A Conceptual Proof of the Kesten-Stigum Theorem for Multi-type Branching Processes

by Thomas G. Kurtz, Russell Lyons, Robin Pemantle, and Yuval Peres

Abstract. We give complete proofs of the theorem of convergence of types and the Kesten-Stigum theorem for multi-type branching processes. Very little analysis is used beyond the strong law of large numbers and some basic measure theory.

Consider a multi-type Galton-Watson branching process with J types. Let $L^{(i,j)}$ be a random variable representing the number of particles of type j produced by one type-i particle in one generation. For $\mathbf{k} := (k_1, \ldots, k_J)$, let $p_{\mathbf{k}}^{(i)} = \mathbf{P}[\forall j\ L^{(i,j)} = k_j]$. Assume that $m^{(i,j)} = \mathbf{E}[L^{(i,j)}]$ is finite for all pairs (i,j). For any J-vector vector $\mathbf{x} = (x_1, \ldots, x_J)$, write $|\mathbf{x}| := x_1 + \cdots + x_J$. Let ρ be the maximum eigenvalue of the mean matrix $M := \big(m^{(i,j)}\big)$ with left unit eigenvector $\mathbf{b}$, where "unit" means that $|\mathbf{b}| = 1$. We assume that the process is supercritical (i.e., $\rho > 1$) and positive regular (i.e., some power of M has all entries positive). Let $Z_n^{(j)}$ be the number of particles of type j in generation n and $\mathbf{Z}_n := (Z_n^{(1)}, \ldots, Z_n^{(J)})$. All vectors are row vectors unless otherwise specified. The Kesten-Stigum theorem says the following (Kesten and Stigum (1966), Athreya and Ney (1972), p. 192):

Theorem 1. *There is a scalar random variable W such that*

$$\lim_{n\to\infty} \frac{\mathbf{Z}_n}{\rho^n} = W\mathbf{b} \quad a.s. \tag{1}$$

and $\mathbf{P}[W > 0] > 0$ iff

$$\mathbf{E}\left[\sum_{i,j=1}^{J} L^{(i,j)} \log^+ L^{(i,j)}\right] < \infty\,. \tag{2}$$

1991 Mathematics Subject Classification. Primary 60J80.
Key words and phrases. Galton-Watson, size-biased distribution.
Research partially supported by two Alfred P. Sloan Foundation Research Fellowships (Lyons and Pemantle), and NSF Grants DMS-9306954 (Lyons), DMS-9300191 (Pemantle), and DMS-9213595 (Peres).

We shall give a proof of this theorem that avoids much analysis, extending the proof of the single-type case given in Lyons, Pemantle and Peres (1995). The multi-type case has an additional difficulty not present in the single-type case: namely, the convergence of the quotient in (1) is no longer automatic. Thus, we begin with an elementary proof of this result simplifying Kurtz (1973).

THEOREM 2. (CONVERGENCE OF TYPES) *Almost surely on nonextinction, we have*

$$\lim_{n\to\infty} \frac{\mathbf{Z}_n}{|\mathbf{Z}_n|} = \mathbf{b}\,.$$

Note that no moment assumptions beyond finite means are made. We shall use the following elementary consequence of the strong law of large numbers.

LEMMA 3. *Suppose that N_k are random variables and that $\left\{X_n^{(k)}\,;\ n \ge 1,\ k \ge 1\right\}$ are i.i.d. mean-zero random variables. On the event $\{\exists d \in \mathbf{N}\,;\ \liminf N_{k+d}/N_k > 1\}$, we have*

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} X_n^{(k)} = 0 \qquad a.s.$$

We also need the following lemma.

LEMMA 4. *Suppose that $\{L_i^{(n)}\,;\ n, i \ge 1\}$ are i.i.d. random variables with values in $\mathbf{N}$ and with mean $m > 1$. If $\{V_n\}$ are $\mathbf{N}$-valued random variables such that $V_{n+1} \ge \sum_{i=1}^{V_n} L_i^{(n)}$ for all n, then $\liminf V_{n+1}/V_n \ge m$ a.s. on the event $E := \{\lim V_n \ne 0\}$.*

Proof. By comparison with a single-type branching process, it follows that V_n grow exponentially a.s. on E. Choose any $m' < m$. By truncating the random variables $L_i^{(n)}$ to a level with mean larger than m', we see by Chebyshev's inequality that there is some constant C such that $\mathbf{P}[V_{n+1} < m'V_n \mid V_n] \le C/V_n$ for all n. The conditional Borel-Cantelli lemma (Durrett (1991), p. 207) then implies that on the event that V_n grows exponentially, $\liminf V_{n+1}/V_n \ge m'$. Since this event occurs a.s. when E does and m' is arbitrary, the result follows. ∎

Proof of Theorem 2. Let $L_{n,k}^{(i,j)}$ be the number of type j children of the kth type-i particle in generation n, so that for all $n \ge 0$ and $1 \le j \le J$,

$$Z_{n+1}^{(j)} = \sum_{i=1}^{J} \sum_{k=1}^{Z_n^{(i)}} L_{n,k}^{(i,j)}\,.$$

Because the process is supercritical and positive regular, for each i, there is some $d \in \mathbf{N}$ such that for each k, the variables $\{Z_{dn+k}^{(i)}\,;\ n \ge 0\}$ dominate a single-type supercritical branching process. Therefore, Lemma 4 shows that the event in Lemma 3 occurs a.s. on nonextinction. Hence we may apply Lemma 3 to obtain that for each (i, j),

$$\lim_{n\to\infty} \frac{1}{Z_n^{(i)}} \sum_{k=1}^{Z_n^{(i)}} \left(L_{n,k}^{(i,j)} - m^{(i,j)} \right) = 0 \quad \text{a.s.}$$

Taking a weighted average of these equations, we see that for each j,

$$\lim_{n\to\infty} \frac{1}{|\mathbf{Z}_n|} \left(Z_{n+1}^{(j)} - \sum_{i=1}^{J} Z_n^{(i)} m^{(i,j)} \right) = \lim_{n\to\infty} \frac{1}{|\mathbf{Z}_n|} \sum_{i=1}^{J} \sum_{k=1}^{Z_n^{(i)}} \left(L_{n,k}^{(i,j)} - m^{(i,j)} \right) = 0 \quad \text{a.s.}$$

For simplicity, write $\mathbf{v}_n := \mathbf{Z}_n/|\mathbf{Z}_n|$, $\mathbf{A} := M/\rho$, and $\gamma_{n+1} := |\mathbf{Z}_{n+1}|/(\rho|\mathbf{Z}_n|)$. Then

$$\lim_{n\to\infty} |\gamma_{n+1}\mathbf{v}_{n+1} - \mathbf{v}_n\mathbf{A}| = 0 \quad \text{a.s.}$$

Since

$$\mathbf{v}_n \prod_{j=0}^{k-1} \gamma_{n-j} - \mathbf{v}_{n-k}\mathbf{A}^k = \sum_{r=0}^{k-1} (\gamma_{n-r}\mathbf{v}_{n-r} - \mathbf{v}_{n-r-1}\mathbf{A})\mathbf{A}^r \prod_{j=r+1}^{k-1} \gamma_{n-j} \, ,$$

the triangle inequality yields that for every $k \geq 1$,

$$\lim_{n\to\infty} |\mathbf{v}_n \prod_{j=0}^{k-1} \gamma_{n-j} - \mathbf{v}_{n-k}\mathbf{A}^k| = 0 \quad \text{a.s.}$$

But $\mathbf{A}^k \to \mathbf{cb}$, where $\mathbf{c}$ is a right column ρ-eigenvector. Choosing k large enough, we can therefore make

$$\limsup_{n\to\infty} |\mathbf{v}_n \prod_{j=0}^{k-1} \gamma_{n-j} - \mathbf{v}_{n-k}\mathbf{cb}|$$

arbitrarily small, which means

$$\limsup_{n\to\infty} |\mathbf{v}_n - \left(\mathbf{v}_{n-k}\mathbf{c} / \prod_{j=0}^{k-1} \gamma_{n-j} \right) \mathbf{b}|$$

can also be made arbitrarily small. Since $\mathbf{v}_n$ and $\mathbf{b}$ are unit vectors, this implies that $\mathbf{v}_n \to \mathbf{b}$ a.s. ∎

Proof of Theorem 1. Let $\mathbf{c}$ be a right column ρ-eigenvector. For any tree t with J possible types of vertices, define

$$W_n(t) := \rho^{-n} \frac{\mathbf{Z}_n(t)\mathbf{c}}{\mathbf{Z}_0(t)\mathbf{c}} \, .$$

For $r = 1, \ldots, J$, let $\mathbf{GW}^{(r)}$ denote multi-type Galton-Watson measure with one initial particle of type r. Then it is easily seen and well known that W_n is a $\mathbf{GW}^{(r)}$-martingale. We shall show that its limit is non-degenerate iff (2) holds.

We first construct some useful measures on trees. Set

$$\widehat{p}_{\mathbf{k}}^{(i)} := \frac{p_{\mathbf{k}}^{(i)} \mathbf{kc}}{\rho c_i} .$$

Given $r_0 \in [1, J]$, start with one particle v_0 of type r_0. Generate offspring according to the probabilities $\widehat{p}_{\mathbf{k}}^{(r_0)}$. Pick one of these children v_1 at random, with children being picked with probabilities proportional to c_j when their type is j. The children other than v_1 get ordinary independent $\mathbf{GW}^{(j)}$ trees, while v_1 gets an independent number of offspring according to the probabilities $\widehat{p}_{\mathbf{k}}^{(r_1)}$, where r_1 is the type of v_1. Again, pick one of the children of v_1 at random, call it v_2, and give the others ordinary independent $\mathbf{GW}^{(j)}$ trees, and so on.

Define the measure $\widehat{\mathbf{GW}}_*^{(r_0)}$ as the joint distribution of the random tree and the random path $(v_0, v_1, v_2, \ldots)$. Let its marginal on the space of trees be $\widehat{\mathbf{GW}}^{(r_0)}$.

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.

Assume that t is a tree of height at least $n + 1$ and that the root of t is of type r and has $\mathbf{k}$ children with descendant trees $t^{(1)}, t^{(2)}, \ldots, t^{(|\mathbf{k}|)}$ having roots of types $r_1, \ldots r_{|\mathbf{k}|}$. Any vertex v in level $n + 1$ of t is in one of these, say $t^{(i)}$. The measures $\widehat{\mathbf{GW}}_*^{(r)}$ clearly satisfy the recursion

$$\begin{aligned}
\widehat{\mathbf{GW}}_*^{(r)}[t; v]_{n+1} &= \widehat{p}_{\mathbf{k}}^{(r)} \cdot \frac{c_i}{\mathbf{kc}} \widehat{\mathbf{GW}}_*^{(r_i)}[t^{(i)}; v]_n \cdot \prod_{j \neq i} \mathbf{GW}^{(r_j)}[t^{(j)}]_n \\
&= \frac{p_{\mathbf{k}}^{(r)}}{\rho} \cdot \widehat{\mathbf{GW}}_*^{(r_i)}[t^{(i)}; v]_n \cdot \prod_{j \neq i} \mathbf{GW}^{(r_j)}[t^{(j)}]_n .
\end{aligned}$$

By induction, we conclude that

$$\widehat{\mathbf{GW}}_*^{(r)}[t; v]_n = \frac{c_i}{\rho^n \mathbf{Z}_0(t) \mathbf{c}} \mathbf{GW}^{(r)}[t]_n$$

for all n and all $[t; v]_n$ as above, where v is of type i. Therefore,

$$\widehat{\mathbf{GW}}^{(r)}[t]_n = W_n(t) \mathbf{GW}^{(r)}[t]_n \tag{3}$$

for all n and all trees t.

Now (2) is equivalent to

$$\sum_{j=1}^{J} \sum_{\mathbf{k}} \widehat{p}_{\mathbf{k}}^{(j)} \log^{+} |\mathbf{k}| < \infty \,. \tag{4}$$

The remaining details of the proof are a straightforward modification of the proof for the single-type case given in Lyons, Pemantle and Peres (1995). Namely, by conditioning on the numbers of children of the vertices v_n, one shows that with respect to the measure $\widehat{\mathbf{GW}}_*^{(r_0)}$, we have that $\limsup W_n < \infty$ a.s. is equivalent to (4). On the other hand, the Radon-Nikodym relation (3) shows that $\limsup W_n < \infty$ $\widehat{\mathbf{GW}}^{(r_0)}$-a.s. is equivalent to $\lim W_n > 0$ with positive $\mathbf{GW}^{(r_0)}$-probability. ∎

REFERENCES

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

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

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

Kurtz, T. G. (1973). Almost sure convergence of the type distribution for a supercritical branching process, *unpublished manuscript.*

Lyons, R., Pemantle, R. and Peres, Y. (1995). Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes, *Ann. Probab.* **23**, 1125–1138.

Department of Mathematics, University of Wisconsin, Madison, WI 53706
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