Niels Keiding

Estimation in the Birth Process

Preprint
March
1973

4

Institute of Mathematical Statistics
University of Copenhagen

Niels Keiding

ESTIMATION IN THE BIRTH PROCESS

Preprint 1973 No. 4

INSTITUTE OF MATHEMATICAL STATISTICS

UNIVERSITY OF COPENHAGEN

March 1973

SUMMARY

Maximum likelihood estimation of the parameter λ of a pure birth process is studied on the assumptions that the process is observed either completely in a time interval $[0,t]$ or at equidistant time points $0,\tau,\ldots,k\tau$.

The exact distribution of the minimal sufficient statistic is given in the first case and for both cases the asymptotic theory as $t \to \infty$, respectively $k \to \infty$, is studied. The related conditional Poisson process discussed recently by D.G. Kendall and W.A. O'N. Waugh is also studied and the results are shown to illustrate the modern theory of exponential families and conditional inference. Some efficiency results comparing the two sampling schemes are also given.

Key words: Pure birth process, Maximum likelihood estimation, estimation in Markov processes, exponential families, conditional inference, conditional Poisson process.

1. Introduction.

Let X_t be the population size at time t of the pure birth process, that is the Markov process in which

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

$i = 1,2,3,\ldots,\ \lambda > 0$. Assume throughout that $P\{X_0=q\} = 1$ where q is a fixed positive integer.

We shall discuss maximum likelihood estimation of the parameter λ from observations in a finite time interval $[0,t]$. Specifically, three different sampling schemes may be considered.

A. Permanent observation in a fixed time interval $[0,t]$.

B. Sampling at equidistant time points $0,\tau,\ldots,k\tau$.

C. Permanent observation until the time at which X_t jumps to n.

The sampling scheme C, which is often called inverse sampling was considered by Moran (1951).

The "direct" sampling schemes A and B were considered briefly by D.G. Kendall (1949) and related results (in effect, for the pure death process) were given by Sverdrup (1965) and Hoem (1971).

All of these authors only considered asymptotic results as $q \to \infty$. In the present paper we apply results by P.S. Puri (1966,1968) to study the exact distribution of the estimator and we concentrate on asymptotic results for $t \to \infty$, that is, for one long realization of the process. The usual asymptotic normal theory no longer holds and an asymptotic "Student" distribution applies for both sampling schemes A (Section 3) and B (Section 5). The results for equidistant sampling in Section 5 are closely connected to recent results by Dion (1972) on estimation in the Galton-Watson process.

Kendall (1966) and Waugh (1970,1972) have recently discussed the conditional distribution of the birth process given $W = \lim_{t\to\infty}$ a.s. X_t/EX_t. They show that the conditional process is an inhomogeneous Poisson process with intensity $\lambda W \exp(\lambda t)$. Section 4 is devoted to a discussion of maximum likelihood estimation in this conditional process and in the process $\{X_u | u \leq t\}$ given X_t. Several interesting aspects are discussed in the light of the modern theory of exponential families and conditional inference (Barndorff-Nielsen 1970,1971), and it is pointed out that the "extra randomness" in the asymptotic Student-distribution is due to the gamma-distributed random variable W. Section 2 states formally the result of Kendall for easy reference.

2. The birth process and the conditional Poisson process.

It is well-known (Harris 1963) that if $\{X_t, t \geq 0\}$ is a birth process with $X_0 = q$, then the expectation $EX_t = q \exp(\lambda t)$ and there exists a random variable W such that $X_t/EX_t \to W$ a.s. as $t \to \infty$. The distribution of W is gamma (q,q^{-1}), that is, with density

$$q^q w^{q-1} e^{-qw}/\Gamma(q), \qquad w > 0$$

and $EW = 1$.

The following result is due to D.G. Kendall (1966) in the case $q = 1$. See further discussion by Waugh (1970), Athreya and Ney (1972, Theorem III.11.2) and Tjur (1973). The generalization to $q > 1$ is straightforward.

Theorem 2.1 Conditioned on W, $X_t - q$ is a time-inhomogeneous Poisson process with intensity $W\lambda \exp(\lambda t)$, that is $E(X_t - q | W) = W(\exp(\lambda t) - 1)$.

Theorem 2.2 Let $\{Z_t, t \geq 0\}$ be a Poisson process with intensity $w\lambda \exp(\lambda t)$ and $Z_0 = q$ and define

$$T_t = \int_0^t Z_u \, du.$$

Then as $t \to \infty$

(a) $Z_t \exp(-\lambda t) \to w$ a.s.

and

(b) $T_t \exp(-\lambda t) \to w/\lambda$ a.s.

Proof. $(Z_t - q)/[\exp(\lambda t)-1]$ is a Markov process with $EU_t = w$, and is thus a nonnegative martingale. It follows that there is a random variable U such that $U_t \to U$ a.s. On the other hand it is easily seen that $U_t \xrightarrow{P} w$ and it follows that $U = w$ a.s. which proves (a). To prove (b), let $\omega \notin N$, the null set where (a) does not hold. To a given ε choose t_0 such that

$$w - \varepsilon < Z_t \exp(-\lambda t) < w + \varepsilon$$

for $t > t_0$.

For $t > t_0$

$$T_t e^{-\lambda t} = e^{-\lambda t}\int_0^{t_0} Z_u du + e^{-\lambda t}\int_{t_0}^{t} Z_u du$$

and since

$$e^{-\lambda t}\int_{t_0}^{t} Z_u du < e^{-\lambda t}\int_{t_0}^{t} (w+\varepsilon)e^{\lambda u}du$$

$$= \frac{w+\varepsilon}{\lambda}(1-e^{-\lambda(t-t_0)})$$

and similarly for the lower boundary, (b) follows by letting $t \to \infty$. This simple but quite general proof was pointed out by Martin Jacobsen (private communication).

Remark. Theorem 2.1 may be used to derive results for the birth process from corresponding results for the inhomogeneous Poisson process by mixing over the gamma-distributed random variable W (cf. the discussion by Waugh (1970)). This procedure will be used repeatedly in the following.

In particular, from Theorem 2.2 we may conclude the a.s.

convergence of

$$q^{-1}e^{-\lambda t}\int_0^t X_u\,du \to W/\lambda$$

in the birth process, which was first given by Puri (1966).

3. Permanent observation: Inference in the birth process.

Consider first sampling scheme A. The distribution of $\{X_u,\ 0 \leqq u \leqq t\}$, is fully determined by X_t and the random times $T_{q+1},\ldots,T_{X_t}$, where $T_i = \inf\{u \mid X_u = i\}$ is the time where the process jumps to state i. The random variable $(X_t, T_{q+1}\ldots, T_{X_t})$ (which is understood as X_t if $X_t = q$) takes values in the set

$$\mathcal{Y} = \{q\} \cup \bigcup_{n=q+1}^{\infty} \{n\} \times [0,\infty)^{n-q}$$

with probability one. If ν_n is Lebesgue measure on the Borel σ-algebra $\mathcal{B}_n$ on $[0,\infty)^n$, an invariant measure κ on $\mathcal{Y}$ is given by

$$\kappa(B_0 \cup \bigcup_{n=q+1}^{\infty} \{n\}\times B_{n-q}) = 1\{B_0=\{q\}\} + \sum_{n=q+1}^{\infty} \nu_{n-q} B_{n-q},$$

$B_0 = \{q\}$ or $\emptyset$, $B_{n-q} \in \mathcal{B}_{n-q}$ for $n > q$.

The likelihood function is the density of the distribution of Y_t with respect to κ.

<u>Theorem 3.1</u> The likelihood function is given by

$$L(\lambda) = (X_t-1)^{(X_t-q)}\lambda^{X_t-q}e^{-\lambda S_t}$$

where $S_t = \int_0^t X_u\,du$ and the factorial $a^{(x)} = a(a-1)\ldots(a-x+1)$.

(X_t, S_t) is minimal sufficient and the maximum likelihood estimator is given by

$$\hat{\lambda} = \frac{X_t-q}{S_t}.$$

<u>Proof</u>. Let $T_q = 0$. For given x, the sojourn times $T_{i+1}-T_i$, $i = q,\dots,x-1$, are independent, exponentially distributed with expectations $(i\lambda)^{-1}$. Hence the density of $(T_{q+1},\dots,T_x)$ with respect to ν_{x-q} is seen to be

$$f(t_{q+1},\dots,t_x) = (x-1)^{(x-q)}\lambda^{x-q}e^{-\lambda\sum_{i=q}^{x-1} i(t_{i+1}-t_i)},$$

$0 \leqq t_{q+1} \leqq \dots \leqq t_n$.

Since $P\{X_t = x,\ T_{q+1} \leqq t_{q+1},\dots,T_{X_t} \leqq t\}$

$$= P\{T_{q+1} \leqq t_{q+1},\dots,T_x \leqq t,\ T_{x+1} > t\},$$

the density of $(X_t,\ T_{q+1},\dots,T_{X_t})$ with respect to κ is

$$g(x,t_{q+1},\dots,t_x) = \int_t^\infty f(t_{q+1},\dots,t_{x+1})dt_{x+1}$$

$$= (x-1)^{(x-q)}\lambda^{x-q}e^{-\lambda(xt-\sum_{i=q+1}^{x} t_i)}.$$

The likelihood function is, thus,

$$L(\lambda) = (X_t-1)^{(X_t-q)}\lambda^{X_t-q}e^{-\lambda(tX_t-\sum_{i=q+1}^{X_t} T_i)}.$$

and it is immediately seen that

$$S_t = \int_0^t X_u du = tX_t - \sum_{i=q+1}^{X_t} T_i$$

which completes the derivation of the likelihood function. The rest of the proof is immediate.

<u>Theorem 3.2</u> (a) The distribution of S_t given that $X_t = x$ has characteristic function

$$E(e^{ivS_t}|X_t=x) = e^{ivqt}\left(\frac{1-e^{(iv-\lambda)t}}{(1-(iv)/\lambda)(1-e^{-\lambda t})}\right)^{x-q}$$

and density

$$(\lambda t)^{x-q} e^{-\lambda(s-qt)} (1-e^{-\lambda t})^{-x+q} g_{x,q}(s), \qquad qt \leq s < \infty.$$

Here $g_{x,q}(s)$ is the density of $qt + Y_1 + \ldots + Y_{x-q}$ where the Y_i are independent, uniformly distributed on $[0,t]$.

(b) The distribution of S_t given W and $X_t = x$ is the same as under (a).

(c) The distribution of the minimal sufficient statistic (X_t, S_t) has characteristic function given by

$$(E e^{iuX_t + ivS_t})^{1/q} = \frac{(1-(iv)/\lambda)e^{iu+(iv-\lambda)t}}{1-(iv)/\lambda - e^{iu} + e^{iu+(iv-\lambda)t}}$$

and density with respect to counting measure on the integers and Lebesgue measure on B

$$\binom{x-1}{q-1}(\lambda t)^{x-q} e^{-\lambda s} g_{x,q}(s), \qquad x = q, q+1, \ldots, \; qt \leq s < \infty.$$

Proof. It follows from the representation of S_t given by (3.1) that S_t is measurable with respect to the σ-algebra A_t spanned by $\{X_u | u<t\}$. That the distributions under (b) and (a) are identical then follows from the Markov property, since W is measurable w.r.t. the tail σ-algebra $\bigcap_t \sigma\{X_u | u>t\}$. (For further discussion see Tjur (1973)).

The derivation of the distributions has been carried out by Puri (1968). In the present setting an approach based on the conditional Poisson process is more direct. We omit the details.

Theorem 3.3 $\hat{\lambda} \to \lambda$ a.s. as $t \to \infty$.

Proof. This is a corollary of Theorems 2.1 and 2.2, cf. the Remark after Theorem 2.2.

Lemma 3.1 Let $\{Z_t, t \geq 0\}$ be a Poisson process with intensity $w\lambda \exp(\lambda t)$ and $Z_0 = q$. Then the distribution of

$$(A_t, \lambda e^{-\lambda t} T_t) = (e^{-(\lambda t)/2}(Z_t - q - \lambda T_t), \lambda e^{-\lambda t} T_t)$$

where $T_t = \int_0^t Z_u du$, converges weakly as $t \to \infty$ towards the distribution of (A,w) where A is normal $(0,w)$.

Proof. The characteristic function of (Z_t,T_t) is given by

$$E(e^{iuZ_t+ivT_t}) = \exp\left[-w(e^{-\lambda t}-1)+\frac{\lambda w e^{\lambda t+iu}(1-e^{(iv-\lambda)t})}{\lambda-iv}\right]$$

as is seen from Theorem 3.2 and by using the Poisson distribution of Z_t-q with parameter $w(\exp(\lambda t)-1)$.

The characteristic function of $(A_t,\lambda e^{-\lambda t}T_t)$ is then easily obtained and the result follows by letting $t \to \infty$.

Theorem 3.4 (a) As $t \to \infty$, $(\lambda S_t)^{1/2}(\hat{\lambda}/\lambda-1)$ is asymptotically normal $(0,1)$.

(b) As $t \to \infty$, $\exp[(\lambda t)/2]q^{1/2}(\hat{\lambda}/\lambda-1)$ converges weakly towards a Student-distribution with $2q$ d.f.

Proof. It is a consequence of Lemma 4.1 and Theorem 2.1 that given W,

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

$$= \frac{A_t}{(\lambda e^{-\lambda t}S_t)^{1/2}} \xrightarrow{w} \text{normal } (0,1).$$

Since this limiting distribution is independent of W, the same result is valid in the marginal distribution, which proves (a). To prove (b), notice that

$$e^{(\lambda t)/2}q^{1/2}(\hat{\lambda}/\lambda-1) = e^{(\lambda t)/2}q^{1/2}\frac{X_t-q-\lambda S_t}{\lambda S_t}$$

$$= \frac{A_t}{(\lambda e^{-\lambda t}S_t)^{1/2}}\left(\frac{e^{\lambda t}q}{\lambda S_t}\right)^{1/2}.$$

As $t \to \infty$, the second factor tends almost surely to $W^{-1/2}$ by Theorems 2.1 and 2.2 and since the first factor, given W, is asymptotically normal $(0,1)$, we infer that the limiting distribution of the product is that of $A/W^{1/2}$ where A and W are independent, A is normal (0,1), and W is (cf. Theorem 2.1) gamma (q,q^{-1}) or χ^2/f, $f = 2q$. (b) then follows from a standard result in the theory of the normal distribution.

Theorem 3.5 For fixed t and $q \to \infty$,

(a) $(\lambda S_t)^{1/2}(\hat{\lambda}/\lambda-1)$ is asymptotically normal (0,1).

(b) $e^{(\lambda t)/2}q^{1/2}(\hat{\lambda}/\lambda-1)$ is asymptotically normal $(0,(1-e^{-\lambda t})^{-1})$.

Proof. The birth process X_t with $X_0 = q$ has the same distribution as $U_t^1 + \ldots + U_t^q$, where U_t^i is a birth process with $U_0^i = 1$ and parameter λ and $U_t^1, \ldots, U_t^q$ are independent.

The results are therefore easily obtained from the central limit theorem.

Remark. (b) was given by D.G. Kendall (1949), see further Section 5 below.

Remark. In theorem 3.4(b), the limiting Student distribution approaches a normal distribution (0,1) as $q \to \infty$. This limit is also obtained by letting $t \to \infty$ in the limiting distribution in Theorem 3.5(b).

4. Permanent observation: Inference in the conditional Poisson process.

For large t, the sample functions $\log X_t$ tend to be linear with "deterministic" slope λ but with a random intercept $\log W$ on the ordinate axis (Waugh 1972). When considering a single long realization of the birth process the random variation "due to W" seems irrelevant so that it becomes warranted to consider estimation of λ and w in the conditional process given that $W = w$.

A related argument due to S.L. Lauritzen (private communication) is as follows. When the sampling situation is one realization, it is intrinsically impossible no matter for how long time this realization is observed, to decide whether the sample function is from a birth process or from the corresponding conditional Poisson process. If both the Poisson processes and the birth process (being a mixture of the Poisson processes) are included in the model, the generalized maximum likelihood solution could never be a mixture and hence is the Poisson process. Lauritzen will publish his general study on "maximum likelihood prediction" elsewhere.

Theorem 4.1 Let $\{X_t, t\geqq 0\}$ be a Poisson process with intensity $w\lambda \exp(\lambda t)$ and $X_0 = q$. For the problem of estimation (λ,w), the likelihood function is

$$L(\lambda,w) = (\lambda w)^{X_t-q} e^{-\lambda(S_t-tX_t)-w(e^{\lambda t}-1)} .$$

(X_t,S_t) is minimal sufficient and the likelihood equations are

$$X_{t-q} = E(X_{t-q}) = w(e^{\lambda t}-1)$$

$$tX_t - S_t = E(tX_t-S_t) = w\,\frac{\lambda te^{\lambda t}+1-e^{\lambda t}}{\lambda}$$

If $2(S_t-qt) < t(X_t-q)$, these equations have a unique solution $(\lambda,w) \in (0,\infty)^2$ and this solution is the maximum likelihood estimator. If $2(S_t-qt) \geqq t(X_t-q)$, the likelihood function has no maximum within $(0,\infty)^2$.

Remark. The canonical exponential family generated by the present estimation problem has canonical statistics $Y = X_t-q$ and $Z = tX_t - S_t$, $0 \leqq Z \leqq tY < \infty$ and canonical parameters $\mu = \log(\lambda w)$ and $\kappa = \lambda$. The canonical parameter domain is R^2. The likelihood equations are obtained as $Y = EY$ and $Z = EZ$, and they have a unique solution for all (Y,Z). However, $\hat{\kappa} > 0$ only when $2Z > tY$, (which is equivalent to $2(S_t-qt) < t(X_t-q)$.) Otherwise $\hat{\kappa} \leqq 0$ and there seems to be no natural way to extend the definition of the maximum likelihood estimator for the initial parameters. For a comprehensive account of the exact (that is, non-asymptotic) theory of exponential families, see Barndorff-Nielsen (1970).

Proof of Theorem 4.1 The proof is framed in the exponential family approach given above. The likelihood equations are

$$Y = e^{\mu}\frac{e^{\kappa t}-1}{\kappa}$$

$$Z = e^{\mu}\frac{\kappa t e^{\kappa t}+1-e^{\kappa t}}{\kappa^2}$$

so that by eliminating μ

$$\frac{Z}{tY} = \frac{1}{1-e^{-\kappa t}} - \frac{1}{\kappa t} = 1 - f(\kappa t)$$

where the function f is decreasing, $f(-\infty) = 1$, $f(0) = \frac{1}{2}$, $f(\infty) = 0$, $Df(0) = 0$ and $Df(x) < 0$ otherwise. The results for $2Z > tY$ follows immediately.

Lemma 4.1 Let

$$f(x) = \frac{1}{x} - \frac{1}{e^x-1}, \qquad x > 0.$$

If $g(t)$ is a positive function such that there exists a $c \in (0,\infty)$ with

$$tf(tg(t)) \to c^{-1} \text{ as } t \to \infty,$$

then

$$g(t) \to c \text{ as } t \to \infty.$$

Proof. To any $\varepsilon > 0$ choose t_0 so that for $t > t_0$

$$|ctf(tg(t)) - 1| < \varepsilon$$

or

$$c^{-1}(1-\varepsilon) < tf(tg(t)) < c^{-1}(1+\varepsilon). \tag{4.1}$$

For all $x > 0$, $f(x) = \frac{1}{x} - \frac{1}{e^x-1} < \frac{1}{x}$ so that for $t > t_0$

$$c^{-1}(1-\varepsilon) < \frac{t}{tg(t)} = g(t)^{-1}.$$

To any $\delta > 0$ choose x_0 so that

$$f(x) > \frac{1}{x(1+\delta)} \quad \text{for } x > x_0.$$

Choose t_1 such that

$$\frac{1+\varepsilon}{ct_1} < f(x_0).$$

Then for $t > t_0 \vee t_1$, by the right inequality in (4.1)

$$f(tg(t)) < f(x_0)$$

that is (since f is monotonically decreasing) $tg(t) > x_0$ or

$$f(tg(t)) < \frac{1}{tg(t)(1+\delta)}$$

and by applying the right inequality in (4.1) once more,

$$c^{-1}(1+\varepsilon) > \frac{t}{tg(t)(1+\delta)} = \frac{1}{g(t)(1+\delta)}$$

The results are summarized in

$$\frac{c}{(1+\varepsilon)(1+\delta)} < g(t) < \frac{c}{1-\varepsilon} \tag{4.2}$$

for $t > t_0 \vee t_1$ and the proof is complete.

Theorem 4.2 For the estimation problem discussed in Theorem 4.1, as $t \to \infty$

(a) $$P\{2(S_t-qt) < t(X_t-q) | X_t = x\} \to 1$$

uniformly in $x = q+1, q+2, \ldots$

and $$P\{X_t = q\} \to 0.$$

It follows that

$$P\{2(S_t-qt) < t(X_t-q)\} \to 1$$

as $t \to \infty$ which by Theorem 4.1 may by expressed by stating that the maximum likelihood estimator (λ^*, w^*) exists almost surely as $t \to \infty$.

(b) As $t \to \infty$, $(\lambda^*, w^*) \xrightarrow{P} (\lambda, w)$.

(c) As $t \to \infty$, $(\lambda S_t)^{1/2}(\lambda^*/\lambda - 1)$

is asymptotically normal (0,1), and $\exp[(\lambda t)/2](\lambda^*/\lambda-1)$ is asymptotically normal (0,1).

Proof. To prove (a) we have by Theorem 3.2(a) and (b) that

$$P\{2(S_t-qt) \geqq t(X_t-q)\,|\,X_t=x\}$$

$$= \left(\frac{\lambda t}{1-e^{-\lambda t}}\right)^{x-q} \int_{t(x-q)/2}^{t(x-q)} e^{-r} h_{x-q}(r)\,dr \tag{4.3}$$

where h_{x-q} is density of a sum of x-q independent uniformly distributed random variables on [0,t]. Hence (4.3) is less than

$$\tfrac{1}{2}(1-e^{-\lambda t})^{-x+q} e^{-(x-q)(\frac{t}{2}-\log(\lambda t))}$$

which tends to 0 as $t \to \infty$ uniformly in $x = q+1, q+2, \ldots$. That $P\{X_t=q\} \to 0$ is well-known.

To prove (b), we first prove that $\lambda^* \to \lambda$ a.s. λ^* is defined as the solution of

$$V_t^{-1} = \frac{S_t-qt}{\bar{X}_t-q} = tf(\lambda^* t) = \frac{1}{\lambda^*} - \frac{t}{e^{\lambda^* t}-1}$$

By Theorem 2.2, $V_t \to \lambda$ a.s. For fixed $\omega \notin$ the null set we may thus use Lemma 4.1 with $g(t) = \lambda^*$ and $c = \lambda$ to show that $\lambda^* \to \lambda$ a.s.

That $w^* \xrightarrow{P} w$ is proved applying (c) which is proved below. (c) implies that

$$t(\lambda^*-\lambda) \xrightarrow{P} 0$$

as $t \to \infty$ and therefore, since $X_t e^{-\lambda t} \to w$ a.s. by Theorem 2.2,

$$w^* = \frac{X_t-q}{e^{\lambda^* t}-1} = \frac{X_t-q}{e^{\lambda t}} \frac{1}{e^{(\lambda^*-\lambda)t}-e^{-\lambda t}} \xrightarrow{P} w.$$

Finally, to prove (c), we use Lemma 3.1 to conclude that $e^{\frac{\lambda t}{2}}(V_t/\lambda-1)$ is asymptotically normal $(0,w^{-1})$. λ^* is the solution of $V_t^{-1} = tf(\lambda^* t)$, so that

$$e^{\frac{\lambda t}{2}}(\lambda^*/\lambda-1)$$

$$= e^{\frac{\lambda t}{2}}(V_t/\lambda-1) + \frac{\lambda^*}{\lambda}\,\frac{\lambda^* t e^{\frac{\lambda t}{2}}}{e^{\lambda^* t}-1}\left(1 - \frac{\lambda^* t}{e^{\lambda^* t}-1}\right)^{-1}$$

In the last term, the first and the last factor converge towards one a.s. by results above. For the middle factor, we use (4.2) to show that for $\omega \notin N$ and t large

$$\frac{\lambda^* t e^{\frac{\lambda t}{2}}}{e^{\lambda^* t}-1} < \frac{\lambda t e^{\frac{\lambda t}{2}}}{(1-\varepsilon)(e^{\lambda t/[(1+\varepsilon)(1+\delta)]}-1)} \to 0$$

if $(1+\varepsilon)(1+\delta) < 2$. It follows that the second term above converges to zero almost surely and hence in probability and the result therefore follows from Lemma 3.1. The last result follows in the same way as above.

Remark. In the estimation problem in Theorem 4.3, the statistic X_t-q is sufficient for the parameter w. Furthermore, to any given value of the other parameter λ and any given value x of X_t-q it is possible to find a w such that the distribution of X_t-q has its mode at x. This property of X_t-q which here (as will often be the case) sharpens the concept of sufficiency with respect to w, is called M-ancillarity with respect to λ by Barndorff-Nielsen (1971). It is suggested to study inference problems regarding λ in the conditional distribution given a statistic which is M-ancillary for λ.

This conditional distribution is identical for the conditional Poisson process and the original birth process by Theorem 3.2(b).

Theorem 4.3 For the problem of estimating λ in the conditional distribution of $\{X_u,\ 0\leq u\leq t\}$ given that $X_t = x$, the likelihood function is

$$L(\lambda) = (X_{t-q})!\,[\lambda t/(1-e^{-\lambda t})]^{x-q} e^{-\lambda(S_t-qt)} \qquad 0 < \lambda < \infty.$$

The likelihood equation D log $L(\lambda) = 0$ is equivalent to

$$\frac{S_t - qt}{x-q} = \frac{1}{\lambda} - \frac{t}{e^{\lambda t}-1} = tf(\lambda t).$$

This equation has one positive root $\bar{\lambda}$ for $2(S_t-qt) < (x-q)t$, and one non-positive root for $2(S_t-qt) \geqq (x-q)t$. The maximum likelihood estimator is given as $\bar{\lambda}$ in the first case. In the second case the likelihood function does not attain its supremum within $(0,\infty)$. A straightforward extension of the parameter set to $[0,\infty)$ results in an estimator of 0 in the second case.

Proof. The likelihood function is obtained as in Theorem 3.1, using the negative binomial distribution of X_t-q and the likelihood equation is easily derived.

It is seen that $\bar{\lambda}$ is formally identical to the estimator λ^* studied above, replacing X_t-q by the conditioned value x-q. The rest of the proof is then immediate.

Remark. The present results may also be discussed in the exponential family framework as given above. It may be shown that Y is M-ancillary with respect to κ so that it is reasonable to draw inference concerning $\kappa(=\lambda)$ in the conditional distribution given $Y(=X_t-q)$.

Remark. It is instructive to notice the intuitive significance of the condition $2(S_t-qt) < (x-q)t$ for a positive estimate of λ. Apart from the unavoidable contribution of qt to the integral $S_t = \int_0^t X_u du$, the integral has to be less than half of the rectangle with edges (0,q) - (0,x) and (0,q) - (t,q). That is, the sample function must be "convex" in the sense just defined.

Theorem 4.4 For the estimation problem discussed in Theorem 4.3, as $t \to \infty$

(a) $$P\{2(S_t-qt) < (x-q)t \mid X_t=x\} \to 1$$

uniformly in $x = q+1, q+2, \ldots$.

It follows that for x > q, the maximum likelihood estimator $\overline{\lambda}$ will be positive a.s. as $t \to \infty$.

(b) If $t \to \infty$ and $x-q \to \infty$, such that $t(x-q)^{-1/2} \to 0$, $(x-q)^{1/2}[(\overline{\lambda}/\lambda)-1]$ will be asymptotically normal (0,1).

Proof. (a) is proved as Theorem 4.2(a) above. (b) is obtained by passing to the limit in the characteristic function given in Theorem 3.2(a).

5. Equidistant sampling

In Sections 3 and 4 we assumed that the complete history of the process was known from time 0 to time t. Assume now that the process is observed at the points $0, \tau, 2\tau, \ldots, k\tau = t$, that is, in the Sampling scheme B in the Introduction. The observations then form a Galton-Watson process $Z_n = X_{n\tau}$ with geometric offspring distribution

$$P\{Z_1=i|Z_0=1\} = e^{-\lambda\tau}(1-e^{-\lambda\tau})^{i-1},$$

$i = 1,2,\ldots$, and $P\{Z_1=0|Z_0=1\} = 0$ as is well-known (Harris 1963). The moments of the offspring distribution are

$$E(Z_1|Z_0=1) = e^{\lambda\tau}$$

$$V(Z_1|Z_0=1) = e^{\lambda\tau}(e^{\lambda\tau}-1).$$

By applying results for estimation of the offspring expectation in the Galton-Watson process (Harris 1948) or directly (Kendall 1949) we obtain

Theorem 5.1 The likelihood function is

$$L(\lambda) = \left[\prod_{n=1}^{k}\binom{X_{n\tau}-1}{X_{n\tau}-X_{(n-1)\tau}}\right] e^{-\lambda\tau \sum_1^k X_{(n-1)\tau}} (1-e^{-\lambda\tau})^{X_{k\tau}-X_0}$$

and the maximum likelihood estimator is

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

Theorem 5.2 As $k \to \infty$,

(a) $$\tilde{\lambda} \to \lambda \qquad \text{a.s.}$$

(b) $$\tau(X_0+\ldots+X_{k\tau})^{1/2}(\tilde{\lambda}-\lambda)$$

is asymptotically normal $(0,\ 1-e^{-\lambda\tau})$.

(c) $$\tau[q(e^{\lambda(k+1)\tau}-1)e^{\lambda\tau}]^{1/2}(e^{\lambda\tau}-1)^{-1}(\tilde{\lambda}-\lambda)$$

converges weakly towards a Student-distribution with 2q d.f.

Proof. (a) follows from general results by Heyde (1970) on estimation in the Galton-Watson process.

To prove (b), we recall a result by Dion (1972, Théorème 3.2.1) for general Galton-Watson processes implying that if $X_0 = q = 1$(and the generalization to general q is immediate)

$$(X_0+\ldots+X_{k\tau})^{1/2}\ \frac{e^{\tilde{\lambda}\tau}-e^{\lambda\tau}}{[e^{\lambda\tau}(e^{\lambda\tau}-1)]^{1/2}}$$

is asymptotically normal (0,1). By Taylor expansion

$$\begin{aligned}
&\tau(X_0+\ldots+X_{k\tau})^{1/2}(\tilde{\lambda}-\lambda)\\
&= (X_0+\ldots+X_{k\tau})^{1/2}\ \frac{e^{\tilde{\lambda}\tau}-e^{\lambda\tau}}{e^{\lambda\tau}}\\
&+ (X_0+\ldots+X_{k\tau})^{1/2}\ \frac{(e^{\tilde{\lambda}\tau}-e^{\lambda\tau})^2}{2\tau e^{2\lambda\tau}}.
\end{aligned}$$

Since $X_0+\ldots+X_{k\tau} \to \infty$ a.s., the last term tends to zero in probability by the above result of Dion and the limiting distribution is obtained by another application of Dion's result.

(c) is proved in a similar way, applying Dion (1972, Théorème 3.2.2) and remarking that

$$W = \lim_{k\to\infty} \text{a.s. } X_{k\tau}q^{-1}e^{-\lambda k\tau}$$

is gamma-distributed (q,q^{-1}) (cf. Theorem 2.1) so that the asymptotic distribution of

$$(1+e^{\lambda\tau}+\ldots+e^{\lambda k\tau})^{1/2} \frac{e^{\tilde{\lambda}\tau}-e^{\lambda\tau}}{[e^{\lambda\tau}(e^{\lambda\tau}-1)]^{1/2}}$$

becomes the Student-distribution with 2q d.f.

<u>Remark</u>. If $\tau \to 0$, $k \to \infty$, $k\tau \to t$, one should expect to obtain the results for permanent observation in $[0,t]$. In fact, Kendall (1949) used this idea to derive $\hat{\lambda}$. An examination of the Theorems of Section 3 and the present Section shows that this does hold true. We collect some of these results below.

(a) As $\tau \to 0$, $k \to \infty$, $k\tau \to t$,

$$\tilde{\lambda} = \frac{1}{\tau} \log \frac{X_\tau+\ldots+X_{k\tau}}{X_0+\ldots+X_{(k-1)\tau}} \to \frac{X_t-X_0}{S_t} = \hat{\lambda}.$$

(b) As $\tau \to 0$, $k \to \infty$, $t \to \infty$, $k\tau/t \to c > 0$, the scale parameter

$$s_{\tilde{\lambda}} = \tau[q(e^{\lambda(k+1)\tau}-1)e^{\lambda\tau}]^{1/2}(e^{\lambda\tau}-1)^{-1}$$

of the asymptotic Student distribution of $\tilde{\lambda}$ (cf. Theorem 5.2 (c)) and the scale parameter

$$s_{\hat{\lambda}} = e^{\frac{\lambda t}{2}} q^{1/2} \lambda^{-1}$$

of the asymptotic Student-distribution of $\hat{\lambda}$ (cf. Theorem 3.4 (a)) are asymptotically equal:

$$s_{\tilde{\lambda}}/s_{\hat{\lambda}} \to 1.$$

It is to be expected that this approximation scheme will work quite generally so that known results for discrete-time processes can be used to derive results for processes with continuous time.

A particularly simple case of equidistant sampling is that of $k = 1$, that is, only X_t is observed.

Theorem 5.3 The maximum likelihood estimator $\check{\lambda}$ based on observation of X_t only is given by

$$\check{\lambda} = \frac{1}{t} \log \frac{X_t}{q}.$$

As $t \to \infty$,

$$t(\check{\lambda}-\lambda) \to \log W \text{ a.s.},$$

where the distribution of W is gamma (q,q^{-1}) so that $E(\log W) = \psi(q) - \log q$ and $V(\log W) = \psi'(q)$. ψ is the digamma function $\psi(x) = D \log \Gamma(x)$.

Proof. The form of the estimator is concluded from Theorem 5.1(a). The rest of the theorem is based on the a.s. convergence

$$X_t q^{-1} e^{-\lambda t} \to W.$$

Notice that $\psi(q) - \log q < 0$ but that $\psi(q) - \log q \to 0$ as $q \to \infty$. In particular, $\psi(1) = -\gamma = -0.577$. Furthermore $q\psi'(q) \to 1$ as $q \to \infty$, $\psi'(1) = \pi^2/6 = 1.645$.

Remark. As $q \to \infty$,

$$q^{1/2}(\check{\lambda}-\lambda)$$

is asymptotically normal $\left(0, \left(\frac{\sinh(\lambda\tau/2)}{\lambda\tau/2}\right)^2 \frac{\lambda^2}{e^{\lambda\tau}-1}\right)$, and $q^{1/2}(\check{\lambda}-\lambda)$ is asymptotically normal with asymptotic variance given by the same expression, taking $\tau = t$.

This was derived by Kendall (1949) from standard asymptotic maximum likelihood theory, using the fact the each of the q ancestors at time 0 starts its own process.

The efficiency of equidistant sampling in comparison to permanent observation may be studied by comparing the rate of convergence of the estimator of λ to the true parameter under the various asymptotic approximations considered. The results are easily obtained from the Theorems. We give some examples:

$\hat{\lambda}_t$, $\check{\lambda}_t$ and $\tilde{\lambda}_{k\tau}$ are the maximum likelihood estimators from the birth process with $X_0 = q$, based on complete observation in $[0,t]$, observation of X_t alone, and sampling at $0,\tau,\ldots,k\tau$, respectively.

A. The efficiency of $\tilde{\lambda}_{k\tau}$ with respect to $\hat{\lambda}_t$, $t = k\cdot\tau$ when $q \to \infty$ was given by Kendall (1949) for the birth process and by Sverdrup (1965) for the death process in the particular case $k = 1$, $\tau = t$, that is, for $\check{\lambda}_t$. From Theorem 3.5 (b) and the Remark above the asymptotic efficiency is

$$e_q(\tilde{\lambda},\hat{\lambda}) = [\tfrac{\lambda\tau}{2}/\sinh\left(\tfrac{\lambda\tau}{2}\right)]^2$$

which is tabulated as a function of $\lambda\tau$ by Kendall and Sverdrup. Obviously the efficiency tends to 1 as $\lambda\tau \to 0$ and to 0 as $\lambda\tau \to \infty$.

B. As $t \to \infty$, the convergence rate of $\hat{\lambda}_t$ to λ is expressed by the scale parameter in the asymptotic Student-distribution as

$$\lambda e^{-(\lambda t)/2} q^{-1/2},$$

cf. Theorem 3.4(a). In contrast, Theorem 5.3 tells that as $t \to \infty$, $t(\check{\lambda}_t-\lambda)$ converges, but not towards zero. It is seen that $\check{\lambda}_t$ is biased to the order of t^{-1} and has asymptotic efficiency 0 as $t \to \infty$ compared to $\hat{\lambda}_t$.

C. The efficiency of $\tilde{\lambda}_{k\tau}$ with respect to $\hat{\lambda}_t$ for $t = k\cdot\tau$ and $k \to \infty$ is obtained from the asymptotic Student-distributions in Theorems 3.4(b) and 5.2(c) as

$$e_t(\tilde{\lambda},\hat{\lambda}) = [(1-e^{-\lambda\tau})/\lambda\tau]^2$$

Some numerical values of e_q and e_t are given in Table 5.1. It is obvious that $e_t(\tilde{\lambda},\hat{\lambda}) \to 1$ as $\lambda\tau \to 0$ and $\to 0$ as $\lambda\tau \to \infty$.

Table 5.1

Asymptotic Efficiency of Equidistant Sampling.

$\lambda\tau$	0	.02	.1	.2	1	2	4	6	8	10
$e_q(\tilde{\lambda},\hat{\lambda})$	1	1.000	.999	.997	.921	.724	.304	.090	.021	.005
$e_t(\tilde{\lambda},\hat{\lambda})$	1	.980	.906	.822	.400	.187	.060	.028	.016	.010

Acknowledgements.

The author has benefited much from discussions with his colleagues at the Institute of Mathematical Statistics in Copenhagen, in particular Søren Johansen and Steffen Lauritzen, and with Dr. W.A. O'N. Waugh. Dr. J.-P. Dion kindly sent a copy of his recent thesis.

References.

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

Barndorff-Nielsen, O. (1970). Exponential Families, Exact Theory. Mathematical Institute, University of Aarhus: Various Publication Series No. 19.

Barndorff-Nielsen, O. (1971). On Conditional Statistical Inference.

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

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

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

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

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

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

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

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

Puri, P.S. (1966). On the homogeneous birth-and-death process and its integral. Biometrika 53, 61-71.

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

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

Tjur, T. (1973). *Conditional probability distributions.* (Forthcoming).

Waugh, W.A. O'N. (1970). Transformation of a birth process into a Poisson process. J.R.Statist.Soc. B **31**, 418-31.

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.

Some corrections.

	For	Read
p. 2^4	λW	$q\lambda W$
p. 2_5	$W\lambda$	$qW\lambda$
p. 2_4	$= W$	$= qW$
p. 3^5	Proof.	Proof. $U_t =$
p. 4^8	variable	variable $Y_t =$
p. 5^6	t_n	t_x
p. 9^9	estimation	estimation of
p. 9^{13}	$X_{t-q} = E(X_{t-q})$	$X_t - q = E(X_t - q)$
p. 13^{11}	Theorem 4.3	Theorem 4.1
p. 13_2	X_{t-q}	$x-q$
p. 15_{5-7}	should be deleted.	