Niels Keiding

Estimation Theory for Branching Processes

Preprint
April
1975

7

Institute of Mathematical Statistics
University of Copenhagen

Niels Keiding

ESTIMATION THEORY FOR BRANCHING PROCESSES

Preprint 1975 No. 7

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

April 1975

ESTIMATION THEORY FOR BRANCHING PROCESSES

SUMMARY

The paper examines Maximum likelihood estimation of important parameters in branching processes with discrete and continuous time, given complete record of the size of the process in a time period. Exponential family theory is used to define a class of statistical models leading to simple explicit maximum likelihood estimators and a survey is given of the known asymptotic distribution theory.

SOMMAIRE

Cet article examine estimation du maximum de vraisemblance des paramètres importants dans les processus de ramification à temps discret et à temps continu, sur la base de connaissance complète du nombre d'individus dans une période du temps. La théorie des familles exponentielles est appliquée pour définer une classe de modèles statistiques qui entraîne des estimateurs à maximum de vraisemblance simples et explicites. Il y a encore donné un résumé de la théorie asymptotique de la distribution des estimateurs.

1. INTRODUCTION

The present paper studies estimation in branching processes based on a complete record of the size of the process in a time period, concentrating on some situations where the maximum likelihood method leads to simple explicit results.

A simple property of a certain class of exponential families of offspring distributions explains the generality of the classical maximum likelihood estimator $\hat{m} = (X_1 + \dots + X_n)/(X_0 + \dots + X_{n-1})$ of the offspring mean in discrete time and the occurrence/exposure rate $\hat{\alpha} = (X_t - x_0)/\int_0^t X_u du$ as maximum likelihood estimator of the Malthusian growth parameter in the continuous-time Markov case.

We give some examples of these offspring distributions and survey briefly the known (almost exclusively asymptotic) results concerning the distribution of the estimators.

The main part of the paper is organized in three sections: Section 2 considers the Galton-Watson process and Section 3 the Markov branching process in continuous time. A short Section 4 discusses how far the results for the Markov branching processes may be generalised to Bellman-Harris (age-dependent) processes.

2. BRANCHING PROCESSES IN DISCRETE TIME

Consider a Galton-Watson process $X_0, X_1, X_2, \dots$ with $X_0 = x_0$ fixed and assume that the offspring distribution with support $S \subseteq \{0,1,2,\dots\}$ belongs to an exponential family

$$(2.1) \qquad p_x(\underset{\sim}{\theta}) = a(\underset{\sim}{\theta})\, b(x)\, e^{\underset{\sim}{\theta}\cdot \underset{\sim}{T}(x)}.$$

We shall first study observation of a fixed number n of generations and assume provisionally that the whole family tree $(\underset{\sim}{Y}_0, \underset{\sim}{Y}_1, \dots, \underset{\sim}{Y}_{n-1})$ is observable, where $\underset{\sim}{Y}_i = (Y_{i1}, \dots, Y_{iX_i})$ and Y_{ij} is the number of offspring of the j'th member of the i'th generation, $j = 1, \dots, X_i$, $i = 0, \dots, n-1$. In particular, $\sum_j Y_{ij} = X_{i+1}$. Then the likelihood is

$$(2.2) \qquad a(\underset{\sim}{\theta})^{X_0 + \dots + X_{n-1}} \exp[\sum_{i=0}^{n-1} \sum_{j=1}^{X_i} \underset{\sim}{\theta}\cdot \underset{\sim}{T}(Y_{ij})],$$

and under certain regularity conditions on the exponential family, the maximum likelihood estimator (mle) of $\underset{\sim}{\theta}$ is determined as the solution to the likelihood equations

$$(X_0 + \cdots + X_{n-1})\, D_k \log a(\underset{\sim}{\theta}) + \sum_{i=0}^{n-1} \sum_{j=1}^{X_i} T_k(Y_{ij}) = 0$$

or

$$\sum_{i=0}^{n-1} \sum_{j=1}^{X_i} T_k(Y_{ij}) = (X_0 + \cdots + X_{n-1})\, E_{\underset{\sim}{\theta}}(T_k),$$

$k = 1,2,\ldots$

It follows from these equations that if the exponential family is such that there exist constants C_k, $k \in S$, so that

$$X = \sum_k C_k\, T_k(X), \tag{2.3}$$

then

$$\sum_{i=0}^{n-1} \sum_{j=1}^{X_i} \sum_k C_k\, T_k(Y_{ij}) = \sum_{i=0}^{n-1} \sum_{j=1}^{X_i} Y_{ij} = X_1 + \cdots + X_n$$

and since

$$\sum_k C_k\, E_{\underset{\sim}{\theta}}(T_k) = E_{\underset{\sim}{\theta}}(X) = m,$$

the offspring mean, we deduce that the maximum likelihood estimator of m is

$$\hat{m} = \frac{X_1 + \cdots + X_n}{X_0 + \cdots + X_{n-1}} \,. \tag{2.4}$$

This derivation was based on the assumption that $(\underset{\sim}{Y}_0, \underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_{n-1})$ is observable, but since $\hat{m}$ depends only on the total generation sizes $X_0,\ldots,X_n$, it is also the mle in the narrower sample $(X_0,\ldots,X_n)$.

As examples of offspring distributions satisfying (2.3) we quote first the <u>power series distributions</u>, defined as the one-parameter exponential families with $T = X$ as canonical statistic. This class includes the important <u>binary splitting</u> case with $p_0 + p_2 = 1$. Another example is the two-parameter <u>modified geometric distributions</u> with $p_0 = \alpha$, $p_x = (1-\alpha)(1-\beta)\,\beta^{x-1}$, $x = 1,2,\ldots$ where $\underset{\sim}{T} = (I\{X=0\}, X)$ so that we may choose $C_1 = 0$, $C_2 = 1$ to satisfy (2.3).

A one-parameter exponential family <u>not</u> satisfying (2.3) is given by

$$p_x = (x+1)^{-\theta}/\zeta(\theta), \quad x = 0,1,2,\ldots, \quad 1 < \theta < \infty \quad ,$$

where ζ is the Riemann zeta function. Here $T = \log X$ and $(X_1 + \cdots + X_n)/(X_0 + \cdots + X_{n-1})$ is <u>not</u> the mle of m.

The fact, first derived by Harris (1948), that (2.4) is true even for the completely general family $(p_0, p_1 \dots)$, $\Sigma p_k = 1$ may be explained by an interpretation of the latter as an exponential family with infinite-dimensional parameter set, setting $\theta_k = \log [p_k/(1 - p_0)]$ and $T_k = I\{X = k\}$, $k = 1,2,\dots$ and then choosing $C_k = k$ to satisfy (2.3). The details of this have yet to be worked out.

The distribution theory for mle in the discrete time case consists almost entirely of asymptotic results. Let first n be fixed. The branching property implies that we have x_0 independent replications of a branching process with $x_0 = 1$, and therefore standard asymptotic mle theory will work for $x_0 \to \infty$, proving consistency and asymptotic normality of $\hat{m}$. If $n \to \infty$ for fixed x_0, $X_n \to 0$ a.s. unless $m > 1$, and even then, there is usually a positive probability of extinction. However, consistency of $\hat{m}$ given nonextinction follows fairly readily from standard branching process results, as shown by Harris (1948) for convergence in probability. Asymptotic normality (with the random normalising factor $X_0 + \dots + X_{n-1}$) may be proved by appealing to central limit theory for sums of a random number of independent identically distributed random variables, as shown by Dion (1974) and Jagers (1973a). Since the normalising factor is random, the standard results on efficiency of the mle are not directly applicable. However, Heyde (1974) has shown the following efficiency result for the power series distributions. Let $X_0 = 1$ and $Z_n = X_0 + \dots + X_n$, so that $\hat{m} = (Z_n - 1)/Z_{n-1}$. Then

$$E(Z_{n-1}(\hat{m} - m)^2) = \sigma^2,$$

the offspring variance, and for any other unbiased estimator U_n of m, based on $(X_0, \dots, X_n)$,

$$\liminf_{n \to \infty} E(Z_{n-1}(U_n - m)^2) \geqq \sigma^2.$$

If the number N of generations to be observed is random but a function of $(X_0, \dots, X_N)$, that is, if N is a stopping time not depending on the parameter $\underset{\sim}{\theta}$, then it is well known that the likelihood function is obtained from (2.2) by replacing n by N. As an example, if the process is observed until extinction (which of course only is a complete prescription for subcritical and critical processes), then, letting $Z = X_0 + \dots + X_{N-1}$ denote the total number of individuals that have lived, and noting that $X_N = 0$, one gets $\hat{m} = 1 - X_0/Z$, provided (2.3) is satisfied. This was derived by Becker (1974b) for the power series distributions.

3. MARKOV BRANCHING PROCESSES (CONTINUOUS TIME)

Let (X_t), $t \geq 0$, be a Markov branching process with split intensity λ and offspring distribution with support $S \subseteq \{0,2,3,\ldots\}$, belonging to an exponential family (2.1). If (X_u) is observed in a prescribed time interval $[0,t]$, the likelihood function becomes

$$(3.1) \qquad L(\lambda,\underset{\sim}{\theta}) = \lambda^{N_t} e^{-\lambda S_t} a(\underset{\sim}{\theta})^{N_t} e^{\sum_{i=1}^{N_t} \underset{\sim}{\theta}\cdot T(Y_i)}$$

where $S_t = \int_0^t X_u du$, N_t is the number of splits (discontinuities of (X_u)) in $[0,t]$, and $Y_i - 1$ the size of the i'th discontinuity.

We assume that the statistical model is specified by $(\lambda,\underset{\sim}{\theta}) \in \Lambda \times \Theta$, $\Lambda = (0,\infty)$.

Then the estimation problem splits into two parts: First, there are the random number N_t of independent replications of observations on the <u>offspring distribution</u>, for which standard mle theory applies.

In particular, if the exponential family satisfies (2.3), we get

$$(3.2) \qquad \hat{m} = \sum_{i=1}^{N_t} Y_i/N_t = (X_t - x_0 + N_t)/N_t .$$

For the simple one- or two-parameter exponential families mentioned above, the extinction probability $q = P\{X_t \to 0 \mid X_0 = 1\}$ will usually be a known simple function of the parameters, and the mle $\hat{q}$ of q will therefore also be given directly. For the completely general family (p_x), $\Sigma\, p_x = 1$, Stigler (1971) observed that $\hat{q}$ is given as the smallest nonnegative solution of the equation

$$q = \sum_{j=0}^{\infty} \hat{p}_j\, q^j = \sum_{j=0}^{\infty} N_t(j)\, q^j/N_t$$

where $N_t(j)$ = the number of Y_i in $[0,t]$ that equal j.

Secondly, the mle of the <u>split intensity</u> λ is given by the occurence/exposure rate $\hat{\lambda} = N_t/S_t$.

As in the discrete-time case, there are virtually no small-sample results concerning the distribution of the estimators. Asymptotic results for the mle's of the parameters of the offspring distribution as $N_t \to \infty$ follow from standard theory, using N_t as sample size. In particular, Stigler (1971) showed how to establish consistency and asymptotic normality of $\hat{q}$ as given by (3.2) above.

These results are thus applicable if $x_0 \to \infty$, for fixed t, which certainly will imply $N_t \to \infty$, and also for $t \to \infty$ for fixed x_0 in the supercritical case given nonextinction.

Standard mle theory and the branching property imply consistency and asymptotic normality of $\hat{\lambda}$ as $x_0 \to \infty$. In the supercritical case, conditioned on nonextinction, consistency and asymptotic normality (with random normalising factor $N_t^{\frac{1}{2}}$) also holds, see Athreya and Keiding (1975).

It is a consequence of the factorization of the likelihood (3.1) and the independent parametrization $\Lambda \times \Theta$ that $\hat{\lambda}$ and $\hat{\underset{\sim}{\theta}}$ will be asymptotically independent.

An important parameter in a Markov branching process is the Malthusian parameter $\alpha = \lambda(m - 1)$ which determines the growth rate of the process. It follows from the results above that provided the offspring exponential family satisfies (2.3), we get $\hat{\alpha} = (X_t - x_0)/S_t$, another classical occurrence/exposure rate. Asymptotic results concerning $\hat{\alpha}$ may easily be derived from the results on $\hat{\lambda}$ and $\hat{m}$ already given.

If observation is stopped at the random time τ, where τ is a stopping time, independent of the parameters $(\lambda, \underset{\sim}{\theta})$, then the likelihood becomes (3.1) with t replaced by τ. A particularly simple situation is that of $\tau = \inf \{t \mid N_t = n\}$ since then $N_\tau = n$ and $\hat{\lambda}$ and $\hat{\underset{\sim}{\theta}}$ become independent for finite t and x_0. (Notice that if $p_0 > 0$, τ is only well-defined if $x_0 \geqq n$.) This situation was discussed by Moran (1951, 1953) for the particular case of the birth-and-death process, for which it becomes possible to obtain results on the exact distribution of the mle's.

An interesting example of a sampling situation leading to a mle given only implicitly was discussed by Becker (1974a), who approximated the initial spread of an epidemic by a Markov branching process (life-times corresponding to the latent periods). It is then reasonable to assume that any transition from latent to infectious period is observable but that the number of susceptibles infected by the particular infective is not observable. This corresponds to assuming that the number of splits N_t but not the sizes $Y_i - 1$ of the splits are known. The important question here is to assess whether the epidemic is minor (the branching process is subcritical) or major (the branching process is supercritical). This means that the interest centers around extracting information on the offspring distribution from the sequence of the split times, and mle does not lead to an explicit solution.

4. BELLMAN-HARRIS PROCESSES

We close the paper with a brief discussion of estimation in Bellman-Harris (age-dependent branching) processes. Assume that the offspring $(p_x(\underset{\sim}{\theta}))$ is an exponential family (2.1) and that the life-length distribution G is parameterised independently of $(p_x(\underset{\sim}{\theta}))$. Then the estimation problem based on complete record of the population size X_u in a fixed time interval $[0,t]$ splits into estimation of the offspring distribution on one hand and of the lifelength distribution on the other. Thus Jagers (1973b), studying tumour cell growth, and Becker (1974b), in an application to epidemics, concentrated on estimating parameters of the offspring distribution, for which the procedure is as in Section 3 above. On the other hand Hoel and Crump (1974), also motivated by cell kinetic studies, assumed $p_2 = 1$ and concentrated on estimation of the generation-size distribution. It turns out, however, that even for life-length distributions as simple as the Erlangian (i.e., gamma with integer form parameter) there is no simple explicit mle of the Malthusian parameter (growth rate) α, based only on the observation of $\{X_u, 0 \leqq u \leqq t\}$. This was discussed by Athreya and Keiding (1975), who then went on to propose that the occurrence/exposure rate $\tilde{\alpha} = (X_t - x_0)/S_t$ (cf. Section 3) be used as estimator of α even in the non-Markovian case. The motivation for this was the following <u>asymptotic Markovian property</u> of the supercritical Bellman-Harris process: Let τ_n be the time at which the n'th split takes place.

The normalised inter-split times $V_n = X_{\tau_n}(\tau_{n+1} - \tau_n)$ are independent and exponentially distributed with mean λ^{-1} in the Markov case, and for the supercritical Bellman-Harris process given nonextinction, the distribution of any finite number of V_n's converges towards independent exponentials with expectation $(m - 1)/\alpha$. It is easily seen that $\tilde{\alpha}$ is consistent and a conjecture concerning asymptotic normality based on the asymptotic Markovian property was also suggested by Athreya and Keiding (1975).

REFERENCES

Athreya, K.B. & N. Keiding (1975): Estimation theory for continuous-time branching processes. Preprint No. 6 , Institute of Mathematical Statistics, University of Copenhagen.

Becker, N. (1974a): On assessing the progress of epidemics. In: Contributions to Probability and Statistics; Essays in Honour of E.J.G. Pitman (E.J. Williams, ed.) Jerusalem Academic Press.

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

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. (1948): Branching processes. Ann. Math. Statist. 19, 474-494.

Heyde, C.C. (1974): Remarks on efficiency in estimation for branching processes. Technical report. Dept. of Statistics, Australian National University.

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

Jagers, P. (1973a): A limit theorem for sums of random numbers of i.i.d. random variables. In: Jagers, P. & L. Råde (ed.) Mathematics and Statistics; Essays in Honour of Harald Bergström. Göteborg, 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, Dept. of Math., Chalmers University of Tech. and Univ. Göteborg.

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

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

Stigler, S.M. (1971): The estimation of the probability of extinction and other parameters associated with branching processes. Biometrika 58, 499-508.