

Remarks on efficiency in estimation for branching processes

BY C. C. HEYDE

Department of Statistics, S.G.S., Australian National University, Canberra

SUMMARY

This paper is concerned with efficiency in the estimation of the mean θ of the offspring distribution of a supercritical Galton–Watson branching process on the basis of a sample of n consecutive generation sizes. First, a direct comparison is made between the maximum likelihood estimator and the simple ratio of generation sizes estimator. Next, a new definition of asymptotic efficiency of an estimator is given, generalizing that of Rao (1973). It is shown that, for offspring distributions belonging to the class of power series distributions, the maximum likelihood estimator is efficient in this new sense. The paper concludes with some remarks on the implications of this theory in the estimation of the growth rate in a pure birth process.

Some key words: Asymptotic efficiency; Birth processes; Branching process estimation; Galton–Watson processes; Maximum likelihood estimation.

1. INTRODUCTION

Let $\{Z_0 = 1, Z_1, \dots, Z_n\}$ be a sample of consecutive generation sizes from a so-called supercritical Galton–Watson process with $1 < \theta = E(Z_1) < \infty$ and $\sigma^2 = \text{var}(Z_1) < \infty$. We shall discuss the estimation of θ on the basis of this sample. Our starting point will be a comparison of two estimators that have previously been proposed for θ , namely the maximum likelihood estimator $\hat{\theta}_n = (Y_n - 1)Y_n^{-1}$, where $Y_n = Z_0 + \dots + Z_n$ (Harris, 1948; Heyde, 1970) and the simple ratio estimator $\hat{\theta}_n = Z_n Z_{n-1}^{-1}$ (Nagaev, 1967; Heyde, 1974). We shall then discuss the efficiency of estimation of θ for a particular class of offspring distributions. We do this by proposing a generalization of Rao's concept of efficiency (Rao, 1973, p. 348). This theory is of special interest since a random norming turns out to be appropriate and this highlights the inadequacy of direct, constant norming, generalizations of classical theory to the Markov case. The paper concludes with some remarks on the estimation of the growth rate in a pure birth process.

2. COMPARISON OF MAXIMUM LIKELIHOOD AND RATIO ESTIMATORS

For the purposes of this section we shall suppose for convenience that $\text{pr}(Z_1 = 0) = 0$, so that $Z_n \rightarrow \infty$ almost surely as $n \rightarrow \infty$. If $\text{pr}(Z_1 = 0) > 0$, it is well known that $Z_n \rightarrow \infty$ almost surely on the nonextinction set and the results which we shall obtain hold conditionally on nonextinction.

The estimate $\hat{\theta}$ has been studied by Jagers (1973) and in an unpublished thesis of J. P. Dion. It is known that

$$Y_{n-1}^{\frac{1}{2}}(\hat{\theta}_n - \theta) \rightarrow N(0, \sigma^2) \quad (1)$$

in distribution, while for $\check{\theta}$ we have, from a result due to Bühler (1969),

$$Z_{n-1}^{\frac{1}{2}}(\check{\theta}_n - \theta) \rightarrow N(0, \sigma^2) \tag{2}$$

in distribution. Further, it is well known that $\theta^{-n}Z_n \rightarrow W$ almost surely with W non-degenerate and almost surely positive (Harris, 1963, p. 13) and, from Theorem 3 of Heyde (1970), $\theta^{-n}Y_n \rightarrow \theta W / (\theta - 1)$, so that $Z_n Y_n^{-1} \rightarrow (\theta - 1)\theta^{-1}$ and (2) can be rewritten as

$$(\theta - 1)^{\frac{1}{2}}\theta^{-\frac{1}{2}}Y_{n-1}^{\frac{1}{2}}(\check{\theta}_n - \theta) \rightarrow N(0, \sigma^2) \tag{3}$$

in distribution. We can thus conclude that in an obvious, albeit not entirely standard, sense the asymptotic efficiency of $\check{\theta}_n$ relative to that of $\hat{\theta}_n$ is $1 - \theta^{-1}$.

The random normings in (1) and (3) can be replaced by constant normings without difficulty. Note that

$$Y_{n-1}(\hat{\theta}_n - \theta) = \sum_{i=1}^n (Z_i - \theta Z_{i-1}) = \sum_{i=1}^{Y_{n-1}} \eta_i,$$

where the η_i 's are independent and identically distributed, each with the distribution of $Z_1 - \theta$. Also,

$$Z_{n-1}(\check{\theta}_n - \theta) = Z_n - \theta Z_{n-1} = \sum_{i=1}^{n-1} \xi_{in},$$

where the ξ_{in} 's are independent and identically distributed, each with the distribution of $Z_1 - \theta$ and are, furthermore, independent of Z_{n-1} . A minor adjustment to the proof of Theorem 17.2 of Billingsley (1968) yields the limits in distribution

$$\sigma^{-1}(\theta - 1)^{-\frac{1}{2}}\theta^{\frac{1}{2}n}(\hat{\theta}_n - \theta) \rightarrow W^{-\frac{1}{2}}N(0, 1), \quad \sigma^{-1}\theta^{-\frac{1}{2}}\theta^{\frac{1}{2}n}(\check{\theta}_n - \theta) \rightarrow W^{-\frac{1}{2}}N(0, 1),$$

where in each case the $W^{-\frac{1}{2}}$ is independent of the $N(0, 1)$; see also Jagers (1973).

The estimators $\hat{\theta}_n$ and $\check{\theta}_n$ of course do not exhaust the possibilities for simple estimators of θ . Another example is $\theta_n^* = Z_n^{1/n}$. From the almost sure limit $\theta^{-n}Z_n \rightarrow W$ we readily deduce that θ_n^* is strongly consistent for θ . Furthermore, from the iterated logarithm analogue given in Theorem 2 of Heyde & Leslie (1971) it is a straightforward matter to show that almost surely $n(\theta_n^* - \theta) \rightarrow \theta \log W$ as $n \rightarrow \infty$. Although the asymptotic properties of θ_n^* are not directly comparable to those of $\hat{\theta}_n$ and $\check{\theta}_n$, it is clearly substantially inferior, providing basically a linear rate of convergence in contrast to a geometric one.

3. THE GENERAL ASYMPTOTIC EFFICIENCY QUESTION

It is, of course, useful if a more objective analysis than that of §2 can be given for the performance of estimators. To that end we propose a definition of asymptotic efficiency which generalizes that of Rao (1973, pp. 348, 349).

Suppose that we have a sample X_1, \dots, X_n of consecutive observations from some stochastic process whose distribution depends on a single parameter θ , $\theta \in \Theta \subset R$. Let $L_n(\theta)$ be the likelihood function associated with X_1, \dots, X_n and suppose that $L_n(\theta)$ is differentiable with respect to θ and that $E\{(d \log L_n / d\theta)^2\} < \infty$ for each n . We write \mathcal{F}_k for the σ -field generated by X_1, \dots, X_k and suppose that, taking $L_0 = 1$, we have almost surely for $n \geq 1$

$$E\left(\frac{d \log L_n}{d\theta} \middle| \mathcal{F}_{n-1}\right) = \frac{d \log L_{n-1}}{d\theta}. \tag{4}$$

A sufficient condition for (4) is that, if $P_n(X_1, \dots, X_n) = L_n(\theta)$ is the joint probability function of X_1, \dots, X_n , then

$$\sum_{X_n} P_n(X_1, \dots, X_n)$$

can be differentiated under the summation sign; there is a corresponding definition and requirement in the continuous case. The condition (4) is precisely that

$$\left\{ d \log L_n / d\theta = \sum_{i=1}^n u_i, \mathcal{F}_n, n \geq 1 \right\}$$

is a zero mean martingale.

Next, we set

$$I_n(\theta) = \sum_{k=1}^n E \left\{ \left(\frac{d \log L_k}{d\theta} - \frac{d \log L_{k-1}}{d\theta} \right)^2 \middle| \mathcal{F}_{k-1} \right\}. \quad (5)$$

This is a form of conditional information which reduces to the standard Fisher information in the case where the X_i 's are independent. We can think of

$$E \left[\left(\frac{d \log L_k}{d\theta} - \frac{d \log L_{k-1}}{d\theta} \right)^2 \middle| \mathcal{F}_{k-1} \right] = E \left[\left(\frac{d \log L_k}{d\theta} \right)^2 \middle| \mathcal{F}_{k-1} \right] - \left(\frac{d \log L_{k-1}}{d\theta} \right)^2 = E(u_k^2 | \mathcal{F}_{k-1})$$

as the information contained in X_1, \dots, X_k which is not contained in X_1, \dots, X_{k-1} , for given X_1, \dots, X_{k-1} .

DEFINITION. We shall say that a consistent estimator T_n of θ is asymptotically efficient if

$$I_n^{\frac{1}{2}}(\theta) \left\{ T_n - \theta - \beta(\theta) I_n^{-1}(\theta) \frac{d \log L_n(\theta)}{d\theta} \right\} \rightarrow 0 \quad (6)$$

in probability as $n \rightarrow \infty$ for some β which does not involve the observations.

In the standard case of independent and identically distributed observations, this definition reduces to that of Rao (1973). Furthermore, the motivation behind the general definition is the same as for the independence case. Under the conditions we have imposed

$$\left\{ d \log L_n(\theta) / d\theta = \sum_{i=1}^n u_i \right\}$$

is a martingale and a Central Limit result for martingales gives that

$$I_n^{-\frac{1}{2}}(\theta) \frac{d \log L_n(\theta)}{d\theta} = \left(\sum_{i=1}^n u_i \right) / \left\{ \sum_{i=1}^n E(u_i^2 | \mathcal{F}_{i-1}) \right\}^{\frac{1}{2}} \quad (7)$$

converges in distribution to $N(0, 1)$ under certain regularity conditions. This result, together with (6), will ensure that

$$I_n^{\frac{1}{2}}(\theta) (T_n - \theta) \rightarrow N\{0, \beta^2(\theta)\}$$

in distribution. Interestingly, the random norming $I_n^{\frac{1}{2}}(\theta)$ in (6) and (7) gives rise to much more general Central Limit results than are provided by the constant norming

$$[E\{I_n(\theta)\}]^{\frac{1}{2}} = \left\{ E \left(\sum_{i=1}^n u_i \right)^2 \right\}^{\frac{1}{2}} = \left\{ \sum_{i=1}^n E(u_i^2) \right\}^{\frac{1}{2}}.$$

If the condition (6) is strengthened to convergence in the mean of order two, then equivalent forms are

$$\text{corr} \left[I_n^{\frac{1}{2}}(\theta) (T_n - \theta), \{I_n(\theta)\}^{-\frac{1}{2}} \frac{d \log L_n(\theta)}{d\theta} \right] \rightarrow 1, \quad (8)$$

or alternatively,

$$E \left\{ (T_n - \theta) \frac{d \log L_n(\theta)}{d\theta} \right\} / \left(E \{I_n(\theta) (T_n - \theta)^2\} E \left[(I_n(\theta))^{-1} \left(\frac{d \log L_n(\theta)}{d\theta} \right)^2 \right] \right)^{\frac{1}{2}} \rightarrow 1. \quad (9)$$

A detailed discussion of the general properties of asymptotically efficient estimators will be given elsewhere.

To check asymptotic efficiency via the definition (6), the vital piece of information is a suitably tractable expression for $d \log L_n(\theta)/d\theta$. For the branching process application we shall consider the class of offspring distributions depending on a single parameter which we can write in the form

$$\text{pr}(Z_1 = k) = p_k(\theta) \quad (k = 0, 1, \dots),$$

where $\theta = E(Z_1)$. We shall further suppose that the following three conditions are satisfied:

- (i) the set S of k 's for which $p_k(\theta)$ is positive is independent of θ , $1 < \theta < \infty$;
- (ii) for all k , $(d/d\theta)p_k(\theta)$ exists for all $1 < \theta < \infty$;
- (iii) $\sum_{k_1 \in S} \dots \sum_{k_n \in S} \prod_{i=1}^n \text{pr}(Z_1 = k_i | Z_0 = k_{i-1})$ ($k_0 = 1$) can be differentiated under the summation signs.

In this situation, since

$$\text{pr}(Z_1 = k_1, \dots, Z_n = k_n) = \prod_{i=1}^n \text{pr}(Z_1 = k_i | Z_0 = k_{i-1}),$$

we have

$$\frac{d \log L_n(\theta)}{d\theta} = \sum_{i=1}^n \frac{d}{d\theta} \log Q(Z_{i-1}, Z_i), \quad (10)$$

where $Q(k_{i-1}, k_i) = \text{pr}(Z_1 = k_i | Z_0 = k_{i-1})$.

Now,

$$Q(k_{i-1}, k_i) = \sum' p_{j_1} p_{j_2} \dots p_{j_{k_{i-1}}},$$

where the sum is over $j_1, \dots, j_{k_{i-1}} \in S$, $j_1 + \dots + j_{k_{i-1}} = k_i$, so that

$$\frac{dQ(k_{i-1}, k_i)}{d\theta} = \sum' \left(\sum_{r=1}^{k_{i-1}} \frac{d \log p_{j_r}}{d\theta} \right) p_{j_1} \dots p_{j_{k_{i-1}}}. \quad (11)$$

To proceed further we consider the class of Galton-Watson processes whose offspring distributions are power series distributions. That is, the class where

$$p_j = a_j \lambda^j \{f(\lambda)\}^{-1} \quad (j = 0, 1, \dots; \lambda > 0),$$

where $a_j \geq 0$ and $f(\lambda) = \sum a_j \lambda^j$. For this class we readily find that

$$\theta = \lambda f'(\lambda) \{f(\lambda)\}^{-1}, \quad \sigma^2 = \{(d/d\theta) \log \lambda\}^{-1}$$

and

$$(d/d\theta) p_j = \sigma^{-2} (j - \theta) p_j. \quad (12)$$

Power series distributions are in fact characterized by the property (12). Details and references concerning the class of power series distributions may be obtained from Ord (1972, Chapter 6) and Johnson & Kotz (1969, Chapter 2). Examples are the geometric, binomial, Poisson, negative binomial and logarithmic distributions. An exception, notable from the point of view of the theory of the Galton-Watson process, is the distribution with linear functional probability generating function; for fixed b , $f(s) = 1 - b(1-p)^{-1} + bs(1-ps)^{-1}$. Here $\theta = b(1-p)^{-2}$.

Using (12) in (11) we find that

$$(d/d\theta) \log Q(Z_{i-1}, Z_i) = \sigma^{-2}(Z_i - \theta Z_{i-1}), \tag{13}$$

so that

$$\begin{aligned} \frac{d \log L_n(\theta)}{d\theta} &= \sigma^{-2} \sum_{i=1}^n (Z_i - \theta Z_{i-1}) \\ &= \sigma^{-2} (Y_n - 1 - \theta Y_{n-1}) = \sum_{i=1}^n u_i, \end{aligned} \tag{14}$$

say. Notice that, for each i ,

$$E(u_i | \mathcal{F}_{i-1}) = E(Z_i | Z_{i-1}) - \theta Z_{i-1} = 0,$$

which is the martingale property and

$$E(u_i^2 | \mathcal{F}_{i-1}) = E\{(Z_i - \theta Z_{i-1})^2 | Z_{i-1}\} = \sigma^2 Z_{i-1},$$

so that

$$I_n(\theta) = \sigma^{-2} Y_{n-1} \tag{15}$$

and hence

$$\frac{d \log L_n(\theta)}{d\theta} = I_n(\theta) (\hat{\theta}_n - \theta). \tag{16}$$

It is then clear that the maximum likelihood estimator $\hat{\theta}_n$ is asymptotically efficient in the case of power series distributions, (6) being satisfied with $\beta(\theta) = 1$.

Of course if there was another asymptotically efficient estimator T_n of θ for which (6) was satisfied with a $\beta(\theta) < 1$, then this estimator would be preferred to $\hat{\theta}_n$ since, in view of (1), we would have $I_n^{\frac{1}{2}}(\theta) (T_n - \theta) \rightarrow N\{0, \beta^2(\theta)\}$, while $I_n^{\frac{1}{2}}(\theta) (\hat{\theta}_n - \theta) \rightarrow N(0, 1)$, both in distribution.

It certainly does not seem possible to establish the nonexistence of such a T_n which is better than $\hat{\theta}_n$ unless $\{I_n(\theta) (\hat{\theta}_n - \theta)^2\}$ is uniformly integrable or equivalently, as $n \rightarrow \infty$,

$$E\{I_n(\theta) (\hat{\theta}_n - \theta)^2\} \rightarrow 1$$

using Theorem 5.4 of Billingsley (1968), since $I_n(\theta) (\hat{\theta}_n - \theta)^2 \rightarrow \chi_1^2$ in distribution as a consequence of (1). In fact, we need even more conditions. If, indeed, $\{I_n(\theta) (\hat{\theta}_n - \theta)^2\}$ is uniformly integrable and $\{T_n\}$ satisfies (6) strengthened to convergence in the mean of order two, then

$$E\{I_n(\theta) (T_n - \theta)^2\} - \beta^2(\theta) \rightarrow 0 \tag{17}$$

and

$$E\{I_n(\theta) (T_n - \theta)^2\} - 2\beta(\theta) E\{(T_n - \theta) I_n(\theta) (\hat{\theta}_n - \theta)\} + \beta^2(\theta) \rightarrow 0. \tag{18}$$

Furthermore, if T_n is unbiased for θ and the right-hand side of

$$\theta = \sum_{k_1 \in S} \dots \sum_{k_n \in S} T_n(k_1, \dots, k_n) \prod_{i=1}^n Q(k_{i-1}, k_i)$$

can be differentiated under the summations, then from (11), (12) and (15),

$$1 = E\{T_n I_n(\theta) (\hat{\theta}_n - \theta)\} = E\{(T_n - \theta) I_n(\theta) (\hat{\theta}_n - \theta)\}. \tag{19}$$

We thus obtain from (17) and (18) that $\beta(\theta) = 1$ and hence $I_n^{\frac{1}{2}}(\theta) (T_n - \hat{\theta}_n) \rightarrow 0$ in the mean of order two.

Notice that the alternative conditions (8) and (9) provide a method of assessing an estimator by computing an appropriate correlation or analogue. For example, when

$$T_n = \check{\theta}_n = Z_n Z_{n-1}^{-1}$$

we have, using (19),

$$\begin{aligned} E\{I_n^{\frac{1}{2}}(\theta) (\check{\theta}_n - \theta) I_n^{\frac{1}{2}}(\theta) (\hat{\theta}_n - \theta)\} / [E\{I_n(\theta) (\check{\theta}_n - \theta)^2\} E\{I_n(\theta) (\hat{\theta}_n - \theta)^2\}]^{\frac{1}{2}} \\ = [E\{I_n(\theta) (\check{\theta}_n - \theta)^2\} E\{I_n(\theta) (\hat{\theta}_n - \theta)^2\}]^{-\frac{1}{2}}, \end{aligned}$$

and, using Theorem 5.3 of Billingsley (1968),

$$\liminf E\{I_n(\theta) (\check{\theta}_n - \theta)^2\} \geq \theta(\theta - 1)^{-1}$$

since $I_n^{\frac{1}{2}}(\theta) (\check{\theta}_n - \theta) \rightarrow N\{0, \theta(\theta - 1)^{-1}\}$ in distribution, while

$$\liminf E\{I_n(\theta) (\hat{\theta}_n - \theta)^2\} \geq 1,$$

since $I_n^{\frac{1}{2}}(\theta) (\hat{\theta}_n - \theta) \rightarrow N(0, 1)$ in distribution, so that

$$\limsup E\{I_n^{\frac{1}{2}}(\check{\theta}_n - \theta) I_n^{\frac{1}{2}}(\theta) (\hat{\theta}_n - \theta)\} / [E\{I_n(\theta) (\check{\theta}_n - \theta)^2\} E\{I_n(\theta) (\hat{\theta}_n - \theta)^2\}]^{\frac{1}{2}} \leq (\theta - 1)^{\frac{1}{2}} \theta^{-\frac{1}{2}}.$$

4. ESTIMATION OF THE GROWTH RATE IN A PURE BIRTH PROCESS

Let X_t be the population size at time t in a pure birth process with rate parameter λ . That is, X_t is a Markov process for which

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

$i = 1, 2, \dots; \lambda > 0$. Suppose that $\text{pr}(X_0 = q) = 1$, where q is a fixed positive integer. Various sampling schemes have been considered for the estimation of λ in this model and the reader is referred to the recent work of Keiding (1974) for references and details. We shall here discuss the case where the process is sampled at equidistant time points $0, \tau, \dots, k\tau$. Then, as is well known (Harris, 1963, pp. 103-4), the observations form a Galton-Watson process $Z_n = X_{n\tau}$ with geometric offspring distribution

$$\text{pr}(Z_1 = i | Z_0 = 1) = e^{-\lambda\tau}(1 - e^{-\lambda\tau})^{i-1} \quad (i = 1, 2, \dots)$$

and $\text{pr}(Z_1 = 0 | Z_0 = 1) = 0$. This is a power series distribution for which we have

$$E(Z_1 | Z_0 = 1) = e^{\lambda\tau}, \quad \text{var}(Z_1 | Z_0 = 1) = e^{\lambda\tau}(e^{\lambda\tau} - 1).$$

Furthermore, it is clear from results quoted above that the maximum likelihood estimator of λ is

$$\hat{\lambda}_{k\tau} = \tau^{-1} \log \{(X_\tau + \dots + X_{k\tau}) / (X_0 + \dots + X_{(k-1)\tau})\}.$$

Also, from Keiding (1974), subject to a minor correction made in (21) below, almost surely

$$\hat{\lambda}_{k\tau} \rightarrow \lambda, \tag{20}$$

$$\tau(X_0 + \dots + X_{(k-1)\tau})^{\frac{1}{2}} (\hat{\lambda}_{k\tau} - \lambda) \rightarrow N(0, 1 - e^{-\lambda\tau}) \tag{21}$$

in distribution.

Writing $L_n(\lambda)$ for the likelihood function, we find in this case that

$$\begin{aligned}\frac{d \log L_k(\lambda)}{d\lambda} &= \frac{\tau e^{-\lambda\tau}}{1 - e^{-\lambda\tau}} (Y_{k\tau} - q - e^{\lambda\tau} Y_{(k-1)\tau}) \\ &= \frac{\tau e^{-\lambda\tau}}{1 - e^{-\lambda\tau}} Y_{(k-1)\tau} (e^{\tau\hat{\lambda}_{k\tau}} - e^{\lambda\tau}),\end{aligned}$$

while $I_k(\lambda) = \tau^2(1 - e^{-\lambda\tau})^{-1} Y_{(k-1)\tau}$. Then,

$$\begin{aligned}I_k^{-1}(\lambda) \left\{ \hat{\lambda}_{k\tau} - \lambda - I_k^{-1}(\lambda) \frac{d \log L_k(\lambda)}{d\lambda} \right\} &= \tau(1 - e^{-\lambda\tau})^{-\frac{1}{2}} Y_{(k-1)\tau}^{\frac{1}{2}} \{ \hat{\lambda}_{k\tau} - \lambda - \tau^{-1}(e^{\tau(\hat{\lambda}_{k\tau} - \lambda)} - 1) \} \\ &\rightarrow 0\end{aligned}$$

in probability, using (20) and (21). This gives the asymptotic efficiency of $\hat{\lambda}_{k\tau}$, for fixed τ . Nevertheless one can, not unexpectedly, obtain a more efficient estimate if the process is observed continuously. Let

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

If the process starts with q individuals at time zero and is observed continuously up to time t , the maximum likelihood estimator of λ is $(X_t - q) S_t^{-1}$ and for $t = k\tau$, the efficiency of this estimator with respect to $\hat{\lambda}_{k\tau}$ as $k \rightarrow \infty$ is $(1 - e^{-\lambda\tau})^2 (\lambda\tau)^{-2} e^{\lambda\tau}$ from Keiding (1974) with modification as in (21) above.

REFERENCES

- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: Wiley.
- BÜHLER, W. (1969). Ein zentraler Grenzwertsatz für verzweigungsprozesse. *Z. Wahr. verw. Geb.* **11**, 139–41.
- HARRIS, T. E. (1948). Branching Processes. *Ann. Math. Statist.* **19**, 474–94.
- HARRIS, T. E. (1963). *The Theory of Branching Processes*. Berlin: Springer-Verlag.
- HEYDE, C. C. (1970). Extension of a result of Seneta for the supercritical Galton–Watson process. *Ann. Math. Statist.* **41**, 739–42.
- HEYDE, C. C. (1974). On estimating the variance of the offspring distribution in a simple branching process. *Adv. Appl. Prob.* **6**, 421–33.
- HEYDE, C. C. & LESLIE, J. R. (1971). Improved classical limit analogues for Galton–Watson processes with or without immigration. *Bull. Aust. Math. Soc.* **5**, 145–55.
- JAGERS, P. (1973). A limit theorem for sums of random numbers of i.i.d. random variables. In *Mathematics. Essays in Honour of Harald Bergström*, pp. 33–9. Gotëborg: Chalmers Institute of Technology.
- JOHNSON, N. L. & KOTZ, S. (1969). *Discrete Distributions*. Boston: Houghton Mifflin.
- KEIDING, N. (1974). Estimation in the birth process. *Biometrika* **61**, 71–80.
- NAGAEV, A. V. (1967). On estimating the number of direct descendents of a particle in a branching process. *Theory Prob. Appl.* **12**, 314–20.
- ORD, J. K. (1972). *Families of Frequency Distributions*. London: Griffin.
- RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.

[Received February 1974. Revised October 1974]