Niels Keiding

Maximum Likelihood Estimation in the Birth-and-Death Process

Preprint
April
1974

4

Institute of Mathematical Statistics
University of Copenhagen

Niels Keiding

MAXIMUM LIKELIHOOD ESTIMATION IN THE BIRTH-AND-DEATH PROCESS

Preprint 1974 No. 4

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

April 1974

This is a reprint of Technical Report No. 13, from Department of Statistics, Stanford University, prepared under the auspices of 'Office of Naval Research Contract # N00014-67-A-0112-0030.'

Maximum likelihood estimation in the birth-and-death process

Niels Keiding

University of Copenhagen*

* Work finished at Stanford University with support from Danish Natural Science Research Council

Summary

Maximum likelihood estimation of the parameters λ and μ of a simple (linear) birth-and-death process observed continuously over a fixed time interval is studied. Asymptotic distributions for large initial populations and for large periods of observation are derived and some nonstandard results appear. The related problem of estimation from the discrete skeleton of the process is also discussed.

American Mathematical Society 1970 subject classifications.

Primary 62F10, 62M05, 60J80

Secondary 60F05, 62B05

Key words and phrases: Birth-and-death process, Maximum likelihood estimation, Estimation in Markov processes, Discrete skeleton of Markov process, Weak convergence under random change of time.

1. Introduction

Let X_t be the population size at time t of the (linear) birth-and-death process, that is, the Markov process in which

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

$i=0,1,2,\ldots$, $\lambda \geqq 0$, $\mu \geqq 0$, and assume throughout that X_0 is degenerate at some $x_0 > 0$. We shall consider maximum likelihood estimation of the parameters λ and μ assuming that the process has been observed continuously over some time interval.

The maximum likelihood estimators are the occurrence-exposure rates $\hat{\lambda} = B_t/S_t$, $\hat{\mu} = D_t/S_t$, B_t and D_t being the number of births and deaths and $S_t = \int_0^t X_u du$ the total time lived by the population in the time interval $[0,t]$. The sampling properties of these estimators have been studied by a number of authors, to be referred to in Section 2, who assumed various stopping rules depending on the number of births and deaths.

In this paper the sampling properties of the estimators will be studied under the assumption of a fixed interval $[0,t]$ of observation. Exact results are as yet scarce; see Section 2. Asymptotic results for large initial population sizes are standard and stated in Section 3. Asymptotic results for large t are studied in Sections 4 and 5. Two aspects of the birth-and-death process cause novel features. First, if $\lambda > \mu$, $X_t \to 0$ ("becomes extinct") with probability $(\mu/\lambda)^{x_0}$ and $X_t \to \infty$ otherwise, whereas if $\lambda \leqq \mu$, $X_t \to 0$ a.s. Accordingly, Section 4 gives asymptotic results conditioned on $X_t \to \infty$ and Section 5 results conditioned on extinction. Secondly, given $X_t \to \infty$, $X_t/E(X_t) \to W$ a.s. where W is nondegenerate. Asymptotic normal theory no longer holds but is replaced by "Student"-distribution results. The proof uses a generalization of Lamperti's random time change transforming the birth-and-death process into a compound Poisson process, and then transforms the time scale back in a Billingsley (1968)-type approach.

Finally, estimation under the assumption that the process is only observed at equidistant points of time (the so-called discrete skeleton) is discussed in Section 6 and the results are shown to improve the early work of Immel (1951) and Darwin (1956). Furthermore, it is pointed out that by making the discrete skeleton infinitesimal, the results for continuous observation are recovered.

The results are related to recent work by Dion (1972) on estimation in the Galton-Watson process and results by Jagers (1973c) on estimation of the offspring distribution of a Bellman-Harris process.

The particular cases $\mu = 0$ (the pure birth process) and $\lambda = 0$ (the pure death process) have been studied previously. Keiding (1974) gave results for the pure birth process, using different proofs, and Beyer, Keiding and Simonsen (forthcoming) give exact and L_p-convergence results for the pure birth process and the pure death process as well as a numerical evaluation of the asymptotic results (for these particular processes), given in Sections 3 through 5 of the present paper.

The literature on estimation in the pure death process is vast, this problem occurring in a variety of life-testing situations. We shall not attempt to review this literature but call attention to the review by Cox (1965) and a paper by Wolff (1965) on estimation in birth-and-death processes of queueing theory. Further specific references are given in Sections 2 and 3.

The asymptotic results in Sections 3 through 6 specialize in an obvious way to the pure birth and pure death processes. We shall not state this specialization explicitly in each case and will therefore assume through those Sections that $\lambda > 0$ and $\mu > 0$.

2. Maximum likelihood estimation from continuous observation

Theorem 2.1. The likelihood function is proportional to

$$L(\lambda,\mu) = \lambda^{B_t}\mu^{D_t}e^{-(\lambda+\mu)S_t},$$

where B_t and D_t are the number of births and deaths, respectively, and $S_t = \int_0^t X_u du$ is the total time lived in the population during $[0,t]$. (B_t, D_t, S_t) is minimal sufficient and the maximum likelihood estimators are given by $\hat{\lambda} = B_t/S_t$ and $\hat{\mu} = D_t/S_t$.

Remark. $N_t = B_t + D_t$ is to be understood as the number of discontinuities of X_u, $0 \leq u \leq t$. At B_t of these, X_u jumps $+1$, at D_t of them, X_u jumps -1. Thus B_t and D_t depend on $\{X_u | 0 \leq u \leq t\}$ only and it is seen that $X_t - x_0 = B_t - D_t$.

Proof. The likelihood function seems to have been derived first by Darwin (1956). The other results are immediately derived from the likelihood function.

Remark. The characteristic function and some other results concerning the distribution of the minimal sufficient statistic were given by Puri (1968). For discussions of the distribution of $\hat{\mu}$ when $\lambda = 0$ see Hoem (1969b) and his references.

Exact and approximate small-sample properties of $\hat{\lambda}$ in the pure birth process $(\mu = 0)$ and of $\hat{\mu}$ in the pure death process $(\lambda = 0)$ are given by Beyer, Keiding and Simonsen (forthcoming).

Alternative Stopping Rules. Stopping rules making the observation time

t a random variable have been studied by a number of authors. Thus Moran (1951, 1953) observed conditional on N_t and on the particular sequence of births and deaths (thereby avoiding untimely extinction). Kendall (1952) observed conditional on $D_t = x_0$ and Bartlett (1955, Sec. 8.3) described briefly observation conditional on N_t or D_t. In all such cases, the likelihood function, and hence the maximum likelihood estimators, properly interpreted, remain the same (which, it seems, was not always realized by these authors) but the sampling properties, of course, differ. Anscombe (1953) studied sequential estimation with the criterion that $\lambda - \mu$ be estimated with a prescribed small standard error a. He obtained the stopping rule: observe until $S_t \geqq N_t^{1/2}/a$. (Some device must be prescribed to avoid extinction before then.) Asymptotically (as $N_t \to \infty$) unbiased estimates of $\lambda + \mu$ and $\lambda - \mu$ were obtained as $aN_t^{1/2}$ and $a(B_t - D_t)N_t^{-1/2}$, respectively. We remark that since under this stopping rule we may substitute S_t for $N_t^{1/2}/a$, these estimators are nothing but N_t/S_t and $(B_t - D_t)/S_t$, or the maximum likelihood estimators once again.

The point is perhaps best illustrated by the reparametrization $(\phi, \theta) = (\lambda + \mu, \mu/\lambda)$. Then the likelihood function will read $\phi^{N_t} e^{-\phi S_t} \theta^{D_t} (1 + \theta)^{N_t}$ so that ϕ and θ are L-independent, and furthermore, given N_t, S_t is S-sufficient for ϕ and D_t S-sufficient for θ (in Barndorff-Nielsen's (1971) terminology). The probabilistic interpretation of this reparametrization is that ϕ governs the "split time" process (Athreya and Ney (1972)) and θ the imbedded random walk, cf. Moran (1951, 1953).

3. Asymptotic results for large populations.

The birth-and-death process with $X_0 = x_0$ can be interpreted as the sum of x_0 independent birth-and-death processes with the same parameters and $x_0 = 1$. The following asymptotic results for x_0 and fixed t may therefore be obtained from standard asymptotic maximum likelihood theory.

Theorem 3.1 *As* $x_0 \to \infty$, $(\hat{\lambda},\hat{\mu}) \to (\lambda,\mu)$ a.s. *and*

$$\left\{\frac{x_0(e^{(\lambda-\mu)t} - 1)}{\lambda - \mu}\right\}^{1/2} \begin{pmatrix} \hat{\lambda} - \lambda \\ \hat{\mu} - \mu \end{pmatrix} \overset{w}{\to} \text{Normal} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}\right),$$

the factor in $\{\ \}$ *being replaced by* $x_0 t$ *when* $\lambda = \mu$.

Proof. It was shown by Puri (1968) that

$$E\begin{pmatrix} B_t \\ D_t \\ S_t \end{pmatrix} = \begin{pmatrix} \lambda \\ \mu \\ 1 \end{pmatrix} \frac{e^{(\lambda - \mu)t} - 1}{\lambda - \mu} x_0$$

(appropriately modified when $\lambda = \mu$), and since by the strong law of large numbers $B_t/x_0 \to EB_t$, $D_t/x_0 \to ED_t$ and $S_t/x_0 \to ES_t$ as $x_0 \to \infty$, the strong consistency follows. Asymptotic normality follows from standard theory, we need only compute the information matrix. But

$$\begin{pmatrix} -D_\lambda^2 \log L & -D_\lambda D_\mu \log L \\ -D_\mu D_\lambda \log L & -D_\mu^2 \log L \end{pmatrix} = \begin{pmatrix} B_t \lambda^{-2} & 0 \\ 0 & D_t \mu^{-2} \end{pmatrix}$$

and the result follows by taking the expectations.

<u>Remark</u>. Sverdrup (1965) gave a careful study of similar properties for related processes, cf. also Kendall (1949) and Hoem (1969a, 1971).

4. Asymptotic results for large periods of observation given non-extinction

In the supercritical case, that is, when $\lambda > \mu$, it is well known that $X_t/E(X_t) = X_t/\{x_0\exp(\lambda - \mu)t\} \to W$ a.s. as $t \to \infty$, where $P\{W = 0\} = P\{X_t \to 0\} = (\mu/\lambda)^{x_0}$ and the distribution of W, given $W > 0$, is gamma (x_0, x_0^{-1}), that is, has the density $x_0^{x_0} w^{x_0-1} e^{-x_0 w}/\Gamma(x_0)$, $w > 0$ (Harris (1963)). Similar results hold for a.s. convergence of the minimal sufficient statistic which implies the following consistency result.

Theorem 4.1

(a) As $t \to \infty$,

$$x_0^{-1} e^{-(\lambda-\mu)t} \begin{pmatrix} B_t \\ D_t \\ S_t \end{pmatrix} \to \begin{pmatrix} \lambda \\ \mu \\ 1 \end{pmatrix} \frac{W}{\lambda-\mu} \quad \text{a.s.}$$

(b) As $t \to \infty$, $(\hat{\lambda}, \hat{\mu}) \to (\lambda, \mu)$ a.s. on the set $\{X_t \to \infty\}$.

Proof. (b) is a corollary of (a). To prove (a), we may use Jagers' (1973a) results on almost sure convergence of random functionals of general branching processes. In fact, the birth-and-death process is a general branching process with Malthusian parameter $\lambda - \mu$, life-length distribution function $L(x) = 1 - e^{-\mu x}$ and expected reproduction process given by the density $\lambda e^{-\mu x}dx$. In the case $x_0 = 1$, (a) is now obtained directly from Jagers' Corollaries 1, 2, and 5, and the generalization to $x_0 > 1$ is immediate.

The asymptotic behavior of the estimators on the set $\{W = 0\} = \{X_t \to 0\}$ is described in Section 5 below.

Theorem 4.2 *As* $t \to \infty$,

$$\begin{pmatrix} S_t^{1/2}(\hat{\lambda} - \lambda)\lambda^{-1/2} \\ S_t^{1/2}(\hat{\mu} - \mu)\mu^{-1/2} \\ (\lambda - \mu)S_t e^{-(\lambda - \mu)t}/x_0 \end{pmatrix} \xrightarrow{w} \begin{pmatrix} A \\ B \\ W \end{pmatrix}$$

in the conditional distribution, given $W > 0$, *where* A, B *and* W *are independent*, A *and* B *are normal* $(0,1)$ *and* W *is gamma* (x_0, x_0^{-1}).

Proof. The almost sure convergence of S_t, properly normalized, was shown in Theorem 4.1.

The asymptotic normal distribution is obtained from asymptotic normality in a certain compound Poisson process which is converted into a birth-and-death process by a random time change, and then applying Billingsley (1968)-type results to verify that the asymptotic normality holds after the random time change.

Let a compound two-dimensional Poisson process (Q_t, R_t) be defined in the following way. At each event of a Poisson process U_t with intensity $\lambda + \mu$ the two-dimensional cluster size $(M,N) = (1,0)$ or $(0,1)$ with probability $\lambda/(\lambda + \mu)$ and $\mu/(\lambda + \mu)$, respectively. Then if (M_i, N_i), $i=1,2,\ldots$ are independent replications of (M,N), $(Q_t, R_t) = (\sum_1^{U_t} M_i, \sum_1^{U_t} N_i)$. Obviously $E(M,N) = (\lambda/(\lambda+\mu), \mu/(\lambda+\mu))$ so that $E(Q_t, R_t) = (\lambda t, \mu t)$ and similarly it is seen that

$V(B_t) = (\lambda + \mu)tE(M^2) = \lambda t$, $V(D_t) = \mu t$ and $Cov(B_t, D_t) = (\lambda + \mu)tE(MN) = 0$. Since (Q_t, R_t) has independent increments, it follows by the central limit theorem that $((Q_t - \lambda t)(\lambda t)^{-1/2}, (R_t - \mu t)(\mu t)^{-1/2})$ is asymptotically two-dimensional normal with mean zero and variance matrix the identity.

Let $K_t = Q_t - R_t + x_0$ and define L_t by $L_t = \int_0^{t \wedge T_0} K_{L_u} du$, where $T_0 = \inf\{t | K_{L_t} = 0\}$. Then K_{L_t} is a birth-and-death process with parameters λ and μ, which may be seen in a similar way as for the Lamperti representation of the pure birth process (Athreya and Ney (1972), Theorem III.11.1). Furthermore Q_{L_t} and R_{L_t} correspond to B_t and D_t as defined above, and L_t to S_t.

Replace now t by L_t in the asymptotic result above to get $((Q_{L_t} - \lambda L_t)(\lambda L_t)^{-1/2}, (R_{L_t} - \mu L_t)(\mu L_t)^{-1/2})$ which in light of the interpretation as birth-and-death process has the same distribution as $(S_t^{1/2}(\hat{\lambda} - \lambda)\lambda^{-1/2}, S_t^{1/2}(\hat{\mu} - \mu)\mu^{-1/2})$. The proof that the asymptotic normality will still hold after the random time change $t \to L_t$ on the set $\{L_t \to \infty\}$ is now similar to Dion's (1972) proof for discrete time, cf. also Jagers (1973b). From this proof we also conclude that the asymptotic normal distribution is independent of W as stated.

Remark. This method of proof will yield a series of central-limit type theorems for Markov branching processes. Such theorems will be useful counterparts to central limit results *e.g.* of the type stated by Athreya and Ney (1972, Sec. III.10).

Corollary. As $t \to \infty$,

$$[x_0(e^{(\lambda-\mu)t}-1)/(\lambda-\mu)]^{1/2} \begin{pmatrix} (\hat{\lambda}-\lambda)\lambda^{-1/2} \\ (\hat{\mu}-\mu)\mu^{-1/2} \end{pmatrix} \overset{w}{\to} \begin{pmatrix} AW^{-1/2} \\ BW^{-1/2} \end{pmatrix}$$

in the conditional distribution, given $X_t \to \infty$. The limiting distribution is bivariate Student with common denominator (cf. Johnson and Kotz, 1972, p. 134), that is, each component is Student with $2x_0$ d.f. and the components are independent for given W.

Remark. As $x_0 \to \infty$ in the Corollary above, the limiting distribution tends towards the two-dimensional standardized normal in accordance with the result in Theorem 3.1.

Remark. The results of Theorem 4.2 and its Corollary will still hold if considered in the conditional distribution given $\{X_t > 0\}$ instead of $\{X_t \to \infty\}$. By this remark, which is parallel to one made by Dion (1972), approximate confidence limits may be obtained.

5. Asymptotic results for large periods of observation given ultimate extinction

If $X_t \to 0$ as $t \to \infty$ which happens with probability $(\mu/\lambda)^{x_0}$ in the supercritical case $\lambda > \mu$ and almost surely otherwise, the consistency of the estimators no longer holds, since the sample will be in effect finite as $t \to \infty$.

Since for a supercritical process with $\lambda > \mu$, the conditional distribution, given that $X_t \to 0$, is identical to that of a birth-and-death process with birth parameter μ and death parameter λ (Waugh 1958), the results in the present section are relevant for supercritical processes, given extinction.

Theorem 5.1 For $\lambda \leq \mu$, $(\hat{\lambda}, \hat{\mu}) \to (B/S, D/S)$ a.s., *as* $t \to \infty$, *where* B *and* D *are the total number of births and deaths until extinction and* $S = \int_0^\infty X_t \, dt$.

The distribution of (B,D,S) *is given by the density*

$$\frac{x_0}{b!d!} \lambda^b \mu^d e^{-(\lambda+\mu)s} s^{b+d-1}$$

$b=0,1,2,\ldots,\ d=b+x_0,\ s \geq 0.$

Proof. Most of the results are immediate. The distributions of N and of S given N were given by Puri (1968). Gani and McNeil (1971) discussed further aspects of the distribution of (N,S).

From the results in the Theorem, various results concerning the limiting distribution of $(\hat{\lambda}, \hat{\mu})$ may be derived. A couple of examples are shown below.

(a) The expected values of $(B/S, D/S)$ quickly become complicated. Given $N = B + D$, $E(S^{-1}) = (\lambda+\mu)/(N-1)$ so that

$$E(\frac{B}{S} \mid N) = \frac{(\lambda+\mu)(N-x_0)}{2(N-1)}, \quad E(\frac{D}{S} \mid N) = \frac{(\lambda+\mu)(N+x_0)}{2(N-1)}$$

and it follows that for $x_0 = 1$, $E(B/S) = (\lambda+\mu)/2$ and for $x_0 = 2$ it may be seen that

$$E(\frac{B}{S}) = \frac{6\mu^3 + 6\mu^2\lambda - 3\lambda^2\mu + \lambda^3}{12\mu^2} .$$

(b) The estimator $\hat{\lambda} - \hat{\mu}$ of the Malthusian parameter $\lambda - \mu$ is asymptotically equal to $-x_0/S$, whose distribution is given by the density

$$-\frac{x_0}{u} e^{\frac{x_0}{u}(\mu+\lambda)} (\frac{\mu}{\lambda})^{\frac{1}{2}x_0} I_{\frac{1}{2}x_0}(-2x_0(\mu\lambda)^{\frac{1}{2}}/u),$$

$u < 0$, where I is a modified Bessel function.

6. Equidistant sampling

Until now it has been assumed that the complete process $\{X_u \mid 0 \leq u \leq t\}$ was observed. It may be more realistic to assume that the process is observed at the equidistant points $0, \tau, 2\tau, \ldots, k\tau = t$. The observations $X_{n\tau}$ (sometimes called "the discrete skeleton") then form a Galton-Watson process as is well-known (Harris, 1963, p. 101), but the transition probabilities are rather messy, and direct maximum likelihood estimation does not seem feasible (cf. Darwin (1956)).

In this Section, however, we shall show that interesting results may be obtained by assuming that the following observations are available. Interpret the discrete skeleton Galton-Watson process in the usual way as a chain of generations of independently reproducing particles, and assume that in addition to $X_{n\tau}$ itself, the number C_n of particles among the $X_{(n-1)\tau}$ that have 0 offspring is known.

Proposition 6.1 Under the sampling scheme described above, the likelihood function is proportional to

$$L(\lambda,\mu) = \alpha^{C_t}\{(1-\alpha)(1-\beta)\}^{\sum_0^{k-1} X_{n\tau} - C_t}\,\beta^{X_t - x_0 + C_t}, \quad \lambda > 0, \mu > 0$$

where

$$\alpha = \frac{\mu\{e^{(\lambda-\mu)\tau} - 1\}}{\lambda e^{(\lambda-\mu)\tau} - \mu}, \quad \beta = \frac{\lambda}{\mu}\alpha \quad \text{and} \quad C_t = \sum_{n=1}^{k} C_n$$

The maximum likelihood estimators of α and β are given by

$$\overline{\alpha} = C_t / \sum_0^{k-1} X_{n\tau} \quad \underline{\text{and}} \quad \overline{\beta} = (X_t - x_0 + C_t) / \sum_1^k X_{n\tau}.$$

Proof. Let $Z_n = X_{n\tau}$ and let $C_0 = 0$. Then clearly $\{(Z_n, C_n) | n=0,1,2,\ldots\}$ is a Markov chain with stationary transition probabilities $q_i(z,c) = P\{Z_n=z,\ C_n=c \mid Z_{n-1}=i,\ C_{n-1}=j\}$ given by $q_i(z,c) = q_1^{*i}(z,c)$, the i'th convolution, $q_1(z,c) = \alpha^c(1-\alpha)^{1-c}\beta^{(z-1)(1-c)}(1-\beta)^{1-c}$, and $q_0(0,0) = 1$. The likelihood function is then derived as $P\{Z_k=z_k,\ C_k=c_k \mid Z_{k-1}=z_{k-1},\ C_{k-1}=c_{k-1}\} \cdots P\{Z_1=z_1, C_1=c_1 \mid Z_0=x_0, C_0=0\}$.

Theorem 6.1 Assuming that only $X_0, X_\tau, \ldots, X_{k\tau} = X_t$ are observed, the maximum likelihood estimator of the Malthusian growth parameter $\lambda - \mu$ is given by

$$\overline{\lambda - \mu} = \frac{1}{\tau} \log \left(\frac{X_\tau + \ldots + X_{k\tau}}{X_0 + \ldots + X_{(k-1)\tau}} \right).$$

Proof. We have $\lambda - \mu = \tau^{-1} \log\{(1-\alpha)/(1-\beta)\}$, and the result is therefore true by Proposition 6.1 if $X_{n\tau}$ and C_n were observed. But since $\overline{\lambda - \mu}$ is a function of the $X_{n\tau}$s only, the result holds in the more narrow sample.

This proof is patterned after Harris' (1948) derivation of the maximum likelihood estimator of the mean in an unrestricted offspring distribution of a Galton-Watson process and settles a question left open by Darwin (1956), who studied $e^{(\overline{\lambda-\mu})\tau}$ as an estimator of $e^{(\lambda-\mu)\tau}$ without proving that it is the maximum likelihood estimator. Darwin obtained results concerning bias and asymptotic variance of this estimator as well as asymptotic efficiency relative to the maximum likelihood estimator $\exp\{(B_t-D_t)\tau/S_t\}$ obtained from continuous observation.

E. R. Immel (1951) also remarked in his unpublished UCLA thesis that estimation directly from the discrete skeleton is unfeasible. Immel then considered the problem of estimating the parametric function $\theta = \mu/\lambda$ in the restricted model with $(\log\mu - \log\lambda)/(\mu-\lambda) = \tau$ and showed that in this situation the maximum likelihood estimator was given by $\overline{\theta} = \sum_0^{k-1} X_{n\tau} / \sum_1^{k} X_{n\tau}$. Immel proved consistency and asymptotic normality of this estimator for large x_0 within the restricted model and proposed to use $\overline{\theta}$ as an approximation in the general case.

In the restricted model $\theta = e^{(\mu-\lambda)\tau}$, so that the maximum likelihood estimator of $e^{(\mu-\lambda)\tau}$ in the restricted model is $\overline{\theta}$. From Theorem 6.1 we see, however, that $\overline{\theta}$ is in fact the maximum likelihood estimator of $e^{(\mu-\lambda)\tau}$ in the unrestricted model, which indicates that Immel's proposal of using $\overline{\theta}$ as an estimator of θ in general should not be followed.

Comparison with permanent observation. The infinitesimal discrete skeleton

By applying Dion's (1972) results for the Galton-Watson process one may give asymptotic results as $k \to \infty$ for the maximum likelihood estimators $(\overline{\lambda},\overline{\mu})$ of (λ,μ) in this sampling situation as well as efficiency results. This was shown in detail by Keiding (1974) for the pure birth process and we shall not give the full details for the birth-and-death process.

We may, however, call attention to the fact that when $k \to \infty$,

$\tau \to 0$, $k\tau \to t$, the likelihood function of the equidistant sampling situation approaches that of continuous observation given in Section 2. (When observation is continuous, the number of deaths in any interval is (almost surely) given by the knowledge of the total population number at each instant in that interval and, in particular, $C_t \to D_t$.) It may be seen that not only will the estimators $(\bar{\lambda},\bar{\mu}) \to (\hat{\lambda},\hat{\mu})$ under this limiting process but that also the asymptotic distributions are asymptotically equal. As was assumed by Keiding (1974), the infinitesimal discrete skeleton again yields an alternative way of deriving the correct results, but it should be emphasized that this success depends on the introduction of the observation of C_n above.

Acknowledgement. My thanks are due to Steffen Lauritzen for discussions through the preparation of this paper.

[1] Anscombe, F. J. (1953). Sequential estimation (with discussion). J. R. Statist. Soc. B 15, 1-29.

[2] Athreya, K. B. and Ney, P. (1972). Branching Processes. Berlin: Springer.

[3] Barndorff-Nielsen, O. (1971). On Conditional Statistical Inference. Århus. Reprinted in: O. Barndorff-Nielsen (1973). Exponential Families and Conditioning. Copenhagen: Wiley.

[4] Bartlett, M. S. (1955). An Introduction to Stochastic Processes. Cambridge: Cambridge University Press.

[5] Billingsley, P. (1968). Convergence of Probability Measures. New York: Wiley.

[6] Cox, D. R. (1965). Some problems of statistical analysis connected with congestion. Congestion Theory (W. L. Smith and W. E. Wilkinson, eds.), University of North Carolina Press, 289-316.

[7] Darwin, J. H. (1956). The behaviour of an estimator for a simple birth and death process. Biometrika 43, 23-31.

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

[9] Gani, J. and McNeil, D. R. (1971). Joint distributions of random variables and their integrals for certain birth-death and diffusion processes. Adv. Appl. Prob. 3, 339-352.

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

[11] Harris, T. E. (1963). The Theory of Branching Processes. Berlin: Springer.

[12] Hoem, J. M. (1969). Fertility rates and reproduction rates in a probabilistic setting. Biom.-Praxim. 10, 38-66. (Correction note: Biom.-Praxim. 11, 20.)

[13] Hoem, J. M. (1969). The sampling distribution of an estimator arising in connection with the truncated exponential distribution. Ann. Math. Statist. 40, 702-703.

[14] Hoem, J. M. (1971). Point estimation of forces of transition in demographic models. J. R. Statist. Soc. B, 33, 275-289.

[15] Immel, E. R. (1951). Problems of Estimation and of Hypothesis Testing Connected with Birth-and-Death Markov Processes. Thesis, University of California, Los Angeles. (Abstract: Ann. Math. Statist. 22, 485.)

[16] Jagers, P. (1973a). Almost sure convergence of general branching processes and functionals thereof. Tech. rep. 1973-13, Dept. of Math., Chalmers Univ. Tech. and Univ. Göteborg.

[17] Jagers, P. (1973b). A limit theorem for sums of random numbers of i.i.d. random variables. Mathematics and Statistics. Essays in Honour of Harald Bergström (P. Jagers and L. Råde, eds.) Göteborg. 33-39.

[18] Jagers, P. (1973c). Maximum likelihood estimation of the reproduction distribution in branching processes and the extent of disintegration in cell proliferation. Tech. rep. 1973-17, Dept. of Math., Chalmers Univ. of Tech. and Univ. Göteborg.

[19] Johnson, N. L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. New York: Wiley.

[20] Keiding, N. (1974). Estimation in the birth process. Biometrika 61.

[21] Kendall, D. G. (1949). Stochastic processes and population growth. J. R. Statist. Soc. B 11, 230-264.

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

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

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

[25] Puri, P. S. (1968). Some further results on the birth-and-death process and its integral. Proc. Camb. Phil. Soc. 64, 141-154.

[26] Sverdrup, E. (1965). Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. Skand. Akt. 52, 184-211.

[27] Waugh, W. A. O'N. (1958). Conditioned Markov processes. _Biometrika_ 45, 241-249.

[28] Wolff, R. W. (1965). Problems of statistical inference for birth and death queuing models. _Op. Res._ 13, 343-357.