The Annals of Statistics
1975, Vol. 3, No. 5, 1183–1187

ESTIMATION OF THE VARIANCE OF A BRANCHING PROCESS[1]

BY JEAN-PIERRE DION

Université du Québec à Montréal

Assume given the $(n+1)$-first generation sizes of a supercritical branching process. An estimator is proposed for the variance σ^2 of this process when the mean is known. It is shown to be unbiased, consistent and asymptotically normal. From that one deduces a consistent and asymptotically normal estimator for σ^2 in the case of an unknown mean. Finally, the maximum likelihood estimator of σ^2, based on a richer sample, is found and asymptotic properties are studied.

1. Introduction. Let $\{Z_n; n = 0, 1, 2, \cdots\}$ be a Galton–Watson process $(Z_0 \equiv 1)$. Let $p_k = P(Z_1 = k)$, $k = 0, 1, \cdots$ be a completely free offspring distribution only subject to the conditions $m > 1$ and $0 < \sigma^2 < \infty$, where m and σ^2 are repectively the mean and the variance of the offspring distribution. The number of individuals of the nth generation is represented by Z_n.

Consider a sample $\{Z_0, Z_1, \cdots, Z_{n+1}\}$ formed by the $(n+1)$ first generation sizes of this process. The problem of point estimation for the mean has been treated by Lotka (1939), Nagaev (1967), T. E. Harris (1948) and Heyde (1970). The asymptotic distribution of these estimators and confidence intervals for the mean were given by Dion (1972, 1974). Results by Jagers (1973) on the age dependent branching process and by Keiding (1974) on the birth process, are related to this. Unfortunately, these asymptotic results are functions of σ^2 and as such are of limited value when σ^2 is also unknown.

Nice estimators for σ^2 are proposed both when the mean is known and unknown and a confidence interval for the mean is deduced when the variance is unknown. The Nagaev estimator (1967) for the mean is used in Section 2 to construct an unbiased, consistent and asymptotically normal estimator for σ^2; this estimator is a function of m. In Section 3, another consistent and asymptotically normal estimator for σ^2 is given; it uses both Nagaev and Harris estimators for the mean but does not involve the mean itself. Some of the results of Sections 2 and 3 were also obtained independently by Heyde (1974). He proposes the same estimator in the case of known mean, and uses the Lotka–Nagaev estimator instead of the Harris estimator when the mean is unknown.

Finally, we consider a richer sample $\{Z_{jk} : j = 0, 1, 2, \cdots, n; k = 0, 1, 2, \cdots\}$, where Z_{jk} is the number of individuals of the jth generation having exactly k

Received March 1974; revised February 1975.

[1] This research was partially supported by the National Research Council of Canada, under grant CNR A 8852.

AMS 1970 *subject classification.* Primary 62.15, 62.70; Secondary 60.67.

Key words and phrases. Branching process, estimation of variance, estimation of mean, asymptotic normality, maximum likelihood estimation.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

direct descendants. The maximum likelihood estimator of σ^2 is then found and shown to be consistent and asymptotically normal.

2. Estimation with known mean. Define, for each $k = 0, 1, 2, \cdots, n$, $\tau_k = ((Z_{k+1}/Z_k) - m)^2 Z_k$, and put $\bar{\sigma}_n^2 = (\sum_{k=0}^n \tau_k)/(n+1)$.

THEOREM 1. *Assume* $p_0 = 0$, $m > 1$, $0 < \sigma^2 < \infty$ *and* $E(Z_1^4) < +\infty$. *Then*

$$E(\bar{\sigma}_n^2) = \sigma^2 \tag{1}$$

and

$$\operatorname{Var}(\bar{\sigma}_n^2) = (n+1)^{-1}[2\sigma^4 + (n+1)^{-1}(\operatorname{Var}(Z_1 - m)^2 - 2\sigma^4) \textstyle\sum_{k=0}^n E(Z_k^{-1})] . \tag{2}$$

PROOF. We will use repeatedly the fact that given Z_k, Z_{k+1} is distributed as the sum of Z_k independent copies of Z_1, call them $\xi_1^{(k)}, \xi_2^{(k)}, \cdots, \xi_{Z_k}^{(k)}$. Assume (since such a space can be constructed) that we are working on a space where the random variables $\xi_i^{(k)}$ exists and are i.i.d. and that $Z_{k+1} = \sum_{i=1}^{Z_k} \xi_i^{(k)}$; we will omit the superscript (k) as a matter of notational convenience.

First, we have

$$\begin{aligned} E(\tau_k \mid Z_k) &= \frac{1}{Z_k} E((Z_{k+1} - mZ_k)^2 \mid Z_k) \\ &= \frac{1}{Z_k} \operatorname{Var}(Z_{k+1} \mid Z_k) = \frac{1}{Z_k} \operatorname{Var}(\textstyle\sum_1^{Z_k} \xi_i \mid Z_k) = \sigma^2 \end{aligned}$$

so that $E(\tau_k) = \sigma^2$; secondly,

$$\begin{aligned} E(\tau_i \tau_j) &= E(E(\tau_i \tau_j \mid Z_i, Z_{i+1}, \cdots, Z_j)) \\ &= E(\tau_i E(\tau_j \mid Z_j)) = E(\tau_i \sigma^2) = \sigma^4 , \quad \forall\, i < j , \end{aligned}$$

which shows that the τ_k's are uncorrelated. Thus

$$\operatorname{Var}(\bar{\sigma}_n^2) = (n+1)^{-2} \textstyle\sum_{k=0}^n \operatorname{Var}(\tau_k) .$$

There remains only to show that $\operatorname{Var}(\tau_k \mid Z_k) = Z_k^{-1} \operatorname{Var}(Z_1 - m)^2 + Z_k^{-1} 2(Z_k - 1)\sigma^4$, since $\operatorname{Var} E(\tau_k \mid Z_k) = 0$. But $\operatorname{Var}(\tau_k \mid Z_k) = Z_k^{-2} \operatorname{Var}((\sum_{i=1}^{Z_k} (\xi_i - m))^2 \mid Z_k) = Z_k^{-2} \operatorname{Var}[(\sum_{i=1}^{Z_k} (\xi_i - m)^2 + 2 \sum_{i<j} (\xi_i - m)(\xi_j - m)) \mid Z_k]$.

Now, $\forall\, i, j, k, j < k$, $(\xi_i - m)^2$ and $(\xi_j - m)(\xi_k - m)$ are uncorrelated. Also, $\forall\, i < j$, $k < l$ and $(i, j) \neq (k, l)$, $(\xi_i - m)(\xi_j - m)$ and $(\xi_k - m)(\xi_l - m)$ are uncorrelated, so $\operatorname{Var}(\tau_k \mid Z_k) = Z_k^{-2}[\operatorname{Var}(Z_1 - m)^2 + 2Z_k(Z_k - 1)\sigma^4]$, from which the conclusion follows.

It is immediate from Theorem 1 and Tchebyschev's inequality that $\bar{\sigma}_n^2 \to_P \sigma^2$. Furthermore, there are some reasons to believe that if k_n is big enough the $\{\tau_k : k > k_n\}$ are asymptotically independent and identically distributed (see [3]). So it would seem natural to have asymptotic normality for $\bar{\sigma}_n^2$. The next theorem, which could easily be obtained using a central limit theorem for martingales due to Billingsley (1961, page 52), states that this is indeed true. The proof is omitted since it is a special case of Heyde's Theorem 2 (1974).

THEOREM 2. *Let* $p_0 = 0$, $m > 1$, $0 < \sigma^2 < \infty$ *and* $E(Z_1^6) < +\infty$. *Then,*

$$\frac{((n+1)/2)^{\frac{1}{2}}(\bar{\sigma}_n^2 - \sigma^2)}{\sigma^2} \to_D N(0, 1) .$$

3. Estimation with unknown mean. Consider now the case where the mean is unknown. The maximum likelihood estimator of m is $\hat{m}_n = (Z_1 + \cdots + Z_{n+1})/(Z_0 + \cdots + Z_n)$. Put $\tilde{\sigma}_n^2 = (n+1)^{-1} \sum_{k=0}^{n} Z_k((Z_{k+1}/Z_k) - \hat{m}_n)^2$.

THEOREM 3. *Let* $p_0 = 0$, $m > 1$, *and* $0 < \sigma^2 < \infty$. *Then,*

$$|\bar{\sigma}_n^2 - \tilde{\sigma}_n^2|(n+1)^{1-\varepsilon} \to_P 0 \quad \forall \varepsilon > 0 .$$

PROOF.

$$\forall \varepsilon > 0$$

$$\begin{aligned}
(\bar{\sigma}_n^2 - \tilde{\sigma}_n^2)&(n+1)^{1-\varepsilon} \\
&= (n+1)^{-\varepsilon} \textstyle\sum_{k=0}^{n} Z_k[((Z_{k+1}/Z_k) - m)^2 - ((Z_{k+1}/Z_k) - \hat{m}_n)^2] \\
&= (n+1)^{-\varepsilon} \textstyle\sum_{k=0}^{n} Z_k[(\hat{m}_n - m)(2(Z_{k+1}/Z_k) - (\hat{m}_n + m))] \\
&= (n+1)^{-\varepsilon}(\hat{m}_n - m)[2(Z_1 + \cdots + Z_{n+1}) - (\hat{m}_n + m)(Z_0 + \cdots + Z_n)] \\
&= (Z_0 + \cdots + Z_n)(\hat{m}_n - m)^2(n+1)^{-\varepsilon}
\end{aligned}$$

and this tends to zero in probability, since $((Z_0 + \cdots + Z_n)/\sigma^2)(\hat{m}_n - m)^2 \to_D \chi_1^2$ by Theorem 3.1 of [3]. □

(A stronger result would follow by using the full generality of Theorem 3.1 of [3]).

NOTE. This proof also showed that $\tilde{\sigma}_n^2 < \bar{\sigma}_n^2$, a.s. Since $\bar{\sigma}_n^2$ is an unbiased estimator of σ^2, $\tilde{\sigma}_n^2$ will tend to underestimate σ^2 (although it is consistent).

As an immediate consequence of Theorem 3 one has the following.

COROLLARY. *Under the hypotheses of Theorem* 2,

$$\frac{((n+1)/2)^{\frac{1}{2}}(\tilde{\sigma}_n^2 - \sigma^2)}{\sigma^2} \to_D X ,$$

where X *stands for a* rv *normal* (0, 1), *and the convergence is in distribution.*

4. Maximum likelihood estimation. In some situations we may have a richer sample. We may know $\{Z_{jk}, j = 0, 1, 2, \cdots, n, k = 0, 1, 2, \cdots\}$ where Z_{jk} represents the number of individuals of the jth generation having exactly k direct descendants. With respect to this sample Harris (1948) showed that $\hat{p}_i = (\sum_{j=0}^{n} Z_{ji})/(Z_0 + \cdots + Z_n)$ is the maximum estimator for p_i, $i = 0, 1, 2, \cdots$. Thus the maximum likelihood estimator for σ^2, say $\hat{\sigma}_n^2$, will be $\sum_{k=0}^{\infty} (k - \hat{m}_n)^2 \hat{p}_k$ when the mean is unknown.

The following theorem suggests that this estimator would perform better than the one given in Section 3. (This is only natural since it uses more information.)

Let A be the set of non-extinction of the process, i.e. $A = (\omega : Z_k(\omega) > 0, k = 1, 2, \cdots\}$.

Suppose $m > 1$ (*p_0 need not be zero*). Then $P(A) > 0$ and $P_A(\cdot) = P(\cdot \mid A)$ is well defined. Put $S(w) = P_A(W < w)$, where W is the almost sure limit of the martingale $\{Z_n/m^n\}$.

THEOREM 4. *Assume $m > 1$, $0 < \sigma^2 < \infty$, and $E(Z_1^4) < +\infty$. Then for any probability measure Q absolutely continuous with respect to P_A,*

$$Q\left\{\omega : \frac{(\hat{\sigma}_n^2 - \sigma^2)(Z_0 + \cdots + Z_n)^{\frac{1}{2}}}{(\mathrm{Var}\,(Z_1 - m)^2)^{\frac{1}{2}}} \leqq x\right\} \to \Phi(x)\,, \qquad n \to \infty \tag{1}$$

and

$$Q\left\{\omega : \frac{(\hat{\sigma}_n^2 - \sigma^2)(1 + m + \cdots + m^n)^{\frac{1}{2}}}{(\mathrm{Var}\,(Z_1 - m)^2)^{\frac{1}{2}}} \leqq x\right\} \to \int_0^\infty \Phi(x(w^{\frac{1}{2}})\, dS(w) \tag{2}$$

with $\Phi(x) = 1/(2\pi)^{\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt$.

Furthermore, the conclusion of the theorem remains valid if $Q(\cdot)$ is replaced by $P(\cdot \mid Z_{n+1} > 0)$.

PROOF. As, by Theorem 3.1 of [3],

$$P_A\left\{\omega : \frac{(Z_0 + \cdots + Z_n)^{\frac{1}{2}}}{\sigma}(\hat{m}_n(\omega) - m) \leqq x\right\} \to \Phi(x)\,,$$

it follows that $|\hat{\sigma}_n^2 - \sum_{k=0}^{\infty}(k - m)^2\hat{p}_k|(Z_0 + \cdots + Z_n)^{\frac{1}{2}} \to_{P_A} 0$. So it is enough to show that the theorem holds if $\hat{\sigma}_n^2$ is replaced by $\sum_{k=0}^{\infty}(k - m)^2\hat{p}_k$.

Put $N_k = \sum_{k=0}^{n} Z_k$; it is the number of individuals among the $(1 + Z_1 + \cdots + Z_n)$-firsts, having exactly k direct descendants. With the help of a sequence of independent rv $\{\xi_i\}$, each distributed as Z_1, it is possible to express the distribution of $\sum_{k=0}^{\infty}(k - m)^2 N_k(\omega)/(Z_0 + \cdots + Z_n)$, as that of $\sum_{i=1}^{1+Z_1+\cdots+Z_n}(\xi_i(\omega) - m)^2/(Z_0 + \cdots + Z_n)$, $\omega \in A$.

Also $(Z_0 + \cdots + Z_n)/(1 + m + \cdots + m^n) \to W$ a.s. (see [5]) and $P(W > 0) = P(A)$ so that $P(\cdot \mid W > 0) = P_A(\cdot)$. The conclusion follows by the results of [3] (namely by Theorem 1, Theorem 2, and the lemma of Section 2).

Acknowledgments. The author has benefited from discussions with Professors A. Joffe and R. Roy. He is also indebted to the referee for many comments and for pointing out the existence of Heyde's forthcoming paper.

REFERENCES

[1] BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes.* Univ. of Chicago Press.

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

[3] DION, J.-P. (1974). Estimation of the mean and the initial probabilities of a branching process. *J. Appl. Probability* **11** 687-694.

[4] HARRIS, T. E. (1948). Branching processess. *Ann. Math. Statist.* **19** 474-494.

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

[6] HEYDE, C. C. (1974). On estimating the variance of the offspring distribution in a simple branching process. *Advances in Appl. Probability* **3** 421, 433.

[7] JAGERS, P. (1973). Maximum likelihood estimation of the reproduction distribution in branching processes and the extent of disintegration in cell proliferation. Technical report 1973-17, Dept. of Math. Chalmers Univ. of Tech. and Univ. Göteborg.

[8] KEIDING, N. (1974). Estimation in the birth process. *Biometrika* **61** 71-80.

[9] LOTKA, A. J. (1939). Théorie analytique des associations biologiques. *Actualités scientifiques et industrielles* **2** 780, 123-136.

[10] NAGAEV, A. V. (1967). On estimating the expected number of direct descendants of a particle in a branching process. *Theor. Probability Appl.* **12** 312-320.

DÉPARTMENT DE MATHÉMATIQUES
UNIVERSITÉ DU QUÉBEC À MONTRÉAL
CASE POSTALE 8888
MONTRÉAL, P., QUÉBEC
CANADA H3C 3P8