dots + Z_n)/n \to \lambda^{-1}$ a.s. and hence (granted the easily shown fact that $Z'_t/X_t \to 0$ a.s.)

$$\hat{\lambda}^{-1} = S_t/(X_t - x) \to \lambda^{-1} \text{ a.s. on } \{X_t \to \infty\}.$$

Also, the asymptotic distribution of

$$\lambda n^{-\frac{1}{2}}(\sum_{i=1}^{n} Z_i - n\lambda^{-1})$$

is standard normal, and by again using random sum central limit theory it may be concluded that

$$S_t(X_t - x)^{-\frac{1}{2}}(\lambda - \hat{\lambda}) = \lambda(X_t - x)^{-\frac{1}{2}}(\sum_{i=1}^{X_t - x} Z_i + Z'_t - (X_t - x))$$

is asymptotically standard normal, from which part (b) follows by noticing that $[S_t/(X_t - x)]^{\frac{1}{2}} \xrightarrow{P} \lambda^{-\frac{1}{2}}$.

Third proof. (Keiding [49]. An approach which does not seem amenable to generalization beyond the pure birth process is the following. Let as usual $W = \lim \text{ a.s. } X_t e^{-\lambda t}$. Then the stochastic process $X_t - x$ has, conditional on $W = w$, the same distribution as an (inhomogeneous) Poisson process with intensity function $x w \lambda e^{\lambda t}$ (Kendall [56]).

The strong consistency and asymptotic normality may now be proved by direct calculation in the Poisson process. - Besides yielding these asymptotic results, this latter approach might suggest possible conditional inference procedures. We refer to Keiding [49] for further discussion.

2.3 Estimation in multitype Markov branching processes.

The simple estimators studied so far for the single-type processes generalize in an obvious way to multitype processes as long as all individual offspring may be observed, which it seems natural to assume in a genuine continuous time context. Details were given by Athreya and Keiding [7] and will not be reproduced here.

2.4 Estimation theory for more general continuous-time branching processes.

The simple occurrence/exposure rate $(X_t - x)/S_t$ is obviously strongly consistent for the Malthusian parameter α in any branching process. To elucidate its rate of convergence, Athreya and Keiding [7] remarked that any Bellman-Harris process will for large t in a certain sense look very much like a Markov branching process. Asymptotic normality was conjectured but later disproved by Asmussen and Keiding [6] who showed that in fact the rate of convergence of the age distribution becomes essential, much in the same way as was the case in the discrete-time, multitype case, cf. Section 1.6 above.

Estimation problems in cell kinetics based on the framework of the general branching process defined by Crump and Mode [20] and Jagers [43] have been studied by Jagers [45] and Jagers and Norrby [47]. Hoel and Crump [42] studied alternative estimators of the parameters of the generation time distribution in a Bellman-Harris process.

It should finally be mentioned that Brown and Hewitt [18] studied inference for the diffusion branching process.

3. THE ROLE OF MARTINGALE THEORY; EFFICIENCY CONCEPTS

3.1 The score function as a martingale

Let $X_1,\dots,X_p$ be i i d random variables from a one-dimensional exponential family with density $a(\theta)b(x)e^{\theta t(x)}$, $\theta\in\Theta$, and let $\tau(\theta) = E_\theta[t(X_i)]$. Then under suitable regularity conditions the statistical model may be parameterized by τ and the score function based on the p replications is

$$S_p(\tau) = D_\tau \log L_p = p\, i(\tau)(\overline{T} - \tau)$$

with L_p the likelihood function, $\overline{T} = \Sigma t(X_i)/p$ and $i(\tau) = -D_\tau^2 \log L_p$ the usual Fisher information in one observation. (See Barndorff-Nielsen [9] for exponential family theory). Notice that the maximum likelihood estimator $\hat{\tau}$ of τ is exactly $\overline{T}$, and that

$$\mathrm{Var}\,(S_p(\tau)) = p\, i(\tau) = [\,\mathrm{Var}\,(\hat{\tau})\,]^{-1}$$

so that each new observation adds an amount of $i(\tau)$ to the information content or "precision" of the experiment. For purposes of later generalization we also remark that the stochastic process $(S_p(\tau),\ p = 1,2,\dots)$ has i i d increments with mean zero and finite variance, and hence is a square integrable (local) martingale. The Doob decomposition of $S_p^2(\tau)$ is $S_p^2(\tau) = I_p(\tau) + \text{martingale}$, where

$$I_p(\tau) = \sum_{j=0}^{p-1} \mathrm{Var}\,(S_{j+1}(\tau) \mid X_1,\dots,X_j) = p\, i(\tau).$$

In this sense the increasing process corresponding to $S_p^2(\tau)$ exactly measures the information content.

For generalizations of these ideas to the branching process context it is useful to concentrate on linear discrete one-parameter exponential families, also called power series families, where the canonical statistic $t(X) = X$. As just stated, such a family may be parameterized by the mean m. It now follows from the discussion in Section 1.1 that if the offspring distribution is assumed to belong to a power series family, then the score function $S_n(m)$ corresponding to observation of the first $n+1$ generation sizes $Z_0,\dots,Z_n$ of a Galton-Watson process is obtained by replacing p by $Z_0+\dots+Z_{n-1}$, so that

$$S_n(m) = (Z_0+\dots+Z_{n-1})\, i(m)(\hat{m}-m),$$

where, as in Chapter 1, $\hat{m} = (Z_1+\dots+Z_n)/(Z_0+\dots+Z_{n-1})$.

It is quite generally true and easily directly verified in this particular case that $(S_n(m),\ n = 0,1,\dots)$ is a square integrable (local) martingale with respect to the "self-exciting" family of σ-algebras $(\mathcal{N}_n = \sigma\{Z_0,\dots,Z_n\},\ n = 0,1,2,\dots)$.

The Doob decomposition of the submartingale $(S_n^2(m), \mathcal{N}_n)$ is given by

$$S_n^2(m) = I_n(m) + \text{martingale},$$

where

$$I_n(m) = \sum_{j=0}^{n-1} [E(S_{j+1}^2(m) \mid \mathcal{N}_j) - S_j^2(m)]$$

is increasing. We may alternatively write

$$I_n(m) = \sum_{j=0}^{n-1} E[(S_{j+1}(m) - S_j(m))^2 \mid \mathcal{N}_j]$$

where the individual term in our case is

$$E[\{i(m)(Z_{j+1} - mZ_j)\}^2 \mid \mathcal{N}_j] = i(m)^2 \operatorname{Var}(Z_{j+1} \mid \mathcal{N}_j) = Z_j\, i(m)$$

so that we have

$$I_n(m) = (Z_0+\dots+Z_{n-1})\, i(m) \tag{1.1}$$

and the score function may be written

(1.2) $S_n(m) = I_n(m)(\hat{m} - m)$.

This very simple form led Heyde [35] and Heyde and Feigin [36] to several definitions and generalizations, interpreting quite generally the increasing process $I_n(m)$ as a measure of information content. The form

$$I_n(\tau) = \sum_{j=0}^{n-1} \mathrm{Var}\,(S_{j+1}(\tau) \mid N_j)$$

illustrates that I_n successively accumulates whatever new information there is in taking an extra observation, given the previous samples.

When as in the present branching process context (cf. (1.1)) the information function $I_n(m)$ factorises into a product of a deterministic factor and an observable quantity, Heyde and Feigin termed the statistical model a "conditional exponential family". We notice that this concept generalizes the role of <u>mean value parameterized</u> exponential families in the i i d theory.

These ideas carry over to continuous time, with the Doob-Meyer decomposition of the squared score function as a key tool, see Feigin [27]. For the Yule process studied in Section 2.2 above, the information content becomes $I_t(\lambda) = S_t/\lambda$ with $S_t = \int_0^t X_u \, du$. A general framework for statistical models for point processes whose conditional intensities factorize into a product of a deterministic function and an observable process was given by Aalen [1,2,3] who also leaned heavily on martingale theory.

Turning next to asymptotic theory, we demonstrated in the previous chapters how it is possible to reduce all asymptotic results for one-type processes to the basic i i d structure inherent in branching processes. The analysis by Asmussen and Keiding [6] indicated, however, that already for the multitype Galton-Watson process, this may no longer be feasible. At least as a starting point for a methodology for more general stochastic process estimation problems, it is useful to point out that a martingale central limit theorem with random norming factors will yield asymptotic normality of

$$I_n^{-\frac{1}{2}}(m)S_n(m) = I_n^{\frac{1}{2}}(m)(\hat{m}-m)$$

which is equivalent to Theorem 1.2.3 (a) for the particular processes here discussed. Of course consistency is even more directly shown. The martingale approach is based only on the rather generally valid martingale property of the score function and will thus have considerably wider applicability.

3.2 Efficiency concepts.

Heyde [35] proposed to call a consistent estimator T_n of θ asymptotically efficient if there exists a (deterministic) function $\beta(\theta)$ such that

$$I_n^{\frac{1}{2}}(\theta)[T_n - \theta - \beta(\theta)S_n(\theta)/I_n(\theta)] \xrightarrow{P} 0$$

as $n \to \infty$. This definition reduces to that of Rao [67, Sec. 5c] in the i i d case. Heyde proved that the maximum likelihood estimator of the offspring mean in power series families (including the geometric distributions discussed above) is asymptotically efficient in this sense.

A complication with the use of the maximum likelihood estimator here is that it is not sufficient. Indeed, from (1.2) it is seen that since $I_n(m)$ is random, the two-dimensional statistic $(\hat{m}, Z_0+\ldots+Z_{n-1})$ is minimal sufficient. Generalizing an argument due to Rao on the local behaviour of the power function near the null hypothesis Basawa and Scott [10] pointed out that asymptotically, hypothesis testing based on the maximum likelihood estimator alone is inferior to tests based on the full sufficient statistic. A different analysis by Feigin [29] based on the concept of contiguity raised some doubts about Basawa and Scott's approach.

A general framework for statistical models with one "curved" (nonlinear) submodels of exponential families was provided by Efron [25] who coined the term "statistical curvature" for a quantity entering into the asymptotic distributions of estimators and test statistics. In the discussion of Efron's paper, Keiding [52] mentioned the possible role of this concept in branching process situations, and further work along these lines, so far unpublished, has been done by I.V. Basawa.

4. APPLICATIONS

The area of statistical inference for branching processes is closely related to classical life testing, certain problems in reliability, and more general statistical problems for point processes, all of which are very important for the applications. However, it is not yet clear how far branching process models such as those discussed in this paper will be directly applicable as statistical models. Although a branching process may be a useful description of the underlying stochastic phenomena, the important statistical problems might often derive from incomplete observation, random temporal variation of parameters, or plain measurement uncertainty, and none of these sources of random variation are included in the present theory (save for the preliminary results concerning BPRE and quoted in Section 1.7). Furthermore, the number of replications will presumably often be so large that the intrinsic random variation discussed in this paper will be negligible.

We have already mentioned some specific applications of methods like the ones discussed here: the estimation of the age of rare human blood types (Section 1.8) and various questions in cell kinetics (Section 2.4). For an example where a branching process model leads to quite classical statistical problems, see Gani and Saunders [30].

We conclude the paper by quoting two further applications of the present theory that we have come across.

4.1 Population dynamics of the whooping crane population of North America

The whooping crane is an extremely rare migratory bird with breeding area in western Canada. Miller et al. [59] reported the annual counts from 1938-1972 of whooping cranes arriving to the wintering grounds in Texas. The birds born the previous spring have a different plumage (they are referred to as "young") and it is therefore possible to obtain the annual number of births and deaths. The population has increased from 14 individuals in 1938 to 57 in 1970. Miller et al. fitted a simple linear birth-death process and used the estimated values to produce prediction intervals concerning future population sizes. A critical discussion and suggestions of other possible models were provided by Keiding [53] and a time series approach is due to Kashyap and Rao [48, p. 296].

4.2 Epidemics: early evaluation of whether an epidemic is minor or major.

In a series of papers Becker [11,12,13] has developed and applied aspects of the theory of statistical inference for branching processes to estimate the initial infection rate of an epidemic. The method is to approximate the number of infectives in a stochastic epidemic process by a suitable Galton-Watson process, and then estimate the offspring mean of the latter. It is also possible to study the impact of vaccination by assuming that a fixed known proportion of the population in question is immunized.

Becker first [11] considered the application of the simple Galton-Watson process to the 49 smallpox epidemics in Europe between 1950 and 1970. These epidemics were all minor (became extinct) but the estimated mean offspring size was not significantly less than one which may raise the question whether major smallpox epidemics were only avoided by luck.

The approximation by a Galton-Watson process involves several assumptions. First, successive "generations" of infectives should be reasonably discrete, that is, there should be a fairly long latent period and a short infective period. Second, the number of susceptibles should be so large that it does not decrease appreciably when individuals get infected. As an example of a situation where this assumption clearly does not hold Becker [12] considered an observed (major) smallpox epidemic in a small, closed Nigerian community. To implement the theory for this small population, he approximated the number of infectives with a size-dependent Galton-Watson process and used a least squares approach to derive an estimator of the initial infection rate.

A third assumption is that of a homogeneously mixing population. It was mentioned in Sections 1.6 and 1.7 above that Becker [13] studied the fate of several estimators of the offspring mean on the assumption that the process is in fact multitype Galton-Watson or develops in random environments.

ACKNOWLEDGMENTS

We are grateful to Paul Feigin for critical comments and to several authors for providing us with prepublication versions of their work.

REFERENCES

[1] Aalen, O. O. Statistical inference for a family of counting processes. Ph. D. Dissertation, Department of Statistics, University of California, Berkeley, 1975.

[2] Aalen, O. O. Weak convergence of stochastic integrals related to counting processes. Z. Wahrscheinlichkeitsth. verw. Geb. 38, 261-277 (1977).

[3] Aalen, O. O. Nonparametric inference for a family of counting processes. Ann. Statist. (to appear 1978).

[4] Albert, A. Estimating the infinitesimal generator of a continuous time, finite state Markov process. Ann. Math. Statist. 33, 727-753 (1962).

[5] Asmussen, S. Almost sure behavior of linear functionals of supercritical branching processes. *Trans. Amer. Math. Soc*. 231, 233-248 (1977).

[6] Asmussen, S. & Keiding, N. Martingale central limit theorems and asymptotic estimation theory for multitype branching processes. *Adv. Appl. Prob*. 10 (to appear 1978).

[7] Athreya, K.B. & Keiding, N. Estimation theory for continuous-time branching processes. *Sankhyā Ser. A* (to appear 1976).

[8] Athreya, K.B. & Ney, P.E. *Branching processes*. Springer, Berlin, 1972.

[9] Barndorff-Nielsen, O. *Information and exponential families in statistical theory*. Wiley, London, 1978.

[10] Basawa, I.V. & Scott, D.J. Efficient tests for branching processes. *Biometrika* 63, 531-536 (1976).

[11] Becker, N. On parametric estimation for mortal branching processes. *Biometrika* 61, 393-399 (1974).

[12] Becker, N. Estimation for an epidemic model. *Biometrics* 32, 769-777 (1976).

[13] Becker, N. Estimation for discrete time branching processes with application to epidemics. *Biometrics* 33, 515-522 (1977).

[14] Beyer, J.E., Keiding, N. & Simonsen, W. The exact behaviour of the maximum likelihood estimator in the pure birth process and the pure death process. *Scand. J. Statist*. 3, 61-72 (1976).

[15] Billingsley, P. *Statistical inference for Markov processes*. University of Chicago Press, 1961.

[16] Billingsley, P. *Convergence of probability measures*. Wiley, New York, 1968.

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

[18] Brown, B.M. & Hewitt, J.I. Inference for the diffusion branching process. *J. Appl. Prob*. 12, 588-594 (1975).

[19] Crump, K.S. & Howe, R.B. Nonparametric estimation of the age of a Galton-Watson branching process. *Biometrika* 59, 533-538 (1972).

[20] Crump, K.S. & Mode, C.J. A general age-dependent branching process. *J. Math. Anal. Appl*. 24, 494-508 (1968).

[21] Dion, J.-P. *Estimation des probabilités initiales et de la moyenne d'un processus de Galton-Watson*. Thèse de doctorat, Université de Montreal, 1972.

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

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

[24] Dion, J.-P. & Esty, W.W. Estimation problems in branching processes with random environments. Manuscript, 1977.

[25] Efron, B. Defining the curvature of a statistical problem (with applications to second order efficiency). (with discussion). Ann. Statist. 3, 1189-1242 (1975).

[26] Eschenbach, W. & Winkler, W. Maximum-Likelihood-Schätzungen beim Verzweigungsprozess von Galton-Watson. Math. Operationsforsch. u. Statist. 6, 213-224 (1975).

[27] Feigin, P.D. Maximum likelihood estimation for continuous-time stochastic processes. Adv. Appl. Prob. 8, 712-736 (1976).

[28] Feigin, P.D. A note on maximum likelihood estimation for simple branching processes. Austr. J. Statist. (to appear 1977).

[29] Feigin, P.D. The efficiency criteria problem for stochastic processes. Stoch. Proc. Appl. (to appear 1977).

[30] Gani, J. & Saunders, I.W. Fitting a model to the growth of yeast colonies. Biometrics 33, 113-120 (1977).

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

[32] Harris, T.E. The theory of branching processes. Springer, Berlin, 1963.

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

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

[35] Heyde, C.C. Remarks on efficiency in estimation for branching processes. Biometrika 62, 49-55 (1975).

[36] Heyde, C.C. & Feigin, P.D. On efficiency and exponential families in stochastic process estimation. In: Statistical distributions in scientific work (ed. G.P. Patil, S. Kotz, J.K. Ord) Reidel, Dordrecht Vol. 1, 227-240, 1975.

[37] Heyde, C.C. & Leslie, J.R. Improved classical limit analogues for Galton-Watson processes with or without immigration. Bull. Austr. Math. Soc. 5, 145-156 (1971).

[38] Heyde, C.C. & Scott, D.J. Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. Ann. Prob. 1, 428-436 (1973).

[39] Heyde, C.C. & Seneta, E. Analogues of classical limit theorems for the supercritical Galton-Watson process with immigration. Math. Biosc. 11, 249-259 (1971).

[40] Heyde, C.C. & Seneta, E. Estimation theory for growth and immigration rates in a multiplicative process. J. Appl. Prob. 9, 235-256 (1972).

[41] Heyde, C.C. & Seneta, E. Notes on "Estimation theory for growth and immigration rates in a multiplicative process". J. Appl. Prob. 11, 572-577 (1974).

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

[43] Jagers, P. A general stochastic model for population development. Skand. Akt. 1969, 84-103 (1969).

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

[45] Jagers, P. Maximum likelihood estimation of the reproduction distribution in branching processes and the extent of disintegration in cell proliferation. Tech. rep. 1973-17, Dep. of Math., Chalmers Univ. of Techn. and Univ. Göteborg, 1973.

[46] Jagers, P. Branching processes with biological applications. Wiley, New York, 1975.

[47] Jagers, P. & Norrby, K. Estimation of the mean and variance of cycle times in cinemicrographically recorded cell populations during balanced exponential growth. Cell Tissue Kinet. 7, 201-211 (1974).

[48] Kashyap, R.L. & Rao, A.R. Dynamic stochastic models from empirical data. Academic Press, New York, 1976.

[49] Keiding, N. Estimation in the birth process. Biometrika 61, 71-80 and 647 (1974).

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

[51] Keiding, N. Maximum likelihood estimation in the birth-and-death process. Ann. Statist. 3, 363-372, (1975). Correction note Vol. 6, No. 2 (1978).

[52] Keiding, N. Contribution to the discussion of Efron [25].

[53] Keiding, N. Population growth and branching processes in random environments. Proc. 9 Int. Biom. Conf. Inv. Papers II, 149-165 (1976).

[54] Keiding, N. & Lauritzen, S. Marginal maximum likelihood estimates and estimation of the offspring mean in a branching process. Scand. J. Statist. 5 (to appear 1978).

[55] Kendall, D.G. Les processus stochastiques de croissance en biologie. Ann. Inst. H. Poincaré 13, 43-108 (1952).

[56] Kendall, D.G. Branching processes since 1873. J. London Math. Soc. 41, 385-406 (1966).

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

[58] Lotka, A.J. Théorie analytique des associations biologiques. Actualités Sci. Indust. 2, no. 780, 123-136 (1939).

[59] Miller, R.S. Botkin, D.B. & Mendelssohn, R. The whooping crane (Grus americana) population of North America. Biol. Conserv. 6, 106-111 (1974).

[60] Moran, P.A.P. Estimation methods for evolutive processes. J. Roy. Statist. Soc. B 13, 141-146 (1951).

[61] Moran, P.A.P. The estimation of parameters of a birth and death process. J. Roy. Statist. Soc. B 15, 241-245 (1953).

[62] Nagaev, A.V. On estimating the expected number of direct descendants of a particle in a branching process. Theor. Prob. Appl. 12, 314-320 (1967).

[63] Pakes, A.G. Nonparametric estimation in the Galton-Watson process. Math. Biosc. 26, 1-18 (1975).

[64] Papangelou, F. Integrability of expected increments of point processes and a related change of scale. Trans. Amer. Math. Soc. 165, 483-506 (1972).

[65] Quine, M.P. Asymptotic results for estimators in a subcritical branching process with immigration. Ann. Prob. 4, 319-325 (1976) Corr. Note Vol. 5, 318 (1977).

[66] Quine, M.P. & Durham P. Estimation for multitype branching processes. Manuscript, University of Sydney, 1977.

[67] Rao, C.R. Linear Statistical Inference and its applications. Wiley, New York, 1965.

[68] Rudemo, M. On a random transformation of a point process to a Poisson process. In: Jagers, P. & Råde, L. (ed.) Mathematics and Statistics. Essays in honour of Harald Bergström. Göteborg, pp. 79-85, 1973.

[69] Stigler, S.M. Estimating the age of a Galton-Watson branching process. Biometrika 57, 505-512 (1970).

[70] Thompson, E.A. Estimation of age and rate of increase of rare variants. Am. J. Hum. Genet. 28, 442-452 (1976).

[71] Thompson, E.A. Estimation of the characteristics of rare variants. In: Measuring selection in natural populations. (F.B. Christiansen and T.M. Fenchel, ed.) Springer Lecture Notes in Biomathematics 19, 531-543 (1977).

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

Statistical Inference in Branching Processes

Jean-Pierre Dion and Niels Keiding

INDEX

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.

No. 2 Jacobsen, Martin: Stochastic Processes with Stationary Increments in Time and Space.

No. 3 Johansen, Søren: Product Integrals and Markov Processes.

No. 4 Keiding, Niels & Lauritzen, Steffen L. : Maximum likelihood estimation of the offspring mean in a simple branching process.

No. 5 Hering, Heinrich: Multitype Branching Diffusions.

No. 6 Aalen, Odd & Johansen, Søren: An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations.

No. 7 Johansen, Søren: The Product Limit Estimator as Maximum Likelihood Estimator.

No. 8 Aalen, Odd & Keiding, Niels & Thormann, Jens: Interaction Between Life History Events.

No. 9 Asmussen, Søren & Kurtz, Thomas G.: Necessary and Sufficient Conditions for Complete Convergence in the Law of Large Numbers.

No. 10 Dion, Jean-Pierre & Keiding, Niels: Statistical Inference in Branching Processes.