Conditional Limit Laws and Inference for Generation Sizes of Branching Processes

P.E. Ney and A. N. Vidyashankar

University of Wisconsin 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 \( R_n \equiv \frac{Z_n}{m^n} > a \), is observed. Furthermore, under a more refined conditioning, the asymptotic a posteriori 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 \( f(s) = E(s^{Z_1}) | Z_0 = 1 \) denote the probability generating function. Let \( m = E(Z_1) = f'(1) \), where \( f'(.) \) denotes the derivative of \( f(.) \). 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 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 \( n \geq 0 \)

\[
Z_{n+1} = \sum_{j=1}^{Z_n} \xi_{n,j}, \tag{1.0.1}
\]

where \( \xi_{n,j} \) is interpreted as the number of children produced by the \( j^{th} \) parent in the \( n^{th} \) generation; and \( P(\xi_{0,1} = j) = p_j \). This implies that the generating function of the \( n^{th} \) generation population size is given by the \( n \)-fold iteration of \( f(.) \); i.e. \( E(s^{Z_n}) = f_n(s) = f(f(f(\ldots(s)))) \), \( 0 \leq s \leq 1 \). 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 \( W_n = \frac{Z_n}{m^n} \). Let \( \mathcal{G}_n \) denote the sigma field generated by the first \( n \) generation sizes, namely, \( \{Z_0, Z_1, \ldots, Z_n\} \). Then it is well-known that \( \{(W_n, \mathcal{G}_n) : n \geq 1\} \) 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 $w(.)$ and $w(x) > 0$ for $x > 0$.

Let $R_n = \frac{Z_n}{Z_{n+1}}$. 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 branching processes 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.

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

\begin{equation}
P(X_1 \in \cdot | \frac{S_n}{n} \in A), \; EX_1 \notin A,
\end{equation}

or more generally, of

\begin{equation}
P\{(X_1, X_2, \ldots, X_{k_n}) \in \cdot | \frac{S_n}{n} \in A\},
\end{equation}

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 $P(Z_1 \in \cdot | R_n > a > m)$ and that of $P(\xi_{n,1} \in \cdot | R_n > a > m)$. 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 $Z_n \sim x m^n$ 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 $0 \leq s < 1$

\begin{equation}
Q_n(s) = \frac{f_n(s) - q}{\gamma^n},
\end{equation}

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, \( Q(.) \) admits a power series representation; that is,

\[
Q(s) = \sum_{k \geq 1} \nu_k s^k. 
\]

When \( p_0 = 0 \), \( \gamma \) reduces to \( p_1 \). It follows from (1.0.5) that (see [3])

\[
\lim_{n \to \infty} \frac{P(Z_n = k)}{\gamma^n} = \nu_k.
\]

The quantities \( \nu_k \) 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 \( P_k(.) = P(|Z_0| = k) \).

**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 \( \mathcal{C}_n = \{ x \in [c,d] : x y_n \text{ is an integer} \} \). Then the following hold:

\[
\begin{align*}
1. \quad & \lim_{n \to \infty} m^n P_k(Z_n = m^n y_n (1 + \frac{x_n}{m^n})) = \delta^k(\Delta_0); \\
2. \quad & \lim_{n \to \infty} m^n \sup_{\{x \in \mathcal{C}_n\}} P_k(Z_n = m^n y_n (1 + \frac{x}{m^n})) = \delta^k(\Delta_0); \\
3. \quad & \lim_{n \to \infty} m^n \inf_{\{x \in \mathcal{C}_n\}} P_k(Z_n = m^n y_n (1 + \frac{x}{m^n})) = \delta^k(\Delta_0).
\end{align*}
\]

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,

\[
\lim_{n \to \infty} P(Z_n = k|R_n > a > m) = \gamma(k) \geq 0,
\]

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 (2.0.4). 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,

\[
\lim_{n \to \infty} P(Z_1 = k|R_n > a > m) = \delta_1(k).
\]

Furthermore, for any \( k(n) \to \infty \) and \( k_n = o(n) \)

\[
\lim_{n \to \infty} P((Z_1 = 1, \ldots, Z_{k_n} = 1)|R_n > a) = 1.
\]
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 $c > 0$,

\begin{equation}
\pi_l(c, d) = \int_{cm}^{dm} w^\star_l(x) dx \int_c^d w(x) dx.
\end{equation}

Then,

\begin{equation}
\lim_{n \to \infty} P(Z_k = l | Z_n \in m^n[c, d]) = \pi_l(c, d) P(Z_k = l),
\end{equation}

where

\begin{equation}
\sum_{l \geq 1} \pi_l(c, d) P(Z_k = l) = 1.
\end{equation}

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

\begin{equation}
\lim_{n \to \infty} P(Z_1 = l | Z_n \in m^n[c, d]) = \pi_l(c, d) p_l,
\end{equation}

and

\begin{equation}
\lim_{n \to \infty} P(Z_1 = l | Z_n = cm^n) = \frac{w^\star_l(mc)}{w(c)} p_l m.
\end{equation}

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 $c > 0$,

\begin{equation}
\lim_{n \to \infty} P(Z_1 = l | R_n > a, Z_n \in m^n[c, d]) = \frac{w^\star_l(mc)}{w(c)} p_l m.
\end{equation}

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 (2.0.11) and (2.0.12) 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 (2.0.10) by

\begin{equation}
\lim_{d \to c} \pi_l(cm, dm) = \frac{w^\star_l(mc)}{w(c)} p_l m.
\end{equation}
Remark 2.5. The case when one conditions on \( Z_n \in v_n[c,d] \) with \( v_n = \mathcal{O}(m^n) \) 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 (2.0.8). In Theorem 2, when the conditioning includes \( R_n > a \), the age and the divergence time again agree. But if \( v_n = \mathcal{O}(m^n) \), 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 \( Z_n = 1 \) we take \( P(Z_0 = k) = \pi(k) \), \( \sum_{k \geq 1} \pi(k) = 1 \), 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 \( \overline{X} = k^{-1} \sum_{j=1}^{k} X_j \), \( X_j \)'s are i.i.d. with distribution \( Z_1 \). Then, using Theorem 1 of [4], and (1.0.5) it follows that

\[
P(Z_n = k | R_n > a) = \frac{P(R_n > a | Z_n = k) P(Z_n = k)}{P(R_n > a)}
= P(\overline{X} > a) \left( \frac{P(Z_n = k) p_1^{-n}}{P(R_n > a) p_1^{-n}} \right)
= P(\overline{X} > a) \frac{v_k}{\lambda_a} = a_k,
\]

where \( v_k \) is as in (1.0.6), and

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

Thus, \( \sum_{k \geq 1} a_k = 1 \).  

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

**Proof of Proposition 2.** Let \( k(n) = \Omega(n) \). Then using Theorem 1 of [4], it follows that

\[
P(Z_k(n) = 1 | R_n > a) = \frac{P(R_n > a | Z_{k(n)} = 1) P(Z_{k(n)} = 1)}{P(R_n > a)}
= \frac{P(R_n > a) - 1}{P(R_n > a) p_1}
= \frac{p_1^{-n(k(n))} P(R_n > a)}{p_1^{-n} P(R_n > a) p_1}
\to 1
\]
implying (2.0.5) and (2.0.6). ■

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

\[
P(Z_k = l | Z_n \in \mathcal{A}_n) = \frac{P(Z_n \in \mathcal{A}_n | Z_k = l)P(Z_k = l)}{P(Z_n \in \mathcal{A}_n)}
\]

\[
= \left\{ \frac{P(Z_{n-k} \in m^k\mathcal{A}_{n-k})}{P(Z_n \in \mathcal{A}_n)} \right\}P(Z_k = l)
\]

\[
\rightarrow \left\{ \frac{\int_{c}^{d} m^k w^*(x)dx}{\int_{c}^{d} w(x)dx} \right\}P(Z_k = l)
\]

\[
= \pi_{l,k}(c, d) P(Z_k = l).
\]

To complete the proof of Proposition 3, we need to establish 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 (2.0.10) 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.** \(\sum_{l \geq 1} P(Z_k = l) \int_{am^k}^{bm^k} w^*(x)dx = \int_{a}^{b} w(x)dx\).

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

\[
w^*(x) = \frac{1}{2\pi} \int_{R} e^{-i\theta x} (\phi(\theta))^l d\theta.
\]

(3.0.1)

Now, integrating the LHS of (3.0.1) between \(am^k\) and \(bm^k\) we get

\[
\int_{am^k}^{bm^k} w^*(x)dx = m^k \int_{a}^{b} w^*(ym^{k})dy,
\]

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

\[
m^k \int_{a}^{b} w^*(ym^{k})dy = \frac{m^k}{2\pi} \int_{a}^{b} \int_{R} e^{-i\eta ym^k} (\phi(\eta))^l d\eta dy
\]

\[
= \frac{1}{2\pi} \int_{a}^{b} \int_{R} e^{-i\eta y} (\phi(\eta))^l d\eta dy,
\]

where the last identity follows upon setting \(\eta m^k = \xi\). Thus,

\[
\sum_{l \geq 1} \int_{am^k}^{bm^k} w^*(x)dx P(Z_k = l) = \int_{a}^{b} \sum_{l \geq 1} \int_{R} e^{-i\eta y} (\phi(\frac{\eta}{m^k}))^l d\eta dy P(Z_k = l)
\]

\[
= \int_{a}^{b} \int_{R} e^{-i\eta x} \phi(\eta) d\eta dx
\]

\[
= \int_{a}^{b} w(x)dx,
\]

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 \( j_n(x) = m^n y_n(1 + x m^{-n}) \) 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

\[
\lim_{n \to \infty} \sup_{x \in \mathcal{C}_n} |m^{-n} j_n(x) - \Delta_0| = 0. \tag{3.0.2}
\]

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 (\([7]\))

\[
P_k(Z_n = j_n(x)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f_n(e^{i\theta}))^{k} e^{-ij_n(x)\theta} d\theta. \tag{3.0.3}
\]

Now, integrating by parts the RHS of (3.0.3) and using \( f_n^k(e^{i\pi}) e^{-it\pi} = f_n^k(e^{-i\pi}) e^{it\pi} \) for all integers \( k \) and \( l \), it follows that

\[
P_k(Z_n = j_n(x)) = \frac{1}{2\pi} \left( \frac{k}{j_n(x)} \right) I(n, k, x), \tag{3.0.4}
\]

where

\[
I(n, k, x) = \int_{-\pi}^{\pi} (f_n(e^{i\theta}))^{k-1} (f_n'(e^{i\theta})) e^{-i(j_n(x)^{-1})\theta} d\theta. \tag{3.0.5}
\]

Next, making a change of variable \( \theta = tm^{-n} \) and setting \( \psi_n(t) = E(e^{itW_n}) \), (3.0.5) reduces to

\[
I(n, k, x) = \int_{-\pi m^n}^{\pi m^n} (\psi_n(t))^{k-1} (m^{-n} f_n'(e^{it m^{-n}})) e^{-itm^{-n}(j_n(x)^{-1})} dt. \tag{3.0.6}
\]

Thus,

\[
m^n P_k(Z_n = j_n(x)) - w^k(\Delta_0) = T_n(1, x) + T_n(2, x) + T_n(3), \tag{3.0.7}
\]

where

\[
T_n(1, x) = \frac{k}{2\pi} ((m^{-n} j_n(x))^{-1} - \Delta_0^{-1}) I(n, k, x), \tag{3.0.8}
\]

\[
T_n(2, x) = \frac{k}{2\pi \Delta_0} (I(n, k, x) - I(n, k, 0)), \tag{3.0.9}
\]

and

\[
T_n(3) = \frac{1}{2\pi \Delta_0} (kI(n, k, 0) - 2\pi \Delta_0 w^k(\Delta_0)). \tag{3.0.10}
\]

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

\[
I(n, k, x) - I(n, k, 0) = \int_{-\pi m^n}^{\pi m^n} ((\psi_n(t))^{k-1} (m^{-n} f_n'(e^{it m^{-n}})) (B(n, x, t) dt
\]

\[
= \left( \int_{0}^{\pi m^n} + \int_{0}^{\pi m^n} \right) ((\psi_n(t))^{k-1} (m^{-n} f_n'(e^{it m^{-n}})) B(n, x, t) dt
\]

\[
= J(n, 1)(x) + J(n, 2)(x), \tag{3.0.11}
\]

\[
\]
where

\begin{equation}
B(n, x, t) = e^{-itm^{-n}(j_n(x)-1)} - e^{-itm^{-n}(j_n(0)-1)}.
\end{equation}

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

\begin{equation}
\lim_{n \to \infty} \sup_{x \in C_n} |B(n, x, t)| \to 0.
\end{equation}

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

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

\begin{equation}
J(n, 2)(x) = \{ \int_0^\pi + \sum_{r=1}^{\pi m^r} \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
\end{equation}

where

\begin{equation}
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,
\end{equation}

and, for \(1 \leq r \leq n\),

\begin{equation}
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.
\end{equation}

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

\begin{equation}
\lim_{n \to \infty} \sup_{x \in C_n} |J(n, 2, 0)(x)| = 0.
\end{equation}

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

\begin{equation}
|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}}))| \leq |m^{-r} f'_r(f_{n-r}(e^{itm^{-n}}))|.
\end{equation}

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 \(f_{n-r}(e^{itm^{-n}}) \in \mathbb{S}\), where

\begin{equation}
S = \bigcup_{j \geq 0} \{ f_j(e^{ium^{-j}}) : \frac{\pi}{m} \leq u \leq \pi \}.
\end{equation}

Define

\begin{equation}
\mu_r = \sup_{s \in S} f_r(s).
\end{equation}

Then, for \(t \in (\pi m^{(r-1)}, \pi m^r)\), it follows from (3.0.18) that

\begin{equation}
|m^{-n} f'_n(e^{itm^{-n}})| \leq |m^{-r} f'_r(f_{n-r}(e^{itm^{-n}}))| \leq m^{-r} \mu_r,
\end{equation}

\begin{equation}
\text{imsart-lmns ver. 2007/04/02 file: Formatconditional_limits_081307_refereesreport_ed1.tex date: October 26, 2008}
\end{equation}
where $\sum_{r \geq 1} \mu_r < \infty$, by Dubuc’s lemma (see Lemma 1, Page 80 of [3]). Also observe that,

\begin{equation}
|J(n, 2, r)(x)| \leq A(n, r),
\end{equation}

where

\begin{equation}
A(n, r) = \int_{\pi m^r}^{\pi m^{r-1}} |m^{-n}f_n(x)| \left( \sup_{x \in \mathcal{C}_n} |B(n, x, t)| \right) dt \leq 2\mu_r,
\end{equation}

where the last inequality follows from (3.0.18)-(3.0.22). Since $\sum_{r \geq 1} \mu_r < \infty$, it follows by the dominated convergence theorem that

\begin{equation}
\lim_{n \to \infty} \sum_{r \geq 1} |J(n, 2, r)| = \sum_{r \geq 1} \lim_{n \to \infty} |J(n, 2, r)|.
\end{equation}

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 (3.0.22) and then use (3.0.23) to take the limit inside the integral to get,

\begin{equation}
0 \leq \lim_{n \to \infty} \sup_{x \in \mathcal{C}_n} |J(n, 2, r)(x)|
\end{equation}

\begin{equation}
\leq \lim_{n \to \infty} \int_{\pi m^r}^{\pi m^{r-1}} |m^{-n}f_n(x)| \left( \sup_{x \in \mathcal{C}_n} |B(n, x, t)| \right) dt
\end{equation}

\begin{equation}
= \int_{\pi m^r}^{\pi m^{r-1}} \lim_{n \to \infty} |m^{-n}f_n(x)| \left( \sup_{x \in \mathcal{C}_n} |B(n, x, t)| \right) dt = 0.
\end{equation}

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 (3.0.5)-(3.0.24) that

\begin{equation}
\sup_{x \in \mathcal{C}_n} |I(n, r, r)| \leq C < \infty
\end{equation}

where $C$ is a positive constant. Thus, it follows from (3.0.2) 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 $\sup_{x \in \mathcal{C}_n} |m^n P_k(Z_n = j_n(x)) - w^k(\Delta_0)| \to 0$. This then also implies, with some further analysis, that $\inf_{x \in \mathcal{C}_n} |m^n P_k(Z_n = j_n(x)) - w^k(\Delta_0)| \to 0$, which is (3).

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

\begin{equation}
\lim_{l \to \infty} \sqrt{l} e^{l\Lambda^*(a)} P(\mathcal{X}_1 > a) = c_a.
\end{equation}

Let us set

\begin{equation}
B(l, a) = \sqrt{l} e^{l\Lambda^*(a)} P(\mathcal{X}_1 > a) - c_a.
\end{equation}
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,

$$P(Z_1 = k|R_n > a, Z_n \in m^n[c, d]) = p_k \frac{\sum_{l=I_{n,1}}^{I_{n,2}} P(X_l > a)P_k(Z_{n-1} = l)}{\sum_{l=I_{n,1}}^{I_{n,2}} P(X_l > a)P(Z_n = l)}$$

(3.0.32)

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

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

(3.0.33)

and

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

(3.0.34)

and where $\eta(n, t) = \frac{(1 + \frac{t}{l_{n,1}})^{\frac{
u}{2}} e^{-l\Lambda^*(a)}}{B(t + l_{n,1})}$, $l = t + l_{n,1}$, and $B$ is as in 3.0.31. Similarly, one can express $J_n$ as a sum of $J_{n,1}$ and $J_{n,2}$ where

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

(3.0.35)

and

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

(3.0.36)

Thus the conditional probability in question becomes

$$P(Z_1 = k|R_n > a, Z_n \in m^n[c, d]) = \frac{I_{n,1}}{J_{n,1}} (1 + \frac{J_{n,2}}{J_{n,1}})^{-1} + \frac{I_{n,2}}{J_{n,2}} (1 + \frac{J_{n,1}}{J_{n,2}})^{-1}.$$
Using this we get that
\[
\lim_{n \to \infty} m^{n-1} \eta(n-1, t + l_{n,1}) = \lim_{n \to \infty} m^{n-1} P_k(Z_{n-1} = m^{n-1} \frac{l_{n,1}}{m^{n-1}} (1 + \frac{t}{l_{n,1}})) \\
= w^k(cm),
\]
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 \( \theta(n, t) \leq \exp(-t \Lambda^*(a)) \), it follows from the dominated convergence theorem that
\[
\lim_{n \to \infty} \tilde{I}_{n,2} = w^k(cm) \Gamma,
\]
where \( \Gamma = \frac{\exp(\Lambda^*(a))}{\exp(\Lambda^*(a)) - 1} \). In a similar manner, one can show that
\[
\lim_{n \to \infty} \tilde{J}_{n,2} = \lim_{n \to \infty} m^n (c_n h(l_{n,1}))^{-1} J_{n,2} \\
= \lim_{n \to \infty} m^n \sum_{t=1}^{d_n} \theta(n, t) \eta(n, 1, t + l_{n,1}) \\
= w(c) \Gamma.
\]
Finally,
\[
\lim_{n \to \infty} \frac{I_{n,1}}{J_{n,1}} = \lim_{n \to \infty} \frac{m^{-(n-1)}(p_k c_n h(l_{n,1}))^{-1} \tilde{I}_{n,2}}{m^{-n} (c_n h(l_{n,1}))^{-1} \tilde{J}_{n,2}} \\
= \frac{m p_k w^k(cm) \Gamma}{w(c) \Gamma}
\]
yielding (3).

Turning to the proof of (1), note that
\[
\frac{J_{n,1}}{J_{n,2}} = \frac{\sum_{t=1}^{d_n} B(t + l_{n,1}) \theta(n, t) \eta(n, 1, t + l_{n,1})}{c_n \sum_{t=1}^{d_n} \theta(n, t) \eta(n, 1, t + l_{n,1})}.
\]
Now, using \( \sup_{1 \leq t \leq d_n} |B(t + l_{n,1})| \to 0 \) 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,
\[
\frac{I_{n,1}}{J_{n,1}} = \frac{m p_k \sum_{t=0}^{d_n} B(t + l_{n,1}) \theta(n, t) m^{n-1} \eta(n - 1, k, t + l_{n,1})}{\sum_{t=0}^{d_n} B(t + l_{n,1}) \theta(n, t) m^n \eta(n, 1, t + l_{n,1})} \\
\leq \frac{m p_k \max_{0 \leq t \leq d_n} m^{n-1} \eta(n - 1, k, t + l_{n,1})}{\min_{0 \leq t \leq d_n} m^n \eta(n, 1, t + l_{n,1})} \\
\leq C < \infty,
\]
where \( C \) is a constant. This completes the proof of Theorem 2. \( \blacksquare \)
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 $P(Z_t = N(t) | Z_t > 0)$ 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 (4.0.1) 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

$$T_1(N) = \frac{\log N(t)}{\log m}.$$ (4.0.1)

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

$$T_2(N) = \inf \{ r : X_r = 1 \}.$$ (4.0.2)

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 $N \rightarrow \infty$. The bias in the estimate of age is given by

$$B(t) = t - T_1(N(t)).$$ (4.0.3)

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 $k(t) = o(t)$ as $t \rightarrow \infty$. Then,

$$\lim_{t \rightarrow \infty} P(B(t) \geq k(t) | R_t > a > m) = 1.$$ (4.0.4)

**Proof.** Let $\epsilon > 0$; then by Proposition 1, there exists a $k_0(\epsilon)$ be 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 $k(t) = o(t)$, it follows that $m^t m^{-k(t)}$ diverges to $\infty$. Thus,

$$P(B(t) \geq k(t) | R_t > a) = P(W_t \leq m^{-k(t)} | R_t > a)$$

$$= P(Z_t \leq m^{-k(t)} | R_t > a)$$

$$\geq P(Z_t \leq k_0 | R_t > a)$$

$$\rightarrow \sum_{j=1}^{k_0} \gamma(j) > 1 - \epsilon,$$ (4.0.5-4.0.8)

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