When Does a Branching Process Grow Like its Mean? Conceptual Proofs of L log L Criteria

By Russell Lyons, Robin Pemantle, and Yuval Peres

Abstract. The Kesten-Stigum Theorem is a fundamental criterion for the rate of growth of a supercritical branching process, showing that an $L\log L$ condition is decisive. In critical and subcritical cases, results of Kolmogorov and later authors give the rate of decay of the probability that the process survives at least n generations. Through a unified approach, we give conceptual proofs of these theorems which are free of manipulation of generating functions. This approach also explains Yaglom's exponential limit law for conditioned critical branching processes via a simple characterization of the exponential distribution.

§1. Introduction.

Consider a Galton-Watson branching process with each particle having probability p_k of generating k children. Let L stand for a random variable with this progeny distribution. Let $m := \sum_k kp_k$ be the mean number of children per particle and let Z_n be the number of particles in the n^{th} generation. The most basic and well-known fact about branching processes is that the extinction probability $q := \lim \mathbf{P}[Z_n = 0]$ is equal to 1 if and only if $m \le 1$. It is also not hard to establish that in the case $m > 1$,

$$\frac{1}{n} \log Z_n \to \log m$$

almost surely on nonextinction, while in the case $m \le 1$,

$$\frac{1}{n} \log \mathbf{P}[Z_n > 0] \to \log m.$$

Finer questions may be asked:

1991 Mathematics Subject Classification. Primary 60J80.

Key words and phrases. Galton-Watson, size-biased distributions.

Research partially supported by two Alfred P. Sloan Foundation Research Fellowships (Lyons and Pemantle), by NSF Grants DMS-9300191 and DMS-9213595 (Pemantle and Peres), by a Presidential Faculty Fellowship (Pemantle), and by the hospitality of Washington University (Pemantle).

- In the case $m > 1$, when does the mean $\mathbf{E}[Z_n] = m^n$ give the right growth rate up to a random factor?
- In the case $m < 1$, when does the first moment estimate $\mathbf{P}[Z_n > 0] \le \mathbf{E}[Z_n] = m^n$ give the right decay rate up to a random factor?
- In the case $m = 1$, what is the decay rate of $\mathbf{P}[Z_n > 0]$?

Of course, the first question is the most important. These questions are answered by the following three classical theorems.

THEOREM 1: THE KESTEN-STIGUM THEOREM ON SUPERCRITICAL PROCESSES (1966). *Suppose that $1 < m < \infty$ and let W be the limit of the martingale Z_n/m^n. The following are equivalent:*

(i) $\mathbf{P}[W = 0] = q$*;*
(ii) $\mathbf{E}[W] = 1$*;*
(iii) $\mathbf{E}[L \log L] < \infty$*.*

THEOREM 2: SUBCRITICAL PROCESSES (HEATHCOTE, SENETA, VERE-JONES (1967)). *The sequence $\{\mathbf{P}[Z_n > 0]/m^n\}$ is decreasing. If $m < 1$, then it has a positive limit if and only if $\mathbf{E}[L \log L] < \infty$.*

The sufficiency of $\mathbf{E}[L \log L] < \infty$ in Theorem 2 was proved under a second moment assumption by Kolmogorov (1938). As is common, we prove this theorem by consideration of the law of Z_n conditioned on $Z_n > 0$, which we denote by μ_n. It is interesting that this sequence $\{\mu_n\}$ always converges in a strong sense, even when its means are unbounded; see Section 6.

THEOREM 3: CRITICAL PROCESSES (KESTEN, NEY AND SPITZER (1966)). *Suppose that $m = 1$ and let $\sigma^2 = \mathrm{Var}(L) = \mathbf{E}[L^2] - 1 \le \infty$. Then we have:*

(i) Kolmogorov's estimate:

$$\lim_{n \to \infty} n \mathbf{P}[Z_n > 0] = \frac{2}{\sigma^2} \,.$$

(ii) Yaglom's limit law: If $\sigma < \infty$, then the conditional distribution of Z_n/n given $Z_n > 0$ converges as $n \to \infty$ to an exponential law with mean $\sigma^2/2$. If $\sigma = \infty$, then $Z_n/n \xrightarrow{\mathcal{D}} \infty$.

Under a third moment assumption, parts (i) and (ii) of Theorem 3 are due to Kolmogorov (1938) and Yaglom (1947), respectively.

For classical proofs of these theorems, the reader is referred to Athreya and Ney (1972), pp. 15–33 and 38–45 or Asmussen and Hering (1983), pp. 23-25, 58–63, and 74–76.

The condition $\mathbf{E}[L \log L] < \infty$ in Theorems 1 and 2 certainly appears technical, and previous proofs are indeed technical (i.e., they rely on detailed analysis of generating functions). However, while studying ergodic theory on Galton-Watson trees (Lyons, Pemantle and Peres (1993)), we found conceptual proofs of these theorems that use only the crudest estimates. In particular, the dichotomies between mean and sub-mean behavior in the first two theorems turn out to arise from the following elementary dichotomy, which is an easy consequence of the Borel-Cantelli lemmas.

LEMMA 1.1. *Let $X, X_1, X_2, \ldots$ be nonnegative i.i.d. random variables. Then*

$$\limsup_{n\to\infty} \frac{1}{n} X_n = \begin{cases} 0 & \textit{if } \mathbf{E}[X] < \infty\,, \\ \infty & \textit{if } \mathbf{E}[X] = \infty\,. \end{cases}$$

Size-biased distributions, which arise in many contexts, play an important role in the present paper. Let X be a nonnegative random variable with finite positive mean. Say that $\widehat{X}$ has the corresponding *size-biased* distribution if

$$\mathbf{E}[g(\widehat{X})] = \frac{\mathbf{E}[Xg(X)]}{\mathbf{E}[X]}$$

for every positive Borel function g. The analogous notion for random trees is the topic of Section 2.

Note that if X is an exponential random variable, then $\widehat{X}$ is the sum of two independent copies of X. It follows that if U is uniform in $[0, 1]$ and independent of $\widehat{X}$, then the product $U \cdot \widehat{X}$ has the same distribution as X. The fact that this property actually characterizes the exponential distributions is used in Section 4 to derive part (ii) of Theorem 3.

The next section is basic for the rest of the paper; Sections 3, 4 and 5, which contain the proofs of Theorems 1, 3 and 2, respectively, may be read independently of each other.

§2. Size-biased Trees.

Our proofs depend on viewing Galton-Watson processes as generating random family trees, not merely as generating various numbers of particles; of course, this goes back at least to Harris (1963). We think of these trees as rooted and labeled, with the (distinguishable) offspring of each vertex ordered from left to right. We shall define another way of growing random trees, called **size-biased Galton-Watson**. (Related ideas occur in Hawkes (1981) and in Joffe and Waugh (1982).) The law of this random tree will be denoted $\widehat{\mathbf{GW}}$, whereas the law of an ordinary Galton-Watson tree is denoted $\mathbf{GW}$. Because we have more than one probability measure, we shall use integral notation for expectations with respect to $\mathbf{GW}$ and $\widehat{\mathbf{GW}}$. Similarly, we shall use the general notation $\int X \, d(\mu \mid \mathcal{F})$ for the conditional expectation of the random variable X with respect to the measure μ and the σ-field $\mathcal{F}$.

The construction of size-biased trees can be motivated by the general principle that in order to study asymptotics, it is useful to construct a suitable limiting object first. In the supercritical case, to determine the almost sure asymptotic behavior of the martingale W_n with respect to $\mathbf{GW}$, it is natural to consider the sequence of measures $W_n \, d\mathbf{GW}$, which will converge weakly to $\widehat{\mathbf{GW}}$. When $m \le 1$, the size-biased tree may be obtained by conditioning a Galton-Watson tree to survive forever. The generation sizes of size-biased Galton-Watson trees are known as a Q-process in the case $m \le 1$; see Athreya-Ney (1972), pp. 56-60. The critical case is also discussed in Kesten (1986). Now since W_n is the normalized size of the nth generation, a probabilistic way of biasing $\mathbf{GW}$ according to it is to choose uniformly a random particle in that generation; the resulting joint distribution will be called $\widehat{\mathbf{GW}}'$. These remarks will not be used explicitly, but they motivate the definitions which follow.

Notation: For a tree t with z_n particles at level n, write $W_n(t) := z_n/m^n$. Let t be a tree of height n and let v be a vertex at the nth level of t. Denote by $[t]$ the set of infinite rooted trees T such that their truncation to the first n levels yields t. Let $[t; v]$ denote the set of **trees with distinguished paths** where the tree is in $[t]$ and the path starts from the root, does not backtrack, and goes through v.

We shall construct a measure $\widehat{\mathbf{GW}}'$ on the set of trees with distinguished paths such that for all n and all $[t; v]$ as above,

$$\widehat{\mathbf{GW}}'[t; v] = \frac{1}{m^n} \mathbf{GW}[t] \,. \tag{2.1}$$

By using the branching property and the fact that the expected number of children of v is m, it is quite easy to verify consistency of these finite-dimensional distributions; however, this verification may be skipped, as we shall give a direct construction of a measure with these marginals in a moment.

Note that *if* a measure $\widehat{\mathbf{GW}}'$ satisfying (2.1) exists, then its projection to the space of trees, which is denoted simply by $\widehat{\mathbf{GW}}$, automatically satisfies

$$\widehat{\mathbf{GW}}[t] = W_n[t]\mathbf{GW}[t]\,. \tag{2.2}$$

for all n and all trees t of height n.

Now the recursive structure of Galton-Watson trees yields a recursion for $\widehat{\mathbf{GW}}'$. Assume that t is a tree of height n and that the root of t has k children with descendant trees $t^{(1)}, t^{(2)}, \ldots, t^{(k)}$ of height $n-1$. Any vertex v in the nth level of t is in one of these, say $t^{(i)}$. Now

$$\mathbf{GW}[t] = p_k \prod_{j=1}^{k} \mathbf{GW}[t^{(j)}] = kp_k \cdot \frac{1}{k} \cdot \mathbf{GW}[t^{(i)}] \cdot \prod_{j\neq i} \mathbf{GW}[t^{(j)}]\,.$$

Thus any measure $\widehat{\mathbf{GW}}'$ which satisfies (2.1) must satisfy the recursion

$$\widehat{\mathbf{GW}}'[t;v] = \frac{kp_k}{m} \cdot \frac{1}{k} \cdot \widehat{\mathbf{GW}}'[t^{(i)};v] \cdot \prod_{j\neq i} \mathbf{GW}[t^{(j)}]\,. \tag{2.3}$$

Conversely, it is clear by induction that a probability measure $\widehat{\mathbf{GW}}'$ on the set of trees with distinguished paths which satisfies this recursion must satisfy (2.1); this observation leads to the following construction of $\widehat{\mathbf{GW}}'$.

Let $\widehat{L}$ be a random variable whose distribution is that of size-biased L; i.e.,

$$\mathbf{P}[\widehat{L} = k] = \frac{kp_k}{m}\,.$$

To construct a size-biased Galton-Watson tree $\widehat{T}$, start with an initial particle v_0. Give it a random number $\widehat{L}_0$ of children, where $\widehat{L}_0$ has the law of $\widehat{L}$. Pick one of these children at random, v_1. Give the *other* children independently ordinary Galton-Watson descendant trees and give v_1 an independent size-biased number $\widehat{L}_1$ of children. Again, pick one of the children of v_1 at random, v_2, give the others ordinary Galton-Watson descendant trees, and so on. Note that size-biased Galton-Watson trees are always infinite (there is no extinction).

Now we can finally **define** the measure $\widehat{\mathbf{GW}}'$ as the joint distribution of the random tree $\widehat{T}$ and the random path $(v_0, v_1, v_2, \ldots)$. This measure clearly satisfies the recursion (2.3), and hence also (2.1).

Note that, given the first n levels of the tree $\widehat{T}$, the measure $\widehat{\mathbf{GW}}'$ makes the vertex v_n in the random path $(v_0, v_1, \ldots)$ uniformly distributed on the nth level of $\widehat{T}$; this is not obvious from the explicit construction of this random path, but it is immediate from the formula (2.1) in which the right-hand side does not depend on v.

The probabilistic content of the assumption $\mathbf{E}[L\log L] < \infty$ will come by applying Lemma 1.1 to the variables $\{\log \widehat{L}_n\}$, since $\mathbf{E}[\log \widehat{L}] = m^{-1}\mathbf{E}[L \log L]$.

§3. Proof of the Kesten-Stigum Theorem.

Recall the following elementary result, whose proof we include for the sake of completeness:

PROPOSITION 3.1. *Either $W = 0$ a.s. or $W > 0$ a.s. on nonextinction. In other words,* $\mathbf{P}[W = 0] \in \{q, 1\}$.

Proof. Let $f(s) := \mathbf{E}[s^L]$ be the probability generating function of L. The roots of $f(s) = s$ in $[0, 1]$ are $\{q, 1\}$. Thus, it suffices to show that $\mathbf{P}[W = 0]$ is such a root. Now the ith individual of the first generation has a descendant Galton-Watson tree with, therefore, a martingale limit, $W_{[i]}$, say. These are independent and have the same distribution as W. Furthermore,

$$W = \frac{1}{m} \sum_{i=1}^{Z_1} W_{[i]},$$

or, what counts for our purposes,

$$W = 0 \iff \forall i \le Z_1 \quad W_{[i]} = 0 .$$

Conditioning on Z_1 now gives immediately the desired fact that $f(\mathbf{P}[W = 0]) = \mathbf{P}[W = 0]$. ∎

Before beginning the proof proper of the Kesten-Stigum Theorem, we need to rewrite (2.2) as follows. Let $\mathcal{F}_n$ be the σ-field generated by the first n levels of trees and $\mathbf{GW}_n$, $\widehat{\mathbf{GW}}_n$ be the restrictions of $\mathbf{GW}$, $\widehat{\mathbf{GW}}$ to $\mathcal{F}_n$. Then (2.2) is the same as

$$\frac{d\widehat{\mathbf{GW}}_n}{d\mathbf{GW}_n}(t) = W_n(t) .$$

It is convenient now to interpret the last expression for *infinite* trees t, where both sides depend only on the truncation of t to the first n levels. In order to define W for every infinite tree t, set

$$W(t) := \limsup_{n \to \infty} W_n(t) .$$

In fact, $1/W_n$ is a $\widehat{\mathbf{GW}}$-martingale and hence $\lim W_n$ exists $\widehat{\mathbf{GW}}$-a.s. From this follows the key dichotomy:

$$W = 0 \quad \mathbf{GW}\text{-a.s.} \iff \mathbf{GW} \perp \widehat{\mathbf{GW}} \iff W = \infty \quad \widehat{\mathbf{GW}}\text{-a.s.}$$

while

$$\int W \, d\mathbf{GW} = 1 \iff \widehat{\mathbf{GW}} \ll \mathbf{GW} \iff W < \infty \quad \widehat{\mathbf{GW}}\text{-a.s.}$$

(see Durrett (1991), p. 210, Exercise 3.6). This is key because it allows us to change the problem from one about the $\mathbf{GW}$-behavior of W to one about the $\widehat{\mathbf{GW}}$-behavior of W.

We now begin the proof. We first show that if $\mathbf{E}[L \log L] = \infty$, then $W = 0 \quad \mathbf{GW}$-a.s. by showing that $W = \infty \quad \widehat{\mathbf{GW}}$-a.s. This assumption, which is the same as $\mathbf{E}[\log \widehat{L}] = \infty$, yields by Lemma 1.1

$$\limsup_{n\to\infty} \frac{1}{n} \log \widehat{L}_n = \infty \qquad \widehat{\mathbf{GW}}'\text{-a.s.}$$

Since $W_n(t) \geq \widehat{L}_n/m^n$, it follows that

$$W(t) = \limsup_{n\to\infty} W_n(t) = \infty \qquad \widehat{\mathbf{GW}}\text{-a.s.},$$

as desired.

Now suppose that $\mathbf{E}[L \log L] < \infty$, i.e., $\mathbf{E}[\log \widehat{L}] < \infty$. We shall show that $W < \infty$ $\widehat{\mathbf{GW}}$-a.s., which will complete the proof. Let $\mathcal{G}$ be the σ-field generated by $\{\widehat{L}_k\,;\ k \geq 1\}$. By the conditional version of Fatou's lemma, it suffices to show that

$$\lim_{n\to\infty} \int W_n \, d(\widehat{\mathbf{GW}}' \mid \mathcal{G}) < \infty \qquad \widehat{\mathbf{GW}}'\text{-a.s.}$$

We calculate this as follows. For fixed n, let Y_k be the number of descendants at level n of the particle v_k that are not descendants of particle v_{k+1}. Thus, the total number of particles at level n is $\sum_{k=0}^{n} Y_k$. This gives

$$\int W_n \, d(\widehat{\mathbf{GW}}' \mid \mathcal{G}) = \int \frac{1}{m^n} \sum_{k=0}^{n} Y_k \, d(\widehat{\mathbf{GW}}' \mid \mathcal{G}) = \sum_{k=0}^{n} \frac{1}{m^k} \int \frac{Y_k}{m^{n-k}} \, d(\widehat{\mathbf{GW}}' \mid \mathcal{G}) \,.$$

Now, for $k < n$, the random variable Y_k/m^{n-k} is the $(n-k)$th element of the ordinary Galton-Watson martingale sequence starting with, however, $\widehat{L}_k - 1$ particles. This has expectation, therefore, just $\widehat{L}_k - 1$. Since $Y_n = 1$, we get

$$\int W_n \, d(\widehat{\mathbf{GW}}' \mid \mathcal{G}) = \sum_{k=0}^{n-1} \frac{\widehat{L}_k - 1}{m^k} + \frac{1}{m^n} \,.$$

Our assumption gives, by Lemma 1.1, that $\widehat{L}_k$ grows subexponentially, whence this series converges $\widehat{\mathbf{GW}}'$-a.s., as desired.

Remark. The fact that the limit $\widehat{\mathbf{GW}}$-law of Z_n/n is that of $\widehat{X}$, i.e., the sum of two independent exponentials with mean $\sigma^2/2$ each, is due to Harris (see Athreya and Ney (1972), pp. 59–60).

§4. Critical Processes: Proof of Theorem 3.

Lemma 4.1. *Let B_n be the event that none of the children of v_0 to the left of v_1 have any descendants in generation n. Then*

$$\lim_{n\to\infty} \widehat{\mathbf{GW}}'(B_n) = 1\,.$$

Proof. This follows from writing the above probability as

$$\widehat{\mathbf{GW}}'(B_n) = \int \sum_{l=0}^{\widehat{L}_0 - 1} \frac{1}{\widehat{L}_0} \big(1 - \mathbf{GW}(Z_{n-1} = 0)\big)^l \, d\widehat{\mathbf{GW}}'\,. \quad \blacksquare$$

Proof of (i): Let A_n be the event that v_n is the leftmost particle in generation n. As observed earlier, $\widehat{\mathbf{GW}}'(A_n \mid Z_n) = 1/Z_n$. From this, it follows that conditioning on A_n reverses the effect of size-biasing. That is, the law of the first n generations of a tree under $(\mathbf{GW} \mid Z_n > 0)$ is the same as under $(\widehat{\mathbf{GW}}' \mid A_n)$. In particular,

$$\int Z_n \, d(\widehat{\mathbf{GW}}' \mid A_n) = \int Z_n \, d(\mathbf{GW} \mid Z_n > 0) = \frac{1}{\mathbf{GW}(Z_n > 0)}.$$

We are thus required to show that

$$\frac{1}{n} \int Z_n \, d(\widehat{\mathbf{GW}}' \mid A_n) \to \frac{\sigma^2}{2}\,. \tag{4.1}$$

For any tree with a distinguished line of descendants $v_0, v_1, \ldots$, decompose the size of the nth generation by writing $Z_n = 1 + \sum_{j=0}^{n-1} Z_{n,j}$, where $Z_{n,j}$ is the number of particles at generation n descended from v_j but not from v_{j+1}. The intuition behind (4.1) is that the unconditional expectation of $Z_{n,j}$ is $\mathbf{E}[\widehat{L}] - 1 = \sigma^2$; half of these fall to the left of v_n and half to the right. Since the chance that any given particle at generation $n-k$ other than v_{n-k} has no descendant in generation n tends to 1 as $k \to \infty$, conditioning on none surviving to the left leaves us with $\sigma^2/2$.

To prove this, define $R_{n,j}$ to be the number of particles in generation n descended from those children of v_j to the right of v_{j+1} and $R_n := \sum_{j=0}^{n-1} R_{n,j}$, the number of particles in generation n to the right of v_n. Let $A_{n,j}$ be the event that $R_{n,j} = Z_{n,j}$. Couple the random variables $R_{n,j}$ to independent random variables $R^*_{n,j}$ having the $(\widehat{\mathbf{GW}}' \mid A_{n,j})$-distribution of $R_{n,j}$ as follows: Define $R^*_{n,j} := R_{n,j}$ on the event $A_{n,j}$ and to be an independent pick from the $(\widehat{\mathbf{GW}}' \mid A_{n,j})$-distribution of $R_{n,j}$ on the complementary event $\neg A_{n,j}$. Then

$R_n^* := \sum_{j=0}^{n-1} R_{n,j}^*$ has the same distribution as the $(\mathbf{GW} \mid Z_n > 0)$-law of Z_n since A_n is the intersection of the independent events $A_{n,j}$. Also, $\int R_{n,j}\, d\widehat{\mathbf{GW}}' = \sigma^2/2$ and

$$\int R_{n,j}\, d(\widehat{\mathbf{GW}}' \mid A_{n,j}) \le \int Z_{n,j}\, d\widehat{\mathbf{GW}}' = \mathbf{E}[\widehat{L}] - 1 = \sigma^2 .$$

Therefore, if $\sigma < \infty$, we have

$$\int \frac{1}{n}|R_n - R_n^*| \le \int \frac{1}{n} \sum_{j=0}^{n-1} |R_{n,j} - R_{n,j}^*| \le \frac{3\sigma^2}{2n} \sum_{j=0}^{n-1} \widehat{\mathbf{GW}}'(\neg A_{n,j}) \to 0 \tag{4.2}$$

as $n \to \infty$ by the above Lemma 4.1 applied to the descendant tree of v_j. In particular, since $\int R_n/n\, d\widehat{\mathbf{GW}}' = \sigma^2/2$, we get $\int R_n^*/n \to \sigma^2/2$. The case $\sigma = \infty$ follows from this by truncating $\widehat{L}_k$. This shows (4.1), as desired.

Proof of (ii): Suppose first that $\sigma < \infty$. Using the same notation as above, suppose that $\{n_k\}$ is a sequence tending to infinity such that the laws of $R_{n_k}^*/n_k$ converge to a probability measure μ and the $\widehat{\mathbf{GW}}$-laws of Z_{n_k}/n_k converge to a probability measure ν. By (4.2), it follows that the $\widehat{\mathbf{GW}}'$-laws of R_{n_k}/n_k tend to μ also. Let X be a random variable having the distribution μ and $\widehat{X}$ be a random variable having the distribution ν. This notation is justified, i.e., $\widehat{X}$ is indeed size-biased X, since the law of R_n^* is the same as the $(\mathbf{GW} \mid Z_n > 0)$-law of Z_n and since $\int x\, d\mu(x) = \sigma^2/2 > 0$. On the other hand, let U be a uniform $[0, 1]$-valued random variable independent of every other random variable encountered so far. Then R_n and $\lfloor U \cdot Z_n \rfloor$ have the same law (with respect to $\widehat{\mathbf{GW}}'$), while

$$\left| \frac{1}{n}\lfloor U \cdot Z_n \rfloor - \frac{1}{n} U \cdot Z_n \right| \le \frac{1}{n} \to 0 .$$

Hence X and $U \cdot \widehat{X}$ have the same distribution. However, an immediate Laplace-transform calculation shows that the only nonnegative random variables X with a positive finite mean such that X and $U \cdot \widehat{X}$ have the same distribution are the exponentials. Since we know that in our case, the mean of X is $\sigma^2/2$, it follows that X is an exponential random variable with this mean. In particular, this is independent of the sequence n_k, and hence we actually have convergence along the whole sequence $R_n^*/n \xrightarrow{\mathcal{D}} X$, as desired.

Now suppose that $\sigma = \infty$. A truncation argument shows that the $\widehat{\mathbf{GW}}$-laws of Z_n/n tend to infinity, whence so do the laws of $\lfloor U \cdot Z_n \rfloor / n$. Thus, the $(\mathbf{GW} \mid Z_n > 0)$-laws of Z_n tend to infinity as well.

§5. Subcritical Processes: Proof of Theorem 2.

Let μ_n be the law of Z_n conditioned on $Z_n > 0$. We shall show that $\{\mu_n\}$ is a stochastically increasing sequence and that its means are bounded if and only if $\mathbf{E}[L \log L] < \infty$. The theorem is then a consequence of the fact that

$$\mathbf{GW}(Z_n > 0) = \frac{\mathbf{E}[Z_n]}{\mathbf{E}[Z_n \mid Z_n > 0]},$$

which, in turn, is equal to m^n divided by the mean of μ_n.

We shall apply the following immediate consequence of measure-theoretic definitions to the laws μ_n.

Lemma 5.1. *Let $\{\nu_n\}$ be a stochastically increasing sequence of distributions on the positive integers, with finite means a_n. Let $\widehat{\nu}_n$ be size-biased, i.e., $\widehat{\nu}_n(k) = k\nu_n(k)/a_n$. If $\{\widehat{\nu}_n\}$ is tight, then $\sup a_n < \infty$, while if $\widehat{\nu}_n \to \infty$ in distribution, then $a_n \to \infty$.*

If t is any tree, let $X_n(t)$ be the leftmost particle in the first generation having at least one descendant in generation n if $Z_n > 0$, and let $Y_n(t)$ be the number of descendants of $X_n(t)$ in generation n, or zero if $Z_n = 0$. It is easy to see that

$$\mathbf{GW}(Y_n = k \mid Z_n > 0) = \mathbf{GW}(Y_n = k \mid Z_n > 0, X_n = x) = \mathbf{GW}(Z_{n-1} = k \mid Z_{n-1} > 0)$$

for all children x of the root. Since $Y_n \le Z_n$, this shows that $\{\mu_n\}$ increases stochastically in n. In the case $\mathbf{E}[L \log L] < \infty$, it therefore suffices to show that $\{\widehat{\mu}_n\}$ is tight. In the case $\mathbf{E}[L \log L] = \infty$, we must show that $\widehat{\mu}_n \to \infty$.

Recall the decomposition $Z_n = 1 + \sum_{j=0}^{n-1} Z_{n,j}$ from Section 4 and the σ-field $\mathcal{G}$ from Section 3. The easy estimate

$$\mathbf{P}[L > 0]^n \le \mathbf{GW}(Z_n > 0) \le m^n,$$

combined with the fact that v_{n-j} has $\widehat{L}_{n-j} - 1$ offspring other than v_{n-j+1}, each with an ordinary Galton-Watson descendant tree, leads via the inclusion-exclusion principle to

$$\text{(5.1)} \qquad \frac{1 \wedge \mathbf{P}[L > 0]^{j-1}(\widehat{L}_{n-j} - 1)}{2} \le \widehat{\mathbf{GW}}'(Z_{n,n-j} > 0 \mid \mathcal{G}) \le m^j \widehat{L}_{n-j}.$$

Assume first that $\mathbf{E}[L \log L] < \infty$. Using (5.1), we may estimate

$$\text{(5.2)} \qquad \widehat{\mathbf{GW}}'(\{\exists j \ge J\,;\ Z_{n,n-j} > 0\} \mid \mathcal{G}) \le \sum_{j=J}^{n} m^j \widehat{L}_{n-j} \overset{\mathcal{D}}{=} \sum_{j=J}^{n} m^j \widehat{L}_{j-J} \le m^J \sum_{k=0}^{\infty} m^k \widehat{L}_k\,.$$

Since $\{\widehat{L}_k\}$ is almost surely subexponential in k by Lemma 1.1, this sum converges, whence the left-hand side of (5.2) goes to zero in probability as $n \geq J \to \infty$. Thus $\widehat{\mathbf{GW}}'(Z_n \neq 1 + Z_{n,n-1} + \cdots + Z_{n,n-J}) \to 0$, which is the same as saying that the total variation distance between $\widehat{\mu}_n$ and $\widehat{\mu}_J$ goes to zero as $n \geq J \to \infty$. Hence $\{\widehat{\mu}_n\}$ is tight.

Assume instead that $\mathbf{E}[L \log L] = \infty$. Since $\{\widehat{L}_j\}$ almost surely exceeds any exponential in j infinitely often, the lower estimate in (5.1) shows that $\sum_{j=0}^{n} \widehat{\mathbf{GW}}'(Z_{n,n-j} \geq 1 \mid \mathcal{G})$ converges in probability to infinity as $n \to \infty$. Borel-Cantelli then shows that Z_n converges in probability to infinity under $\widehat{\mathbf{GW}}$, i.e., $\widehat{\mu}_n \to \infty$.

§6. Strong Convergence of the Conditioned Process in the Subcritical Case.

Yaglom (1947) showed that when $m < 1$ and Z_1 has a finite second moment, the conditional distribution μ_n of Z_n given $\{Z_n > 0\}$ converges to a proper probability distribution as $n \to \infty$. This was proved without the second moment assumption by Joffe (1967) and by Heathcote, Seneta and Vere-Jones (1967). The following stronger convergence result was proved, in an equivalent form, by Williamson (cf. Athreya-Ney (1972), pp. 64–65.)

Theorem 6.1. *The sequence $\{\mu_n\}$ always converges in a strong sense: if $\|\cdot\|$ denotes total variation norm, then*

$$\sum_n \|\mu_n - \mu_{n-1}\| < \infty\,.$$

For the sake of completeness, we include a proof here.

Remark. Note that this is strictly stronger than weak convergence to a probability measure, even for a stochastically increasing sequence of distributions; for example, consider the distributions ν_n on the positive integers defined by

$$\nu_n = \left(\frac{1}{1\cdot 2}, \frac{1}{2\cdot 3}, \ldots, \frac{1}{(n-1)n}, \frac{1}{n}\right)\,.$$

Proof of Theorem 6.1. Recalling the notation of the previous section and the events $A_{n,j}$ from Section 4, we see that

$$\frac{1}{2}\|\mu_n - \mu_{n-1}\| \leq \mathbf{GW}(Y_n \neq Z_n \mid Z_n > 0) = \widehat{\mathbf{GW}}'(Y_n \neq Z_n \mid A_n) = \widehat{\mathbf{GW}}'(Y_n \neq Z_n \mid A_{n,0})\,.$$

Let λ be the number of children of the root to the left of v_1 and let $s_n = \mathbf{GW}(Z_n > 0)$. Now condition on $\widehat{L}_0$ and λ and use the fact that $\inf_{n \ge 2} \widehat{\mathbf{GW}}'(A_{n,0}) =: \delta > 0$ to estimate

$$\begin{aligned}\widehat{\mathbf{GW}}'(Y_n \ne Z_n \mid A_{n,0}) &\le \delta^{-1} \widehat{\mathbf{GW}}'(\{Y_n \ne Z_n\} \cap A_{n,0}) \\ &= \delta^{-1} \sum_{k=1}^{\infty} \sum_{l=0}^{k-1} \widehat{\mathbf{GW}}'(\widehat{L} = k, \lambda = l, Y_n \ne Z_n, A_{n,0}) \\ &= \delta^{-1} \sum_{k=1}^{\infty} \frac{kp_k}{m} \sum_{l=0}^{k-1} \frac{1}{k} (1 - s_{n-1})^l [1 - (1 - s_{n-1})^{k-1-l}] .\end{aligned}$$

Sum this in n by breaking it into two pieces: those n for which $s_{n-1}^{-1} \le k$ and those for which $s_{n-1}^{-1} > k$. For the first piece, use $\sum_{l=0}^{k-1} (1 - s_{n-1})^l \le s_{n-1}^{-1}$, and for the second piece, use $\sum_{l=0}^{k-1} [1 - (1 - s_{n-1})^{k-1-l}] \le \sum_{l=0}^{k-1} (k - 1 - l) s_{n-1} \le k^2 s_{n-1}/2$. These estimates yield

$$\sum_{n=1}^{\infty} \|\mu_n - \mu_{n-1}\| \le 2\delta^{-1} \sum_k \frac{p_k}{m} \left[\sum_{s_{n-1}^{-1} \le k} s_{n-1}^{-1} + \sum_{s_{n-1} < 1/k} k^2 s_{n-1}/2 \right] .$$

By virtue of Theorem 2, we have $s_j \le m s_{j-1}$, so that each of these two inner sums is bounded by a geometric series, and the total sum is at most

$$\frac{3}{\delta(1-m)} \sum \frac{kp_k}{m} = \frac{3}{\delta(1-m)} ,$$

which is finite.

REFERENCES

Asmussen, S. and Hering, H. (1983). Branching Processes. Birkhäuser, Boston.

Athreya, K. B. and Ney, P. (1972). Branching Processes. Springer, New York.

Durrett, R. (1991). Probability: Theory and Examples. Wadsworth, Pacific Grove, California.

Harris, T. E. (1963). The Theory of Branching Processes. Springer, Berlin.

Hawkes, J. (1981). Trees generated by a simple branching process, *J. London Math. Soc.* **24**, 373–384.

Heathcote, C. R., Seneta, E., and Vere-Jones, D. (1967). A refinement of two theorems in the theory of branching processes, *Theory of Probability and its Applications* **12**, 297–301.

Joffe, A. (1967). On the Galton-Watson branching processes with mean less than one, *Ann. Math. Statist.* **38**, 264–266.

Joffe, A. and Waugh, W. A. O'N. (1982). Exact distributions of kin numbers in a Galton-Watson process, *J. Appl. Prob.* **19**, 767–775.

KESTEN, H. (1986). Subdiffusive behavior of random walk on a random cluster, *Ann. Inst. Henri Poincaré Probab. Statist.* **22**, 425–487.

KESTEN, H., NEY, P. and SPITZER, F. (1966). The Galton-Watson process with mean one and finite variance, *Theory of Probability and its Applications* **11**, 513–540.

KESTEN, H. and STIGUM, B. (1966). A limit theorem for multidimensional Galton-Watson processes, *Ann. Math. Statist.* **37**, 1211–1223.

KOLMOGOROV, A. N. (1938). On the solution of a problem in biology, *Izv. NII Matem. Mekh. Tomskogo Univ.* **2**, 7–12.

LYONS, R., PEMANTLE, R. and PERES, Y. (1993). Ergodic theory on Galton-Watson trees, I: speed of random walk and dimension of harmonic measure, *preprint.*

SENETA, E. (1970). On the supercritical branching process with immigration, *Math. Biosci.* **7**, 9–14.

YAGLOM, A. M. (1947). Certain limit theorems of the theory of branching processes, *Dokl. Acad. Nauk. SSSR* **56**, 795–798.

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY, BLOOMINGTON, IN 47405-5701

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN, MADISON, WI 53706

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, BERKELEY, CA 94720