K. B. Athreya
Niels Keiding

Estimation Theory for Continuous-Time Branching Processes

Preprint
March
1975
6

Institute of Mathematical Statistics
University of Copenhagen

K. B. Athreya[*] and Niels Keiding

ESTIMATION THEORY FOR CONTINUOUS-TIME BRANCHING PROCESSES

Preprint 1975 No. 6

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

March 1975

[*] Department of Mathematics, Indian Institute of Science, Bangalore, India

ABSTRACT

Results are given on estimation theory for some continuous-time branching processes assumed observed continuously in a fixed time interval [0,t]. Maximum likelihood theory works without problems for Markov branching processes. For Bellman-Harris (age-dependent) processes only the offspring distribution is easily estimable.

An "asymptotic Markovian property" of Bellman-Harris processes is observed. This property, which may have some independent interest, is then used as motivation for a study of a simple occurrence/exposure estimator for the Malthusian parameter.

Some remarks are also given on previous results, in particular concerning cell growth studies.

1. INTRODUCTION

This paper presents some results on estimation theory for continuous-time branching processes, cf. Harris (1963, chap. V and VI) or Athreya and Ney (1972, chap. III-V), assuming that a full record of the population size X_u in some fixed interval [0,t] is available. This sampling situation was also assumed in the basic work on estimation in Markov chains with continuous time by Albert (1962) and Billingsley (1961 a,b). In fact, the present work may be seen as a partial answer to the request made by Billingsley (1961a) in his closing remarks for a systematic investigation of statistical inference for non-recurrent processes.

The main part of the paper is divided into three Sections. Section 2 reports results for the one-dimensional Markov branching process, where explicit maximum likelihood estimators are derived and their limiting sampling properties studied for large t as well as for large initial population sizes. Similar results hold for the multitype Markov branching process, assuming that the vector $\underset{\sim}{X}_u$ of different types is observed continuously in [0,t]. These results are presented in Section 3. It is a common feature that the asymptotic behaviour for large t follows the extinction/explosion dichotomy of the process. Also the asymptotic normality requires a random normalising factor. The results may be transformed into statements with deterministic normalising factors, but then the asymptotic distributions are no longer normal, but mixtures of normals with zero means and random variances inversely proportional to $W = \lim_{t\to\infty} \text{a.s. } X_t e^{-\alpha t}$.

For the Bellman-Harris (or age-dependent) branching process that we study in Section 4, explicit maximum likelihood results exist only as regards the offspring distribution. A simple example indicates the difficulties of maximum likelihood methods for estimating the Malthusian parameter α. We then show that the sample functions of the population size of any Bellman-Harris process are asymptotically much like those of a corresponding Markov branching process in the sense that any finite number of times between adjacent discontinuities multiplied by the current population sizes are asymptotically independent, identically distributed exponentials, just as is the case in the Markov branching process. This observation, which may have some independent interest, leads us to the study of the simple occurrence-exposure estimator $\tilde{\alpha} = (X_t - x_0)/\int_0^t X_u du$ for α, $\tilde{\alpha}$ being the maximum likelihood estimator in the Markovian case.

Previous work on estimation theory for branching processes in continuous time under the proposed observation scheme is rather limited. Keiding (1974,1975) and Beyer, Keiding and Simonsen (1975) considered the simple (linear) birth-and-death process and gave further references. Keiding also considered estimation theory in the situation where only a discrete skeleton $\{X_{i\tau}, i = 1,...,n\}$ is observed, and discussed the random variation "due to W" as mentioned above and its implication for principles of statistical inference as applied to this particular problem.

Hoel and Crump (1974) proposed some estimators for the parameters of a multiphase birth process, that is a binary splitting Bellman-Harris process with gamma distributed life lengths. These estimators were also based on observation of only a few values of the process, and we give some comparisons with complete record in Section 4.

Estimation of the offspring distribution of a Bellman-Harris process has been considered by Jagers (1973b), who gave the maximum likelihood estimators based on observation of the full family tree without assumptions on the form of the offspring distribution. Jagers applied the results to problems of tumour cell growth. Becker (1974) assumed that observation is continued until extinction and gave results for one-parameter exponential family offspring distributions with application to smallpox epidemics.

The mathematical tools used in this paper are heavily based on the recent monograph by Athreya and Ney (1972) and a recent result by Athreya and Kaplan (1975) on the almost sure convergence of the relative age distribution for supercritical Bellman-Harris processes.

2. ONE-DIMENSIONAL MARKOV BRANCHING PROCESSES

Let $\{X_t, t \geq 0\}$ be a Markov branching process with split intensity $\lambda > 0$ and offspring distribution $\{p_i, i \geq 0\}$ where without loss of generality we assume $p_1 = 0$. The process was discussed by Harris (1963, Chapter V) and Athreya and Ney (1972, Chapter III) and is a time-homogeneous Markov process with state space $\{0,1,2,...\}$ and infinitesimal transition probabilities

$$P\{X_{t+h} = j | X_t = i\} = \begin{cases} i\lambda p_{j-i+1}h + o(h), & j = i-1,\ i+1,\ i+2,... \\ 1 - i\lambda h + o(h), & j = i \\ o(h) & \text{otherwise.} \end{cases}$$

We shall assume $X_0 = x_0$ degenerate throughout.

a. The likelihood function and maximum likelihood estimators.

Assume that the process has been observed continuously over a fixed time interval $[0,t]$. Albert (1962) constructed for the case of a continuous-time, finite state Markov process a measure dominating the probabilities of realizations of the process in $[0,t]$ and obtained the likelihood function as the Radon-Nikodym derivative with respect to this measure. It is readily seen that Albert's derivation of the likelihood function may be generalised to countable-state Markov processes which with probability one have only finitely many transitions in any finite interval. A sufficient condition for this to hold for a Markov branching process is (Harris 1963, p. 107) that the mean $m = \Sigma ip_i$ of the offspring distribution be finite, which will be assumed throughout. The following Theorem is now an easy consequence of Albert's results cf. also Billingsley (1961b).

Theorem 2.1 There exists a measure on the space of realizations $\{X_u | 0 \leq u \leq t\}$ such that the Radon-Nikodym derivative, that is, the likelihood function, is

$$L(\lambda, p_0, p_2, p_3, \ldots) = \lambda^{N_t} e^{-\lambda S_t} \prod_{i=0,2}^{\infty} p_i^{N_t(i)}$$

where $N_t(i) = \# \{u \in [0,t] | X_u - X_{u-} = i - 1\}$ is the number of transitions (or splits) of size $i - 1$ in $[0,t]$, $N_t = \sum_{i=0,2}^{\infty} N_t(i)$ is the total number of splits, and

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

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

Further, S_t and $\{N_t(i) | i = 0,2,3,\ldots\}$ are jointly sufficient, and if the model is parameterized by $(\lambda, p_0, p_2, \ldots)$, $\lambda > 0$, $0 \leq p_i \leq 1$, $\Sigma p_i = 1$, they are minimally sufficient. In the latter case the maximum likelihood estimators are given when $N_t > 0$ as

$$\hat{\lambda} = N_t / S_t, \quad \hat{p}_i = N_t(i)/N_t.$$

When $N_t = 0$ they are undefined, although a natural extension is $\hat{\lambda} = 0$.

For use in the following study we need the expectations of the sufficient statistics, which are given in the Lemma below.

Lemma 2.1 Assume $X_0 = x_0$ and let $\alpha = \lambda(m-1)$ be the Malthusian parameter. Then

$$E(X_t) = x_0 e^{\alpha t}$$

$$E(S_t) = \begin{cases} x_0(e^{\alpha t} - 1)/\alpha & \text{if } \alpha \neq 0, \\ x_0 t & \text{if } \alpha = 0, \end{cases}$$

$$E(N_t) = \begin{cases} x_0(e^{\alpha t} - 1)/(m-1) & \text{if } m \neq 1, \\ x_0 \lambda t & \text{if } m = 1 \end{cases}$$

and

$$E\{N_t(i)\} = p_i E(N_t).$$

Proof. The expectation $E(X_t)$ is well-known and $E(S_t) = E(\int_0^t X_u du) = \int_0^t E(X_u)du$. Finally $E(N_t)$ and $E\{N_t(i)\}$ may be obtained from the integral equations that they satisfy (Athreya and Karlin (1967, p. 270)). Alternatively, a Wald-identity $E(X_t) - x_0 = E(N_t)(m-1)$ may be derived by using the optional sampling theorem on the martingale $X(\tau_n) - x_0 - n(m-1)$, stopped at the random time N_t, where τ_n is the n'th split time. However, this only gives the result for $m \neq 1$.

b. Asymptotic results for large populations.

We shall in this paper be concerned mainly with the asymptotic properties of the estimators as $t \to \infty$. However, for the sake of completeness it may be remarked that since a branching process with $X_0 = x_0$ may be interpreted as the sum of x_0 independent processes with the same parameters and $x_0 = 1$, the following asymptotic results for large X_0 and fixed t may be obtained from standard maximum likelihood theory.

Theorem 2.2 *As* $x_0 \to \infty$, $(\hat{\lambda}, \hat{p}_0, \hat{p}_2, \ldots) \to (\lambda, p_0, p_2, \ldots)$ a.s. *Let* $\mu_i = \lambda p_i$, *then* $\hat{\mu}_i = \hat{\lambda}\hat{p}_i$. *The joint distribution of*

$$\left(\frac{x_0(e^{\alpha t}-1)}{\alpha}\right)^{1/2} \left(\hat{\mu}_0 - \mu_0, \hat{\mu}_2 - \mu_2, \hat{\mu}_3 - \mu_3, \ldots\right)$$

converges towards independent normals with means 0 *and variances* $\mu_0, \mu_2, \mu_3, \ldots$ (*All results have to be modified in the obvious way when* $\alpha = 0$).

Proof. By the strong law of large numbers and Lemma 2.1 it follows that as $x_0 \to \infty$, $N_t(i)/x_0 \to p_i(e^{\alpha t} - 1)/(m-1)$, $N_t/x_0 \to (e^{\alpha t}-1)/(m-1)$ and $S_t/x_0 \to (e^{\alpha t}-1)/\alpha$ almost surely, thus proving the consistency. The asymptotic normality follows from standard theory by considering an arbitrary finite subset of $\hat{\mu}_i$'s and computing the information matrix. Now we may write

$$\log L = \Sigma N_t(i) \log \mu_i - \Sigma \mu_i S_t$$

so that

$$D^2_{\mu_i} \log L = -N_t(i)/\mu_i^2, \quad D^2_{\mu_i\mu_j} = 0 \text{ for } i \neq j,$$

and thus the information matrix has 0's outside the diagonal and diagonal elements

$$-E(D^2_{\mu_i} \log L) = x_0(e^{\alpha t} - 1)/\alpha\mu_i.$$

c. Asymptotic results for large periods of observation.

In the subcritical and critical cases ($m \leq 1$), it is well known that $X_t \to 0$ a.s. as $t \to \infty$. In the supercritical case this happens with probability $q^{x_0} < 1$, where q is the smallest nonnegative solution of the equation $q = \Sigma p_i q^i$. Then the estimators will have almost sure, nondegenerate limits, because in effect only a finite sample was ever taken. It is possible to spell out various aspects of the distribution of these variables but it is not essential to our present purpose and we omit the details here. We finally notice that these remarks are also relevant for the conditional supercritical process given ultimate extinction (cf. Keiding (1975, Section 5)). It is obvious that accurate inferences in these cases will have to be based on large initial population sizes, that is, replicated experiments, cf. Becker (1974).

In the supercritical case $m > 1$, $X_t \to \infty$ with probability $1 - q^{x_0}$. The following consistency results then hold. We assume $x_0 = 1$ in the rest of this Subsection.

Theorem 2.3. *As $t \to \infty$, with probability one on the set $\{X_t \to \infty\}$, $\hat{\lambda} \to \lambda$ and $\hat{p}_i \to p_i$, $i = 0,2,3,\ldots$.*

Proof. Let $0 = \tau_0, \tau_1, \tau_2, \ldots$ be the split times and define $Y_i = X_{\tau_{i-1}}(\tau_i - \tau_{i-1})$, $Y_t' = X_{t-}(t - \tau_{N_t})$. Then (cf. Athreya and Ney (1972, p. 127)),

$$S_t = \sum_{i=1}^{N_t} Y_i + Y_t'$$

where $Y_1, Y_2, \ldots$ are independent and exponential with intensity λ. It follows from Lemma 2.2 below that $Y_t'/N_t \to 0$ a.s. on $\{X_t \to \infty\}$. By the strong law of large numbers

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \to \lambda^{-1}$$

almost surely, and it follows that on the set $\{X_t \to \infty\}$, where $N_t \to \infty$ a.s., $S_t/N_t = \hat{\lambda}^{-1} \to \lambda^{-1}$ a.s.

To show the second result we use the representation

$$N_t(i) = \sum_{j=1}^{N_t} Z_j(i),$$

where $Z_j(i)$ is the indicator random variable for the event that the j'th split resulted in a jump of size i-1. Clearly, $Z_1(i), Z_2(i)$ are independent and identically distributed with $P\{Z_j(i) = 1\} = p_i$. Then use the strong law of large numbers once more.

Lemma 2.2 *Let* $Y_t' = X_{\tau_{N_t}}(t - \tau_{N_t})$ *as defined in the proof of Theorem 2.3. Then* $Y_t'/N_t^{1/2} \to 0$ a.s. *on* $\{X_t \to \infty\}$.

Proof. Obviously,

$$Y_t' \leqq X_{\tau_{N_t}}(\tau_{N_t+1} - \tau_{N_t}).$$

Now $Y_n = X_{\tau_{n-1}}(\tau_n - \tau_{n-1})$ is exponential with mean λ^{-1}, and so

$$P\{Y_n n^{-1/2} > \varepsilon\} = P\{Y_n > \varepsilon n^{1/2}\} \leqq E(Y_n^4)/(n^2\varepsilon^4) \leqq 24/(\lambda^4 \varepsilon^4 n^2).$$

By the Borel-Cantelli lemma $Y_n/n^{1/2} \to 0$ a.s. Hence also

$$Y_t'/N_t^{1/2} \leqq Y_{N_t+1}/N_t^{1/2} \to 0$$

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

<u>Theorem 2.4.</u> <u>Let again</u> $\mu_i = \lambda p_i$, <u>then</u> $\hat{\mu}_i = N_t(i)/S_t$. <u>As</u> $t \to \infty$, <u>the donditional distribution, given</u> $\{X_t \to \infty\}$ <u>(or</u> $\{X_t > 0\}$) <u>of</u>

$$S_t^{1/2}\begin{pmatrix} \hat{\mu}_0 - \mu_0 \\ \hat{\mu}_2 - \mu_2 \\ \hat{\mu}_3 - \mu_3 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

<u>converges to a set of independent normals with parameters</u> $(0,\mu_i)$, $i = 0,2,3,\ldots$.

<u>Proof.</u> We study an arbitrary finite subset of random variables

$$S_t N_t^{-1/2}(\hat{\mu}_i - \mu_i)$$

$$= N_t^{-1/2}\{N_t(i) - \mu_i S_t\}$$

$$= N_t^{-1/2} \sum_{j=1}^{N_t} [Z_j(i) - \mu_i X_{\tau_{j-1}}(\tau_j - \tau_{j-1})] - N_t^{-1/2}\mu_i X_t(t-\tau_{N_t}) \; .$$

First notice that by Lemma 2.2, the last term goes to 0 a.s. on $\{X_t \to \infty\}$.

The $Z_j(i)$ are the same as in the proof of Theorem 2.3 above, and $0 = \tau_0 < \tau_1 < \ldots$ are the split times. The representation is obvious from the minimal construction of the process, cf. Athreya and Ney (1972, p. 119 and 127). Generalising their analysis, it is now seen that the random infinite dimensional vectors $\underset{\sim}{U}_j = \{U_j(i)\}$, $U_j(i) = Z_j(i) - \mu_i X_{\tau_{j-1}}(\tau_j - \tau_{j-1})$, $j = 1,2,\ldots$ are independent and identically distributed. In fact, let for each j $\acute{F}_j$ denote the σ-algebra associated with the stopping time τ_j. Then the conditional distribution of $\underset{\sim}{U}_{j+1}$ given $\acute{F}_j$ is that of $\underset{\sim}{A} - \underset{\sim}{B}$, where $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are independent, $\underset{\sim}{A} = \{A(i)\}$ is a vector of indicators with probabilities p_i and $\underset{\sim}{B} = (\mu_0 C, \mu_2 C, \mu_3 C, \ldots)$ where C is exponential with mean λ^{-1}. Hence $\underset{\sim}{U}_{j+1}$ is independent of F_j, and the independence of all $\underset{\sim}{U}_j$ follows by induction. We also notice that $E\{U_j(i)\} = p_i - p_i = 0$, $\mathrm{Var}\{U_j(i)\} = p_i(1-p_i) + \mu_i^2/\lambda^2 = p_i$, and $\mathrm{Cov}\{U_j(i), U_j(k)\} = E\{U_j(i)U_j(k)\} = E\{A(i)A(k)\} - E\{A(i)\mu_k C\} - E\{A(k)\mu_i C\} + E(\mu_i\mu_k C^2) = 0 - p_i p_k - p_k p_i + \mu_i\mu_k(\lambda^{-2} + \lambda^{-2}) = 0$. It then follows by the central limit theorem that as $n \to \infty$, the distribution of

$$n^{-1/2} \sum_{j=1}^{n} \underset{\sim}{U}_j$$

converges to that of a vector of independent normals $(0, p_i)$. Now it is known (Athreya and Ney (1972, p. 113)) that as $t \to \infty$, it is always possible to find a set $\{c_t\}$ of normalising constants such that $X_t/c_t \to W$ a.s., where $P\{W > 0\} = P\{X_t \to \infty\}$. Since

$$X_t/N_t = \frac{1}{N_t} \sum_{i=1}^{N_t} Y_i$$

where $Y_1 + 1$, $Y_2 + 1, \ldots$ are independent and all distributed according to $\{p_j\}$, it is concluded, using the strong law of large numbers, that $X_t/N_t \to m - 1$ a.s. on $\{X_t \to \infty\}$, which is a.s. the same set as $\{N_t \to \infty\}$. Therefore $N_t/c_t \to W/(m-1)$ a.s. on $\{N_t \to \infty\}$. We may now use an analogue of the central limit theorem for sums of a random number of independent random variables as stated by Billingsley (1968, Theorem 17.2) and modified by Dion (1972,1974) and Jagers (1973a) to conclude the asymptotic normality of $S_t N_t^{-1/2}(\hat{\underset{\sim}{\mu}} - \underset{\sim}{\mu})$, $\underset{\sim}{\mu} = (\mu_0, \mu_2, \mu_3, \ldots)$. The Theorem follows by Slutsky's theorem since $N_t/S_t \to \lambda$ a.s. on $\{N_t \to \infty\}$ by Theorem 2.3.

<u>Remark</u>. It also follows from the above mentioned theorem on random sums that the asymptotic normal distribution is independent of W. This may be used to modify the Theorem to yield results on the deterministic normalising factors $c_t^{1/2}$,

cf. Keiding's (1974, 1975) analysis of the birth process and birth-and-death process. As an example, if $\Sigma p_j \, j \log j < \infty$, $c_t = e^{\alpha t}$, (Athreya and Ney (1972, p. 111-112)) and then the conditional distribution of

$$\{(e^{\alpha t}-1)/\alpha\}^{1/2} \begin{pmatrix} \hat{\mu}_0 - \mu_0 \\ \hat{\mu}_2 - \mu_2 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

given $\{X_t \to \infty\}$ (or $\{X_t > 0\}$) is asymptotically that of $\underset{\sim}{T}/W^{1/2}$, where $\underset{\sim}{T}$ is the set of independent normals referred to in the Theorem and $\underset{\sim}{T}$ and W are independent.

Any individual component $\{(e^{\alpha t}-1)/\alpha\}^{1/2}(\hat{\mu}_i-\mu_i)$ thus has the asymptotic distribution function

$$\int_0^\infty \Phi(x(w/\mu_i)^{1/2})F(dw)$$

where F is the conditional distribution function of W, given W > 0. The Laplace transform of this distribution may in principle be determined from the offspring distribution, cf. Athreya and Ney (1972, Theorem III. 8.3).

However, only in the case of the birth-and-death process does the distribution of W take the particular simple exponential form (Athreya and Ney (1972, p. 136, Problem 2)) discussed by Keiding (1974, 1975) and leading to the asymptotic Student-distribution.

d. Estimators of particular functionals.

By the theorem on transformation of maximum likelihood estimators it follows that the maximum likelihood estimator of the offspring mean $m = \Sigma i \, p_i$ is

$$\hat{m} = \Sigma \, i \, \hat{p}_i = \frac{1}{N_t} \Sigma \, i \, N_t(i) = \frac{X_t + N_t - x_0}{N_t}$$

as also pointed out by Jagers (1973b). Therefore the maximum likelihood estimator of the Malthusian parameter $\alpha = \lambda(m-1)$ is $\hat{\lambda}(\hat{m}-1) = (X_t - x_0)/S_t$. From the previous results we thus deduce that all of these estimators are strongly consistent both for $x_0 \to \infty$ and $t \to \infty$, and that the following results on asymptotic distribution hold.

<u>Theorem 2.5</u> <u>As</u> $x_0 \to \infty$ <u>for fixed</u> t,

$$(x_0(e^{\alpha t}-1)/\alpha)^{1/2}(\hat{\lambda}-\lambda) \xrightarrow{D} \text{normal } (0,\lambda)$$

<u>and if the offspring variance</u> $\sigma^2 = \Sigma(i-m)^2 p_i < \infty$,

$$(x_0(e^{\alpha t}-1)/\alpha)^{1/2}(\hat{m}-m) \xrightarrow{D} \text{normal } (0,\sigma^2/\lambda)$$

<u>and</u>

$$(x_0(e^{\alpha t}-1)/\alpha)^{1/2}(\hat{\alpha}-\alpha) \xrightarrow{D} \text{normal } (0,\lambda(\sigma^2+(m-1)^2)).$$

<u>Furthermore</u> $\hat{\lambda}$ <u>and</u> $\hat{m}$ <u>are asymptotically independent and the asymptotic correlations with</u> $\hat{\alpha}$ <u>are</u>

$$\rho(\hat{\alpha},\hat{m}) \sim \sigma^2/[\sigma^2(\sigma^2+(m-1)^2)]^{1/2},$$

$$\rho(\hat{\alpha},\hat{\lambda}) \sim (m-1)/[\sigma^2+(m-1)^2]^{1/2}.$$

<u>As</u> $t \to \infty$ for $x_0 = 1$, <u>the conditional distributions given</u> $\{X_t \to \infty\}$ (or $\{X_t > 0\}$) <u>converge as shown</u>:

<u>Without further conditions on the offspring distributions</u>

$$S_t^{1/2}(\hat{\lambda}-\lambda) \xrightarrow{D} \text{normal } (0,\lambda),$$

<u>and if</u> $\sigma^2 < \infty$,

$$S_t^{1/2}(\hat{m}-m) \xrightarrow{D} \text{normal } (0,\sigma^2/\lambda)$$

<u>and</u>

$$S_t^{1/2}(\hat{\alpha}-\alpha) \xrightarrow{D} \text{normal } (0,\lambda(\sigma^2+(m-1)^2))$$

<u>Parallel to above</u>

$$S_t^{1/2}(\hat{\lambda}-\lambda) \text{ \underline{and} } S_t^{1/2}(\hat{m}-m)$$

are asymptotically independent and the asymptotic correlations are again the same.

Remark. The analysis leading to the above Theorem shows that in fact $\{\hat{p}_i \mid i=0,2,3,\ldots\}$ is asymptotically independent of $\hat{\lambda}$, both as $t \to \infty$ and as $x_0 \to \infty$.

Remark. The results for $t \to \infty$ may be transformed into statements with deterministic normalising factors as discussed in the Remark of the end of the last Subsection.

3. MULTITYPE MARKOV BRANCHING PROCESSES

Let $\{\underset{\sim}{X}_t = (X_t(1),\ldots,X_t(k)) \mid t \geq 0\}$ be a k-type Markov branching process with split intensities $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_k)$ and offspring distributions $\underset{\sim}{p}(\underset{\sim}{s}) = (p_1(\underset{\sim}{s}),\ldots,p_k(\underset{\sim}{s}))$, $\underset{\sim}{s} \in N_0^k = \{0,1,2,\ldots\}^k$. A survey of the definition and main properties of such processes was given by Athreya and Ney (1972, V. 7-8). We assume all offspring means finite, *i.e.*

$$m_{ij} = \sum_{s_1=0}^{\infty} \cdots \sum_{s_k=0}^{\infty} s_j p_i(s_1,\ldots,s_k) < \infty .$$

This assumption guarantees non-explosion in finite time with probability one.

a. The likelihood function and maximum likelihood estimators.

As in Section 2, we may now exhibit the likelihood function and the maximum likelihood estimators under freely varying parameters.

Theorem 3.1 There exists a measure on the space of realizations of $\{\underset{\sim}{X}_u \mid 0 \leq u \leq t\}$ such that the Radon-Nikodym derivative, or the likelihood function, is

$L(\underset{\sim}{\lambda}, \underset{\sim}{p}(\underset{\sim}{s}))$

$$= \prod_{j=1}^{k} \lambda_j^{N_t(j)} e^{-\lambda_j S_t(j)} \prod_{s_1=0}^{\infty} \cdots \prod_{s_k=0}^{\infty} p_j(\underset{\sim}{s})^{N_t(j,\underset{\sim}{s})}$$

where $\underset{\sim}{s} = (s_1,\ldots,s_k)$, $N(t,j,\underset{\sim}{s})$ is the number of splits of particles of type j that produce s_1 particles of type $1,\ldots,$ s_k particles of type k, $N_t(j) = \sum_{s_1=0}^{\infty} \cdots \sum_{s_k=0}^{\infty} N_t(j,\underset{\sim}{s})$ the total number of splits of particles of type j and $S_t(j) = \int_0^t X_u(j)du$ the total time lived by particles of type j, all in $[0,t]$.

Sufficient statistics are given by the quantities above. If the statistical model is that all intensities vary freely over $(0,\infty)$ and all probabilities vary freely over $[0,1]$, subject to $\sum_{\underset{\sim}{s}} p_j(\underset{\sim}{s}) = 1, j=1,\ldots,k$, then these are also minimally sufficient and the maximum likelihood estimators are given by

$$\hat{\lambda}_j = \frac{N_t(j)}{S_t(j)}, \quad \hat{p}_j(\underset{\sim}{s}) = \frac{N_t(j,\underset{\sim}{s})}{N_t(j)} .$$

We shall prove consistency on the set of nonextinction below. For this we need a simple lemma.

Let $\underset{\sim}{S}_t = (S_t(1),\ldots,S_t(k))'$ and $\underset{\sim}{N}_t = (N_t(1),\ldots,N_t(k))'$ be as in Theorem 3.1 and define

$$S_t = \sum_{i=1}^{k} S_t(i) \text{ and } N_t = \sum_{i=1}^{k} N_t(i).$$

Let further (following Athreya and Ney (1972,pp.202 ff.)) $\underset{\sim}{A}$ be the matrix with elements $a_{ij} = \lambda_i b_{ij}$, where $b_{ij} = m_{ij} - \delta_{ij}$.

Then $\underset{\sim}{A}$ generates the semigroup $\underset{\sim}{M}_t$ of means of $\underset{\sim}{X}_t$. It is known that $\underset{\sim}{A}$ has a positive real eigenvalue α which strictly dominates the real parts of all other eigenvalues. The corresponding left eigenvector is called $\underset{\sim}{v}$, and assume $\Sigma v_i = 1$. Then if

$$(*) \sum_{\underset{\sim}{s}} p_i(\underset{\sim}{s}) s_j \log s_j < \infty$$

for all i,

$$\underset{\sim}{X}_t \, e^{-\alpha t} \to \underset{\sim}{v} W$$

almost surely as $t \to \infty$, where the random variable W is positive with probability $1 - q$, q being the extinction probability.

Lemma 3.1 Assume (*). Then, as $t \to \infty$, the following hold with probability one:

$$\underset{\sim}{S}_t e^{-\alpha t} \to \underset{\sim}{v} W/\alpha$$

and

$$\underset{\sim}{N}_t e^{-\alpha t} \to \underset{\sim}{z} W/\alpha,$$

where

$$\underset{\sim}{z} = (\lambda_1 v_1, \ldots, \lambda_k v_k)'.$$

Proof. The convergence of $S_t(i)e^{-\alpha t}$ follows easily for almost all ω by choosing t_0 large enough that $|X_t(i)e^{-\alpha t} - v_i W(\omega)| < \varepsilon$ for $t > t_0$.

Then

$$S_t(i)e^{-\alpha t} = e^{-\alpha t} \int_0^{t_0} X_u(i)\,du + e^{-\alpha t} \int_{t_0}^{t} X_u(i)\,du.$$

In the last integral, the integrand may be approximated by $v_i W(\omega) e^{\alpha u}$, so that the last term is less than ε/α different from $v_i W(\omega)(1-e^{-\alpha(t-t_0)})$. Let then $t \to \infty$.

To prove convergence of $\underset{\sim}{N}_t e^{-\alpha t}$ we write

$$X_t(j) = \sum_{i=1}^{N_t(1)} U_{1j}^{(i)} + \ldots + \sum_{i=1}^{N_t(k)} U_{kj}^{(i)}$$

where $U_{hj}^{(i)} + \delta_{hj}$ is the number of particles of type j created at the i'th split of particles of type h. Obviously $U_{hj}^{(i)}$, $i = 1, 2, \ldots$ are independent and identically distributed, so by the strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} U_{hj}^{(i)} \to E(U_{hj}^{(i)}) = m_{hj} - \delta_{hj} = b_{hj}.$$

Furthermore it follows from Athreya and Ney (1972, Theorem V. 7.3) that as $t \to \infty$, and supposing (*) to hold,

$$\frac{N_t(j)}{N_t} \to \frac{\lambda_j v_j}{\Sigma\lambda_i v_i}, \quad j = 1,\ldots,k$$

a.s. on $\{N_t \to \infty\}$, where $N_t = \Sigma N_t(j)$. Hence as $t \to \infty$, remembering that $\underset{\sim}{v}$ is left eigenvector of $\underset{\sim}{A}$ corresponding to the eigenvalue α,

$$\frac{X_t(j)}{N_t} = \sum_{h=1}^{k} \frac{N_t(h)}{N_t}\left(\frac{1}{N_t(h)}\sum_{i=1}^{N_t(h)} U_{hj}^{(i)}\right) \to \sum_{h=1}^{k}\frac{\lambda_h v_h b_{hj}}{\Sigma\lambda_i v_i}$$

$$= \frac{1}{\Sigma\lambda_i v_i}\sum_{h=1}^{k} v_h a_{hj} = \frac{\alpha v_j}{\Sigma\lambda_i v_i}$$

a.s. on $\{N_t \to \infty\}$. It follows that, again a.s. on $\{N_t \to \infty\}$,

$$N_t\, e^{-\alpha t} \to \alpha^{-1}\Sigma\lambda_i v_i W.$$

Applying the known results about $\lim N_t(j)/N_t$ once more the Lemma is proved since $W = 0$ a.s. on $\{N_t \not\to \infty\}$.

<u>Theorem 3.2</u> <u>As</u> $t \to \infty$,

$$\hat{p}_i(\underset{\sim}{s}) \to p_i(\underset{\sim}{s}),\quad i = 1,\ldots,k,\quad \underset{\sim}{s} \in N_0^k$$

<u>and assuming</u> (*)

$$\hat{\lambda}_i \to \lambda_i,\quad i = 1,\ldots,k,$$

<u>where the convergence is a.s. on</u> $\{X_t \to \infty\}$.

<u>Proof</u>. As in the proof of Theorem 2.3, the result concering $\hat{p}_i(\underset{\sim}{s})$ follows from a representation

$$N_t(i,\underset{\sim}{s}) = \sum_{j=1}^{N_t(i)} Z_{ij}(\underset{\sim}{s})$$

where $Z_{i1}(\underset{\sim}{s})$, $Z_{i2}(\underset{\sim}{s}),\ldots$ are independent indicators with $P\{Z_{ij}(\underset{\sim}{s}) = 1\} = p_i(\underset{\sim}{s})$. Since $N_t(i) \to \infty$ a.s. on $\{N_t \to \infty\}$ (because all elements v_i of the eigenvector $\underset{\sim}{v}$ are positive), the strong law of large numbers may again be invoked.

The consistency of $\hat{\lambda}_i$ is a corollary of Lemma 3.1.

Remark. Results on asymptotic normality analogous to those of Theorems 2.2 and 2.4 may be proved along similar lines as in Section 2. Notice that the maximum likelihood estimator of the Malthusian parameter α may be obtained as the largest eigenvalue of the matrix $\hat{\underset{\sim}{A}} = (\hat{a}_{ij}) = (\hat{\lambda}_i(\hat{m}_{ij}-\delta_{ij}))$.

4. BELLMAN-HARRIS PROCESSES.

Let $\{X_t, t \geq 0\}$ be a Bellman-Harris process (or age-dependent branching process) specified by the life-length distribution function G and offspring distribution $\{p_i, i = 0,1,2,\ldots\}$, see Harris (1963, chapter VI) or Athreya and Ney (1972, chapter IV).

a. Estimation of the offspring probabilities

Assume as usual that the process X_u has been observed in some interval $[0,t]$. Since the event of death of one individual and birth of one offspring is not directly observable (the sample function will have no discontinuity), the inference on p_1 will be confounded with that on G. In the statement of the following Theorem we shall therefore assume $p_1 = 0$, but the results are easily interpretable for the conditional probabilities $p_i/(1-p_1)$ in the general case.

Theorem 4.1 Let $p_1 = 0$. The estimators $N_t(i)/N_t$ of p_i are consistent and asymptotically normal as $t \to \infty$:

$$N_t(i)/N_t \to p_i \text{ a.s. on } \{X_t \to \infty\}$$

and the conditional distribution of

$$N_t^{1/2}\{N_t(i)/N_t - p_i\},$$

given $\{X_t \to \infty\}$ (or $\{X_t > 0\}$) is asymptotically normal $(0,p_i(1-p_i))$ and with asymptotic covariances $-p_i p_j$.

If the parametrization is that the p_i's vary freely subject to $p_1 = 0$, $\Sigma p_i = 1$ and G is parametrized independently of $\{p_i\}$, then these are the maximum likelihood estimators.

Proof. Jagers (1973b) studied estimation in Bellman-Harris processes where the whole family tree is observed. Since the estimators here considered only depend on the observed trajectory $\{X_u, 0 \le u \le t\}$, the results follow from his analysis, except that Jagers assumes finite reproduction variance in the consistency proof. We notice, however, that the proof of the second part of Theorem 2.3 is equally applicable here.

Remark. As mentioned in the introduction, Becker (1974) discussed estimation of one-parameter exponential family offspring distributions given extinction of the process.

b. Estimation of the Malthusian parameter: Kendall's multiphase process.

The Malthusian parameter α is given as the solution, if there is any, of the equation

$$m \int_0^\infty e^{-\alpha x} G(dx) = 1,$$

where $m = \Sigma i p_i$ is the offspring mean. We shall illustrate the difficulties of estimating α by the maximum likelihood method by choosing as G a gamma distribution with parameters (k,λ^{-1}) where k is an integer and for simplicity of discussion we also choose $p_2 = 1$, (binary splitting), so that $m = 2$. The process is then the multiphase birth process proposed by Kendall (1948) and it is useful to observe that if $\underset{\sim}{Y}_t$ is a k-type Markov branching process with $\lambda_i = \lambda$ for all i, $p_i(\underset{\sim}{e}_{i+1}) = 1$ for $i = 1,\dots,k-1$ and $p_k(2\underset{\sim}{e}_1) = 1$, where $\underset{\sim}{e}_i$ is the i'th unit vector, then $X_t = \Sigma Y_t(i)$. The Malthusian parameter is $\alpha = \lambda(2^{1/k}-1)$.

The likelihood function given that $\{\underset{\sim}{Y}_u\}$ is observed in $[0,t]$ is obtained as

$$\lambda^{M_t} e^{-\lambda S_t}$$

where $M_t = \Sigma M_t(i)$ is the total number of splits of all types of particles in the $\underset{\sim}{Y}$-process and

$$S_t = \int_0^t [Y_u(1) + \ldots + Y_u(k)]du = \int_0^t X_u \, du.$$

Thus the likelihood equation is

$$M_t - \lambda S_t = 0$$

and it follows from general exponential family theory (see e.g. Sundberg (1974)) that the likelihood equation, given that only $\{X_u | 0 \leq u \leq t\}$ is observed, is

$$E(M_t - \lambda S_t | X_u, \ 0 \leq u \leq t) = 0.$$

In the present case S_t depends only on $\{X_u\}$ so that we obtain $\hat{\lambda}$ as the solution of

$$E(M_t | X_u, \ 0 \leq u \leq t) = \lambda S_t.$$

As an example let $k = 2$, $t = 2$ and assume that the process is known to start with one individual in the first phase ($\underset{\sim}{Y}_0 = (1 \ 0)$) and that $X_u = 1$ for $0 \leq u < 1$, $X_u = 2$ for $1 \leq u \leq 2$. Then $S_t = 3$ and the conditional expectation $E(M_t | X_u, 0 \leq u \leq t)$ is computed as follows. The initial particle changes phase once and then splits at time 1. Each of the two offspring may change phase once but none of them splits. The probability that a particle in phase 1 changes, given that it does not split in an interval of length x is

$$P\{U_1 < x | U_1 + U_2 > x\} = \lambda x/(1+\lambda x)$$

where U_1 and U_2 are independent and exponentially distributed with intensity λ. In the present case

$$M_t = 2 + i, \; i = \begin{cases} 0 \text{ w.pr. } [\lambda/(1+\lambda)]^2 \\ 1 \text{ w.pr. } 2\lambda/(1+\lambda)^2 \\ 2 \text{ w.pr. } 1/(1+\lambda)^2 \end{cases}$$

and thus the likelihood equation is

$$2\lambda^2 + 6\lambda + 4 = \lambda(1+\lambda)^2 S_t$$

which for $S_t = 3$ has $\lambda = 1$ as its only positive solution. In general the likelihood equation will be an (X_t+1)st order equation. We have not succeeded in finding a general formula for the solution even in the present simple case and we therefore think that explicit derivation of maximum likelihood estimators is unfeasible here and much more so in more general circumstances.

c. The asymptotic exponential distribution of the inter-split times.

It is well-known that the relative age distribution in a Bellman-Harris process converges towards the (deterministic) stable age distribution. More specifically, let X_t^a be the number of individuals alive and of age at most a at time t in a Bellman-Harris with non-lattice life-length distribution G, offspring distribution $\{p_i\}$ and positive Malthusian parameter α. Then Harris (1963, p. 154), Jagers (1968) showed that if the offspring variance is finite,

$$X_t^a/X_t \to A(a)$$

as $t \to \infty$ for all a, with probability one on $\{X_t \to \infty\}$, where

$$A(a) = \frac{\int_0^a e^{-\alpha x} G(dx)}{\int_0^\infty e^{-\alpha x} G(dx)} = \frac{\alpha m}{m-1} \int_0^a e^{-\alpha x} G(dx),$$

m as usual denoting the offspring mean.

This result was recently sharpened by Athreya and Kaplan (1975) to hold without any other conditions than $\Sigma p_j j \log j < \infty$, and the convergence in probability (conditional on $\{X_t \to \infty\}$) was shown to hold as soon as $m < \infty$ and a mild condition on G is imposed.

As we shall see presently, it follows from the above result that the split time process, defined as the point process of discontinuities of $\{X_t\}$, is asymptotically, as $t \to \infty$, of the same form as for a Markov branching process with split time intensity $\alpha/(m-1)$. It is thus intrinsically impossible to draw inferences on any other functionals of G than α from the later stages of a sample path of a Bellman-Harris process without further hypotheses or employing a different sampling scheme.

For ease of exposition we assume $p_0 = 0$ so that $X_t \to \infty$ a.s. We also assume $p_1 = 0$ as above since the events of birth of one individual do not imply a discontinuity in $\{X_t\}$.

Theorem 4.2 Let X_t be a Bellman-Harris process with $p_0 = p_1 = 0$, so that $m = \Sigma j p_j > 1$, and assume $\Sigma p_j\, j \log j < \infty$. Then in particular $m < \infty$, the Malthusian parameter α exists and $0 < \alpha < \infty$. Assume that $H = -\log(1-G)$ has a bounded and continuous density h. Let $\tau_0 = 0$, τ_n = the time of the n'th discontinuity of $\{X_t\}$.

(a) Let $\mathcal{F}_n$ be the σ-algebra spanned by the knowledge of the whole family tree up to (the random) time τ_n and let $Y_n = X_{\tau_n}(\tau_{n+1}-\tau_n)$. Then

$$P\{Y_n > y \mid \mathcal{F}_n\} \to \exp\{-\alpha y/(m-1)\}$$

a.s. as $n \to \infty$.

(b) Let $\mathcal{G}_t$ be the σ-algebra spanned by the knowledge of the full family tree in $[0,t]$ and let $Z_t = X_t(\tau_{N_t+1}-t)$. Then

$$P\{Z_t > z \mid \mathcal{G}_t\} \to \exp\{-\alpha z/(m-1)\}$$

as $t \to \infty$, a.s. on $\{X_t \to \infty\}$.

Remark. If G has a continuous density g, the density h is $g/(1-G)$, which is known as the hazard rate corresponding to G. Thus the hypothesis on G states that the hazard rate should be bounded, which is satisfied e.g. for gamma distributions ($h(x) = O(1)$ for $x \to \infty$) and for Pareto distributions ($h(x) = O(x^{-1})$).

Proof. We only give the proof of (a), since the proof of (b) is analogous. Let $G_x(y) = \{G(x+y) - G(x)\}/\{1-G(x)\}$ be the conditional distribution of the residual life-length given that the particle has survived to age x.

When $\acute{F}_n$ is given, the ages a_j, $j = 1,\ldots,X_{\tau_n}$ of the X_{τ_n} particles alive at τ_n are given. Denote the empirical distribution of these by $A_{\tau_n}(x)$. Now $\tau_{n+1} - \tau_n$ is the time until the first death among the X_{τ_n} particles, and hence

$$P\{Y_n > y|\acute{F}_n\} = \prod_{j=1}^{X_{\tau_n}} \{1-G_{a_j}(y/X_{\tau_n})\}$$

and the logarithm becomes

$$-\sum_{j=1}^{X_{\tau_n}}\{H(a_j+y/X_{\tau_n})-H(a_j)\} = -y\int_0^\infty \frac{1}{\delta}\, H(x+\delta)-H(x)\}A_{\tau_n}(dx)$$

$$= -y\int h(x)A_{\tau_n}(dx) + R_n$$

where $\delta = y/X_{\tau_n}$ and the residual R_n is discussed below.

Since $P\{\tau_n \to \infty\} = 1$, it follows from the above mentioned result of Athreya and Kaplan (1975) that $A_{\tau_n} \to A$ vaguely, almost surely as $n \to \infty$. Therefore, a.s. as $n \to \infty$,

$$\int h(x)A_{\tau_n}(dx) \to \int h(x)A(dx)$$

$$= \frac{\alpha m}{m-1}\int_0^\infty \frac{g(x)}{1-G(x)}\{1-G(x)\}e^{-\alpha x}dx = \frac{\alpha}{m-1},$$

To discuss the residual term R_n, choose, to an arbitrary $\varepsilon > 0$, K large enough to ensure that $1 - A(K) < \varepsilon$. On the set $[0,K]$, h is uniformly continuous so that we may choose γ to ensure that

$$\sup_{\substack{|x-v| < \gamma \\ 0 \leqq x \leqq K}} |h(x) - h(v)| < \varepsilon.$$

It follows from the mean value theorem that

$$\sup_{|u| < \gamma} \left|\frac{1}{u}\{H(x + u) - H(x)\} - h(x)\right| \leqq \sup_{v:|x-v|<\gamma} |h(v) - h(x)|$$

which we denote by $\phi(x,\gamma)$. By the results above, $\phi(x,\gamma) < \varepsilon$ for $0 \leqq x \leqq K$ and obviously $\phi(x,\gamma) \leqq 2 \sup h$ everywhere.

Therefore, on $\{y/X_{\tau_n} < \gamma\}$,

$$\begin{aligned}|R_n| &\leqq \int_0^K \phi(X,\gamma)A_{\tau_n}(dx) + \int_K^\infty \phi(X,\gamma)A_{\tau_n}(dx) \\ &\leqq \varepsilon + 2 \quad (1 - A_{\tau_n}(K)) \sup h \\ &\to \varepsilon + 2\varepsilon \sup h\end{aligned}$$

as $n \to \infty$ by the vague convergence, since A is continuous and $A(\infty) = 1$.

In conclusion, we have

$$\limsup_{n\to\infty} |P\{Y_n > y|\acute{F}_n\} - \exp\{-\alpha y/(m-1)\}|$$

$$\leqq \limsup_{n\to\infty} |P\{Y_n > y|\acute{F}_n\} I_{\{y/X_{\tau_n} < \gamma\}} - \exp\{-\alpha y/(m-1)\}| + \limsup_{n\to\infty} I_{\{y/X_{\tau_n} \geqq \gamma\}}.$$

The last term is zero a.s. and the first term is bounded by $\varepsilon(1 + 2 \sup h)$.

<u>Corollary</u>. <u>For fixed k and $n \to \infty$, the distribution of $(Y_{n+1},\ldots,Y_{n+k})$ converges to that of k independent exponentially distributed random variables with common intensity</u> $\alpha/(m-1)$.

Proof. We study the Laplace transform: Let for $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) \in R^k$

$$A_{nk}(\underset{\sim}{\theta}) = E\{\exp(-\sum_{i=1}^{k} \theta_i Y_{n+i})\}$$

$$= E[E\{\exp(-\sum_{i=1}^{k} \theta_i Y_{n+i}) | \acute{F}_{n+k}\}]$$

$$= E[\exp(-\sum_{i=1}^{k-1} \theta_i Y_{n+i}) E\{\exp(-\theta_k Y_{n+k}) | \acute{F}_{n+k}\}]$$

$$= E[\exp(-\sum_{i=1}^{k-1} \theta_i Y_{n+i})(E\{\exp(-\theta_k Y_{n+k}) | \acute{F}_{n+k}\} - a(\theta_k))] + a(\theta_k) A_{n,k-1}(\underset{\sim}{\theta}')$$

$$= B_{nk}(\underset{\sim}{\theta}) + a(\theta_k) A_{n,k-1}(\underset{\sim}{\theta}')$$

where $\underset{\sim}{\theta}' = (\theta_1, \ldots, \theta_{k-1})$ and

$$a(\theta_k) = \frac{\alpha}{(m-1)\theta_k + \alpha}$$

is the Laplace transform of an exponential distribution with intensity $\alpha/(m-1)$. By (a) of the Theorem, $B_{nk}(\underset{\sim}{\theta}) \to 0$ as $n \to \infty$ so by repeating the argument

$$A_{nk}(\underset{\sim}{\theta}) \to \prod_{i=1}^{k} a(\theta_j)$$

which completes the proof of the Corollary.

d. A simple estimator of the Malthusian parameter.

For the Markov branching process, the maximum likelihood estimator of α is $(X_t - x_0)/S_t$, $S_t = \int_0^t X_u du$, or the intuitive occurrence-exposure rate. We notice that this estimator, which we shall denote by $\tilde{\alpha}$, is not in general the maximum likelihood estimator of α in a Bellman-Harris process: Consider the example of Subsection 4b above where $\hat{\alpha} = \lambda(2^{1/2}-1)$, $\hat{\lambda} = 1$, but where S_t is seen to be 3 and therefore $(X_t - x_0)/S_t = \tilde{\alpha} = \frac{1}{3} \neq 2^{1/2}-1$.

Nevertheless, the analysis in the preceding Subsection shows that the sample functions of a Bellman-Harris process for large t look very much like the sample functions of a Markov branching process with the same offspring distribution and split intensity $\alpha/(m-1)$. One might therefore expect $\tilde{\alpha}$ to be asymptotically equivalent to the maximum likelihood estimator.

We present in this Subsection a proof of the consistency of $\tilde{\alpha}$ and a conjecture concerning the asymptotic normality. In the next Subsection $\tilde{\alpha}$ is compared to some other estimators proposed in the literature.

Theorem 4.3 *As $t \to \infty$, and supposing $\Sigma p_j\, j \log j < \infty$,*

$$\tilde{\alpha} = \frac{X_t - x_0}{S_t} \to \alpha$$

a.s. on $\{X_t \to \infty\}$. As usual $S_t = \int_0^t X_u du$.

Proof. This follows at once from the almost sure convergence $X_t e^{-\alpha t} \to W$, where $0 < W < \infty$ a.s. on $\{X_t \to \infty\}$, which was proved by Athreya and Kaplan (1975), cf. the proof of Lemma 3.1.

Conjecture *As $t \to \infty$, and under suitable regularity assumptions, including that of finite reproduction variance $\sigma^2 = \Sigma(i-m)^2 p_i$, the distribution of $S_t^{1/2}(\tilde{\alpha}-\alpha)$ is asymptotically normal* $(0, \{\sigma^2 + (m-1)^2\}\alpha/(m-1))$.

Motivation. The conjecture is the obvious generalisation of the analogous statement of Theorem 2.5 and it should be possible to obtain a proof by extending the argument that proves Theorem 2.4. Assume for convenience $p_1 = 0$, cf. the discussion above. First, it follows from Jagers (1974) or by an extension of Athreya and Kaplan's (1975) method, that as $t \to \infty$, using N_t = the number of discontinuities in $[0,t]$

$$N_t / S_t \to \alpha/(m-1) \text{ a.s. on } \{X_t \to \infty\},$$

and it therefore suffices to study

$$\begin{aligned} & S_t N_t^{-1/2}(\hat{\alpha}-\alpha) \\ &= N_t^{-1/2}(X_t - x_0 - \alpha S_t) \\ &= N_t^{-1/2} \sum_{j=1}^{N_t} (Z_j - Y_j) - N_t^{-1/2} \alpha X_t (t - \tau_{N_t}) \end{aligned}$$

where Z_j is the size of the j'th discontinuity and Y_j =

$\alpha X_{\tau_{j-1}}(\tau_j - \tau_{j-1})$, τ_j being the time of the j'th discontinuity. The second term will be negligible in the limit. To study the first term, we first notice that Z_{j+1} is independent of the σ-algebra F_j defined in Theorem 4.2(a) and by that same Theorem the conditional distribution of Y_{j+1} given F_j converges towards an exponential with mean m-1.

Therefore, for large j, $Z_{j+k} - Y_{j+k}$, $k = 1,2,\ldots$, are approximately independent random variables distributed as $Z - Y$, where $Z + 1$ has the offspring distribution (p_i) and Y is exponential with mean $m - 1$, and Z and Y are independent, so that $E(Z-Y) = 0$ and $\mathrm{Var}(Z-Y) = \sigma^2 + (m-1)^2$.

Let now

$$\phi_j(\theta) = E[\exp\{i\theta(Z_j - Y_j)\}|F_{j-1}]$$

and

$$\psi(\theta) = E[\exp\{i\theta(Z-Y)\}],$$

then

$$\left|E[\exp\{i\theta\, n^{-1/2}\sum_{j=1}^{n}(Z_j - Y_j)\}] - \{\psi(\theta n^{-1/2})\}^n\right|$$

$$\leqq \sum_{j=1}^{n}|\phi_j(\theta n^{-1/2}) - \psi(\theta n^{-1/2})|.$$

This will converge to zero, so that the conjecture may be proved by appealing to central limit theory for random sums of random variables as above, if there exist constants c_j, $j = 1,2,\ldots$ such that

$$|\phi_j(\theta) - \psi(\theta)| \leq |\theta| c_j$$

and

$$n^{-1/2}\sum_{j=1}^{n} c_j \to 0,$$

for instance, if $E(Y_j) = m - 1 + O(j^{-\delta})$ for some $\delta > 1/2$. This seems plausible but we have no proof so far.

e. Hoel and Crump's estimator of the Malthusian parameter.

A recent study by Hoel and Crump (1974) is concerned with estimation of the life-length distribution in a Bellman-Harris process with binary splitting, that is, offspring distribution degenerate at 2. This model has been applied to cell growth studies, cf. Hoel and Crump's references or Jagers (1975, Chapter 9), who gives a comprehensive survey of these models.

Hoel and Crump consider mainly Kendall's multiphase process discussed in Subsection 4b above. For this binary splitting process X_t with gamma (k,λ^{-1}) life-length distribution it is known that as $t \to \infty$

$$E(X_t) \sim X_0 \beta e^{\alpha t}$$

where $\alpha = \lambda(2^{1/k}-1)$ and $\beta^{-1} = 2(1-2^{-1/k})$, and Hoel and Crump reformulate the problem as that of estimating α and β from one realisation of $\{X_u | 0 \leqq u \leqq t\}$. Specifically, they propose to fit a straight line to $\log X_t$, thus "approximating" the "asymptotic true" line $\alpha t + \log \beta + \log X_0$ and subsequently using the slope as an estimate of α and the intercept on the ordinate axis to estimate β. Now there are several problems of using the intercept to estimate β, and Hoel and Crump do point out that no consistent (as the observation time gets large) estimate results. The basic difficulty, as in particular discussed by Waugh (1972), is that it follows from the general convergence theorems that for large t, $X_t \sim W\beta e^{\alpha t}$ or

$$\log X_t \sim \alpha t + \log \beta + \log W$$

where W is a random variable.

As an example, in the pure birth process with $X_0 = 1$, W is exponential with mean 1, so that $E(\log W) = -\gamma = -0.577$ and $\mathrm{Var}(\log W) = \pi^2/6 = 1.645$.

It is a consequence of the invariance considerations in

Subsection 4c above that asymptotically as $t \to \infty$, the sample function will only contain information on α and not on other functionals of the life-time distribution. In this light, it is obvious that the particular procedure suggested by Hoel and Crump will not lead to a consistent estimate. Their proposal to overcome this is to estimate β from replicated experiments.

Let us then turn to the question of estimating the Malthusian parameter α. Hoel and Crump study in particular the estimator

$$\bar{\alpha} = \frac{\log X_t - \log X_s}{t-s}$$

where $s < t$ are some time points.

For the case of $k = 1$, that is, the linear birth process, exact means and standard deviations of the maximum likelihood estimator $\hat{\alpha} = (X_t - x_0)/S_t$ were given by Beyer, Keiding and Simonsen (1975), cf. Beyer (1974). In this case, direct computation is also possible for $\bar{\alpha}$ and some values were kindly supplied by K.S. Crump in a personal communication. As an example, if $E(X_t) = 20$, corresponding to $t = 3.00$, one gets $E(\hat{\alpha}) = 0.851$, $\sigma(\hat{\alpha}) = 0.337$, or a mean square error (mse) of 0.136. Choosing $s = 2.30$, that is, $E(X_t) = 10$, which is the case treated in Hoel and Crump's (1974) Table 2, one gets $E(\hat{\alpha}) = 0.921$ with $\sigma(\hat{\alpha}) = 0.472$, or an mse of 0.229. But it is possible to do better by selecting other values for s. Thus if $s = 0.75$ corresponding to $E(X_t) = 20^{1/4}$, $E(\hat{\alpha}) = 0.869$ and $\sigma(\hat{\alpha}) = 0.366$, giving mse = 0.151 and this clearly compares favourably with the maximum likelihood estimate, which usually will be much harder to compute. Further comparisons of the maximum likelihood estimator to alternatives based on less information seem to be desirable.

Acknowledgements. Our thanks are due to K.S. Crump for providing supplementary information concerning Hoel and Crump (1974) and to Norman Kaplan for helpful suggestions.

K.B. Athreya's work was done at University of Copenhagen with support from Danish Natural Sciences Research Council.

REFERENCES

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

Athreya, K.B. & Kaplan, N. (1975) Convergence of the age distribution in the one dimensional supercritical age-dependent branching process. Preprint No.3 , Institute of Mathematical Statistics, University of Copenhagen.

Athreya, K.B. & Karlin, S. (1967) Limit theorems for the split times of branching processes. J.Math.Mech. 17, 257-277.

Athreya, K.B. & Ney, P. (1972) Branching Processes. Berlin: Springer.

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

Beyer, J.E. (1974) The exact behaviour of the maximum likelihood estimator in the pure birth process and the pure death process. Tables and graphs. Technical report, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

Beyer, J.E., Keiding, N. & Simonsen, W. (1975) The exact behaviour of the maximum likelihood estimator in the pure birth process and the pure death process. Scand. J. Statist. 2

Billingsley, P. (1961a) Statistical methods in Markov chains. Ann.Math.Statist. 32, 12 - 40.

Billingsley, P. (1961b) Statistical inference for Markov processes. University of Chicago Press.

Billingsley, P. (1968) Convergence of probability measures. New York: Wiley.

Dion, J.-P. (1972) Estimation des probabilités initiales et de la moyenne d'un processus de Galton-Watson. Thèse de doctorat, Université de Montréal.

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

Harris, T.E. (1963) The theory of branching processes. Berlin, Springer.

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

Jagers, P. (1968) Renewal theory and the almost sure convergence of branching processes. Ark. Mat. 7, 495 - 504.

Jagers, P. (1973a) 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.

Jagers, P. (1973b) 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.

Jagers, P. (1974) Convergence of general branching processes and functionals thereof. J. Appl.Prob. 11, 471-478.

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

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

Keiding, N. (1975) Maximum likelihood estimation in the birth-and-death process. Ann.Statist. 3.

Kendall, D.G. (1948) On the generalized "birth-and-death" process. Ann.Math.Statist. 19, 1-15.

Sundberg, R. (1974) Maximum likelihood theory for incomplete data from an exponential family. Scand.J.Statist. 1, 49-58.

Waugh, W.A. O'N. (1972) The apparent "lag phase" in a stochastic population model in which there is no variation in the conditions of growth. Biometrics 28, 329-336.