Conceptual Proofs of $L \log L$ Criteria for Mean Behavior of Branching Processes

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. We give conceptual proofs of these theorems based on comparisons of Galton-Watson measure to another measure on the space of trees. 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 offspring 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 $m \le 1$ and $p_1 < 1$. 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:

- 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 $\mathbf{P}[Z_n > 0]$?

These questions are answered by the following three classical theorems.

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-9306954 (Lyons), DMS-9300191 (Pemantle), and DMS-9213595 (Peres), by a Presidential Faculty Fellowship (Pemantle), and by the hospitality of Washington University (Pemantle).

THEOREM A: SUPERCRITICAL PROCESSES (KESTEN AND STIGUM (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 B: SUBCRITICAL PROCESSES (HEATHCOTE, SENETA AND VERE-JONES (1967)).
The sequence $\{\mathbf{P}[Z_n > 0]/m^n\}$ *is decreasing. If* $m < 1$, *then the following are equivalent:*
(i) $\lim_{n\to\infty} \mathbf{P}[Z_n > 0]/m^n > 0$;
(ii) $\sup \mathbf{E}[Z_n \mid Z_n > 0] < \infty$;
(iii) $\mathbf{E}[L \log^+ L] < \infty$.

The fact that (i) holds if $\mathbf{E}[L^2] < \infty$ was proved by Kolmogorov (1938). It is interesting that the law of Z_n conditioned on $Z_n > 0$ always converges in a strong sense, even when its means are unbounded; see Section 6.

THEOREM C: 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 this conditional distribution converges to infinity.*

Under a third moment assumption, parts (i) and (ii) of Theorem C 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. A very short proof of the Kesten-Stigum theorem, using martingale truncation, is in Tanny (1988).

By using simple measure theory, we reduce the dichotomies between mean and sub-mean behavior in the first two theorems to easier known dichotomies concerning the growth of branching processes with immigration. These, in turn, arise from the following dichotomy, which is an immediate consequence of the Borel-Cantelli lemmas.

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

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

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 and U is uniform in $[0,1]$ and independent of $\widehat{X}$, then the product $U \cdot \widehat{X}$ has the same distribution as X. (One way to see this is by considering the first and second points of a Poisson process.) The fact that this property actually characterizes the exponential distributions (Pakes and Khattree (1992)) is used in Section 4 to derive part (ii) of Theorem C.

The next section is basic for the rest of the paper; Sections 3, 4 and 5, which contain the proofs of Theorems A, C and B, respectively, may be read independently of each other. An extension of Theorem A to branching processes in a random environment, due to Tanny (1988), is discussed in the final section.

§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**. The law of this random tree will be denoted $\widehat{\mathbf{GW}}$, whereas the law of an ordinary Galton-Watson tree is denoted $\mathbf{GW}$.

Let $\widehat{L}$ be a random variable whose distribution is that of size-biased L, i.e., $\mathbf{P}[\widehat{L} = k] = kp_k/m$. To construct a size-biased Galton-Watson tree $\widehat{T}$, start with an initial particle v_0. Give it a random number $\widehat{L}_1$ of children, where $\widehat{L}_1$ 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 $\widehat{L}_2$ of children. Again, pick one of the children of v_1 at random, call it v_2, and give the others ordinary Galton-Watson descendant trees. Continue in this way indefinitely. (See Figure 1.) Note that size-biased Galton-Watson trees are always infinite (there is no extinction).

Define the measure $\widehat{\mathbf{GW}}_*$ as the joint distribution of the random tree $\widehat{T}$ and the random path $(v_0, v_1, v_2, \ldots)$. Let $\widehat{\mathbf{GW}}$ be its marginal on the space of trees.

For a tree t with Z_n vertices at level n, write $W_n(t) := Z_n/m^n$. For any rooted tree t and any $n \geq 0$, denote by $[t]_n$ the set of rooted trees whose first n levels agree with those of t. (In particular, if the height of t is less than n, then $[t]_n = \{t\}$.) If v is a vertex at the nth level of t, then let $[t; v]_n$ denote the set of **trees with distinguished paths** such that the tree is in $[t]_n$ and the path starts from the root, does not backtrack, and goes through v.

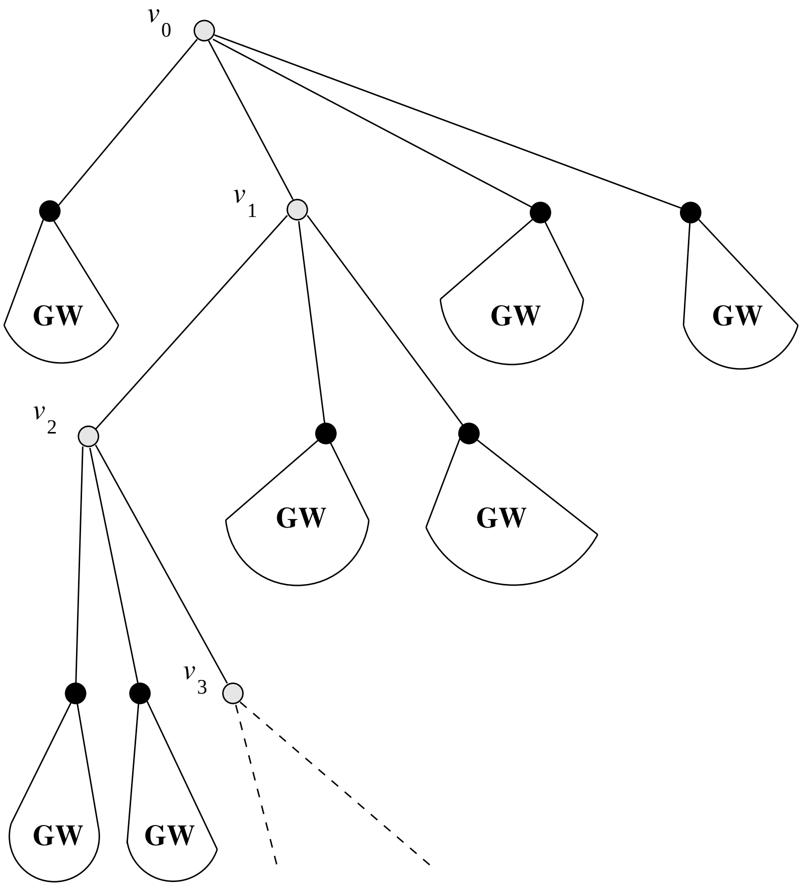

Figure 1. Schematic representation of size-biased Galton-Watson trees.

Assume that t is a tree of height at least $n+1$ and that the root of t has k children with descendant trees $t^{(1)}, t^{(2)}, \ldots, t^{(k)}$. Any vertex v in level $n+1$ of t is in one of these, say $t^{(i)}$. The measure $\widehat{\mathbf{GW}}_*$ clearly satisfies the recursion

$$\widehat{\mathbf{GW}}_*[t; v]_{n+1} = \frac{kp_k}{m} \cdot \frac{1}{k} \cdot \widehat{\mathbf{GW}}_*[t^{(i)}; v]_n \cdot \prod_{j \neq i} \mathbf{GW}[t^{(j)}]_n \,.$$

By induction, we conclude that

$$\widehat{\mathbf{GW}}_*[t; v]_n = \frac{1}{m^n} \mathbf{GW}[t]_n \tag{2.1}$$

for all n and all $[t; v]_n$ as above. Therefore,

$$\widehat{\mathbf{GW}}[t]_n = W_n(t) \mathbf{GW}[t]_n \,, \tag{2.2}$$

for all n and all trees t. From (2.1) we see 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 $\widehat{T}$.

The vertices off the "spine" $(v_0, v_1, \ldots)$ of the size-biased tree form a **branching process with immigration**. In general, such a process is defined by two distributions, an offspring distribution and an immigration distribution. The process starts with no particles, say, and at every generation $n \geq 1$, there is an immigration of Y_n particles, where Y_n are i.i.d. with the given immigration law. Meanwhile, each particle has, independently, an ordinary Galton-Watson descendant tree with the given offspring distribution.

Thus, the $\widehat{\mathbf{GW}}$-law of $Z_n - 1$ is the same as that of the generation sizes of an immigration process with $Y_n = \widehat{L}_n - 1$. The probabilistic content of the assumption $\mathbf{E}[L \log^+ L] < \infty$ will arise in applying Lemma 1.1 to the variables $\{\log^+ Y_n\}$, since $\mathbf{E}[\log^+(\widehat{L} - 1)] = m^{-1}\mathbf{E}[L \log^+(L-1)]$.

The construction of size-biased trees is not new. It and related constructions in other situations occur in Kahane and Peyrière (1976), Kallenberg (1977), Hawkes (1981), Rouault (1981), Joffe and Waugh (1982), Kesten (1986), Chauvin and Rouault (1988), Chauvin, Rouault and Wakolbinger (1991), and Waymire and Williams (1993). The paper of Waymire and Williams (1993) is the only one among these to use such a construction in a similar way to the method we use to prove Theorem A; their work was independent of and contemporaneous with ours. None of these papers use methods similar to the ones we employ for the proofs of Theorems B and C. An *a priori* motivation for the use of size-biased trees in our context comes from the general principle that in order to study asymptotics, it is useful to construct a suitable limiting object first. In the supercritical case, to study the asymptotic behavior of the martingale $W_n := Z_n/m^n$ with respect to $\mathbf{GW}$, it is natural to consider the sequence of measures $W_n \, d\mathbf{GW}$, which converge weakly to $\widehat{\mathbf{GW}}$. As pointed out by the editor, this can also be viewed as a Doob h-transform. When $m \leq 1$, 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 $m \leq 1$; see Athreya-Ney (1972), pp. 56-60. One may also view $\widehat{\mathbf{GW}}_*$ as a Campbell measure and $\widehat{\mathbf{GW}}$ as the associated Palm measure.

§3. Supercritical Processes: Proof of Theorem A.

Theorem A will be an immediate consequence of the following theorem on immigration processes.

THEOREM 3.1. (SENETA (1970)) *Let Z_n be the generation sizes of a Galton-Watson process with immigration Y_n . Let $m := \mathbf{E}[L] > 1$ be the mean of the offspring law and let Y have the same law as Y_n. If $\mathbf{E}[\log^+ Y] < \infty$, then $\lim Z_n/m^n$ exists and is finite a.s., while if $\mathbf{E}[\log^+ Y] = \infty$, then $\limsup Z_n/c^n = \infty$ a.s. for every constant $c > 0$.*

Proof. (Asmussen and Hering (1983), pp. 50–51) Assume first that $\mathbf{E}[\log^+ Y] = \infty$. By Lemma 1.1, $\limsup Y_n/c^n = \infty$ a.s. Since $Z_n \geq Y_n$, the result follows.

Now assume that $\mathbf{E}[\log^+ Y] < \infty$. Let $\mathcal{Y}$ be the σ-field generated by $\{Y_k\, ;\ k \geq 1\}$. Let $Z_{n,k}$ be the number of descendants at level n of the vertices which immigrated in generation k. Thus, the total number of vertices at level n is $\sum_{k=1}^n Z_{n,k}$. This gives

$$\mathbf{E}[Z_n/m^n \mid \mathcal{Y}] = \mathbf{E}\left[\frac{1}{m^n}\sum_{k=1}^n Z_{n,k} \,\middle|\, \mathcal{Y}\right] = \sum_{k=1}^n \frac{1}{m^k}\mathbf{E}\left[\frac{Z_{n,k}}{m^{n-k}} \,\middle|\, \mathcal{Y}\right].$$

Now, for $k < n$, the random variable $Z_{n,k}/m^{n-k}$ is the $(n-k)$th element of the ordinary Galton-Watson martingale sequence starting with, however, Y_k particles. Therefore, its expectation is just Y_k and so

$$\mathbf{E}[Z_n/m^n \mid \mathcal{Y}] = \sum_{k=1}^n \frac{Y_k}{m^k}.$$

Our assumption gives, by Lemma 1.1, that Y_k grows subexponentially, whence this series converges a.s. Since $\{Z_n/m^n\}$ is a submartingale when conditioned on $\mathcal{Y}$ with bounded expectation (given $\mathcal{Y}$), it converges a.s. ∎

To prove Theorem A, recall the following elementary result, whose proof we include for the sake of completeness:

PROPOSITION 3.2. *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,

$$W = 0 \iff \forall i \leq Z_1 \ \ W^{(i)} = 0\,.$$

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

Proof of Theorem A. 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

$$\frac{d\widehat{\mathbf{GW}}_n}{d\mathbf{GW}_n}(t) = W_n(t)\,. \tag{3.1}$$

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

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

From (3.1) follows the key dichotomy:

$$W = 0 \quad \mathbf{GW}\text{-a.s.} \iff \mathbf{GW} \perp \widehat{\mathbf{GW}} \iff W = \infty \quad \widehat{\mathbf{GW}}\text{-a.s.} \tag{3.2}$$

while

$$\int W\, d\mathbf{GW} = 1 \iff \widehat{\mathbf{GW}} \ll \mathbf{GW} \iff W < \infty \quad \widehat{\mathbf{GW}}\text{-a.s.} \tag{3.3}$$

(see Durrett (1991), p. 210, Exercise 3.6). This is the 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. Indeed, since the $\widehat{\mathbf{GW}}$-behavior of W is described by Theorem 3.1, the theorem is immediate: if $\mathbf{E}[L\log^+ L] < \infty$, i.e., $\mathbf{E}[\log^+ \widehat{L}] < \infty$, then $W < \infty$ $\widehat{\mathbf{GW}}$-a.s. by Theorem 3.1, whence $\int W\, d\mathbf{GW} = 1$ by (3.3); while if $\mathbf{E}[L\log^+ L] = \infty$, then $W = \infty$ $\widehat{\mathbf{GW}}$-a.s. by Theorem 3.1, whence $W = 0$ $\mathbf{GW}$-a.s. by (3.2). ∎

§4. Critical Processes: Proof of Theorem C.

LEMMA 4.1. *Consider a critical Galton-Watson process with a random number $Y \geq 1$ of initial particles in generation 0. Choose one of the initial particles, v, at random. Let B_n be the event that at least one of the particles to the left of v has a descendant in generation n and let r_n be the number of descendants in generation n of the particles to the right of v. Let $\beta_n := \mathbf{P}[B_n]$ and $\alpha_n := \mathbf{E}[r_n \mathbf{1}_{B_n}]$. Then $\lim_{n\to\infty} \beta_n = 0$ and, if $\mathbf{E}[Y] < \infty$, then $\lim_{n\to\infty} \alpha_n = 0$.*

Proof. The fact that $\beta_n \to 0$ follows from writing $\mathbf{P}[B_n] = \mathbf{E}\big[\mathbf{P}[B_n \mid Y, v]\big]$ and applying the bounded convergence theorem. Now

$$\alpha_n = \mathbf{E}\big[\mathbf{E}[r_n \mathbf{1}_{B_n} \mid Y, v]\big] = \mathbf{E}\big[\mathbf{E}[r_n \mid Y, v]\mathbf{P}[B_n \mid Y, v]\big] \leq \mathbf{E}\big[Y\mathbf{P}[B_n \mid Y, v]\big] = \mathbf{E}[Y\mathbf{1}_{B_n}]$$

by independence of r_n and B_n given Y and v. Thus, $\mathbf{E}[Y] < \infty$ implies that $\alpha_n \to 0$. ∎

Proof of Theorem C (i). Let A_n be the event that v_n is the leftmost vertex in generation n. By definition, $\widehat{\mathbf{GW}}_*(A_n \mid Z_n) = 1/Z_n$. 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 $(\widehat{\mathbf{GW}}_* \mid A_n)$ is the same as under $(\mathbf{GW} \mid Z_n > 0)$. 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 $Z_n = 1 + \sum_{j=1}^n Z_{n,j}$, where $Z_{n,j}$ is the number of vertices at generation n descended from v_{j-1} but not from v_j. The intuition behind (4.1) is that the unconditional $\widehat{\mathbf{GW}}_*$-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 vertex 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 vertices in generation n descended from those children of v_{j-1} to the right of v_j and $R_n := 1 + \sum_{j=1}^n R_{n,j}$, the number of vertices in generation n to the right of v_n, *inclusive*. Let $A_{n,j}$ be the event that $R_{n,j} = Z_{n,j}$. Let $R'_{n,j}$ be independent random variables with respect to a probability measure $\mathbf{Q}'$ such that $R'_{n,j}$ has the $(\widehat{\mathbf{GW}}_* \mid A_{n,j})$-distribution of $R_{n,j}$. Let $\mathbf{Q} := \widehat{\mathbf{GW}}_* \times \mathbf{Q}'$. Define

$$R^*_{n,j} := R_{n,j} \mathbf{1}_{A_{n,j}} + R'_{n,j} \mathbf{1}_{\neg A_{n,j}},$$

where $\neg$ denotes complement. Then the random variable $R^*_n := 1 + \sum_{j=1}^n R^*_{n,j}$ has the same distribution as the $(\widehat{\mathbf{GW}}_* \mid A_n)$-law of Z_n since the event $A_n = \{R_n = Z_n\}$ is the intersection of the independent events $A_{n,j}$. Also,

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

where the second inequality is due to $Z_{n,j}$ and the indicator of $A_{n,j}$ being negatively correlated. Now, for each j, we apply Lemma 4.1 with $Y = \widehat{L}_j$ to the descendant trees of the children of v_{j-1}, with $\neg A_{n,j}$ playing the role of B_{n-j}. We conclude that if $\sigma < \infty$, then

$$\begin{aligned} \int \frac{1}{n} |R_n - R^*_n| \, d\mathbf{Q} &\le \int \frac{1}{n} \sum_{j=1}^n |R_{n,j} - R^*_{n,j}| \, d\mathbf{Q} \\ &\le \frac{1}{n} \sum_{j=1}^n \int_{\neg A_{n,j}} (R_{n,j} + R^*_{n,j}) \, d\mathbf{Q} \le \frac{1}{n} \sum_{j=1}^n (\alpha_{n-j} + \sigma^2 \beta_{n-j}) \to 0 \end{aligned} \tag{4.2}$$

as $n \to \infty$. In particular, since $\int R_{n,j} \, d\widehat{\mathbf{GW}}_* = \sigma^2/2$ and hence $\int R_n/n \, d\widehat{\mathbf{GW}}_* = \sigma^2/2$, we get $\int R^*_n/n \, d\mathbf{Q} \to \sigma^2/2$. The case $\sigma = \infty$ follows from this by truncating $\widehat{L}_k$ while leaving unchanged the rest of the size-biased tree. This shows (4.1), as desired. ∎

The following simple characterization of the exponential distributions is used to prove part (ii) of the theorem.

LEMMA 4.2. (PAKES AND KHATTREE (1992)) *Let X be a nonnegative random variable with a positive finite mean and let $\widehat{X}$ have the corresponding size-biased distribution. Denote by U a uniform random variable in $[0,1]$ which is independent of $\widehat{X}$. Then $U \cdot \widehat{X}$ and X have the same distribution iff X is exponential.*

Proof. By linearity, we may assume that $\mathbf{E}[X] = 1$. For any $\lambda > 0$, we have

$$\mathbf{E}\left[e^{-\lambda U \cdot \widehat{X}}\right] = \mathbf{E}\left[\int_0^1 X e^{-\lambda u \cdot X}\, du\right] = \frac{1}{\lambda}\mathbf{E}\left[1 - e^{-\lambda X}\right],$$

which equals $\mathbf{E}[e^{-\lambda X}]$ iff $\mathbf{E}[e^{-\lambda X}] = 1/(\lambda + 1)$. By uniqueness of the Laplace transform, this holds for every $\lambda > 0$ iff X is exponential with mean 1. ∎

The following lemma is elementary.

LEMMA 4.3. *Suppose that X, X_n are nonnegative random variables with positive finite means such that $X_n \to X$ in law and $\widehat{X}_n \to Y$ in law. If Y is a proper random variable, then Y has the law of $\widehat{X}$.*

Proof of Theorem C (ii). Suppose first that $\sigma < \infty$. This ensures that the $\widehat{\mathbf{GW}}$-laws of Z_n/n have uniformly bounded means and, *a fortiori*, are tight. Let R_n and R_n^* be as in the proof of part (i). Then R_n^*/n also have uniformly bounded means and hence are tight. Therefore, there is a sequence $\{n_k\}$ tending to infinity such that $R_{n_k}^*/n_k$ converges in law to a (proper) random variable X and the $\widehat{\mathbf{GW}}$-laws of Z_{n_k}/n_k converge to the law of a (proper) random variable Y. Note that the law of $R_{n_k}^*$ is the $(\mathbf{GW} \mid Z_n > 0)$-law of Z_{n_k}. Thus, from Lemma 4.3 combined with (4.1), the variables Y and $\widehat{X}$ are identically distributed. Also, by (4.2), the $\widehat{\mathbf{GW}}_*$-laws of R_{n_k}/n_k tend to the law of X.

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 $\lceil U \cdot Z_n \rceil$ have the same law (with respect to $\widehat{\mathbf{GW}}_*$), while

$$\left| \frac{1}{n}\lceil U \cdot Z_n \rceil - \frac{1}{n} U \cdot Z_n \right| \le \frac{1}{n} \to 0\,.$$

Hence X and $U \cdot \widehat{X}$ have the same distribution. It follows from Lemma 4.2 and (4.1) that X is an exponential random variable with mean $\sigma^2/2$. In particular, the limiting distribution of $R_{n_k}^*/n_k$ is independent of the sequence n_k, and hence we actually have convergence in law of the whole sequence R_n^*/n to 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 $\lceil U \cdot Z_n \rceil/n$. Thus, the $(\mathbf{GW} \mid Z_n > 0)$-laws of Z_n/n tend to infinity as well. ∎

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). The above proof allows us to identify these two exponentials as normalized counts of the vertices to the left and right of the "spine" $(v_0, v_1, \ldots)$.

§5. Subcritical Processes: Proof of Theorem B.

Let μ_n be the law of Z_n conditioned on $Z_n > 0$. For any tree t, let $\xi_n(t)$ be the leftmost vertex in the first generation having at least one descendant in generation n if $Z_n > 0$. Let $H_n(t)$ be the number of descendants of $\xi_n(t)$ in generation n, or zero if $Z_n = 0$. It is easy to see that

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

for all children x of the root. Since $H_n \le Z_n$, this shows that $\{\mu_n\}$ increases stochastically in n. Now

$$\mathbf{GW}(Z_n > 0) = \frac{\mathbf{E}[Z_n]}{\mathbf{E}[Z_n \mid Z_n > 0]} = \frac{m^n}{\int x \, d\mu_n(x)} .$$

Therefore, $\mathbf{GW}(Z_n > 0)/m^n$ is decreasing and (i) $\Leftrightarrow$ (ii). The equivalence of (ii) and (iii) is an immediate consequence of the following routine lemma applied to the laws μ_n and the following theorem on immigration. ∎

LEMMA 5.1. *Let $\{\nu_n\}$ be a sequence of probability measures 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 $a_n \to \infty$.*

THEOREM 5.2. (HEATHCOTE (1966)) *Let Z_n be the generation sizes of a Galton-Watson process with offspring random variable L and immigration Y_n . Suppose that $m := \mathbf{E}[L] < 1$ and let Y have the same law as Y_n. If $\mathbf{E}[\log^+ Y] < \infty$, then Z_n converges in distribution to a proper random variable, while if $\mathbf{E}[\log^+ Y] = \infty$, then Z_n converges in probability to infinity.*

The following proof is a slight improvement on Asmussen and Hering (1983), pp. 52–53.

Proof. Let $\mathcal{Y}$ be the σ-field generated by $\{Y_k ;\ k \ge 1\}$. For any n, let $Z_{n,k}$ be the number of descendants at level n of the vertices which immigrated in generation k. Thus, the total number of vertices at level n is $\sum_{k=1}^n Z_{n,k}$. Since the distribution of $Z_{n,k}$ depends only on $n-k$, this total Z_n has the same distribution as $\sum_{k=1}^n Z_{2k,k}$, which is an increasing process with limit Z'_∞. By Kolmogorov's zero-one law, Z'_∞ is a.s. finite or a.s. infinite. Hence, we need only to show that $Z'_\infty < \infty$ iff $\mathbf{E}[\log^+ Y] < \infty$.

Assume that $\mathbf{E}[\log^+ Y] < \infty$. Now $\mathbf{E}[Z'_\infty \mid \mathcal{Y}] = \sum_{k=1}^\infty Y_k m^{k-1}$. Since $\{Y_k\}$ is almost surely subexponential in k by Lemma 1.1, this sum converges a.s. Therefore, Z'_∞ is finite a.s.

Now assume that $Z'_\infty < \infty$ a.s. Writing $Z_{2k,k} = \sum_{i=1}^{Y_k} \zeta_k(i)$, where $\zeta_k(i)$ are the sizes of generation $k-1$ of i.i.d. Galton-Watson branching processes with one initial particle, we have

$Z'_\infty = \sum_{k=1}^\infty \sum_{i=1}^{Y_k} \zeta_k(i)$ written as a random sum of independent random variables. Only a finite number of them are at least one, whence by the Borel-Cantelli lemma conditioned on $\mathcal{Y}$, we get $\sum_{k=1}^\infty Y_k \mathbf{GW}(Z_{k-1} \geq 1) < \infty$ a.s. Since $\mathbf{GW}(Z_{k-1} \geq 1) \geq \mathbf{P}[L > 0]^{k-1}$, it follows by Lemma 1.1 that $\mathbf{E}[\log^+ Y] < \infty$. ∎

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

REMARK. Note that this is strictly stronger than weak convergence to a probability measure, even for a stochastically increasing sequence of distributions.

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

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

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 $\widehat{L}_1$ and λ and use the fact that $\inf_{n\geq 2} \widehat{\mathbf{GW}}_*(A_{n,1}) =: \delta > 0$ to estimate

$$\begin{aligned}
\widehat{\mathbf{GW}}_*(H_n \neq Z_n \mid A_{n,1}) &\leq \delta^{-1}\widehat{\mathbf{GW}}_*(\{H_n \neq Z_n\} \cap A_{n,1}) \\
&= \delta^{-1} \sum_{k=1}^\infty \sum_{l=0}^{k-1} \widehat{\mathbf{GW}}_*(\widehat{L}_1 = k, \lambda = l, H_n \neq Z_n, A_{n,1}) \\
&= \delta^{-1} \sum_{k=1}^\infty \frac{kp_k}{m} \sum_{l=0}^{k-1} \frac{1}{k}(1 - s_{n-1})^l [1 - (1 - s_{n-1})^{k-1-l}]\,.
\end{aligned}$$

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

$$\sum_{l=0}^{k-1}[1 - (1 - s_{n-1})^{k-1-l}] \leq \sum_{l=0}^{k-1}(k - 1 - l)s_{n-1} \leq k^2 s_{n-1}/2\,.$$

These estimates yield

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

By virtue of Theorem B, we have $s_j \leq m s_{j-1}$, so that each of these two inner sums is bounded by a geometric series, whence the total is finite. ∎

§7. Stationary Random Environments.

The analogue of the Kesten-Stigum theorem for branching processes in random environments (BPRE's) is due to Tanny (1988). Our method of proof applies to this situation too. Furthermore, the probabilistic construction of size-biased trees makes apparent how to remove a technical hypothesis in Tanny's extension of the Kesten-Stigum theorem. The analytic tool needed for this turns out to be an ergodic lemma due to Tanny (1974).

In this set-up, the fixed "environment" f is replaced by a stationary ergodic sequence f_n of random environments. Vertices in generation $n-1$ have offspring according to the law with p.g.f. f_n. Assume that the process is supercritical with finite growth, i.e., $0 < \mathbf{E}[\log f_1'(1)] < \infty$. Write $M_n := \prod_{k=1}^n f_k'(1)$; this is the conditional mean of Z_n given the environment sequence $\mathbf{f} := \langle f_k \rangle$. The quotients Z_n/M_n still form a martingale with a.s. limit W. We shall also use our notation that $\widehat{L}_n$ are random variables which, given $\mathbf{f}$, are independent and have size-biased distributions, so that $f_n'(s) = f_n'(1) \sum_{k \ge 1} \mathbf{P}[\widehat{L}_n = k] s^{k-1}$. Tanny's (1988) theorem with the technical hypothesis removed is as follows.

THEOREM 7.1. *If for some $a > 0$, the sum $\sum_n \mathbf{P}[\widehat{L}_n > a^n \mid \mathbf{f}\,]$ is finite a.s., then $\mathbf{E}[W] = 1$ and $W \neq 0$ a.s. on nonextinction, while if this sum is infinite with positive probability for some $a > 0$, then $W = 0$ a.s. In case the environments f_n are i.i.d., the a.s. finiteness of this sum for some a is equivalent to the finiteness of* $\mathbf{E}[\log^+ \widehat{L}_1]$.

In the proof of Theorem 7.1, for the case of i.i.d. environments, Lemma 1.1 is used in the same way as it is for a fixed environment. However, the general case requires the following extension of Lemma 1.1.

LEMMA 7.2. (TANNY (1974)) *Let τ be an ergodic measure-preserving transformation of a probability space (X, μ), and let f be a nonnegative measurable function on X. Then* $\limsup f(\tau^n(x))/n$ *is either 0 a.s. or ∞ a.s.*

The following argument, indicated to us by Jack Feldman, is considerably shorter than the proofs in Tanny (1974) and Woś (1987). It is similar to but somewhat shorter than the proof in O'Brien (1982).

Proof of Lemma 7.2. By ergodicity, $\limsup f(\tau^n(x))/n$ is a.s. a constant $c \le \infty$. Suppose that $0 < c < \infty$. Then we may choose L so that $A := \{x\,;\ \forall k \ge L\ f(\tau^k(x))/k < 2c\}$ has $\mu(A) > \frac{9}{10}$. The ergodic theorem applied to $\mathbf{1}_A$ implies that for a.e. x, if n is sufficiently large, then there exists $k \in [L,\, n/5]$ such that $y := \tau^{n-k}(x) \in A$, and therefore

$$\frac{f(\tau^n(x))}{n} = \frac{f\left(\tau^k(y)\right)}{k} \cdot \frac{k}{n} < \frac{2c}{5}\,.$$

This contradicts the definition of c. ∎

The role of Lemma 7.2 in the proof of Theorem 7.1 is not large: By Lemma 7.2, we know that $\limsup(1/n)\log^+ \widehat{L}_n$ is a.s. 0 or a.s. infinite. Since the random variables $\widehat{L}_n$ are independent given the environment $\mathbf{f}$, the Borel-Cantelli lemma still shows that which of these alternatives holds is determined by whether the sum $\sum_n \mathbf{P}[\widehat{L}_n > a^n \mid \mathbf{f}]$ is finite for every $a > 0$ or not. In particular, this sum is finite with positive probability for some a if and only if it is finite a.s. for every a.

The proof that if $\limsup(1/n)\log^+ \widehat{L}_n = \infty$ a.s., then $W = 0$ a.s. applies without change to the case of random environments. The proof in the other direction needs only conditioning on $\mathbf{f}$ in addition to conditioning on the σ-field $\mathcal{Y}$.

The fact that if $\mathbf{E}[W] > 0$, then $W \neq 0$ a.s. on nonextinction for BPRE's can be proved by virtually the same method as that of Proposition 3.2, using the uniqueness of the conditional extinction probability (Athreya and Karlin 1971).

Acknowledgement: We are grateful to Jack Feldman for permitting the inclusion of his proof of Lemma 7.2 in this paper.

REFERENCES

ASMUSSEN, S. and HERING, H. (1983). Branching Processes. Birkhäuser, Boston.

ATHREYA, K. B. and KARLIN, S. (1971). On branching processes with random environments I: Extinction probabilities, *Ann. Math. Statist.* **42**, 1499–1520.

ATHREYA, K. B. and NEY, P. (1972). Branching Processes. Springer, New York.

CHAUVIN, B. and ROUAULT, A. (1988). KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees, *Probab. Theory Relat. Fields* **80**, 299–314.

CHAUVIN, B., ROUAULT, A., and WAKOLBINGER, A. (1991). Growing conditioned trees, *Stochastic Process. Appl.* **39**, 117–130.

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. (1966). Corrections and comments on the paper "A branching process allowing immigration", *J. Royal Statist. Soc.* **B 28**, 213–217.

HEATHCOTE, C. R., SENETA, E., and VERE-JONES, D. (1967). A refinement of two theorems in the theory of branching processes, *Theory Probab. Appl.* **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.

KAHANE, J.-P. and PEYRIÈRE, J. (1976). Sur certaines martingales de Benoit Mandelbrot, *Adv. in Math.* **22**, 131–145.

KALLENBERG, O. (1977). Stability of critical cluster fields, *Math. Nachr.* **77**, 7–43.

KESTEN, H. (1986). Subdiffusive behavior of random walk on a random cluster, *Ann. Inst. H. 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 Probab. Appl.* **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.

O'Brien, G. L. (1982). The occurrence of large values in stationary sequences, *Z. Wahrsch. Verw. Gebiete* **61**, 347–353.

Pakes, A. G. and Khattree, R. (1992). Length-biasing, characterization of laws and the moment problem, *Austral. J. Statist.* **34**, 307–322.

Rouault, A. (1981). Lois empiriques dans les processus de branchement spatiaux homogènes supercritiques, *C. R. Acad. Sci. Paris Sér. I. Math.* **292**, 933–936.

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

Tanny, D. (1974). A zero-one law for stationary sequences, *Z. Wahrsch. Verw. Gebiete* **30**, 139–148.

Tanny, D. (1988). A necessary and sufficient condition for a branching process in a random environment to grow like the product of its means, *Stochastic Process. Appl.* **28**, 123–139.

Waymire, E. C. and Williams, S. C. (1993). A cascade decomposition theory with applications to Markov and exchangeable cascades, *preprint.*

Woś, J. (1987). The filling scheme and the ergodic theorems of Kesten and Tanny, *Colloq. Math.* **52**, 263–276.

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