A self-similar invariance of critical binary Galton−Watson trees

GREGORY A. BURD[1], EDWARD C. WAYMIRE[2*] and RONALD D. WINN[2**]

[1]*Department of Mathematics, University of Washington, Seattle WA 98195, USA.*
E-mail: gburd@math.washington.edu
[2]*Department of Mathematics, Oregon State University, Corvallis OR 97331, USA.*
E-mail: *waymire@math.orst.edu; **winn@math.orst.edu

A coding of hierarchical structure in finite directed tree graphs was introduced by Robert Horton in 1945; after a modification by Strahler in 1952, this has become a standard river network code defined as follows. Edges along a path adjoining a source (valence 1 vertex) to a junction (valence 3 or higher vertex) are coded as order 1. Such a path of order 1 edges is also referred to as an order 1 stream. Having defined edges and streams of order i, one now recursively defines edges and streams of order $i + 1$ by the rule that the 'order 1' streams of the tree obtained by pruning the streams of order i (and lower) are assigned order $i + 1$. This code has helped to identify a number of naturally occurring patterns in river network structure as well as in other naturally occurring dendritic structures. Let $T_{i,j}(s)$ denote the number of order $j < i$ junctions in a stream s of order i and let $\overline{T}_{i,j}$ denote the sample average over all such streams s of order i. Empirically it has been observed that $\overline{T}_{i,j}$ is, allowing small-sample fluctuations, approximately a function of $i - j$ for large classes of river networks. A calculation of Ronald Shreve published in 1969 revealed that $\mathrm{E}T_{i,j} = \frac{1}{2}2^{i-j}$ for the critical binary Galton−Watson distribution. This is our starting point. In particular, we introduce a more general notion of *stochastic self-similarity* and show that within a class of Galton−Watson trees both this and the consequent mean Toeplitz property are characteristic of the critical binary offspring distribution. In addition, we obtain interesting conditioned limit theorems corresponding to the invariance of the critical binary branching process under a pruning dynamic in the space of finite rooted trees.

Keywords: conditioned limits; Galton−Watson trees; pruning; self-similarity

1. Introduction and statement of results

A *labelled tree graph* τ rooted at ϕ may be coded as a *set* of finite sequences of positive integers $\langle i_1, i_2, \ldots, i_n \rangle \in \tau$ such that:

   (i) $\phi \in \tau$ is coded as the empty sequence;
   (ii) if $\langle i_1, \ldots, i_k \rangle \in \tau$ then $\langle i_1, \ldots i_j \rangle \in \tau \, \forall 1 \leqslant j \leqslant k$;
   (iii) if $\langle i_1, i_2, \ldots, i_n \rangle \in \tau$ then $\langle i_1, \ldots i_{n-1}, j \rangle \in \tau \, \forall 1 \leqslant j \leqslant i_n$.

If $\langle i_1, \ldots, i_n \rangle \in \tau$ then $\langle i_1, \ldots, i_{n-1} \rangle \in \tau$ is referred to as the *parent* vertex to $\langle i_1, \ldots i_n \rangle$. A pair of vertices is connected by an edge (adjacent) if and only if one of them is *parent* to the other. In this way edges may also be identified with the (unique) non-parental or *descendant*

vertex. A ghost edge is attached to the root ϕ to complete this identification. Also, the tree is naturally *directed* by the 'headward growth' convention that for $u, v \in \tau$ we write $u \to v$ if u is the parent of v. This specifies the graph structure of τ and makes τ a rooted connected graph without cycles. For simplicity, we shall refer to any such graph as a *tree*.

Let $\mathbb{T}$ be the space of labelled tree graphs rooted at ϕ. $\mathbb{T}$ may be viewed as a metric space with metric $d(\tau, \gamma) = 1/(1 + \sup\{n : \gamma|n = \tau|n\})$, where $\tau|n = \{\langle i_1, \ldots, i_k \rangle \in \tau : k \leqslant n\}$. The countable dense subset $\mathbb{T}_0$ of finite labelled tree graphs rooted at ϕ makes $\mathbb{T}$ a Polish space. The distribution of a *random tree* $\mathcal{T}$ is a probability measure on $(\mathbb{T}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-field of $\mathbb{T}$. An important class of probability distributions on $\mathbb{T}$ for this paper is the Galton–Watson distribution with single progenitor and offspring distribution p_k, $k = 0, 1, \ldots$, for which the probability assigned to a closed ball $B(\tau, 1/N)$, $\tau \in \mathbb{T}_0$, $N \in \{1, 2, \ldots\}$ is

$$P\left(B\left(\tau, \frac{1}{N}\right)\right) = GW_{\{p_k\}}\left(B\left(\tau, \frac{1}{N}\right)\right) = \prod_{v \in \tau|(N-1)} p_{l(v)}, \tag{1.1}$$

where $l(v) = \#\{j : \langle v, j \rangle \in \tau|N\}$. The river network statistics considered in this paper are based on observations of the bifurcation structure of the network. From this point of view one may assume $p_1 = 0$. In the case of a binary offspring distribution, $p_2 = p$, $p_0 = q = 1 - p$, we shall write GW_p in place of $GW_{\{p_k\}}$. The distribution of a critical or subcritical tree is supported on the denumerable set $\mathbb{T}_0$. We shall write $\mathcal{T} \in (\mathbb{T}_0, GW)$ to denote that $\mathcal{T}$ is a finite (subcritical or critical) random tree with Galton–Watson distribution. The condition $p_1 = 0$ will be in place throughout without explicit further mention. The random variable L will represent a generic offspring number distributed as $P(L = j) = GW_{\{p_k\}}(L = j) = p_j$, $j \geqslant 0$.

Let τ be a finite tree. The *Horton order* of τ is a non-negative integer-valued function W defined on the set of edges of τ as follows: If $\tau = \phi$ then define the order of the vertex ϕ to be 1. If u is a vertex of $\tau \neq \phi$ and either $l(u) = 0$ or $l(u) = 1$ and $\langle uv \rangle \in \tau$, $u \neq \phi$, $\Rightarrow v = \langle 1, 1, \ldots, 1 \rangle$, then call u a source-path vertex of order $W(u) = 1$. While $l(u) = 1$ is ruled out from randomly occurring by our condition $p_1 = 0$, the pruning operation leads naturally to such effects. So certain aspects of trees having vertices u such that $l(u) = 1$ need to be included in our preliminary definitions. An *order 1 stream* is a maximal path of source-path vertices or, equivalently, edges. Having defined edges and streams of order i, one inductively defines edges and streams of order $i + 1$ by the rule that the 'order 1' streams of the tree obtained by pruning the streams of order i (and lower) are assigned order $i + 1$; see Figure 1.1. The Horton order assigned to each vertex v of τ may also be viewed as the Horton order of the subtree of τ rooted at v. The *order* of the tree $\tau \in \mathbb{T}_0$, also denoted by W, is then given by $W = W(\phi)$. This code has helped to identify a number of naturally occurring patterns in river network structure as well as in other naturally occurring dendritic structures; see, for example, Berry and Bradley (1976), Borchert and Slade (1981), Ershov (1958), Flajolet and Prodinger (1986).

The following is an easily implemented algorithm to compute Horton orders of the edges and paths of an arbitrary given tree τ. For a vertex $u \in \tau$ let e_u denote the corresponding edge; that is, e_u is the edge connecting u to its parent. Recall that it is often convenient to assign ϕ a ghost edge. If u has no offspring then $W(e_u) = 1$, else
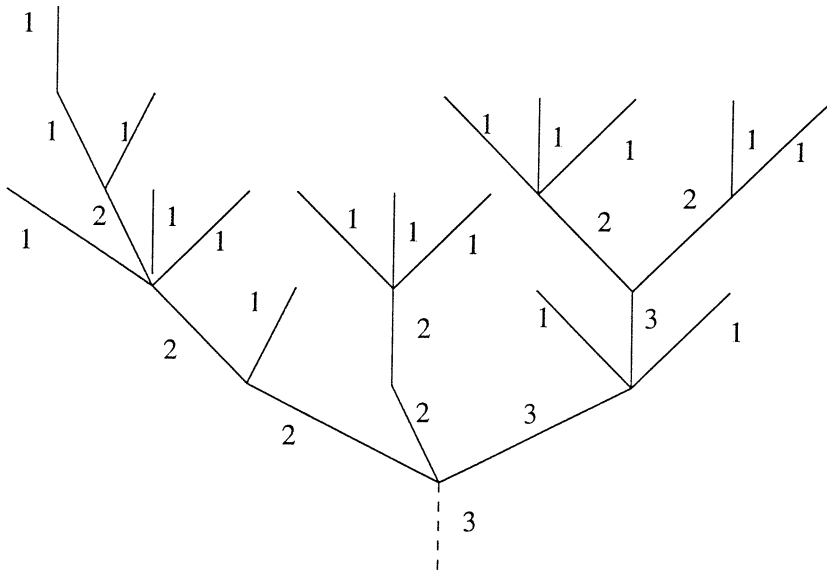
Figure 1.1. Edge, stream and network orders.

$$W(e_u) = \begin{cases} \omega & \text{if only one child of } u \text{ has (maximal) order } \omega \\ \omega + 1 & \text{if at least two children of } u \text{ have (maximal) order } \omega, \end{cases} \qquad (1.2)$$

where $\omega = \max_{v:u \to v} W(e_v)$.

Remark. While the random variables defined above depend on the parameter $\tau \in \mathbb{T}_0$ as a sample point, we follow the usual probability convention of suppressing this dependence.

Each stream s of order i is a maximal path which either consists of a single *terminal vertex* of order i, or contains an *initial vertex* of the path (that closest to the root, with a parent of order $i+1$ or higher and at least one offspring of order i and others of lower order) and a distinct *terminal vertex* (that farthest from the root, having at least two offspring of order $i-1$). Let $T_{i,j}(s)$ denote the number of order $j < i$ subtrees rooted at the non-terminal vertices along a stream s in τ of order i, and let $\overline{T}_{i,j}$ denote the arithmetic average of $T_{i,j}(s)$s over all such streams s of order i. Empirically it has been observed that $\overline{T}_{i,j}$ is, allowing small-sample fluctuations, approximately a function of $i - j$ for large classes of river networks. Table 1.1 provides sample data on generators for a river network in Kentucky; see Ossiander *et al.* (1997), Peckham (1995), Shreve (1969), Gupta and Waymire (1998) for other sample data and further river network properties. With this as our motivation, in this paper we introduce a notion of *stochastic self-similarity* which we consider for the class of Galton–Watson trees. This leads to an extension of a calculation of Shreve (1969) showing that $\mathrm{E}T_{i,j} = \frac{1}{2}2^{i-j}$ for the critical binary Galton–Watson distribution. To be precise, $T_{i,j}$ will be defined by $T_{i,j} = T_{i,j}(s_\phi)$, where s_ϕ denotes the leftmost order i stream closest to the root of a tree of order $i+1$ or larger. In view of the regenerative

Table 1.1 Sample generators for a Kentucky River subnetwork

1.13	0.00	0.00	0.00	0.00	0.00	0.00
2.87	1.13	0.00	0.00	0.00	0.00	0.00
6.78	2.70	1.05	0.00	0.00	0.00	0.00
14.71	5.89	2.78	1.16	0.00	0.00	0.00
51.12	20.00	10.75	3.88	1.88	0.00	0.00
84.50	28.50	14.50	6.00	2.50	1.50	0.00
86.00	24.00	8.00	3.00	2.00	1.00	0.00

with $((\overline{T}_{i+1,j}))_{1 \leqslant i,j \leqslant W} =$ labelling the rows.

*The sample values are computed from digital elevation maps using RiverTools software developed by Scott Peckham, US Geological Survey.

structure of Galton–Watson trees this definition is adequate for the problems concerning $\mathrm{E}T_{i,j}$ under the Galton–Watson distribution. For notational simplicity we make the convention that $\mathrm{E}T_{i,j}$ refers to the conditional expectation $\mathrm{E}[T_{i,j}|W \geqslant i+1]$ understood by this definition without explicit mention, unless required for a particular calculation.

Remark. While our choice of a stream of order i is to make a 'fixed but arbitrary' selection, we condition on trees of order $i+1$ or larger so that all streams of order i will have the same statistical structure. In particular, this definition ensures that there will not be an artificial truncation of the length of an order i stream by the root. In data analysis truncation at the outlet presents an additional source of fluctuation in the largest-order stream generator values.

On $\mathbb{T}_0$, the space of finite trees, define a pruning function $\overline{\pi}: \mathbb{T}_0 \to \mathbb{T}_0$, by $\overline{\pi}(\phi) = \phi$; and for $\tau \in \mathbb{T}_0$, $\tau \neq \phi$, let $\overline{\pi}(\tau)$ be the tree left after pruning off the order 1 streams from τ and then replacing each higher-order stream by a single edge. This last compression of streams to edges is necessary in order to isolate the effect of bifurcation in the definition of order; see Figure 1.2.

The definition of the order of $\tau \in \mathbb{T}_0$ can be restated in terms of the pruning function as a (hitting) *time to absorption* as follows:

Pruning   Compression
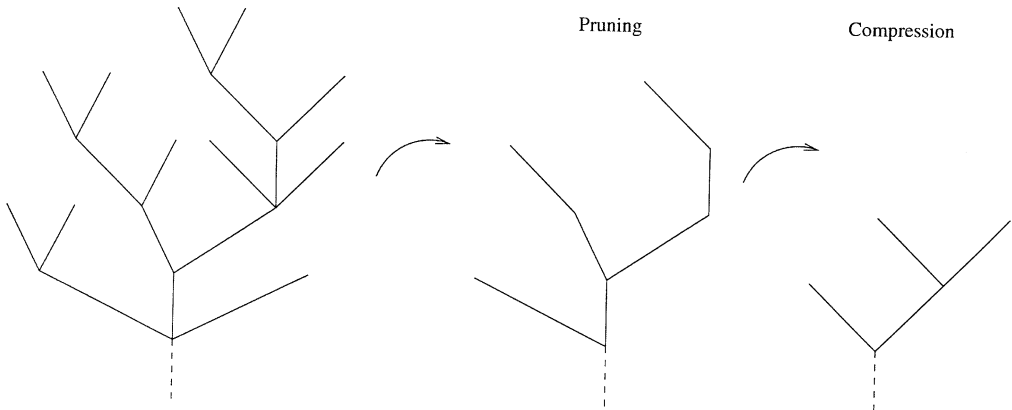
Figure 1.2. Pruning/compression dynamics.

$$W(\phi) = 1 + \inf\{n \geqslant 0 : \overline{\pi}^{(n)}(\tau) = \phi\}. \qquad (1.3)$$

In particular, one may view the successive application of pruning (with compression) as a discrete-time Markov process $\{\mathscr{T}_n\}$ evolving deterministically in the space of finite trees, starting at a random initial tree, by $\mathscr{T}_{n+1} = \overline{\pi}(\mathscr{T}_n)$, $n = 0, 1, \ldots$. The particular case of interest to us in this paper is that in which $\mathscr{T}_0$ has a Galton–Watson distribution.

Remark. Götz Kersting (personal communication) has pointed out that a similar operation was considered in Neveu (1986) for a case of continuous-time Galton–Watson processes. The corresponding discrete-time process induced by constant-rate erasure starting from the tips corresponds to a random pruning dynamics on the space of trees. It will be of interest to compare the invariant distributions for these two classes of random versus deterministic tree evolutions. The example (3.26) cited at the end of this paper shows that there is overlap in the respective classes of invariant distributions.

Although the measure concentrated at ϕ, δ_ϕ, is obviously an invariant distribution for $\{\mathscr{T}_n\}$, analogously to the case of the generation sizes conditioned on non-extinction (Athreya and Ney 1972; Yaglom 1947), conditioning $\mathscr{T}_n \neq \phi$ leads to another form of invariance which we introduce in the form of a *stochastic self-similarity* as follows.

Definition. *We say that a random tree $\mathscr{T}$ with distribution P on $\mathbb{T}_0$ is stochastically self-similar if and only if*

$$P(\cdot|\mathscr{T} \neq \phi) \circ \overline{\pi}^{-1} = P(\cdot).$$

In particular for a stochastically self-similar initial distribution one has $P(\mathscr{T}_{n+1} = \tau|\mathscr{T}_n \neq \phi) = P(\mathscr{T}_n = \tau)$, $\tau \in \mathbb{T}_0$. In Sections 2 and 3 we will show that this property is equivalent to the Toeplitz symmetry of $((\mathrm{E}T_{i,j}))$ in the following sense.

Theorem 1.1. *Let $\mathscr{T} \in (\mathbb{T}_0, GW)$. Then $\mathscr{T}$ is stochastically self-similar if and only if $((\mathrm{E}T_{i,j}))$ is Toeplitz.*

While this result establishes the relevance of stochastic self-similarity to the observed Toeplitz structure in the case of the Galton–Watson trees, the following result identifies the privileged role of the critical binary offspring distribution from this point of view.

Theorem 1.2. *Let $\mathscr{T} \in (\mathbb{T}_0, GW)$. Then $((\mathrm{E}T_{i,j}))$ is Toeplitz if and only if $\mathscr{T}$ has the critical binary offspring distribution.*

Although subcritical and critical Galton–Watson trees are almost surely (a.s.) finite, interesting conditioned limit theorems given non-extinction are well known, dating back to Yaglom (1947). The invariance under pruning suggests a new conditioned limit theorem which distinguishes the critical binary offspring distribution. For this we restrict to *a.s. bounded offspring number L* in the sense that $P(L \leqslant b) = 1$ for some $b \geqslant 2$.

Theorem 1.3. *Let $\mathscr{T}_0$ be a critical Galton–Watson tree with bounded offspring number. Let $\mathscr{T}_1, \mathscr{T}_2, \ldots$ be the pruning sequence of $\mathscr{T}_0$. Then*

$$\lim_{n \to \infty} P(\mathscr{T}_{n+1} = \tau \mid \mathscr{T}_n \neq \phi) = GW_{1/2}(\tau), \ \tau \in \mathbb{T}_0.$$

To see the tendency toward binary trees under pruning in the subcritical case, we change the conditioning of Theorem 1.3 accordingly as follows.

Theorem 1.4. *Let $\mathscr{T}_0$ be a subcritical Galton–Watson tree with bounded offspring number. Let $\mathscr{T}_1, \mathscr{T}_2, \ldots$ be the pruning sequence of $\mathscr{T}_0$. Then*

$$\lim_{n \to \infty} P(\mathscr{T}_{n+1} = \phi \mid \mathscr{T}_n \neq \phi) = 1$$

but

$$\lim_{n \to \infty} P(\mathscr{T}_{n+1} = \beta \mid \mathscr{T}_{n+1} \neq \phi) = 1,$$

where $\beta = \{\phi, \langle 0 \rangle, \langle 1 \rangle\}$ is the binary fork rooted at ϕ.

Before proving these general theorems in Section 3, we shall first consider the problem in the context of binary trees in Section 2. The advantage to this is that, first, we prove our conditioned limit theorem by a reduction to binary trees in the limit and, second, we can give a simple probabilistic argument to explain the role of criticality in the binary case.

2. Critical and subcritical binary Galton–Watson trees

We begin by considering the distribution of a non-degenerate critical and subcritical binary Galton–Watson tree under pruning. Throughout this section $\overline{\pi}$ is viewed as a map restricted to the space of finite binary trees.

Proposition 2.1. *Let $\mathscr{T} \in (\mathbb{T}_0, \ GW_p)$, $p \leqslant \frac{1}{2}$. Let $\mathscr{T}_{n+1} = \overline{\pi}(\mathscr{T}_n)$, $n \geqslant 0$, $\mathscr{T}_0 = \mathscr{T}$. Then*

$$GW_p(\mathscr{T}_{n+1} = \cdot \mid \mathscr{T}_n \neq \phi) = GW^{(n+1)}(\cdot) = GW_{p(n+1)}(\cdot),$$

where $GW^{(n+1)}$ is the Galton–Watson distribution with binary offspring distribution given by

$$p^{(n+1)} = \frac{p^{(n)2}}{p^{(n)2} + q^{(n)2}}, \qquad p^{(0)} = p, \quad q^{(0)} = q.$$

The proof of Proposition 2.1 will be based on the following lemma and an induction argument that easily generalizes to higher branching numbers. To simplify notation, let us write $\tau = (\gamma_1, \gamma_2)$ to mean there exist γ_1 and $\gamma_2 \in \mathbb{T}_0$ such that $\tau = \{\phi, \langle 1, v_1 \rangle, \langle 2, v_2 \rangle, v_i \in \gamma_i, i = 1, 2\}$, where we continue the convention that identifies $\langle i \rangle$ with $\langle \phi, i \rangle$.

Lemma 2.2. *Suppose $\tau = (\gamma_1, \gamma_2)$. Then, restricting $\bar{\pi}$ to the space of binary trees, one has*

$$\bar{\pi}^{-1}(\tau) = \{(s_1, s_2) \colon s_i \in \bar{\pi}^{-1}(\gamma_i), \ i = 1, 2\} \cup \{(s_1, s_2) \colon \exists! i, \ s_i \in \bar{\pi}^{-1}(\tau), \ s_j = \phi, \ j \neq i\}.$$

Proof of Proposition 2.1. It is sufficient to consider $P(\mathscr{T}_1 = \phi | \mathscr{T}_0 \neq \phi)$, as the general assertion will then follow by induction. Since

$$\bar{\pi}^{-1}(\{\phi\}) = \{\{\phi\}\} \cup \{\{\phi, \langle 1 \rangle, \langle 2 \rangle\}\} \cup \{\{\phi, \langle 1 \rangle, \langle 2 \rangle, \langle 1, 1 \rangle, \langle 1, 2 \rangle\}\}$$

$$\cup \{\{\phi, \langle 1 \rangle, \langle 2 \rangle, \langle 2, 1 \rangle, \langle 2, 2 \rangle\}\} \cup \dots,$$

we have

$$P(\mathscr{T}_1 = \phi | \mathscr{T}_0 \neq \phi) = \frac{1}{p} \sum_{k=1}^{\infty} 2^{k-1} p^k q^{k+1} = \frac{q^2}{p^2 + q^2}.$$

Define

$$q^{(1)} = \frac{q^2}{p^2 + q^2}, \qquad p^{(1)} = \frac{p^2}{p^2 + q^2}.$$

Next let $\beta = \{\phi, \langle 1 \rangle, \langle 2 \rangle\}$ and consider $p_\beta = P(\mathscr{T}_1 = \beta | \mathscr{T}_0 \neq \phi)$. Using Lemma 2.2, one obtains

$$p_\beta = \frac{1}{p} 2 p^2 q p_\beta + \frac{1}{p} p^3 (q^{(1)})^2$$

and, solving for p_β, it follows that

$$P(\mathscr{T}_1 = \beta | \mathscr{T}_0 \neq \phi) = p_\beta = p^{(1)} (q^{(1)})^2.$$

Now we will induct on the height of the tree. The theorem has been verified if the height of the tree is 0 or 1. Suppose that every tree of height less than or equal to h satisfies the theorem. Let $\tau \in \mathbb{T}_0$ be a binary tree of height $h + 1$. Necessarily $\tau = (\gamma_1, \gamma_2)$ where one of the two branches has height h and the other has height h or less. Again by Lemma 2.2 and a conventional abuse of notation,

$$GW_p(\bar{\pi}(\gamma) = \tau | \gamma \neq \phi) = \frac{1}{p} GW_p(\bar{\pi}^{-1}(\gamma_1)) GW_p(\bar{\pi}^{-1}(\gamma_2)) + \frac{2}{p} GW_p(\phi) GW_p(\bar{\pi}^{-1}(\tau))$$

Applying the induction hypothesis and solving gives

$$GW_p(\bar{\pi}(\gamma) = \tau | \gamma \neq \phi) = \frac{p^2}{p^2 + q^2} GW^{(1)}(\gamma_1) GW^{(1)}(\gamma_2)$$

$$= p^{(1)} GW^{(1)}(\gamma_1) GW^{(1)}(\gamma_2).$$

This completes the induction argument. $\qquad\square$

At this point it is natural to ask which finite binary Galton–Watson trees remain (conditionally) invariant under $\bar{\pi}(\cdot)$. One readily has the following corollary.

Corollary 2.3. $\mathscr{T} \in (\mathbb{T}_0, GW_p)$ *is stochastically self-similar if and only if* $p = q = \frac{1}{2}$.

In this direction we wish also to consider the asymptotic behaviour of the $\overline{\pi}$ dynamics on subcritical binary trees.

Proposition 2.4. *Suppose* $\mathscr{T}_0$ *has a subcritical binary Galton–Watson distribution with bifurcation probability p. Then*

$$\lim_{n \to \infty} P(\mathscr{T}_{n+1} = \tau | \mathscr{T}_n \neq \phi) = 0, \qquad \tau \in \mathbb{T}_0, \quad \tau \neq \phi.$$

Proof. Observe that $p^{(n)}$ is a bounded decreasing sequence and $r = \lim_{n \to \infty} p^{(n)} \leqslant \frac{1}{2}$ satisfies the quadratic equation

$$r = \frac{r^2}{1 - 2r(1 - r)}.$$

$\square$

Proposition 2.5. *Suppose* $\mathscr{T}_0 \in (\mathbb{T}_0, GW_p)$; *then*

$$GW_p(W = n + 1) = q^{(n)} \prod_{k=0}^{n-1} p^{(k)}, \qquad n \geqslant 0,$$

where the parameters $q^{(n)}$, $p^{(n)}$ *are defined inductively by*

$$q^{(n)} = \frac{q^{(n-1)2}}{p^{(n-1)2} + q^{(n-1)2}}, \qquad p^{(n)} = \frac{p^{(n-1)2}}{p^{(n-1)2} + q^{(n-1)2}}.$$

In particular, the distribution of Horton order W is geometric if and only if $p = q = \frac{1}{2}$.

Proof. Observe first that $GW_p(W = 1) = q = q^{(0)}$. Now, for $m \geqslant 2$, one has

$$GW_p(W = m) = GW_p(\overline{\pi}(\mathscr{T})^{(m-1)} = \phi, \overline{\pi}(\mathscr{T})^{(k)} \neq \phi, k = 0, 1, \ldots, m - 2)$$

$$= GW^{(m-1)}(\mathscr{T} = \phi) GW^{(m-2)}(\mathscr{T} \neq \phi) \cdots GW^{(0)}(\mathscr{T} \neq \phi)$$

$$= p^{(0)} \cdots p^{(m-2)} q^{(m-1)}.$$

If $p = q = \frac{1}{2}$ then, using this and Corollary 2.3, it follows that W has a geometric distribution. Conversely if W has geometric distribution then it follows that $p^{(0)} = p^{(1)}$ by considering probability ratios. $\square$

The following theorem provides a new derivation and generalization of Shreve's formula; see Shreve (1969). It will follow from the results in Section 3 that the critical binary offspring distribution is the only case in which $\mathrm{E}T_{i,j}$ is Toeplitz among all subcritical or critical Galton–Watson trees; cf. Corollary 3.14.

Theorem 2.6. *Let $\mathscr{T} \in (\mathbb{T}_0, \; GW_p)$. Then*

$$\mathrm{E}T_{i,j} = \frac{P(W = j)}{P(W \geqslant i)}.$$

In particular, $\mathrm{E}T_{i,j} = f(i-j)$ *if and only if* $p = \frac{1}{2}$ *and, in this case,*

$$\mathrm{E}T_{i,j} = f(i-j) = \tfrac{1}{2}2^{i-j}.$$

We will use the following simple lemmas in our proof of Theorem 2.6.

Lemma 2.7. *For $\mathscr{T} \in (\mathbb{T}_0, \; GW)$, let $e_v \in \mathscr{T}$ be an edge of $\mathscr{T}$. Then $W(e_v)$ and $W = W(\phi)$ are identically distributed under GW.*

Lemma 2.8. *Let ρ_i denote the probability that an edge of order i in τ initiates a stream s of order i. Let N_i denote the length of a stream s of order i. Then N_i has a geometric distribution with parameter $\rho_i = P(W \geqslant i)$.*

Proof. N_i may be viewed as the number of order i edges to inspect in s until termination of the stream. If one begins inspection of the vertices with the terminal vertex of the stream s, as defined in Section 1, then the inspection continues independently from vertex to vertex depending on whether the order of the joining subnetwork is strictly smaller than i or not. Thus one has $1 - \rho_i = P(W < i)$. ∎

Proof of Theorem 2.6. Since $\mathscr{T}$ is binary GW it follows that the total number of side tributaries to a stream s coincides with a geometrically distributed 'inspection length' of s having parameter ρ_i computed in Lemma 2.8. Now let $s \in \mathscr{T}$ be the leftmost stream of order i closest to the root. Then the number of side tributaries to s having order j ($j < i$) is the same as the number of side tributaries of order 1 to a stream $s' \in \overline{\pi}(\mathscr{T})^{j-1}$ of order $i - (j-1)$, where s' is the image of s under $\overline{\pi}(\cdot)^{j-1}$. Now observe that, in view of the compression in $\overline{\pi}$, if one prunes one more time then $\mathrm{E}T_{i,j}$ is the difference in the expected inspection heights. That is,

$$\mathrm{E}(T_{i,j} \mid W \geqslant i+1) = \frac{1}{\rho_{i-j+1}^{(j-1)}} - \frac{1}{\rho_{i-j}^{(j)}}, \tag{2.1}$$

and the result follows from Proposition 2.5 and the following calculations:

$$GW^{(k)}(W \geqslant i) = \frac{P(W \geqslant i+k)}{P(W \geqslant k+1)}, \tag{2.2}$$

$$\rho_{i-j+1}^{(j-1)} := GW^{(j-1)}(W \geqslant i-j+1) = \frac{P(W \geqslant i)}{P(W \geqslant j)}.$$

In particular, observe from Proposition 2.5 that

$$q^{(j-1)} = \frac{P(W = j)}{P(W \geqslant j)}. \tag{2.4}$$

To see that critical binary is the only Toeplitz solution, simply note the separation of variables in the form $\mathrm{E}T_{i,j} = a_i b_j$ and note that ratios of the form $\mathrm{E}T_{i+k+1,i}/\mathrm{E}T_{i+k,i}$ and $\mathrm{E}T_{i,i-k-1}/\mathrm{E}T_{i,i-k}$ are constant functions of i. This leads to geometric solutions from which one determines that $p = \frac{1}{2}$.   □

There are two simple modifications of Proposition 2.4 which, while trivial in the binary case, lead to interesting conditioned limit theorems in the general case to be considered in the next section. Observe that if one replaces subcritical binary with critical binary in Proposition 2.4 then, in view of Corollary 2.3, the 'limit distribution' is $GW_{\frac{1}{2}}$. On the other hand, in the subcritical binary case but changing the conditioning in Proposition 2.4 to $\mathscr{T}_{n+1} \neq \phi$, one easily observes that the limit tree is a.s. a simple binary fork rooted at ϕ. We will see that each of these simple observations in the binary case continues to hold in the generality of the next section.

3. General case

In this section $\mathscr{T}$ will have a Galton–Watson distribution with subcritical or critical offspring distribution $p_k = P(L = k)$, $k = 0, 1, 2, \ldots$, with $p_1 = 0$. We say that $\mathscr{T}$ has *a.s. bounded offspring number* if and only if $P(L \leqslant b) = 1$ for some $b \geqslant 2$. This latter assumption will be made in connection with the conditioned limit theory. However, we will begin with the mean Toeplitz problem.

Theorem 3.1. *Let $\mathscr{T} \in (\mathbb{T}_0, GW)$ and let W be the order of τ, and $G(s) = \mathrm{E}s^L$ be the probability generating function of L. Then*

$$\mathrm{E}T_{i,j} = \frac{1 - \rho_i}{\rho_i} \frac{P(W = j)}{P(W < i)} (\mathrm{E}[L|L > 0] - 1),$$

where

$$1 - \rho_i = \sum_{l=2}^{\infty} \frac{p_l}{1 - p_0} P(W < i)^{l-1} = \frac{G(P(W < i)) - p_0}{(1 - p_0)P(W < i)}.$$

The proof of Theorem 3.1 will be carried out via a series of simple lemmas which also illuminate the content of some related calculations which occur later. Recall the definition of initial and terminal edges (or equivalently, vertices) from Section 1.

Remark. Note that in the non-binary case there may be contributions to $T_{i,j}$ at the initial vertex of the stream of lower order j. Conditional on $W \geqslant i + 1$, such contributions cannot occur in the binary case since the initial vertex must be an order i edge joined by an edge of order i or higher. In order to be consistent with the identification of vertices and edges established at the outset, these lower order $j < i$ contributions to $T_{i,j}$ at the initial vertex in the non-binary case are not counted since this vertex is of order $i + 1$.

Lemma 3.2. *Let N_i denote the number of edges in the leftmost stream s of order i closest to the root in $\mathcal{T}$. Also let $\hat{T}_{i,j}(s)$ be the number of order j subtrees to a non-terminal edge of order i in s.*

(i) *$\hat{T}_{i,j}(s)$ and N_i are independent.*
(ii) *$\mathrm{E}[T_{i,j}|N_i] = (N_i - 1)\mathrm{E}\hat{T}_{i,j}(s)$.*
(iii) *$\mathrm{E}T_{i,j} = \mathrm{E}[N_i - 1]\mathrm{E}\hat{T}_{i,j}(s)$.*

Proof. Note that the $\hat{T}_{i,j}(s)$ are independent and identically distributed for vertices v interior to s, but for the initial vertex at least one offspring must be order i and at least two must be order at least i. After this, assertions (ii) and (iii) follow directly from (i). To prove (i), simply observe that the offspring subtrees generated at each vertex of $\mathcal{T}$ are independent and identically distributed under the Galton–Watson distribution. □

Lemma 3.3. *Let $a_{i,j} = P(W = j)/P(W < i)$. Then*

$$\mathrm{E}\hat{T}_{i,j}(s) = a_{i,j}(\mathrm{E}[(1 - p_0)^{-1}L] - 1) = a_{i,j}(\mathrm{E}[L|L > 0] - 1).$$

Proof. To obtain $\mathrm{E}\hat{T}_{i,j}(s)$, write $\mathrm{E}\hat{T}_{i,j}(s) = \mathrm{E}((\mathrm{E}\hat{T}_{i,j}(s)|L))$ and note that, for $l \geqslant 2$,

$$\mathrm{E}\hat{T}_{i,j}(s)|L = l) = (l - 1)a_{i,j}.$$

Thus

$$\mathrm{E}(\hat{T}_{i,j}(s)|W \geqslant i + 1) = \sum_{l=2}^{\infty}(l - 1)a_{i,j}\frac{p_l}{1 - p_0} = a_{i,j}(\mathrm{E}[L|L > 0] - 1).$$

□

Lemma 3.4. *For $\tau \in \{W \geqslant i + 1\}$ let s denote the leftmost stream of order i closest to the root. Let ρ_i denote the probability that an edge of order i in s initiates s, and let N_i denote the length of s. Then N_i has a geometric distribution with parameter*

$$\rho_i = 1 - \frac{G(P(W < i)) - p_0}{(1 - p_0)P(W < i)}.$$

Proof. As in the proof of Lemma 2.8 begin the inspection with the terminal (head) vertex of the stream. Then the inspection continues independently until the occurrence of a subnetwork of order i or larger. Thus

$$1 - \rho_i = \sum_{l=2}^{\infty}\frac{p_l}{1 - p_0}P(W < i)^{l-1} = \frac{G(P(W < i)) - p_0}{(1 - p_0)P(W < i)}$$

□

Putting the above lemmas together we obtain Theorem 3.1, namely

$$\mathrm{E}T_{i,j} = \mathrm{E}\hat{T}_{i,j}(s)\mathrm{E}(N_i - 1)$$

$$= \frac{1 - \rho_i}{\rho_i} a_{i,j}(\mathrm{E}[L|L > 0] - 1). \tag{3.1}$$

Note that in the binary case this reduces to

$$\mathrm{E}T_{i,j} = \frac{1 - \rho_i}{\rho_i} a_{i,j} = \frac{P(W = j)}{P(W \geqslant i)}, \tag{3.2}$$

as proved by another method in Theorem 2.6.

The separation of variables in Theorem 3.1 makes the following result obvious.

Theorem 3.5. *Let $\mathscr{T} \in (\mathbb{T}_0, GW)$. If $\mathrm{E}T_{i,j} = f(i - j)$ then the order W has a geometric distribution.*

We will see below that Theorem 3.5 implies $p_0 = \frac{1}{2}$ in the case of bounded offspring number (Lemma 3.15, p. 19). This in turn will be enough to show that in this case $\mathrm{E}T_{i,j} = f(i - j)$ if and only if the offspring distribution is critical binary (cf. Theorem 3.16). For now, we turn our attention to our general notion of stochastic self-similarity and conditioned limits.

Theorem 3.6. *Let $\mathscr{T} \in (\mathbb{T}_0, GW_{\{p_k\}})$, $\mathscr{T}_1 = \overline{\pi}(\tau)$, $\mathscr{T}_0 = \mathscr{T}$. Let $G(t) = \sum_{j=0}^{\infty} p_j t^j$, $G^{(k)}(t) = \sum_{j=k}^{\infty} j(j - 1) \cdots (j - k + 1) p_j t^{j-k}$, $k \geqslant 1$. Then*

$$GW_{\{p_k\}}(\mathscr{T}_1 = \cdot | \mathscr{T}_0 \neq \phi) = GW^{(1)}(\cdot) = GW_{\{p_k^{(1)}\}}(\cdot),$$

where

$$p_0^{(1)} = \frac{G(p_0) - p_0}{(1 - p_0)(1 - G'(p_0))}, \qquad p_1^{(1)} = 0,$$

$$p_k^{(1)} = \frac{(1 - p_0)^{k-1} G^{(k)}(p_0)}{k!(1 - G'(p_0))}, \qquad k \geqslant 2.$$

The proof of Theorem 3.6 is an extension of the proof of Proposition 2.1. However, the following lemma is used to handle the possibility of non-binary branches. Some notation is needed to simplify matters. If $\sigma = (\sigma_1, \ldots, \sigma_k)$ is a tree in $\mathbb{T}_0$, then let $V_\phi(\sigma) = \{i \in \{1, \ldots, k\} : \sigma_i = \{\phi\}\}$. Then, define the subsequence $i_j(\sigma)$ by $i_1 = \min\{i \in \{1, \ldots, k\} : i \notin V_\phi(\sigma)\}$ and $i_j = \min\{i \in \{i_{j-1}, \cdots k\} : i \notin V_\phi(\sigma)\}$. Note that i_1 may not exist, and more generally, i_{j+1} may not exist if i_j does.

Lemma 3.7. *Suppose $\gamma = (\gamma_1, \ldots, \gamma_N)$. Then*

$$\overline{\pi}^{-1}(\gamma) = \bigcup_{k=2}^{\infty} \{(\sigma_1, \ldots, \sigma_k): \exists! i, \sigma_i \in \overline{\pi}^{-1}(\gamma), \sigma_j = \phi, j \neq i\}$$

$$\bigcup_{m=N}^{\infty} \{\sigma = (\sigma_1, \ldots, \sigma_m): \#(V_\phi(\sigma)) = m - N, \sigma_{i_j(\sigma)} \in \overline{\pi}^{-1}(\gamma_j) \forall j \in \{1, \cdots N\}\}.$$

Proof of Theorem 3.6. First, we will derive the conditional probabilities for the root and height 1 trees. Lemma 3.7 does not apply directly to the root; however,

$$\overline{\pi}^{-1}(\phi) = \{\{\phi\}\} \bigcup_{k=2}^{\infty} \{\{\phi, \langle 1 \rangle, \ldots, \langle k \rangle\}\} \bigcup_{k=2}^{\infty} \{\{(\sigma_1, \ldots, \sigma_k): \exists! i, \sigma_i \in \overline{\pi}^{-1}(\phi), \sigma_j = \phi, j \neq i\}\}.$$

Thus,

$$P(\mathcal{T}_1 = \phi | \mathcal{T}_0 \neq \phi) = \frac{1}{1 - p_0} \left[\sum_{k=2}^{\infty} p_k(p_0)^k + \sum_{k=2}^{\infty} kp_k(p_0)^{k-1}(1 - p_0)P(\mathcal{T}_1 = \phi | \mathcal{T}_0 \neq \phi) \right].$$

Note that $G(p_0) = \sum_{k=0}^{\infty} p_k(p_0)^k$ and $G'(p_0) = \sum_{k=2}^{\infty} kp_k(p_0)^{k-1}$. Now solve for the desired probability to obtain

$$p_0^{(1)} = P(\mathcal{T}_1 = \phi | \mathcal{T}_0 \neq \phi) = \frac{G(p_0) - p_0}{(1 - p_0)(1 - G'(p_0))}.$$

If γ is a height 1 tree, then $\gamma = \{\phi, \langle 1 \rangle, \ldots, \langle N \rangle\}$ for some N. Using the lemma gives

$$P(\mathcal{T}_1 = \tau | \mathcal{T}_0 \neq \phi) = \frac{1}{1 - p_0} \left[\sum_{k=2}^{\infty} kp_k(p_0)^{k-1}(1 - p_0)P(\mathcal{T}_1 = \tau | \mathcal{T}_0 \neq \phi) \right.$$

$$\left. + \sum_{m=N}^{\infty} \binom{m}{N} (p_0)^{m-N} p_m(1 - p_0)^N (p_0^{(1)})^{m-N} \right].$$

Solving for $p_N^{(1)} = P(\mathcal{T}_1 = \tau | \mathcal{T}_0 \neq \phi)$ gives the desired result. The proposition is now verified for the root and height 1 trees. Again, we proceed by induction on tree height to complete the proof. Assume that all trees in $\mathbb{T}_0$ of height h satisfy the proposition. Let γ be a tree of height $h + 1$. Then $\gamma = (\gamma_1, \ldots, \gamma_N)$ for some N, there is an i such that γ_i has height h, and the height of $\gamma_j \leqslant h$ for all $j \in \{1, \ldots, N\}$. The lemma gives

$$P(\mathscr{T}_1 = \gamma | \mathscr{T}_0 \neq \phi) = \frac{\sum_{m=N}^{\infty} \binom{m}{N} p_m (p_0)^{m-N} (1-p_0)^{N-1}}{1 - \sum_{k=2}^{\infty} k p_k (p_0)^{k-1}} \prod_{j=1}^{N} P(\overline{\pi}(\sigma) = \gamma_j | \sigma \neq \phi)$$

$$= p_N^{(1)} \prod_{j=1}^{N} GW^{(1)}(\gamma_j).$$

$\square$

Iteration of the result in Theorem 3.6 yields the conditional distribution of $\mathscr{T}_{n+1}$ given $\mathscr{T}_n \neq \phi$ as Galton–Watson with offspring distribution given by

$$p_0^{(n+1)} = \frac{G_n(p_0^{(n)}) - p_0^{(n)}}{(1 - p_0^{(n)})(1 - G_n'(p_0^{(n)}))}, \qquad p_1^{(n)} = 0, \tag{3.3}$$

$$p_k^{(n+1)} = \frac{(1 - p_0^{(n)})^{k-1} G_n^{(k)}(p_0^{(n)})}{k!(1 - G_n'(p_0^{(n)}))}, \qquad k \geq 2, \tag{3.4}$$

where

$$G_n(t) = \sum_{j=0}^{\infty} p_j^{(n)} t^j. \tag{3.5}$$

As an alternative check that (3.3), (3.4) define a probability distribution, note that non-negativity is clear by inspection, and normalization follows by expanding $G(t)$ in a Taylor series about $t = p_0$ and induction on n. That is,

$$G(t) = G(p_0) + G'(p_0)(t - p_0) + \sum_{k=2}^{\infty} \frac{G^{(k)}(p_0)}{k!}(t - p_0)^k, \tag{3.6}$$

so that

$$1 = G(1) = G(p_0) + G'(p_0)(1 - p_0) + (1 - p_0)(1 - G'(p_0)) \sum_{k=2}^{\infty} p_k^{(1)}. \tag{3.7}$$

Similarly, by expanding $G'(t)$ in a Taylor series about $t = p_0$, one has

$$G'(t) = G'(p_0) + G''(p_0)(t - p_0) + \sum_{k=2}^{\infty} \frac{G^{(k+1)}(p_0)}{k!}(t - p_0)^k, \tag{3.8}$$

so that

$$m = G'(1) = G'(p_0) + (1 - G'(p_0)) \sum_{k=2}^{\infty} k p_k^{(1)}. \tag{3.9}$$

Proposition 3.8. *If* $\{p_k\}$ *is critical then* $\{p_k^{(n)}\}$ *is also critical. Moreover, the means of the offspring distributions* $m^{(n)} = \sum_{k=2}^{\infty} k p_k^{(n)}$ *are strictly decreasing in n if and only if* $\{p_k\}$ *is subcritical.*

Proof. Simply apply (3.9) together with induction. □

Theorem 3.9. *If* $\mathscr{T} \in (\mathbb{T}_0, GW)$ *is stochastically self-similar with* $\sum_{k=0}^{\infty} k^2 p_k < \infty$*, then* $p_0 = p_2 = \frac{1}{2}$*.*

Proof. If $\mathscr{T} \in (\mathbb{T}_0, GW)$ is stochastically self-similar then, as noted earlier, it follows from (3.9) that $\{p_k\}$ is critical. It also follows that $\frac{1}{2} \leqslant p_0 < 1$ since

$$1 = \sum_{k=2}^{\infty} k p_k \geqslant 2 \sum_{k=2}^{\infty} p_k = 2(1 - p_0). \tag{3.10}$$

Expanding $G(t)$ in a Taylor series about $t = p_0$, one has

$$G(t) = G(p_0) + G'(p_0)(t - p_0) + \sum_{k=2}^{\infty} \frac{G^{(k)}(p_0)}{k!} (t - p_0)^k. \tag{3.11}$$

Thus, using self-similarity (3.3) and (3.4), it follows that

$$G(t + p_0) = C + G'(p_0)t + (1 - p_0)(1 - G'(p_0))G\left(\frac{t}{1 - p_0}\right), \tag{3.12}$$

where C does not depend on t. Now, from criticality we have $G'(1) = 1$ and thus, differentiating $G(t + p_0)$ with respect to t, we have

$$G'(t + p_0) = \lambda G'(1) + (1 - \lambda)G'\left(\frac{t}{1 - p_0}\right), \tag{3.13}$$

where $0 < \lambda = G'(p_0) < 1$. Note that $t/(1 - p_0) < t + p_0 < 1$ for $t < 1 - p_0$. Thus, since $G'(0) = 0$, $G'(1) = 1$, $G(1) = 1$, and $G''(1) < \infty$, we have, upon taking another derivative in (3.13) at $t = 1 - p_0$, that $G'(p_0) = p_0$. Thus

$$1 = \frac{1}{p_0} G'(p_0) = \sum_{k=2}^{\infty} k p_k p_0^{k-2} \leqslant \sum_{k=2}^{\infty} k p_k = 1,$$

with strict inequality unless $p_0 = p_2 = \frac{1}{2}$. □

For the convergence problem we restrict our attention to the case of bounded number of offspring. In particular, assume

$$b = \max\{j : p_j > 0\} < \infty. \tag{3.14}$$

Writing $a_n(i) = p_{i-1}^{(n)}$, $i = 1, \dots b + 1$, equations (3.3), (3.4) may be expressed as a nonlinear dynamical system of the form

$$a_{n+1} = \frac{1}{\Gamma(a_n)} F(a_n(1))a_n, \qquad n = 0, 1, 2, \ldots, \tag{3.15}$$

where the matrix $F(t)$, defined and continuous for $t \in [0, 1]$, is given by

$$f_{ij}(t) = \begin{cases} \binom{j-1}{i-1}(1-t)^{i-1} t^{j-i} & \text{if } j \geqslant i \geqslant 3, \text{ or } i = 1 \text{ and } j \geqslant 3 \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

and the scalar $\Gamma(c)$, defined and continuous for c, is given by

$$\Gamma(c) = (1 - c(1))(1 - c \cdot \gamma(c(1))) \tag{3.17}$$

and

$$\gamma(t) = (0, 0, 2t, 3t^2, \ldots, (b+1)t^b). \tag{3.18}$$

In particular, we have

$$a_{n+1} = \frac{1}{\prod\limits_{i=0}^{n} \Gamma(a_i)} F(a_n(1)) \cdots F(a_0(1))a_0. \tag{3.19}$$

Note that the eigenvalues of $F(t)$ are given by $\lambda_1(t) = \lambda_2(t) = 0$, $\lambda_i(t) = (1-t)^{i-1}$, $i = 3, \ldots, b+1$. In particular, $\lambda_3(t)$ is the largest eigenvalue. Now observe that the stochastically self-similar distributions coincide with eigenvectors of $F(t)$, $t = p_0(1)$, with eigenvalues of the form (3.17). From this perspective our previous calculations (Corollary 2.3) show for the binary case that the only (normalized) non-negative eigenvectors have $p_0 = 1$ and $p_0 = p_2 = \frac{1}{2}$. These two eigenvectors may be distinguished by subcriticality and criticality, respectively. We will reduce the general calculation to showing that whenever the limits $\lim_{n \to \infty} p_k^{(n)}$ exist then these limits yield such an eigenvector, the selection being determined by subcriticality or criticality.

Remark. While it is obvious that one has tightness and therefore selection of convergent subsequences in the case of bounded offspring number, it is not obvious that the limit is an eigenvector in view of the shift on the left-hand side of (3.15). Thus a simple compactness and uniqueness argument seems to be ruled out.

   While for the critical case we shall see that

$$p_0^{(n)} \to \tfrac{1}{2} \qquad \text{as } n \to \infty, \tag{3.20}$$

the next results indicate a delicate difference between this and the subcritical case. In particular, in the subcritical case we will show, as in the binary case (cf. Proposition 2.4), that

$$p_0^{(n)} \to 1 \qquad \text{as } n \to \infty. \tag{3.21}$$

Unlike the conditioned limit theorems of Yaglom (1947) in which one can identify a certain monotonicity, numerical Matlab calculations show that the convergence need not be

monotonic (see Tables 3.1, 3.2), though it seems to be *eventually monotonic*. In any case, we will see that, conditional on $\mathcal{T}_{n+1} \neq \phi$, convergence to a fixed binary fork is obtained.

Remark. One can easily check by inspection of the non-negative triangular matrix $F(t)$ that regardless of the initial zeros between the first and last terms, after one iteration all of the terms, with the exception of the second, will become positive. Moreover, one can also check by examining the ratio $p_{k+1}^{(n)}/p_k^{(n)}$ that once the decreasing order of terms occurs then it will be preserved throughout the evolution.

As noted above, it is not enough to extract convergent subsequences, so the plan is to reduce the limits to the binary case by direct calculation. The calculation rests on a convexity inequality for the *size-biased* distributions $jp_j^{(n)}/m^{(n)}$, $j = 0, 1, \ldots, b$.

Table 3.1. Critical example ($m = 1$)

n	$p_0^{(n)}$	$p_1^{(n)}$	$p_2^{(n)}$	$p_3^{(n)}$	$p_4^{(n)}$
0	0.7273	0	0	0.0909	0.1818
1	0.1564	0	0.3671	0.0800	0.0064
2	0.5196	0	0.4422	0.0370	0.0011
3	0.5092	0	0.4728	0.0178	0.0002
4	0.5044	0	0.4868	0.0088	0.0001
5	0.5022	0	0.4935	0.0043	0.0000
6	0.5011	0	0.4968	0.0022	0.0000
7	0.5005	0	0.4984	0.0011	0.0000
8	0.5001	0	0.4998	0.0001	0.0000
9	0.5000	0	0.5000	0.0000	0.0000

Table 3.2. Subcritical example ($m = 0.7$)

n	$p_0^{(n)}$	$p_1^{(n)}$	$p_2^{(n)}$	$p_3^{(n)}$	$p_4^{(n)}$
0	0.8000	0	0	0.1000	0.1000
1	0.7693	0	0.3671	0.0800	0.0013
2	0.8945	0	0.1027	0.0028	0.0000
3	0.9856	0	0.0144	0.0000	0.0000
4	0.9998	0	0.0002	0.0000	0.0000
5	1.0000	0	0.0000	0.0000	0.0000

Theorem 3.10. *For bounded offspring number we have, for each $j = 3, \ldots, b$, that*

$$\lim_{n \to \infty} p_j^{(n)} = 0.$$

Proof. First let us observe that

$$1 - G_n'(p_0^{(n)}) = 1 - m^{(n)} \sum_{j=2}^{b} \frac{j p_j^{(n)}}{m^{(n)}} (p_0^{(n)})^{j-1} \geqslant 1 - m^{(n)} p_0^{(n)},$$

since $(p_0^{(n)})^{2-1}$ is the largest term in the average under the size-biased distribution. Now, in view of the calculations (3.3) and (3.9), we have $m^{(n)} \leqslant 1$ and, as in (3.10), $p_0^{(n)} \geqslant \frac{1}{2}$, so that for $k \geqslant 2$ we have from (3.4) that

$$p_k^{(n+1)} = \frac{(1 - p_0^{(n)})^{k-1} G_n^{(k)}(p_0^{(n)})}{k!(1 - G_n'(p_0^{(n)}))}$$

$$\leqslant \frac{(1 - p_0^{(n)})^{k-2}}{k!} G_n^{(k)}(p_0^{(n)})$$

$$\leqslant (\tfrac{1}{2})^{k-2} \sum_{j=k}^{b} \binom{j}{k} p_j^{(n)}, \qquad 3 \leqslant k \leqslant b.$$

Now start with $k = b$, and do a backward induction as follows. First, we have by iteration on n that

$$p_b^{(n)} \leqslant 2^{-n(b-2)} = o(1) \qquad \text{as } n \to \infty.$$

Now if $p_j^{(n)} = o(1)$ as $n \to \infty$ for $j = k+1, \ldots, b$, then since there are no more than a fixed number b of $o(1)$ terms in the sum, we have

$$p_k^{(n+1)} \leqslant (\tfrac{1}{2})^{k-2} p_k^{(n)} + (\tfrac{1}{2})^{k-2} \sum_{j=k+1}^{b} \binom{j}{k} p_j^{(n)} = (\tfrac{1}{2})^{k-2} p_k^{(n)} + o(1).$$

Iteration on n now yields for $3 \leqslant k \leqslant b$ that

$$p_k^{(n+1)} \leqslant (\tfrac{1}{2})^{(k-2)n} p_k + o(1) \sum_{i=0}^{n} (\tfrac{1}{2})^{k-2})^i = o(1).$$

$\square$

Having reduced the calculation to binary distributions, the following results are obtained.

Theorem 3.11. *Suppose $\mathcal{T}$ has a critical Galton–Watson distribution with bounded offspring number. Then*

$$\lim_{n \to \infty} P(\mathcal{T}_{n+1} = \tau | \mathcal{T}_n \neq \phi) = GW_{\frac{1}{2}}(\tau), \qquad \tau \in \mathbb{T}_0.$$

Proof. Using Theorem 3.10 and Proposition 3.8, we see in the critical case that any convergent subsequence is binary critical so that there is only one limit point. □

Theorem 3.12. *Suppose $\mathscr{T}$ has a subcritical Galton–Watson distribution with bounded offspring number. Then*

$$\lim_{n\to\infty} P(\mathscr{T}_{n+1} = \tau | \mathscr{T}_{n+1} \neq \phi) = \delta_{\{\beta\}}(\tau), \qquad \tau \in \mathbb{T}_0,$$

where $\beta = \{\phi, \langle 1 \rangle, \langle 2 \rangle\}$ is the binary fork rooted at ϕ.

Proof. In view of Theorem 3.10 it is enough to show that

$$\lim_{n\to\infty} \frac{p_2^{(n)}(p_0^{(n)})^2}{1 - p_0^{(n)}} = 1. \tag{3.22}$$

Observe from Proposition 3.8 that the means are strictly decreasing and bounded below by zero. Thus $\lim_{n\to\infty} m^{(n)}$ exists and therefore, in view of Theorem 3.10, so does $\lim_{n\to\infty} p_2^{(n)}$. So it also follows that $\lim_{n\to\infty} p_0^{(n)} = \lim_{n\to\infty}(1 - p_2^{(n)})$ exists. So the full limit distribution exists and, in view of (3.15), is a subcritical binary eigenvector with eigenvalue of the form (3.17). The result follows since there are only two such eigenvectors, $a_1 = 1$, $a_j = 0$, $2 \leq j \leq b+1$ and $a_1 = a_3 = \frac{1}{2}$, $a_2 = a_j = 0$, $4 \leq j \leq b+1$. □

The following result generalizes Proposition 2.5.

Proposition 3.13. *Suppose $\mathscr{T} \in (\mathbb{T}_0, GW_{\{p_k\}})$; then*

$$GW(W = n + 1) = p_0^{(n)} \prod_{k=0}^{n-1}(1 - p_0^{(k)}), \qquad n \geq 0.$$

Proof. Observe first that $GW_{\{p_k\}}(W = 1) = p_0 = q^{(0)}$. Now, for $m \geq 2$, we have

$$GW_{\{p_k\}}(W = m) = GW_{\{p_k\}}(\overline{\pi(\mathscr{T})}^{(m-1)} = \phi, \overline{\pi(\mathscr{T})}^{(k)} \neq \phi, k = 0, 1, \ldots, m-2)$$

$$= GW_{\{p_k\}}^{(m-1)}(\mathscr{T} = \phi)GW_{\{p_k\}}^{(m-2)}(\mathscr{T} \neq \phi) \ldots GW_{\{p_k\}}^{(0)}(\mathscr{T} \neq \phi)$$

$$= p^{(0)} \ldots p^{(m-2)}q^{(m-1)}$$

□

Corollary 3.14. *If $\mathscr{T} \in (\mathbb{T}_0, GW_{\{p_k\}})$ is stochastically self-similar then the order W has a geometric distribution.*

Lemma 3.15. *If $\mathscr{T} \in (\mathbb{T}_0, GW)$ has bounded offspring number and the order W is geometrically distributed then the offspring distribution is critical and $p_0^{(n)} = p_0 = \frac{1}{2}$.*

Proof. If $P(W = j) = (1 - \delta)^{j-1}\delta$, $j \geqslant 1$, $0 < \delta < 1$, then $\delta = P(W = 1) = p_0$. In view of Proposition 3.13, we have by induction that $p^{(n)} = p_0$, $n \geqslant 0$. Now use Theorems 3.11 and 3.12 to obtain that the offspring distribution must be critical and $p_0 = \lim_{n\to\infty} p_0^{(n)} = \frac{1}{2}$. □

Theorem 3.16. *For $\mathscr{T} \in (\mathbb{T}_0, GW)$ with bounded offspring distribution,*

$$\mathrm{E}T_{i,j} = f(i - j)$$

if and only if $p_0 = p_2 = \frac{1}{2}$.

Proof. If $\mathscr{T}$ is the critical binary tree then the Toeplitz property was shown in Section 2. Conversely, if the Toeplitz property holds then Theorem 3.5 and Lemma 3.15 yield that $p_0^{(n)} = p_0 = \frac{1}{2}$ for $n = 0, 1, 2, \ldots$. In particular, we have

$$\sum_{k\geqslant 2} p_k = \tfrac{1}{2} \tag{3.23}$$

and

$$\sum_{k\geqslant 2} kp_k = 1. \tag{3.24}$$

Multiply (3.23) by 2 and subtract from (3.24) to obtain

$$\sum_{k\geqslant 3}(k - 2)p_k = 0. \tag{3.25}$$

By non-negativity of the terms in (3.25), it now follows that $p_k = 0$ for all $k \geqslant 3$. □

While some of our results are proven more generally than for bounded offspring number, the following theorem is a summarizing statement of the main results of this paper within this setting.

Theorem 3.17. *Let $\mathscr{T} \in (\mathbb{T}_0, GW_{\{p_k\}})$ with bounded offspring number. Then $\mathscr{T}$ is stochastically self-similar if and only if $p_0 = p_2 = \frac{1}{2}$ if and only if the order W is geometrically distributed if and only if $\mathrm{E}T_{i,j}$ is Toeplitz.*

In closing, let us remark that while it is possible to construct deterministic self-similar trees naturally for arbitrarily prescribed generators (see Peckham 1995), the corresponding construction of stochastic self-similar trees is more difficult. Examples indicate that one may obtain stochastic self-similar solutions by relaxed conditions on the moments of the offspring distribution. For example, the distributions of Zolotarev (1957), defined by the family of probability generating functions

$$G_\theta(t) = t + \frac{1}{\theta}(1 - t)^\theta, \qquad 0 \leqslant t \leqslant 1, \tag{3.26}$$

for fixed parameters $\theta \in (1, 2)$, are easily checked using Theorem 3.6 to be infinite-variance critical offspring distributions of stochastically self-similar Galton–Watson distributions.

However, it is of greatest interest to river network applications to see what other possible stochastic self-similar trees with bounded offspring number may be constructed as conditioned limits by considering random initializations other than by Galton–Watson trees.

Acknowledgements

References

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

Berry, M. and Bradley, P.M. (1976) The application of network analysis to the study of branching patterns of large dendritic fields. *Brain Res.*, **109**, 111–132.

Borchert, R. and Slade, N.A. (1981) Bifurcation ratios and adaptive geometry of trees. *Botanical Gaz.*, **142**, 394–401.

Ershov, A.P. (1958) On programming of arithmetic structures. *Comm. ACM*, **1**, 3–5.

Flajolet, P. and Prodinger, H. (1986) Register allocation for unary-binary trees. *SIAM J. Comput.*, **15**, 629–640.

Gupta, V.K. and Waymire, E. (1998) Some mathematical aspects of rainfall, landforms and floods. In O. Barndorff-Nielsen, V.K. Gupta, V. Perez-Abreu, E. Waymire (eds) *Rainfall, Landforms and Floods*. Singapore: World Scientific.

Neveu, J. (1986) Erasing a branching tree. *Ann. Appl. Probab.*, (Suppl.), **18**, 101–108.

Ossiander, M, Waymire, E. and Zhang, Q. (1997) Some width function asymptotics for weighted trees. *Ann. Appl. Probab.*, **7**, 972–995.

Peckham, S. (1995) New results for self-similar trees with applications to river networks. *Water Resources Res.*, **31**, 1023–1029.

Shreve, R.D. (1969) Stream lengths and basin area in topologically random channel networks. *J. Geology*, **77**, 397–414.

Yaglom, A.M. (1947) Certain limit theorems of the theory of branching random processes. *Doklady* **56**, 795–798.

Zolotarev, V. (1957) More exact statements of several theorems in the theory of branching processes. *Theory Probab. Appl.*, **2**, 256–266.