IMS Collections
Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz
Vol. 4 (2008) 17–30

DOI: 10.1214/074921708000000273

Conditional Limit Laws and Inference for Generation Sizes of Branching Processes

P. E. Ney[1] **and A. N. Vidyashankar**[2,*]

University of Wisconsin–Madison and Cornell University

Abstract: Let $\{Z_n : n \geq 0\}$ denote a single type supercritical branching process initiated by a single ancestor. This paper studies the asymptotic behavior of the history of generation sizes conditioned on different notions of information about the "current" population size. A "suppression property" under the large deviation conditioning, namely that $R_n \equiv Z_{n+1}/Z_n > a$, is observed. Furthermore, under a more refined conditioning, the asymptotic aposteriori distribution of the original offspring distribution is developed. Implications of our results to conditional consistency property of age is discussed.

1. Introduction

The purpose of this note is to provide information on the history of the generation sizes given some "present" information concerning the branching process. We begin with a description of the process. Let $\{Z_n : n \geq 1\}$ denote a single type branching process initiated by a single ancestor. Let $\{p_j : j \geq 1\}$ denote the offspring distribution, that is $P(Z_1 = j) = p_j$. For $0 \leq s \leq 1$, let $f(s) = E(s^{Z_1} \mid Z_0 = 1)$ denote the probability generating function. Let $m = E(Z_1) = f'(1-)$, where $f'(\cdot)$ denotes the derivative of $f(\cdot)$. We denote by q the probability of extinction; then it is well-known that q satisfies the fixed point equation $f(s) = s$. It is also well-known that the process $\{Z_n : n \geq 1\}$ can be defined recursively, using a collection $\{\xi_{k,j}, k \geq 1, j \geq 1\}$ of independent and identically distributed (i.i.d) non-negative integer valued random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ as follows: $Z_0 = 1$ and for $n \geq 0$

$$Z_{n+1} = \sum_{j=1}^{Z_n} \xi_{n,j}, \tag{1.1}$$

where $\xi_{n,j}$ is interpreted as the number of children produced by the jth parent in the nth generation; and $P(\xi_{0,1} = j) = p_j$. This implies that the generating function of the nth generation population size is given by the n-fold iteration of $f(\cdot)$; i.e., $E(s^{Z_n}) = f_n(s) = f(f(f...(s)))$, $0 \leq s \leq 1$. Let S denote the survival set of the process; i.e., $S = \{\omega : Z_n(\omega) \to \infty\}$. Then $P(S) = 1 - q$. We will assume in this paper that the process is supercritical; that is $m > 1$ and for the sake of exposition, that $p_0 = 0$. This implies that $P(S) = 1$.

Let $W_n = Z_n/m^n$. Let $\mathcal{G}_n$ denote the sigma field generated by the first n generation sizes, namely, $\{Z_0, Z_1, \cdots Z_n\}$. Then it is well-known that $\{(W_n, \mathcal{G}_n) : n \geq 1\}$

*Research Supported in part by a grant from NSF DMS 000-03-07057 and also by grants from the NDCHealth Corporation.
[1]Department of Mathematics, University of Wisconsin, Madison, WI 53706-1388.
[2]Department of Statistical Science, Cornell University, Ithaca, NY 14853-4201.
AMS 2000 subject classifications: Primary 60J80; secondary 60F10
Keywords and phrases: branching processes, functional equation, large deviations, uniform local limit theorem, conditional limit laws, age, conditional consistency, a posteriori distribution.

is a non-negative martingale sequence and hence converges with probability one to a random variable W. By the Kesten-Stigum theorem (see [3]), a necessary and sufficient condition for W to be non-trivial is that $E(Z_1 \log Z_1) < \infty$. Furthermore, W has density $w(\cdot)$ and $w(x) > 0$ for $x > 0$.

Let $R_n = Z_{n+1}/Z_n$. The quantity R_n is called the Nagaev estimator of the mean of the branching process and is its maximum likelihood estimator when (Z_n, Z_{n+1}) are observed. Large deviations of R_n (which will be relevant) have been studied in [1], [4], [12], [13], [9]. It is known from these papers that the large deviation behavior of R_n is different depending on whether $p_1 + p_0 > 0$ or $p_1 + p_0 = 0$. The case when $p_1 + p_0 > 0$ is called the Schröder case while $p_1 + p_0 = 0$ is called the Böttcher case.

Recent work in the area of evolutionary biology is concerned with statistically estimating the age of the last common ancestor using the fossil record ([11] and [15]). Such data are modeled using either discrete or continuous time branching processes or variants thereof. In these problems, an important difference between the *age* and the *divergence time* (to be defined below) have been observed. Furthermore, in the context of branching processes, an interesting recent work of [10] attempted to recreate the past based on the "present" observed generation size in order to determine the age of a population. One of the motivations for our study was to understand both these phenomena from the perspective of the conditional limit distributions. It turns out that, when viewed from the viewpoint of conditional limits, the difference between the age and the divergence time occurs if the population size is "smaller than expected" (see Remark 5 in Section 2). Now, "smaller than expected growth" is caused due to small values of Z_k for various values of k. This phenomenon is peculiar to the Schröder case. For this reason, we deal with the Schröder case in this paper and treat the Böttcher case in a different publication.

Gibbs conditioning principle in the context of i.i.d. random variables $\{X_n : n \geq 1\}$ defined on $\mathcal{R}$ is concerned with the asymptotic behavior of

$$P\left(X_1 \in \cdot \;\middle|\; \frac{S_n}{n} \in A\right), \qquad EX_1 \notin A, \tag{1.2}$$

or more generally, of

$$P\left\{(X_1, X_2, \ldots, X_{k_n}) \in \cdot \;\middle|\; \frac{S_n}{n} \in A\right\}, \tag{1.3}$$

where A is a Borel subset of $\mathcal{R}$, $S_n = \sum_{i=1}^{n} X_i$, and $k_n \to \infty$. In the context of branching processes, one approach is to replace $\frac{S_n}{n}$ by R_n; or by the joint event $\{R_n \in (\cdot), Z_n \in (\cdot)\}$. Now, unlike the i.i.d. case, two situations arise; namely the large n behavior of $P(Z_1 \in (\cdot) \mid R_n > a > m)$ and that of $P(\xi_{n,1} \in (\cdot) \mid R_n > a > m)$. We call the former case, a "global" conditional limit law while the latter a "local" conditional limit law. This paper is concerned with the global conditional limit laws.

The main technical tools needed in this paper are a uniform local limit theorem in the range of $Z_n \sim xm^n$ where x belongs to a bounded interval, and rates of convergence of generating functions. To facilitate our discussions in the next sections, we introduce more notation concerning the rate of decay of generating functions. Let, for $0 \leq s < 1$

$$Q_n(s) = \frac{f_n(s) - q}{\gamma^n}, \tag{1.4}$$

where $\gamma = f'(0)$. It is known that (see [3]) $\lim_{n\to\infty} Q_n(s) = Q(s)$ exists, with $Q(1) = \infty$. Furthermore, $Q(\cdot)$ admits a power series representation; that is,

$$Q(s) = \sum_{k\geq 1} \nu_k s^k. \tag{1.5}$$

When $p_0 = 0$, γ reduces to p_1. It follows from (1.5) that (see [3])

$$\lim_{n\to\infty} \frac{P(Z_n = k)}{\gamma^n} = \nu_k. \tag{1.6}$$

The quantities ν_k will show up at several places in the future sections.

The rest of the paper is organized as follows: Section 2 contains statements and discussion of the main results while Section 3 contains proofs. Section 4 deals with the limit laws concerning the age of a branching process.

2. Statement and Discussion of Results

We begin with the uniform local limit theorem which will be needed in the proof of Theorem 2 below. This is a uniform version of Theorem 4.1, Chapter II of [3]. Before we state the theorem, we need a definition. A sequence y_n of real numbers is said to be regular if $y_n m^n$ is an integer for all $n \geq 1$. In the following let $0 < c < d < \infty$ and $P_k(\cdot) = P(\cdot \mid Z_0 = k)$.

Theorem 2.1. *Assume that $E(Z_1 \log Z_1) < \infty$ and that $y_n \to \Delta_0$ is a regular sequence. Assume further that for every $n \geq 1$ there exists an $x_n \in [c, d]$ such that $x_n y_n$ is an integer. Let $\mathcal{C}_n = \{x \in [c, d] : xy_n \text{is an integer}\}$. Then the following hold:*

$$1. \lim_{n\to\infty} m^n P_k\left(Z_n = m^n y_n \left(1 + \frac{x_n}{m^n}\right)\right) = w^{\star k}(\Delta_0); \tag{2.1}$$

$$2. \lim_{n\to\infty} m^n \sup_{\{x\in\mathcal{C}_n\}} P_k\left(Z_n = m^n y_n \left(1 + \frac{x}{m^n}\right)\right) = w^{\star k}(\Delta_0); \tag{2.2}$$

$$3. \lim_{n\to\infty} m^n \inf_{\{x\in\mathcal{C}_n\}} P_k\left(Z_n = m^n y_n \left(1 + \frac{x}{m^n}\right)\right) = w^{\star k}(\Delta_0). \tag{2.3}$$

Turning to conditional limits, we have

Proposition 2.1. *Assume that $E(\exp(\theta Z_1)) < \infty$ for some $\theta > 0$ and that $p_1 > 0$. Then,*

$$\lim_{n\to\infty} P(Z_n = k \mid R_n > a > m) = \gamma(k) \geq 0, \tag{2.4}$$

where $\sum_{k\geq 1} \gamma(k) = 1$.

This suggests that the main contribution to $P(R_n > a)$ comes from “small” values of Z_n, which implies that the usual large deviation estimates and Cramer-type rate functions do not come into the calculation of (2.4). We refer to this as the *suppression property*, and it will manifest itself more subtly in future results. This leads at once to a “degeneracy” property on the early history, namely

Proposition 2.2. *Assume that $E(\exp(\theta Z_1)) < \infty$ for some $\theta > 0$ and that $p_1 > 0$. Then,*

$$\lim_{n\to\infty} P(Z_1 = k \mid R_n > a > m) = \delta_1(k). \tag{2.5}$$

Furthermore, for any $k(n) \to \infty$ and $k_n = \mathbf{o}(n)$

$$\lim_{n\to\infty} P((Z_1 = 1, \dots Z_{k_n} = 1) \mid R_n > a) = 1. \tag{2.6}$$

Remark 2.1. *The small values of Z_n are caused due to $p_0 + p_1$ being positive. Thus, the suppression property is inherent in the Schröder case.*

Our next proposition is concerned with the behavior of the distribution of Z_k when conditioned on $Z_n \in v_n[c, d]$, $c > 0$. In this note we consider the case where $v_n \sim m^n$.

Proposition 2.3. *Assume that $E(Z_1 \log_+ Z_1) < \infty$. Let, for $c > 0$,*

$$\pi_l(c, d) = \frac{\int_{cm^k}^{dm^k} w^{\star l}(x)dx}{\int_c^d w(x)dx}. \tag{2.7}$$

Then,

$$\lim_{n\to\infty} P(Z_k = l \mid Z_n \in m^n[c, d]) = \pi_l(c, d)P(Z_k = l), \tag{2.8}$$

where

$$\sum_{l\geq 1} \pi_l(c, d)P(Z_k = l) = 1. \tag{2.9}$$

Remark 2.2. *Specializing when $k = 1$ we get from the above proposition that, for any $c > 0$,*

$$\lim_{n\to\infty} P(Z_1 = l \mid Z_n \in m^n[c, d]) = \pi_l(c, d)p_l, \tag{2.10}$$

and

$$\lim_{n\to\infty} P(Z_1 = l \mid Z_n = cm^n) = \frac{w^{\star l}(mc)}{w(c)} p_l m. \tag{2.11}$$

Remark 2.3. *Note that the conditional limit mentioned above can be viewed as a change of measure of $P(Z_k = l)$, which is reminiscent of the change of measure in the classical Gibbs conditioning principle.*

The more subtle and interesting result comes from the combined conditioning, namely that $R_n > a, Z_n \in m^n[c, d]$;

Theorem 2.2. *Assume that $E(\exp(\theta Z_1)) < \infty$ for some $\theta > 0$. Then, for any $c > 0$,*

$$\lim_{n\to\infty} P(Z_1 = l \mid R_n > a, Z_n \in m^n[c, d]) = \frac{w^{\star l}(mc)}{w(c)} p_l m. \tag{2.12}$$

Remark 2.4. *Here the interaction between the events $R_n > a$ and $Z_n \in m^n[c, d]$ require estimates of R_n in the large deviation range, and uniform estimates of Z_n as in Theorem 1. Note that the limits in (2.11) and (2.12) are the same even though the conditioning sets are different. This result follows from the fact that the addition of $R_n > a$ to the conditioning $Z_n \in m^n[c, d]$ and the previously mentioned suppression property, forces the limit to be "as small as possible," i.e., $\pi_l(cm, dm)$ is replaced in (2.10) by*

$$\lim_{d\to c} \pi_l(cm, dm) = \frac{w^{\star l}(mc)}{w(c)} p_l m. \tag{2.13}$$

Remark 2.5. *The case when one conditions on $Z_n \in v_n[c,d]$ with $v_n = \mathbf{o}(m^n)$ leads to a different behavior. It turns out that the conditioned limit forces $Z_1 = \ldots = Z_{k_n} = 1$ up to $k_n < n - \log_m v_n$, and then Z_k starts to increase for $k > k_n$. One refers to this time as the divergence time (this is not a random time). Thus, under the conditioning in Propositions 1 and 2, divergence time is close to the "present," i.e., there is no growth until last few generations. The age of the branching process is defined to be the number of generations of the process at the time of observation. In Proposition 3, when the conditioning is $Z_n \in v_n[c,d]$ with $v_n \sim m^n$, the process starts to grow immediately, so the age and the divergence time agree, but the growth distribution changes according to the distribution in (2.8). In Theorem 2, when the conditioning includes $R_n > a$, the age and the divergence time again agree. But if $v_n = \mathbf{o}(m^n)$, the divergence time is of order $n - \log_m v_n$. This result is treated elsewhere. Divergence time is important in several biological and population models as mentioned in Section 1.*

Remark 2.6. *If instead of assuming $Z_0 = 1$ we take $P(Z_0 = k) = \pi(k)$, where $\sum_{k\geq 1} \pi(k) = 1$, then under the conditioning carried out above, the initial distribution $\pi(\cdot)$ will undergo a change of measure along similar lines to Propositions 2 and 3 and Theorem 2.*

3. Proofs

In this section we provide the proofs of our results in Section 2.

Proof of Proposition 1. Let $\overline{X}_k = k^{-1}\sum_{j=1}^{k} X_j$, X_j's are i.i.d. with distribution Z_1. Then, using Theorem 1 of [4], and (1.5) it follows that

$$\begin{aligned} P(Z_n = k \mid R_n > a) &= \frac{P(R_n > a \mid Z_n = k)P(Z_n = k)}{P(R_n > a)} \\ &= P(\overline{X}_k > a)\left\{\frac{P(Z_n = k)p_1^{-n}}{P(R_n > a)p_1^{-n}}\right\} \\ &\to P(\overline{X}_k > a)\frac{\nu_k}{L_a} = a_k, \end{aligned}$$

where ν_k is as in (1.6), and

$$\lim_{n\to\infty} p_1^{-n} P(R_n > a) = L_a = \sum_{k\geq 1} \nu_k P(\overline{X}_k > a).$$

Thus, $\sum_{k\geq 1} a_k = 1$. $\square$

Remark 3.1. *The exponential moment hypothesis in Proposition 1 (or Proposition 2 below) is not necessary. If $E(Z_1^r) < \infty$ and $p_1 m^r > 1$ then the above argument also goes through.*

Proof of Proposition 2. Let $k(n) = \mathbf{o}(n)$. Then using Theorem 1 of [4], it follows that

$$\begin{aligned} P(Z_{k(n)} = 1 \mid R_n > a) &= \frac{P(R_n > a \mid Z_{k(n)=1})P(Z_{k(n)} = 1)}{P(R_n > a)} \\ &= \frac{P(R_{n-k(n)} > a)}{P(R_n > a)} p_1^{k(n)} \\ &= \frac{p_1^{-(n-k(n))}P(R_{n-k(n)} > a)}{p_1^{-n}P(R_n > a)} p_1^{-k(n)} p_1^{k(n)} \\ &\to 1 \end{aligned}$$

implying (2.5) and (2.6). □

Proof of Proposition 3. Let $\mathcal{A}_k = m^k[c,d]$. Then, by Theorem II.4.1 in ([3])

$$\begin{aligned} P(Z_k = l \mid Z_n \in \mathcal{A}_n) &= \frac{P(Z_n \in \mathcal{A}_n \mid Z_k = l)P(Z_k = l)}{P(Z_n \in \mathcal{A}_n)} \\ &= \left\{ \frac{P_l(Z_{n-k} \in m^k \mathcal{A}_{n-k})}{P(Z_n \in \mathcal{A}_n)} \right\} P(Z_k = l) \\ &\to \left\{ \frac{\int_{cm^k}^{dm^k} w^{\star l}(x)dx}{\int_c^d w(x)dx} \right\} P(Z_k = l) \\ &= \pi_{l,k}(c,d)P(Z_k = l). \end{aligned}$$

To complete the proof of Proposition 3, we need to show that $\sum_{l\geq 1} \pi_l(c,d)P(Z_k = l) = 1$. This follows from Lemma 1 below. □

Remark 3.2. *One could take $c = 0$ in (2.10) in Proposition 3. In this case, the proof follows directly from the convergence in distribution of W_n to W and does not use Theorem II.4.1 from [3].*

Lemma 3.1. $\sum_{l\geq 1} P(Z_k = l) \int_{am^k}^{bm^k} w^{\star l}(x)dx = \int_a^b w(x)dx.$

Proof. Let $\phi(\theta) = E(e^{i\theta W})$. Then, by the inversion theorem ([7])

$$w^{\star l}(x) = \frac{1}{2\pi} \int_R e^{-i\theta x} (\phi(\theta))^l d\theta. \tag{3.1}$$

Now, integrating the LHS of (3.1) between am^k and bm^k we get

$$\int_{am^k}^{bm^k} w^{\star l}(x)dx = m^k \int_a^b w^{\star l}(ym^k)dy,$$

where the RHS of the above equation follows from the substitution $x = ym^k$. Now,

$$\begin{aligned} m^k \int_a^b w^{\star l}(ym^k)dy &= \frac{m^k}{2\pi} \int_a^b \int_R e^{-i\theta ym^k} (\phi(\theta))^l d\theta dy \\ &= \frac{1}{2\pi} \int_a^b \int_R e^{-i\eta y} (\phi(\eta/m^k))^l d\eta dy, \end{aligned}$$

where the last identity follows upon setting $\theta m^k = \eta$. Thus,

$$\begin{aligned} \sum_{l\geq 1} \int_{am^k}^{bm^k} w^{\star l}(x)dx P(Z_k = l) &= \int_a^b \sum_{l\geq 1} \int_R e^{-i\eta y} (\phi(\eta/m^k))^l d\eta dy P(Z_k = l) \\ &= \int_a^b \int_R e^{-i\eta x} \phi(\eta) d\eta dx \\ &= \int_a^b w(x)dx, \end{aligned}$$

where we used the identity (which is a consequence of the branching property)

$$\sum_{l\geq 1} \phi^l(\eta/m^k)P(Z_k = l) = \phi(\eta).$$

This completes the proof of the lemma. □

Proof of Theorem 1. By Theorem 4.2 in Chapter II of [3], it is sufficient to establish (ii) and (iii). We will establish (ii) as the proof of (iii) is similar. Let us set $j_n(x) = m^n y_n(1 + xm^{-n})$ and recall that, $\mathcal{C}_n = \{x \in [c,d] : xy_n \text{is an integer}\}$. Then, it follows from the assumptions of the theorem that

$$\lim_{n\to\infty} \sup_{x\in\mathcal{C}_n} |m^{-n} j_n(x) - \Delta_0| = 0. \tag{3.2}$$

Since $j_n(x)$ is an integer for all n and some $x \in [c,d]$, $P_k(Z_n = j_n(x))$ is not identically zero for all $x \in [c,d]$. Now, by the inversion theorem ([7])

$$P_k(Z_n = j_n(x)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f_n(e^{i\theta}))^k e^{-ij_n(x)\theta} d\theta. \tag{3.3}$$

Now, integrating by parts the RHS of (3.3) and using $f_n^k(e^{i\pi})e^{-il\pi} = f_n^k(e^{-i\pi})e^{il\pi}$ for all integers k and l, it follows that

$$P_k(Z_n = j_n(x)) = \frac{1}{2\pi}\left(\frac{k}{j_n(x)}\right) I(n,k,x), \tag{3.4}$$

where

$$I(n,k,x) = \int_{-\pi}^{\pi} (f_n(e^{i\theta}))^{k-1}(f_n'(e^{i\theta}))e^{-i(j_n(x)-1)\theta} d\theta. \tag{3.5}$$

Next, making a change of variable $\theta = tm^{-n}$ and setting $\psi_n(t) = E(e^{itW_n})$, (3.5) reduces to

$$I(n,k,x) = \int_{-\pi m^n}^{\pi m^n} (\psi_n(t))^{k-1}(m^{-n} f_n'(e^{itm^{-n}}))e^{-itm^{-n}(j_n(x)-1)} dt. \tag{3.6}$$

Thus,

$$m^n P_k(Z_n = j_n(x)) - w^{\star k}(\Delta_0) = T_n(1,x) + T_n(2,x) + T_n(3), \tag{3.7}$$

where

$$T_n(1,x) = \frac{k}{2\pi}((m^{-n} j_n(x))^{-1} - \Delta_0^{-1}) I(n,k,x), \tag{3.8}$$

$$T_n(2,x) = \frac{k}{2\pi\Delta_0}(I(n,k,x) - I(n,k,0)), \tag{3.9}$$

and

$$T_n(3) = \frac{1}{2\pi\Delta_0}(kI(n,k,0) - 2\pi\Delta_0 w^{\star k}(\Delta_0)). \tag{3.10}$$

We will now show that $\sup_{x\in\mathcal{C}_n} |T_n(2,x)| \to 0$ as $n \to \infty$. To this end, note that

$$\begin{aligned} &I(n,k,x) - I(n,k,0) \\ &= \int_{-\pi m^n}^{\pi m^n} ((\psi_n(t))^{k-1})(m^{-n} f_n'(e^{itm^{-n}}))(B(n,x,t)dt \\ &= \left(\int_{-\pi m^n}^{0} + \int_0^{\pi m^n}\right)((\psi_n(t))^{k-1})(m^{-n} f_n'(e^{itm^{-n}}))B(n,x,t)dt \\ &= J(n,1)(x) + J(n,2)(x), \end{aligned} \tag{3.11}$$

where

$$B(n,x,t) = e^{-itm^{-n}(j_n(x)-1)} - e^{-itm^{-n}(j_n(0)-1)}. \tag{3.12}$$

Notice that $|B(n,x,t)| \le 2$ and that

$$\lim_{n\to\infty} \sup_{x\in\mathcal{C}_n} |B(n,x,t)| \to 0. \tag{3.13}$$

We now establish that $J(n,2)(x)$ converges uniformly to 0. Similar arguments yield that $J(n,1)(x)$ converges uniformly to 0, thus establishing that $\sup_{x\in\mathcal{C}_n} |T_n(2,x)|$ converges to 0.

Returning to $J(n,2)(x)$, we express it as

$$\begin{aligned} J(n,2)(x) &= \left\{ \int_0^{\pi} + \sum_{r=1}^{n} \int_{\pi m^{r-1}}^{\pi m^r} \right\} ((\psi_n(t))^{k-1}(m^{-n} f_n'(e^{itm^{-n}}))B(n,x,t)dt \\ &= \sum_{r=0}^{n} J(n,2,r)(x), \end{aligned} \tag{3.14}$$

where

$$J(n,2,0)(x) = \int_0^{\pi} ((\psi_n(t))^{k-1}(m^{-n} f_n'(e^{itm^{-n}}))B(n,x,t)dt, \tag{3.15}$$

and, for $1 \le r \le n$,

$$J(n,2,r)(x) = \int_{\pi m^{r-1}}^{\pi m^r} ((\psi_n(t))^{k-1}(m^{-n} f_n'(e^{itm^{-n}}))B(n,x,t)dt. \tag{3.16}$$

Next, we observe that $|(m^{-n} f_n'(e^{itm^{-n}}))| \le 1$. Hence, using the bounded convergence theorem it follows that

$$\lim_{n\to\infty} \sup_{x\in\mathcal{C}_n} |J(n,2,0)(x)| = 0. \tag{3.17}$$

Now, let $1 \le r \le n$. Then for $t \in (\pi m^{(r-1)}, \pi m^r)$,

$$|m^{-n} f_n'(e^{itm^{-n}})| = |m^{-r} f_r'(f_{n-r}(e^{itm^{-n}}))||(m^{-(n-r)} f_{n-r}'(e^{itm^{-n}}))| \tag{3.18}$$

$$\le |m^{-r} f_r'(f_{n-r}(e^{itm^{-n}}))|. \tag{3.19}$$

Now, since $t \in (\pi m^{(r-1)}, \pi m^r)$, it follows that $tm^{-n} \in (\pi m^{-(n-r-1)}, \pi m^{-(n-r)})$; which implies that $f_{n-r}(e^{itm^{-n}}) \in \overline{S}$, where

$$S = \bigcup_{j\ge 0} \left\{ f_j(e^{ium^{-j}}) : \frac{\pi}{m} \le u \le \pi \right\}. \tag{3.20}$$

Define

$$\mu_r = \sup_{s\in\overline{S}} f_r'(s). \tag{3.21}$$

Then, for $t \in (\pi m^{(r-1)}, \pi m^r)$, it follows from(3.18) that

$$|m^{-n} f_n'(e^{itm^{-n}})| \le |m^{-r} f_r'(f_{n-r}(e^{itm^{-n}}))| \le m^{-r} \mu_r, \tag{3.22}$$

where $\sum_{r\geq 1}\mu_r < \infty$, by Dubuc's lemma (see Lemma 1, Page 80 of [3]). Also observe that,

$$|J(n,2,r)(x)| \leq A(n,r), \tag{3.23}$$

where

$$A(n,r) = \int_{\pi m^{r-1}}^{\pi m^r} |m^{-n} f_n'(e^{itm^{-n}}))(|\sup_{x\in\mathcal{C}_n}|B(n,x,t)|)dt \leq 2\mu_r, \tag{3.24}$$

where the last inequality follows from (3.18)-(3.22). Since $\sum_{r\geq 1}\mu_r < \infty$, it follows by the dominated convergence theorem that

$$\lim_{n\to\infty}\sum_{r\geq 1}|J(n,2,r)(x)|I_{[0,n]}(r) = \sum_{r\geq 1}\lim_{n\to\infty}|J(n,2,r)(x)|. \tag{3.25}$$

Now, to evaluate the $\lim_{n\to\infty}|J(n,2,r)(x)|$ we again apply the dominated convergence theorem. To this end, we first use (3.22) and then use (3.23) to take the limit inside the integral to get,

$$0 \leq \lim_{n\to\infty}\sup_{x\in\mathcal{C}_n}|J(n,2,r)(x)| \tag{3.26}$$

$$\leq \lim_{n\to\infty}\int_{\pi m^{r-1}}^{\pi m^r}|m^{-n} f_n'(e^{itm^{-n}}))(|\sup_{x\in\mathcal{C}_n}|B(n,x,t)|)dt \tag{3.27}$$

$$= \int_{\pi m^{r-1}}^{\pi m^r}\lim_{n\to\infty}|m^{-n} f_n'(e^{itm^{-n}}))(|\sup_{x\in\mathcal{C}_n}|B(n,x,t)|)dt = 0. \tag{3.28}$$

This proves the uniform convergence of $|J_n(2,x)|$ to 0 as $n\to\infty$. Similar arguments yield uniform convergence of $|J_n(1,x)|$ to 0 as $n\to\infty$. Combining these two we get $\sup_{x\in\mathcal{C}_n}|T_n(2,x)| \to 0$ as $n\to\infty$. To complete the proof of the theorem, we need to establish the uniform convergence of $|T(n,1,x)|$ to 0 and the convergence of $|T_n(3)|$ to 0 as $n\to\infty$. However, it also follows from the calculations (3.5)-(3.24) that

$$\sup_{x\in\mathcal{C}_n}|I(n,r,x)| \leq C < \infty \tag{3.29}$$

where C is a positive constant. Thus, it follows from (3.2) that $\sup_{x\in\mathcal{C}_n}|T(n,1,x)| \to 0$ as $n\to\infty$. Finally, convergence of $|T_n(3)|$ to zero follows from Theorem 2 on page 81 of [3]. This completes the proof of (2). In fact, we have proved that $\sup_{x\in\mathcal{C}_n}|m^n P_k(Z_n = j_n(x)) - w^{\star k}(\Delta_0)| \to 0$. This then also implies, with some further analysis, that $\inf_{x\in\mathcal{C}_n}[m^n P_k(Z_n = j_n(x)) - w^{\star k}(\Delta_0)] \to 0$, which is (3). □

Proof of Theorem 2. Let $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n Z_{1,k}$, where $\{Z_{1,k}, k\geq 1\}$ are i.i.d. with distribution same as that of Z_1. Let $\Lambda(\theta) = \log E(\exp(\theta Z_1))$ denote the cumulant generating function and $\Lambda^\star(a) = \sup_\theta[\theta a - \Lambda(\theta)]$ denote the Legendre–Fenchel transform of $\Lambda(\theta)$. By the Bahadur–Rao theorem (see [5]),

$$\lim_{l\to\infty}\sqrt{l}e^{l\Lambda^\star(a)}P(\overline{X}_l > a) = c_a. \tag{3.30}$$

Let us set

$$B(l,a) = \sqrt{l}e^{l\Lambda^\star(a)}P(\overline{X}_l > a) - c_a. \tag{3.31}$$

Since a is fixed, we will suppress the dependence on a and write $B(l)$ for $B(l,a)$. Now, by definition of conditional probability,

$$P(Z_1 = k \mid R_n > a, Z_n \in m^n[c,d]) = p_k \left\{ \frac{\sum_{l=l_{n,1}}^{l_{n,2}} P(\overline{X}_l > a) P_k(Z_{n-1} = l)}{\sum_{l=l_{n,1}}^{l_{n,2}} P(\overline{X}_l > a) P(Z_n = l)} \right\}$$

$$\equiv \frac{I_n}{J_n}, \tag{3.32}$$

where $l_{n,1} = \lfloor cm^n \rfloor + 1$ and $l_{n,2} = \lfloor dm^n \rfloor$. Let us set $\eta(n,k,l) = P_k(Z_n = l)$, $h(l) = (1/\sqrt{l})\exp(-l\Lambda^\star(a))$, and $d_n = l_{n,2} - l_{n,1}$. Hence, we can express $I_n = I_{n,1} + I_{n,2}$, where

$$I_{n,1} = p_k h(l_{n,1}) \sum_{t=0}^{d_n} B(t + l_{n,1}) \theta(n,t) \eta(n-1, k, t + l_{n,1}) \tag{3.33}$$

and

$$I_{n,2} = p_k c_a h(l_{n,1}) \sum_{t=0}^{d_n} \theta(n,t) \eta(n-1, k, t + l_{n,1}); \tag{3.34}$$

and where $\theta(n,t) = (1 + t/l_{n,1})^{\frac{-1}{2}} e^{-t\Lambda^\star(a)}$, $l = t + l_{n,1}$, and B is as in 3.31. Similarly, one can express J_n as a sum of $J_{n,1}$ and $J_{n,2}$ where

$$J_{n,1} = h(l_{n,1}) \sum_{t=0}^{d_n} B(t + l_{n,1}) \theta(n,t) \eta(n, 1, t + l_{n,1}), \tag{3.35}$$

and

$$J_{n,2} = c_a h(l_{n,1}) \sum_{t=0}^{d_n} \theta(n,t) \eta(n, 1, t + l_{n,1}). \tag{3.36}$$

Thus the conditional probability in question becomes

$$P(Z_1 = k \mid R_n > a, Z_n \in m^n[c,d]) = \frac{I_{n,1}}{J_{n,1}} \left(1 + \frac{J_{n,2}}{J_{n,1}}\right)^{-1} + \frac{I_{n,2}}{J_{n,2}} \left(1 + \frac{J_{n,1}}{J_{n,2}}\right)^{-1}.$$

We will now establish the following:

1. $\lim_{n\to\infty} J_{n,1}/J_{n,2} = 0$,
2. $\limsup_{n\to\infty} I_{n,1}/J_{n,1} \le C < \infty$,
3. $\lim_{n\to\infty} I_{n,2}/J_{n,2} = \frac{mw^{\star k}(mc)}{w(c)} p_k$.

These facts will imply the theorem. We start with the proof of (3). Consider,

$$\tilde{I}_{n,2} = m^{n-1}(p_k c_a h(l_{n,1}))^{-1} I_{n,2} = m^{n-1} \sum_{t=0}^{d_n} \theta(n,t) \eta(n-1, t + l_{n,1}).$$

By the local limit theorem (see Chapter 2, Section 4.1 in [3]), if $v_n \to \Delta$ then

$$\lim_{n\to\infty} m^n P_k(Z_n = m^n v_n) = w^{\star k}(\Delta).$$

Using this we get that

$$\begin{aligned}\lim_{n\to\infty} m^{n-1}\eta(n-1,t+l_{n,1}) &= \lim_{n\to\infty} m^{n-1}P_k\left(Z_{n-1}=m^{n-1}\frac{l_{n,1}}{m^{n-1}}\left(1+\frac{t}{l_{n,1}}\right)\right)\\ &= w^{\star k}(cm),\end{aligned}$$

since $\frac{l_{n,1}}{m^{n-1}} \to cm$. Now applying the uniform bound in Theorem 1 to $m^{n-1}\eta(n-1,k,t+l_{n,1})$, and noting that $\theta(n,t)\le \exp(-t\Lambda^\star(a))$, it follows from the dominated convergence theorem that

$$\lim_{n\to\infty}\tilde{I}_{n,2} = w^{\star k}(cm)\Gamma,$$

where $\Gamma = \exp(\Lambda^\star(a))/(\exp(\Lambda^\star(a))-1)$. In a similar manner, one can show that

$$\begin{aligned}\lim_{n\to\infty}\tilde{J}_{n,2} &\equiv \lim_{n\to\infty} m^n(c_a h(l_{n,1}))^{-1}J_{n,2}\\ &= \lim_{n\to\infty} m^n\sum_{t=1}^{d_n}\theta(n,t)\eta(n,1,t+l_{n,1})\\ &= w(c)\Gamma.\end{aligned}$$

Finally,

$$\begin{aligned}\lim_{n\to\infty}\frac{I_{n,2}}{J_{n,2}} &= \lim_{n\to\infty}\frac{m^{-(n-1)}(p_kc_ah(l_{n,1}))^{-1}\tilde{I}_{n,2}}{m^{-n}(c_ah(l_{n,1}))^{-1}\tilde{J}_{n,2}}\\ &= \frac{mp_kw^{\star k}(cm)\Gamma}{w(c)\Gamma}\end{aligned}$$

yielding (3).

Turning to the proof of (1), note that

$$\frac{J_{n,1}}{J_{n,2}} = \frac{\sum_{t=1}^{d_n}B(t+l_{n,1})\theta(n,t)\eta(n,1,t+l_{n,1})}{c_a\sum_{t=0}^{d_n}\theta(n,t)\eta(n,1,t+l_{n,1})}.$$

Now, using $\sup_{1\le t\le d_n}|B(t+l_{n,1})|\to 0$ as $n\to\infty$ we have that

$$\lim_{n\to\infty}\frac{J_{n,1}}{J_{n,2}} = 0.$$

Finally, turning to (2), by Theorem 1,

$$\begin{aligned}\frac{I_{n,1}}{J_{n,1}} &= \frac{mp_k\sum_{t=0}^{d_n}B(t+l_{n,1})\theta(n,t)m^{n-1}\eta(n-1,k,t+l_{n,1})}{\sum_{t=0}^{d_n}B(t+l_{n,1})\theta(n,t)m^n\eta(n,1,t+l_{n,1})}\\ &\le mp_k\frac{\max_{0\le t\le d_n}m^{n-1}\eta(n-1,k,t+l_{n,1})}{\min_{0\le t\le d_n}m^n\eta(n,1,t+l_{n,1})}\\ &\le C<\infty,\end{aligned}$$

where C is a constant. This completes the proof of Theorem 2. $\square$

4. Age of a Branching Process

As explained in the introduction, statistical estimation of the age of a simple branching process is an important problem arising in several scientific contexts. It was first studied by Stigler ([14]) who estimated the age using maximum likelihood methods, i.e., by maximizing $P(Z_t = N(t) \mid Z_t > 0)$ with respect to t. In this context, the population age t is treated as an unknown parameter and is estimated using the current population size $N(t)$. Stigler derived the estimator $T_1(N)$ in (4.1) for offspring distributions with fractional linear generating functions, and suggested this as an estimator of the age for general offspring distributions. Stigler's estimate is given by

$$T_1(N) = \frac{\log N(t)}{\log m}. \tag{4.1}$$

Stigler established that $T_1(N(t))$ is β-consistent for t in the sense that $\frac{T_1 N(t)-t}{t^\beta} \to 0$ a.s. for every $\beta > 0$ as $t \to \infty$. More recently, [10] studied age by constructing a backward process X_j and defined the estimate of age as

$$T_2(N) = \inf\{r : X_r = 1\}. \tag{4.2}$$

In this formulation, if the offspring distribution is geometric then the reverse process is a Galton-Watson process with immigration starting with N ancestors. Using the duality between the forward and the backward process, [10] obtained detailed results concerning $T_2(N) - T_1(N)$ as $N \to \infty$. The bias in the estimate of age is given by

$$B(t) = t - T_1(N(t)). \tag{4.3}$$

Our next result shows that the bias $B(t)$ conditioned on $R_t > a > m$ diverges to infinity and is a corollary to Proposition 1.

Corollary 4.1. *Let $k(t)$ be a sequence of constants converging to infinity such that $k(t) = \mathbf{o}(t)$ as $t \to \infty$. Then,*

$$\lim_{t\to\infty} P(B(t) \geq k(t) \mid R_t > a > m) = 1. \tag{4.4}$$

Proof. Let $\epsilon > 0$; then by Proposition 1, there exists $k_0(\epsilon)$ such that $\sum_{j\leq k_0} \gamma(j) > 1-\epsilon$. Now we observe, by simplifying, that $[B(t) \geq k(t)] = [W_t \leq m^{-k(t)}]$. Since $k(t) = \mathbf{o}(t)$, it follows that $m^t m^{-k(t)}$ diverges to ∞. Thus,

$$\begin{aligned}
P(B(t) \geq k(t) \mid R_t > a) &= P(W_t \leq m^{-k(t)} \mid R_t > a) && (4.5)\\
&= P(Z_t \leq m^{t-k(t)} \mid R_t > a) && (4.6)\\
&\geq P(Z_t \leq k_0 \mid R_t > a) && (4.7)\\
&\to \sum_{j=1}^{k_0} \gamma(j) > 1-\epsilon, && (4.8)
\end{aligned}$$

by the choice of k_0. Thus, $\liminf_{t\to\infty} P(B(t) \geq k(t) \mid R_t > a) \geq 1-\epsilon$. Since ϵ is arbitrary, the corollary follows. $\square$

5. Concluding Remarks

In this paper, we studied the evolutionary structure of a branching process through the behavior of conditional limits under various notions of "information" about the current population size. We observed a "suppression property" which is a consequence of the assumption $p_0 + p_1 > 0$. This implies that conditionally on the large deviation type information, the bias in the estimate of the age diverges to infinity; or in other words, the estimator is *conditionally inconsistent*. A natural next question concerns the conditional consistency of the estimator of age under other notions of "information." These and other related issues are studied in a subsequent paper.

Acknowledgements

Authors would like to thank the anonymous referee for a careful reading of the manuscript and several useful suggestions.

References

[1] Athreya, K. B. (1994). Large deviation rates for branching processes-I, the single type case. *Annals of Applied Probability* **4** 779–790.

[2] Athreya, K. B. and Ney, P. E. (1970). The local limit theorem and some related aspects of supercritical branching processes. *Transactions of the American Mathematical Society* **2** 233–251.

[3] Athreya, K. B. and Ney, P. E. (1972). *Branching Processes.* Springer-Verlag, Berlin.

[4] Athreya, K. B. and Vidyashankar, A. N. (1993). Large deviation results for branching processes. In *Stochastic Processes. A Festschrift in Honour of Gopinath Kallianpur* (S. Cambanis, J. K. Ghosh, R. L. Karandikar and P. K. Sen, eds.) 7–12. Springer, New York.

[5] Bahadur, R. R. and Ranga Rao, R. (1960). On deviations of the sample mean. *Annals of Mathematical Statistics* **31** 1015–1027.

[6] Bingham, N. H. (1988). On the limit of a supercritical branching process. *Journal of Applied Probability* **25A** 215–228.

[7] Chow, Y. S. and Teicher, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales*. Springer Verlag, Berlin.

[8] Dubuc, S. and Senata, E. (1976). The local limit theorem for Galton–Watson Process. *Annals of Probability* **3** 490–496.

[9] Fleischmann, K. and Wachtel, V. (2005). Lower deviation probabilities for supercritical Galton–Watson Processes. Preprint.

[10] Klebaner, F. C. and Sagitov, S. (2002). The age of a Galton–Watson population with a geometric offspring distribution. *Journal of Applied Probability* **4** 816–828.

[11] Leman, C. S., Chen, Y., Stajich, J. E., Noor, M. A. F. and Uyenoyama, M. K. (2005). Likelihoods from summary statistics: Recent divergence between species. *Genetics* **171** 1419–1436.

[12] Ney, P. E. and Vidyashankar, A. N. (2003). Harmonic moments and large deviation rates for supercritical branching processes. *Annals of Applied Probability* **13** 475–489.

[13] Ney, P. E. and Vidyashankar, A. N. (2004). Local limit theory and large deviation rates for supercritical branching processes. *Annals of Applied Probability* **14** 1135–1166.

[14] Stigler, S. M. (1970). Estimating the age of a Galton–Watson branching process. *Biometrika* **57** 507–512.

[15] Tavaré, S., Marshall, C. R., Will, O., Soligo, C. and Martin, R. (2002). Using the fossil record to estimate the age of the last common ancestor. *Nature* **416** 726–729.