RECURSIVE FUNCTIONS ON CONDITIONAL GALTON–WATSON TREES

NICOLAS BROUTIN, LUC DEVROYE, AND NICOLAS FRAIMAN

Abstract. A recursive function on a tree is a function in which each leaf has a given value, and each internal node has a value equal to a function of the number of children, the values of the children, and possibly an explicitly specified random element U. The value of the root is the key quantity of interest in general. In this first study, all node values and function values are in a finite set S. In this note, we describe the limit behavior when the leaf values are drawn independently from a fixed distribution on S, and the tree T_n is a random Galton–Watson tree of size n.

1. The probabilistic model

A recursive function on a tree is a function in which each leaf has a given value, and each internal node has a value equal to a function of the number of children, the values of the children, and possibly an explicitly specified random element U. The value of the root is the key quantity of interest in general. In this first study, all node values and function values are in a finite set S. In this note, we describe the limit behavior when the leaf values are drawn independently from a fixed distribution on S, and the tree T_n is a random Galton–Watson tree of size n.

A Galton–Watson (or Galton–Watson–Bienaymé) tree (see Athreya and Ney, 1972) is a rooted random ordered tree. Each node independently generates a random number of children drawn from a fixed offspring distribution ξ. The distribution of ξ defines the distribution of T, a random Galton–Watson tree. We define

$$p_i = \mathbf{P}(\xi = i)\,, i \geq 0.$$

The results sometimes are described in terms of the generating function g of ξ:

$$g(s) \stackrel{\text{def}}{=} \mathbf{E}\left(s^{\xi}\right) = \sum_{i=0}^{\infty} p_i s^i, 0 \leq s \leq 1.$$

In what follows, we are mainly interested in critical Galton–Watson trees, i.e., those having $\mathbf{E}(\xi = 1)$, and $\mathbf{P}(\xi = 1) < 1$. In addition, we assume that the variance of ξ is finite (and hence, nonzero). We denote by T_n a Galton–Watson tree conditional on its size $|T_n|$ being n. These trees encompass many known models of random

Date: June 8, 2018.

2010 *Mathematics Subject Classification.* 60J80,60J85,60G99.

Key words and phrases. Random Galton–Watson tree, probabilistic analysis of algorithms, recursive functions, branching process.

Luc Devroye was sponsored by NSERC Grant A3456. Nicolas Broutin thanks the FRQNT-CRM and Simons-CRM programmes as well as the support from Grant ANR-14-CE25-0014 (ANR GRAAL).

trees, including random Catalan trees (all binary trees of size n being equally likely), random planted plane trees (all ordered trees being equally likely), and random rooted labeled free trees or Cayley trees, thanks to an equivalence property first established by Kennedy (1975). Let $h = \gcd\{i \ge 1 : p_i > 0\}$ be the span of ξ. It is easy to see that $|T_n| \mod h = 1$, so when we provide asymptotic results on T_n, it is understood that $n \mod h = 1$ as $n \to \infty$.

Nodes in a tree are denoted by u, v and w, while their values are denoted by $V(u)$, $V(v)$ and $V(w)$. Without loss of generality, we assume that our state space is

$$S = \{1, \dots, k\}.$$

We associate independently with each node a copy of a generic uniform $[0,1]$ random variable U. Thus, $U(v)$ denotes the copy associated with node v. We are given a possibly infinite family of functions

$$f_0, f_1, f_2, \dots,$$

where f_i maps $S^i \times [0,1]$ to S. The first i arguments refer to the values of the i children of a node, while the last argument refers to the generic random variable associated with a node. In particular, each leaf v, we have

$$V(v) \stackrel{\mathcal{L}}{=} f_0(U(v)).$$

Thus, the leaf values are independent and we denote the distribution of $f_0(U)$ on S by q:

$$\mathbf{P}(f_0(U) = i) = q_i, \quad i \in S.$$

If v is an internal node with children $v_1, \dots, v_\ell$, then

$$V(v) = f_\ell(V(v_1), \dots, V(v_\ell), U(v)).$$

The value of the root node is denoted by V_n.

For a chain, with the root having value V_n and the other nodes having values $V_{n-1}, V_{n-2}, \dots, V_1, V_0$, we have $V_0 = f_0(U_0)$, $V_1 = f_1(V_0, U_1)$, $V_2 = f_1(V_1, U_1)$, and so forth. This is a purely Markovian structure. The limit behavior is entirely known and well-documented in standard texts on Markov chains such as Meyn and Tweedie (1993). The decomposition of the transition matrix graph (which places a directed edge for every transition from i to j in S that has nonzero probability) is of prime importance. The most interesting case is that of the existence of just one irreducible strongly connected component. In that case, V_n either tends to a stationary limit random variable or exhibits a periodic behavior if the period of the irreducible set is more than one.

We exclude chains throughout by requiring that $p_1 \neq 1$ (or, equivalently, $\mathbf{Var}(\xi) > 0$).

2. Recursive functions on random Galton–Watson trees

As a warm-up, we need to study the behavior of the value of the root of T, an unconditional critical Galton–Watson tree. This case has been treated thoroughly by Aldous and Bandyopadhyay [2]; we will come back to their contribution shortly. Since $|T| < \infty$ with probability one, the root's value, W, is a properly defined random variable. What matters is its support set, that is, the set of all possible values W can take. This support set includes the support set of the leaf values.

Note that the support set of V_n is a subset of the support set of W. As we see later, it can be a proper subset.

Since there is no use for values of S that are not in the support set of W, without loss of generality we define S as the support set of W.

We are not concerned with the precise derivation of the law of W. It suffices to say that it is *a solution* of the distributional identity

$$W \stackrel{\mathcal{L}}{=} f_\xi(W_1, \ldots, W_\xi, U),$$

where $W, W_1, W_2, \ldots$ are i.i.d. , and U, ξ and $W_1, W_2, \ldots$ are independent (indeed, without any additional condition, this equation is may admit more than one solutions). Worked out examples follow later.

Remark. In their paper [2], Aldous and Bandyopadhyay investigated this very fixed point equation, and it is in this context that the question of the representation of the solution as an *unconditioned* Galton–Watson tree arose: if one one expands the distributional fixed-point equation into a tree, the tree obtained is a Galton–Watson tree and the fixed-point can be represented by such a tree. Now, one of the main questions they address is the following: when is the value at the root measurable with respect to the sigma-algebra generated by the random variables in the tree ? When this is case, the system is called endogenous. This question of endogeny is only interesting when the tree is infinite, and in the present case of a critical Galton–Watson, the answer is trivial. However, we shall see soon that some of the conditions they had for endogeny are intimately related to the condition for convergence in the context of Galton–Watson trees conditioned on being infinite.

3. Coalescent Markov chains

We have to deal with an explicit Markov chain governed by

$$X_t = f(X_{t-1}, U_t),$$

where the U_t's are independent random elements, and f is a function that maps to the finite state space S. We write $x \xrightarrow{u_1,\ldots,u_k} y$ when $x_i = f(x_{i-1}, u_i)$ with $x_0 = x$ and $x_k = y$.

We call this Markov chain coalescent if it satisfies two conditions:

(i) There is only one irreducible component C in this Markov chain, and it is aperiodic.
(ii) The double Markov chain

$$(X_t, Y_t) = (f(X_{t-1}, U_t), f(Y_{t-1}, U_t))$$

has one irreducible component. (Note that the *same* U_t is used in both maps in this transition.)

Part (i) implies that regardless of the starting value X_0, X_t tends in law to the unique stationary distribution with support on the irreducible component. Part (ii) insures that for any initial pair (X_0, Y_0), we eventually have, with probability tending to one, that $X_t = Y_t$, i.e., we have coalescence if we use the same random elements to define the maps.

The group of automorphisms of C that preserve f is given by

$$\mathrm{Aut}(f) = \big\{\varphi : C \to C \text{ such that } \varphi f(x,u) = f(\varphi x, u)\big\}\,.$$

Remark. Condition (ii) is a version of what Aldous and Bandyopadhyay call bivariate uniqueness; see Section 2.4 there.

Proposition 1. *Condition (ii) holds if and only if* $\mathrm{Aut}(f)$ *is trivial.*

Proof. Suppose there is more than one irreducible component for the double chain. Since the C is irreducible for the single chain and the diagonal $\Delta = \{(x,x) : x \in C\}$ is invariant under the double chain, there must be an irreducible component outside the diagonal. Let $\{(x_1,y_1),\dots,(x_k,y_k)\}$ be such component and define the map $\varphi x_i = y_i$ for $i = 1,\dots,k$ and $\varphi x = x$ for $x \neq x_i$. Since $x_i \neq y_i$ then $\varphi \neq \mathrm{id}$. For all $i = 1,\dots,k$ and u there exists $j \in \{1,\dots,k\}$ such that $f(x_i,u) = x_j$ and $f(y_i,u) = y_j$. Therefore,

$$\varphi f(x_i,u) = \varphi x_j = y_j = f(y_i,u) = f(\varphi x_i, u).$$

Thus, $\varphi \in \mathrm{Aut}(f)$.

Now assume there exists $\varphi \in \mathrm{Aut}(f)$, $\varphi \neq \mathrm{id}$. Pick $x \in C$ such that $\varphi x \neq x$. We claim that the irreducible component containing $(x,\varphi x)$ is outside the diagonal. First note that if $f(x,u) = y$ then $f(\varphi x,u) = \varphi f(x,u) = \varphi y$. Assume that $(x,\varphi x) \xrightarrow{u_1,\dots,u_k} (z,z)$. Then, by iteration we have $\varphi z = z$. Since C is irreducible there exists $z \xrightarrow{v_1,\dots,v_m} x$. This implies $\varphi z \xrightarrow{v_1,\dots,v_m} \varphi x$. Since $\varphi z = z$ we have $\varphi x = x$ which is a contradiction. ☐

It is easy to construct Markov chains that satisfy (i) but not (ii), so both conditions are essential in what follows. However, as we will see below, we need to study X_0 when the chain is run from $-\infty$ to 0.

4. Kesten's tree

It is helpful to recall convergence of T_n under a finite variance condition to Kesten's infinite tree T^∞ (Kesten, 1986; see also Lyons and Peres, 2016). Let us first recall the definition of T^∞. In every generation, starting with the 0-th generation that contains the root, one node is marked. These marked nodes form an infinite path called the spine. The number of children of the node v_i on the spine in generation i is denoted by ζ_i. The sequence $(\zeta_0,\zeta_1,\dots)$ is i.i.d. with common distribution ζ having the size-biased law:

$$\mathbf{P}(\zeta = i) = ip_i = i\mathbf{P}(\xi = i)\,, \quad i \geq 1.$$

Observe that $\mathbf{E}(\zeta) = 1 + \sigma^2$. Furthermore, of the ζ_i children of v_i, we select a uniform random node to mark as v_{i+1}. The unmarked children of v_i are all roots of independent unconditional Galton–Watson trees distributed as T.

Convergence of T_n to T^∞ takes place in the following sense. Let (T_n,k) denote the truncation of T_n to generations $0,1,\dots,k$. Let t_k denote an arbitrary finite ordered tree whose last generation is at most k. Then for all k and t_k,

$$\lim_{n\to\infty} \mathbf{P}((T_n,k) = t_k) = \mathbf{P}((T^\infty,k) = t_k)\,.$$

The total variation distance between (T_n, k) and (T^∞, k) is given by

$$\frac{1}{2}\sum_{t_k} \left|\mathbf{P}((T_n, k) = t_k) - \mathbf{P}((T^\infty, k) = t_k)\right| .$$

It is easy to see that this tends to zero as well.

Let us first analyze the root value of T^∞. It is not at all clear that it is even properly defined since T^∞ has an infinite path. However, the root value is with probability one properly defined under a Markovian condition. To set this up, we consider a Markov chain on S that runs from $-\infty$ up the spine to time 0 (the root), where "time" refers to minus the generation number in the Galton–Watson tree. Let us call X_{-t} the value of node v_t on the spine. Furthermore, we have

$$X_{-t} = f(X_{-t-1}, U_{-t}),$$

where U_{-t} gathers all random elements necessary to compute the value of v_t from that of v_{t+1}, i.e., ζ_t (the number of children), M (the index of the marked child), the random element U, and $W_1, W_2, \ldots$ (the values of the non-marked children, which are i.i.d. and distributed as the value of the root of an unconditional Galton–Watson tree T). This is called the *spine's Markov chain.* The Markov chain of condition (ii),

$$(X_{-t}, Y_{-t}) = (f(X_{-t-1}, U_{-t}) f(Y_{-t-1}, U_{-t}))$$

is called the *spine's double Markov chain.*

Theorem 1 (LIMIT FOR KESTEN'S TREE.). *Assume that the spine's Markov chain is coalescent. Then, the value of the root of T^∞ is with probability one properly defined. Furthermore, it is exactly distributed as the stationary distribution of the spine's Markov chain. In addition, all values on the spine have the same distribution.*

Proof. The proof follows immediately from the coalescent condition along the lines of the proof of Propp and Wilson's theorem (1996) on coupling from the past for explicit Markov chains. See also [2], who have a genuine tree version; here it suffices to follow the infinite spine, so the classical Markov chain setting suffices. □

We use the notation W^∞ for a random variable that is distributed as the stationary distribution of the spine's random chain.

5. Simulating the root value in tree-based Markov chains.

Theorem 1 has an important algorithmic by-product. Assume that we wish to generate on a computer a random variable that is distributed as W^∞. As a first step, we can write a simple procedure that generates an unconditional Galton–Watson tree T, associates with all nodes the uniform elements, and computes the root value, W. The time taken by this method is proportional to $|T|$, which is almost surely finite. In some cases, one can generate W more efficiently if one knows the distribution on S that solves the distributional identity

$$W \overset{\mathcal{L}}{=} f(W, U) \overset{\text{def}}{=} f_\xi(W_1, \ldots, W_\xi, U),$$

where $W_1, \ldots, W_\xi$ are i.i.d. and distributed as W, and U is the random element. To simulate the root value of Kesten's tree under the condition of Theorem 1, we proceed by generating T_∞ iteratively along the spine. As we process v_i, the node on the spine's level i, we generate its random element (U_i), its number of children (ζ_i), its marked child's index (M_i, uniformly distributed between 1 and ζ_i), and the values $W_{j,i}$ for $1 \le j \le \zeta_i, j \ne M_i$ (which are i.i.d. and distributed as W). As we also have these values for all the ancestors of v_i, we can check the root's value given that the marked node takes all possible values in $S = \{1, \ldots, k\}$. If the root's value is unique, then coalescence has taken place, and thus, the root's value is precisely distributed as W^∞. Note that all the random elements generated for each node stay with the node forever. Because our Markov chain is coalescent, this procedure halts with probability one. This is, in fact, a tree-based version of coupling from the past (Propp and Wilson, 1996; Fill, 1998).

6. The main theorem.

We are now ready for the main theorem.

Theorem 2 (LIMIT FOR T_n)**.** *Assume that the spine's Markov chain is coalescent. Then, the value of the root of T_n tends in distribution to W^∞ as $n \to \infty$.*

Proof. We show that for given $\epsilon > 0$, the total variation distance between W^∞ and the value of the root of T_n is less than ϵ. First, we invoke Kesten's theorem: for any fixed k, there exists an n_k such that for all $n \ge n_k$ the total variation distance between (T_∞, k) and (T_n, k) is less than $\epsilon/2$. By Doeblin's coupling theorem (Doeblin, 1937), we can find coupled trees T_n and T_∞ for which

$$\mathbf{P}((T_n, k) \ne (T_\infty, k)) \le \frac{\epsilon}{2}$$

for such n. Let $A_{n,k}$ be the bad event, $(T_n, k) \ne (T_\infty, k)$. Furthermore, on the complement $A_{n,k}^c$, we populate all nodes in (T_∞, k) with the missing random values, i.e., the U's associated with the nodes. Nodes in (T_n, k) receive the same random values values as their counterparts in (T_∞, k). Those that live at or past level k are given independent values.

While T_∞ can be thought of as constructed using a spine $v_0, v_1, \ldots$, this is by no means necessary. If T_∞ is shown, following Kesten (1986), we can just select a uniform random node at level k, and call it v_k. Define $\ell = \lfloor k^{1/3} \rfloor$. Let H be the maximal height of any subtree rooted at any non-marked child of $v_0, \ldots, v_\ell$. Let ζ_i be the number of children of v_i. Then,

$$\begin{aligned}
\mathbf{P}(H \ge k - \ell) &\le \mathbf{P}\left(\sum_{i=0}^{\ell}(\zeta_i - 1) \ge \ell^2\right) + \ell^2 \mathbf{P}(\text{height}(T) \ge k - \ell) \\
&\le \mathbf{P}\left(\sum_{i=0}^{\ell}(\zeta_i - 1) \ge \ell^2\right) + \ell^2 \times \frac{2 + o(1)}{\sigma^2(k - \ell)}
\end{aligned}$$

where in the last step, we used Kolomogorov's estimate (Kolmogorov, 1938; Kesten, Ney and Spitzer, 1966)[1]. By the strong law of large numbers, and since $\ell \sim k^{1/3}$, we see that the limit of the upper bound is zero as $k \to \infty$.

---

[1]In the case that $\sigma^2 = \infty$, the second term should be replaced by $o(1/(k - \ell))$ (see Kesten, Ney and Spitzer, 1966 and Seneta 1969)

Consider the values of the nodes $v_0, \ldots, v_\ell$ for both trees, T_n and T_∞, provided that $A_{n,k}$ holds. Call these $W_{n,0}, \ldots, W_{n,\ell}$ and $W_0, \ldots, W_\ell$, respectively. We observe that if $H < k-\ell$, then $W_{n,0} = W_0$ if $W_{n,\ell} = W_\ell$. If $W_{n,\ell} \neq W_\ell$, then the root values are nevertheless identical if the spine Markov chain, started at level ℓ coalesces before level 0. By our condition, this happens with probability $1 - o(1)$ as $k \to \infty$. Thus, the probability that the root values of T_n and T_∞ are different is less than

$$\mathbf{P}\left(A_{n,k}^c\right) + \mathbf{P}(H \geq k-\ell) + \mathbf{P}\left(A_{n,k}, H < k-\ell, W_{n,\ell} \neq W_\ell, W_{n,0} \neq W_0\right).$$

We first choose k large enough to make each of the last two terms less than $\epsilon/3$. Having fixed k, the first term is smaller than $\epsilon/3$ for all n large enough. Since W_0 has the sought limit distribution, we see hat the total variation distance between $W_{n,0}$ and W_0 is not more than $\mathbf{P}(W_{n,0} \neq W_0) < \epsilon$. □ □

7. Applications

7.1. Negative example 1: The counting function. When

$$f_\ell(w_1, \ldots, w_\ell, \cdot) \equiv 1 + \sum_{i=1}^{\ell} w_i,$$

then the root value of T_n is $|T_n| = n$. The "mod k" version of this function can be considered to force a finite state space: When

$$f_\ell(w_1, \ldots, w_\ell, \cdot) \equiv 1 + \sum_{i=1}^{\ell} w_i \mod k,$$

then the root value of T_n is $n \bmod k$. The spine's double Markov chain is not coalescent: when it is started with values $(i,j) \in \{0,1,\ldots,k-1\}^2$, then all its future values are of the form $(i+\lambda \bmod k, j+\lambda \bmod k)$, so that there are indeed at least k irreducible components in the chain.

7.2. Negative example 2: The leaf counter function. When $f_\ell \equiv \ell$ for $\ell > 1$ and

$$f_\ell(w_1, \ldots, w_\ell, \cdot) \equiv \min\left(1, \sum_{i=1}^{\ell} w_i\right),$$

then the root value of T_n counts L_n, the number of leaves in the tree. Here, the spine's double Markov chain is not coalescent because it has at least k irreducible components, just as in the first example. Even though Theorem 2 does not apply, we know from elsewhere (e.g., Aldous, 1991) that $L_n/n \to p_0$ in probability. What we are saying here is that the much more refined result about the asymptotic limit law of $L_n \bmod k$ for fixed k cannot be obtained from Theorem 2. In particular, when $p_0 = p_2 = 1/2$ (a Catalan tree), T_n is not defined unless n is even. In that case, $L_n = (n+1)/2$, and thus, $L_n \bmod k = (n+1)/2 \mod k$, which cycles through the values of $S = \{0,1,\ldots,k-1\}$.

7.3. **Example 3: Length of a random path.** A random path in a tree is defined by starting at the root and going to a random child until a leaf is reached. The (edge) length of a random path in T_n is called L_n. One can once again consider all computations $\bmod k$, for some arbitrary natural number $k \geq 2$, but we do not write this explicitly. The recursive function can be viewed as follows:

$$f_\ell(w_1, \ldots, w_\ell, u) = \begin{cases} 1 + w_{1+\lfloor u\ell \rfloor} & \text{if } \ell > 0, \\ 0 & \text{if } \ell = 0. \end{cases}$$

Here u is a uniform $[0, 1]$ random variable. If $f(\cdot, u)$ is the Kesten tree version of this, then there is coalescence in one step in the Markov chain if the number of children (recall that it is denoted by ζ on the spine) is more than one, and $\lfloor u\ell \rfloor$ (the child chosen for the random path) is not equal to the marked node. The probability of this is

$$\mathbf{E}\left((1 - 1/\zeta)\right) = 1 - \sum_{i=1}^{\infty} p_i = p_0.$$

The probability of no coalescence in t steps is smaller than

$$(1 - p_0)^t,$$

and thus tends to zero. Thus, Theorem 2 applies to the length of a random path $\bmod k$. Since the expected length of a random path in an unconditional Galton–Watson tree is $1/p_0$ and in a Kesten tree is $2/p_0$, we see that the $\bmod k$ can safely be omitted[2]. The length of a random path in T_n tends in distribution to the root value of the Kesten tree.

It is easy to see that for an unconditional Galton–Watson tree T, the random path length (W) is geometric with parameter p_0, i.e.,

$$\mathbf{P}(W = i) = p_0(1 - p_0)^i, i \geq 0.$$

Also, in Kesten's tree, the number of edges traversed on the spine is geometric with parameter

$$\mathbf{E}\left((1 - 1/\zeta)\right) = 1 - \sum_{i=1}^{\infty} p_i = p_0.$$

Thus, $L_n \xrightarrow{\mathcal{L}} W + W'$, where W, W' are independent geometric(p_0) random variables.

7.4. **Example 4: Existence of a transversal in a pruned tree.** A given tree is pruned by marking, independently and with probability p each node in the tree. This marking corresponds to a defective node. A transversal of a tree is a collection of nodes such that each path from root to leaf must encounter at least one node of the tranversal. We say that a transversal exists in a pruned tree if all nodes in it were marked. This has been used as a model of breaking up terrorist cells (see Chvatal et al, 2013). A node that is marked has the value one. A non-marked node has value one if its subtree contains a transversal, i.e., if all the subtrees corresponding

[2]What we mean here is that, since the sequence $(L_n)_{n\geq 1}$ is tight, the convergence of $\mathbf{P}(L_n \bmod k = i)$, for arbitrary k imply the convergence of $\mathbf{P}(L_n = i)$.

to its children contain transversals. The basic recursion for a node with child values $w_1, \ldots, w_\ell$ and uniform element U (which is used for marking) is

$$w = f_\ell(w_1, \ldots, w_\ell, U) = \begin{cases} 1 & \text{if } U < p, \\ \prod_{i=1}^{\ell} w_i & \text{if } U > p,\ \ell > 0, \\ 0 & \text{if } U > p \text{ and } \ell = 0. \end{cases}$$

If coalescence does not occur in one step, then we must have $U > p$. Therefore, the probability of no coalescence in t steps is not more than $(1-p)^t$, and we have indeed a coalescent Markov chain to which Theorem 2 applies. When the limit law of W^∞ is worked out, i.e., $\mathbf{P}(W^\infty = 1)$, one rediscovers the result of Devroye (2011).

7.5. **Example 5: The random child function.** We define $f_0(U) = U$, thereby attaching an independent random variable, U to each leaf. For internal nodes with ℓ children, we let V be a uniform $[0,1]$ random variable and have the recursion

$$w = f_\ell(w_1, \ldots, w_\ell, V) = w_{1+\lfloor \ell V \rfloor},$$

the value is that of a uniformly at random chosen child. This map percolates one of the leaf values up to the root. In the spine's Markov chain, coalescence occurs in one step if, as in the random path length example, a node does not select its sole marked child. Thus, as in that example, the probability of not having coalesced in t steps is not more than $(1-p_0)^t$, and thus, Theorem 2 also applies to this case. It should be obvious that $W^\infty \stackrel{\mathcal{L}}{=} U$. (In this case, W is not discrete; the results of Theorem 2 still apply because the coalescence actually does not depend on the actual values at the leaves.)

7.6. **Example 6: The minimax function.** This example follows a model studied by Broutin and Mailler (2017). For each node, we flip a Bernoulli(p) coin to determine whether the node is a max-node (with probability p) or a min-node (with probability $1-p$). Max nodes take the maximum of the child values, and min nodes take the minimum. In addition, leaf nodes are given a Bernoulli(q) value. For an unconditional critical Galton–Watson tree, Avis and Devroye (unpublished work, 2017) showed that the root value is Bernoulli(p^*) where p^* is the unique solution of the equation

$$p^* = pp_0 + q(1 - g(1-p^*)) + (1-q)(g(p^*) - p_0),$$

where we recall that $g(s) = \mathbf{E}\left(s^\xi\right)$.

When p and q are both in $(0,1)$, then $p^* \in (0,1)$. For a max (min) node with ζ children, we have coalescence in one step if $\zeta > 1$ and the leftmost non-marked child of the node has the value one (zero). So, the probability of avoiding coalescence in t steps is not more than

$$\left(1 - (1-p_1)(pp^* + (1-p)(1-p^*))\right)^t,$$

and hence we have a coalescent Markov chain when $p, q \in (0,1)$ and $p_1 \neq 1$ (a special case we excluded in the introduction). Note that this result does not require a finite variance for ξ.

If T_n is a critical Galton–Watson tree with $p_1 < 1$, conditioned to be of size n, and if the variance of ξ is finite, then Theorem 2 applies. One can compute the limit law of the Markov chain (see Avis et al, 2017). In particular, the root value is Bernoulli(p_n^*) where

$$\lim_{n\to\infty} p_n^* = \frac{q(1-g'(1-p^*))}{1-qg'(1-p^*)-(1-q)g'(p^*)}.$$

7.7. **Example 7: Random Boolean functions.** This is a "functional version" of the previous example, which also shows that Theorem 2 also applies to objects that are richer than merely integers.

Assume for simplicity that ξ is 0 or 2 with equal probability, so that T_n is binary. For each node, one flips and independent Bernoulli(p) coin to determine wether it is a AND-node (with probability p) or an OR-node (with probability $1-p$). Additionally, the leaves receive one of the k Boolean variables $x_1, x_2, \dots, x_k$ independently and uniformly at random. Here, rather than looking at real or Boolean values, we let S be the set of Boolean functions on the variables $x_1, x_2, \dots, x_k$ (so the value of each node is a Boolean function). Then, the value of an AND-node is the Boolean AND of the values of its children, while an OR node takes the Boolean OR of the values of its children. The value at the root is the random Boolean function of $x_1, x_2, \dots, x_k$ that is computed by this "AND/OR tree".

Note first that, since AND/OR is a complete set of Boolean connectives, every Boolean function of $x_1, x_2, \dots, x_k$ can be computed by some finite binary tree. To see that the spine's Markov chain is coalescent, observe that the chain coalesce in one step if the spine node is a AND node, and the Boolean function computed by the finite tree equals the $\bar{f}$ ("not f"), where f is the function computed by the spine child. Since there are finitely many such functions f, this happens with positive probability, and as a consequence, coalescence does not happen in t steps with probability exponentially small in t. It follows that Theorem 2 applies, which proves that the random Boolean function computated at the root converges in distribution. Note further that, since the Markov chain is irreducible, every Boolean function is charged with positive probability. It thus completes results in Broutin and Mailler (2017).

7.8. **Example 8: Random binary subtree.** One chooses a random binary subtree of T_n, which contains the root as follows. If the root as two children or less, we keep all of them; otherwise, it has at least three children and we select two uniformly at random without replacement. One then continues in this fashion at the selected nodes, therefore constructing a subtree $T^\star$ of T_n whose nodes all have at most two children. If ξ has support contained in $\{0,1,2\}$, the tree $T^\star$ constructed is just T_n, so we suppose that $\mathbf{P}(\xi > 2) > 0$. Then, the size (number of nodes) of the subtree $T^\star$ converges in distribution.

This fits in our framework. Consider first the "mod k" version by setting $f_0(U) = 1$, $f_1(w_1, U) = w_1$, $f_2(w_1, w_2, U) = w_1 + w_2 \bmod k$ and, for $\ell \geq 3$,

$$f_\ell(w_1, w_2, \dots, w_\ell, u) = w_{\sigma(u)} + w_{\tau(u)} \mod k,$$

where $(\sigma(u), \tau(u)) = (i,j)$ if $u \in A_{i,j}$ for some partition $(A_{i,j})_{1\leq i<j\leq \ell}$ of $[0,1]$ into intervals of equal length. Observe that, if $\mathbf{P}(\xi = 1) = 0$, then the size of $T^\star$ is odd with probability one; otherwise it may take any integer value at least three.

The spine Markov chain is coalescent: it coalesces in one step if a node does not select its unique child lying on the spine; this happens with probability $p > 0$, so that there is coalescence after t steps with probability at least $1 - (1 - p)^t$. This implies in particular that $T^\star$ is actually almost surely finite, so that (see Example 3) there is convergence in distribution of the size without the need for the $\bmod k$.

7.9. **Example 9: The majority function.** We consider the much studied majority function model. We associate with the leaves Bernoulli(p) random variables. For fixed $k > 0$, we consider a tree in which all nodes have either 0 or $2k+1$ children. By criticality of the Galton–Watson tree we are studying, this forces $p_{2k+1} = 1/(2k+1)$, $p_0 = 1 - 1/(2k + 1)$ and $p_i = 0$ for $i \notin \{0, 2k + 1\}$. At each internal node with $2k + 1$ children, we take a majority vote among the children. In other words, if $x_1, \ldots, x_{2k+1}$ are the binary child values, then the value at the node is

$$\operatorname{sign}\left(2\sum_{i=1}^{2k+1} x_i - (2k+1)\right).$$

Let us first consider the value W of the root of an unconditional Galton–Watson tree. If W is Bernoulli(p^*), then a simple recursion shows that p^* is the solution of the following recursive equation:

$$p^* = \frac{1}{2k+1}\mathbf{P}(2\text{Binomial}(2k+1, p^*) > 2k+1) + \frac{2k}{2k+1}p.$$

This yields an equation of degree $2k + 1$. The solution p^* increases monotonically from 0 (at $p = 0$) to 1/2 (at $p = 1/2$) and 1 (at $p = 1$).

Let W_n be the value of the root of T_n, a conditional Galton–Watson tree of size n. For $p \in \{0, 1\}$, we have $W_n \in \{0, 1\}$ accordingly. So, we assume $p \in (0, 1)$. For an internal node with $2k + 1$ children, we have coalescence in one step if the $2k$ non-marked children are all one. The probability of this is at least $(p^*)^{2k} > 0$. So, the probability of avoiding coalescence in t steps is not more than

$$\left(1 - (p^*)^{2k}\right)^t,$$

and hence we have a coalescent Markov chain when $p \in (0, 1)$. By Theorem 2, W_n tends to a limit random variable In fact, along the spine, we have a simple Markov chain on $\{0, 1\}$ with transition probabilities $p(0, 1)$ and $p(1, 0)$ explicitly computable:

$$\begin{aligned} p(0,1) &= \mathbf{P}(\text{Binomial}(2k, p^*) > k)\,, \\ p(1,0) &= \mathbf{P}(\text{Binomial}(2k, p^*) < k)\,. \end{aligned}$$

Thus, by well-known results on Markov chains,

$$\lim_{n\to\infty} \mathbf{P}(W_n = 1) = \frac{p(0,1)}{p(0,1) + p(1,0)} = \frac{\mathbf{P}(\text{Binomial}(2k, p^*) > k)}{1 - \mathbf{P}(\text{Binomial}(2k, p^*) = k)}.$$

7.10. **Example 10: The median function.** Assume that ξ is with probability one either 0 or odd, so ζ is odd. The leaves receive uniform values in a finite set S. Internal nodes take the median of the values of their children. It is a good exercise to show that the spine's Markov chain is coalescent, and that Theorem 2 applies.

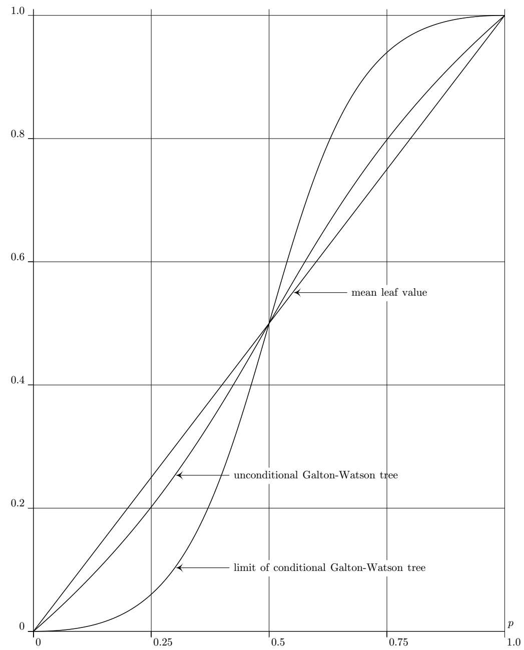

FIGURE 1. Consider the case of a ternary Galton–Watson tree (case $k = 1$ in the text). We show the Bernoulli parameters for the leaf values (the diagonal), the unconditional Galton-Watson tree, and the limit of the conditional Galton-Watson tree.

8. Remarks and open questions

i) We have assumed that the progeny distribution ξ has finite variance for the sake of convenience. The local convergence of T_n towards Kesten's tree actually also holds in the case when $\mathbf{Var}(\xi) = \infty$ (provided that $\mathbf{E}[\xi] = 1$), see for instance, Theorem 7.1 of Janson (2012). In this situation, one still has that the size of an unconditioned tree T satisfies $|T| < \infty$ almost surely, and the proofs can be extended to this case.

ii) We have stated our results for conditioned Galton–Watson trees for the sake of simplicity. One should easily be convinced that the results remain true under the weaker condition that T_n converges locally to an infinite tree such that (1) there is a unique infinite path, and (2) the trees hanging from the spine that are independent and identically distributed.

iii) It would be interesting to investigate the more general setting where the set S may be countably infinite, or an interval of $\mathbb{R}$. It seems believable that, if S is only countably infinite the result might remain true under an additional condition on the positive recurrence of the spine Markov chain. The continuous stage space offers more possibilities for odd behaviours.

References

[1] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, The Annals of Applied Probability **1** (1991), 228–266.

[2] D. Aldous and A. Bandyopadhyay, *A survey of max-type recursive distributional equations*, The Annals of Applied Probability, vol. 15, 1047–1110, 2005.

[3] D. Aldous, and J. Pitman, *Tree-valued Markov chains derived from Galton–Watson processes*, Annals of the Institute Henri Poincaré **34** (1998), 637–686.

[4] K. B. Athreya, and P. E. Ney, *Branching Processes*, Springer Verlag, Berlin (1972).

[5] D. Avis, and L. Devroye, *Minimax Galton–Watson trees*, Unpublished manuscript (2018).

[6] I. J. Bienaymé, *De la loi de multiplication et de la durée des familles*, Soc. Philomath. Paris Extraits **5** (1845), 37–39.

[7] N. Broutin, and C. Mailler, *And/or trees: a local limit point of view*, Random Structures and Algorithms (2017) to appear.

[8] V. Campos, V. Chvatal, L. Devroye, and P. Taslakian, *Transversals in trees*, Journal of Graph Theory **73** (2013), 32–43.

[9] L. Devroye, *On the probability of cutting a Galton–Watson tree*, Electronic Journal of Probability **16** (2011), 2001–2019.

[10] W. Doeblin, *Exposé de la théorie des chaines simples constantes de Markov à un nombre fini d'états*, Revue Mathématique de l'Union Interbalkanique **2** (1937), 77–105.

[11] T. Duquesne, *A limit theorem for the contour process of conditioned Galton–Watson trees*, Annals of Probability **31** (2003), 996–1027.

[12] J. Fill, *An interruptible algorithm for perfect sampling via Markov chains*, The Annals of Applied Probability **8** (1998), 131–162.

[13] F. Galton, and H. W. Watson, *On the probability of extinction of families*, J. Anthropol. Inst. **4** (1874), 138–144.

[14] P. Jagers, *Some Notes on the History of Branching Processes, from my Perspective*, Technical Report, Mathematical Institute, Oberwolfach (2009).

[15] S. Janson, *Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation*, Probability Surveys **9** (2012), 103–252.

[16] S. Janson, *Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees*, Random Structures and Algorithms **48** (2016), 57–101.

[17] D. P. Kennedy, *The Galton–Watson process conditioned on the total progeny*, Journal of Applied Probability **12** (1975), 800–806.

[18] H. Kesten, P. Ney, and F. Spitzer, *The Galton–Watson process with mean one and finite variance*, Teor. Verojatnost. i Primenen. **11** (1975), 579–611.

[19] H. Kesten, *Subdiffusive behavior of a random walk on a random cluster*, Annales de l'Institut Henri Poincaré Probability and Statistics **22** (1986), 425–487.

[20] V. F. Kolchin, *Branching Processes and random trees*, Problems in Cybernetics, Combinatorial Analysis and Graph Theory (in Russian), Nauka, Moscow (1980), 85–97.

[21] V. F. Kolchin, *Random Mappings*, Optimization Software Inc, New York (1986).

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

[23] J.-F. Le Gall, *Random trees and applications*, Probability Surveys **2** (2005), 245–311.

[24] R. Lyons, and Y. Peres, **Probability on Trees and Networks**, Cambridge University Press, New York (2016).

[25] J. F. Marckert, and A. Mokkadem, *The depth first processes of Galton–Watson trees converge to the same Brownian excursion*, Annals of Probability **31** (2003), 1655–1678.

[26] S. P. Meyn, and R. L. Tweedie *Markov Chains and Stochastic Stability*, Springer Verlag, New York (1993).

[27] J. W. Moon, *Counting Labelled Trees*, Canadian Mathematical Congress, Montreal (1970).

[28] J. Pearl, *Asymptotic properties of minimax trees and game-searching procedures*,' Artificial Intelligence **14** (1980), 113–138.

[29] J. Pearl, *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, Addison-Wesley, Reading, MA (1984).

[30] J. G. Propp, and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms **9** (1996), 223–252.

[31] E. Seneta, *Functional Equations and the Galton-Watson Process*, Advances in Applied Probability **1** (1969), 1–42.

Sorbonne Université, Campus Pierre et Marie Curie Case courrier 158, 4, place Jussieu 75252 Paris Cedex 05 France.

Email address: nicolas.broutin@upmc.fr

School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6

Email address: lucdevroye@gmail.com

Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 337 Hanes Hall, Chapel Hill, NC 27599, USA

Email address: fraiman@email.unc.edu