A random recursive tree model with doubling events

Stochastic Processes and their Applications 192 (2026) 104790

Contents lists available at ScienceDirect

Stochastic Processes and their Applications

journal homepage: www.elsevier.com/locate/spa

A random recursive tree model with doubling events

Jakob E. Björnberg [a,*], Cécile Mailler [b]

[a] Department of Mathematics, Chalmers University of Technology and the University of Gothenburg, Sweden
[b] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom

ARTICLE INFO

Keywords:
Random tree
Random recursive tree

ABSTRACT

We introduce a new model of random tree that grows like a random recursive tree, except at some exceptional "doubling events" when the tree is replaced by two copies of itself attached to a new root. We prove asymptotic results for the size of this tree at large times, its degree distribution, and its height profile. We also prove a lower bound for its height. Because of the doubling events that affect the tree globally, the proofs are all much more intricate than in the case of the random recursive tree in which the growing operation is always local.

1. Introduction

1.1. Model and motivation

In this paper we consider a variant of the random recursive tree, with what we call *doubling events*. Recall that the random recursive tree is a process of growing random trees where, at each step, a new leaf is added to a node chosen uniformly at random, starting from a single root node. In our process, we also randomly grow a tree by selecting a uniformly random node at each step, and similarly to the random recursive tree, a leaf is added to that node *unless* it is the root of the tree. If the randomly chosen node is the root, however, a doubling event occurs, which means that we replace the entire subtree of the root with two copies of itself. See Fig. 1 for an illustration; a more formal description of the process, using the Ulam–Harris framework, is given below.

This model was introduced to us by Olivier Bodini, who asked whether we could get some information on the size of the tree at large times (one of our main results is to show that it is linear in the number of steps). Bodini sees this model as a simplification for a more intricate model in which, at every time step, we pick a node v uniformly at random and replace it by a new node whose two subtrees are two copies of the tree rooted at v. In other words, doubling events happen not only at the root, but everywhere in the tree. Bodini would eventually like to understand the size of this doubling tree after n steps: in our last result we prove that, in expectation, this size is superlinear in n (see Proposition 5.1).

We now recall the Ulam–Harris notation for trees, and define our process using this framework. A tree τ is a set of finite words using the alphabet $\{1, 2, 3, \dots\}$ such that, for all $w \in \tau$, all prefixes of w are also in τ. That is to say, if $w = w_1 w_2 \dots w_m \in \tau$, with each $w_i \in \{1, 2, 3, \dots\}$, then $w_1 w_2 \dots w_k \in \tau$ for all $k \in \{0, 1, \dots, m\}$. Each element $w \in \tau$ is called a *node* or *vertex* of τ, and the empty word $\varnothing$ is called the *root*. The number of nodes in a tree τ is denoted $|\tau|$, while if $w_1 w_2 \dots w_m \in \tau$ is a node then $|w| := m$ denotes its length as a word, and is also called its *height*. One can see a tree as a genealogical structure: the prefixes of a word are its ancestors, the longest of its prefixes is its parent, the other children of its parents are its siblings, etc.

We now formally define the random recursive tree with doubling events. We define the sequence of random trees $(\tau_n)_{n\geq 1}$ by setting $\tau_0 = \{\varnothing\}$ and for all $n \geq 0$, given τ_n,

* Corresponding author.
E-mail addresses: jakob.bjornberg@gu.se (J.E. Björnberg), c.mailler@bath.ac.uk (C. Mailler).

https://doi.org/10.1016/j.spa.2025.104790
Received 14 February 2025; Received in revised form 11 August 2025; Accepted 8 October 2025
Available online 21 October 2025
0304-4149/

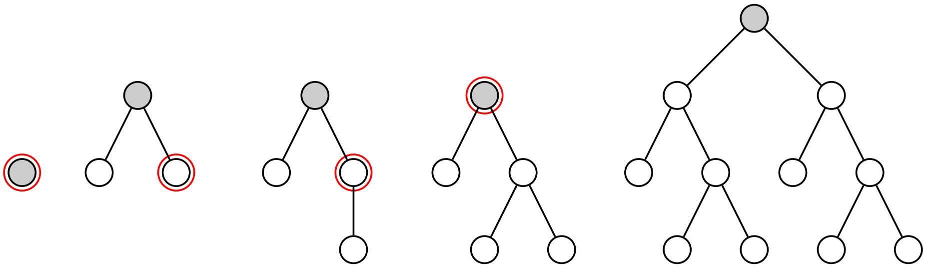

Fig. 1. Steps in the construction of a random doubling tree. The root is drawn grey and, at each step, the randomly selected node is circled in red. In the first and fourth steps, the selected node is the root and a doubling event occurs. At the other steps, a non-root node is selected and a leaf is added to that node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- we pick a node (i.e. word) v_n uniformly at random among the nodes of τ_n,
- if $v_n = \varnothing$, then we set $\tau_{n+1} = \{\varnothing\} \cup \{1w : w \in \tau_n\} \cup \{2w : w \in \tau_n\}$,
- if $v_n \neq \varnothing$, then we set $\tau_{n+1} = \tau_n \cup \{v_n j\}$, where $j = \min\{i \geq 1 : v_n i \notin \tau_n\}$.

In words, this means that each time the randomly chosen node v_n is the root $\varnothing$, then τ_{n+1} is the tree whose root has two children at which are attached two copies of τ_n, while if v_n is not the root then τ_{n+1} equals τ_n with one child added to v_n.

1.2. Main results

Our first result gives an estimate for the size of the tree as time goes to infinity: for all $n \geq 0$, we let B_n be the number of non-root nodes in the tree ($B_0 = 0$) at time n. Note that $|\tau_n| = B_n + 1$, for all $n \geq 0$. Parts (b), (c) and (d) in the following result are simple consequences of part (a); we only state them for clarity.

Proposition 1.1 (*Asymptotic Size*).

(a) For all $k \geq 1$,

$$\mathbb{E}[(B_n/n)^k] \to m_k := \prod_{i=1}^{k}\left(1 - \frac{1}{i} + \frac{2^i}{i}\right) = \frac{2^{k(k-1)/2}}{k!}\prod_{i=1}^{k}\left(1 + \frac{i-1}{2^i}\right). \tag{1.1}$$

(b) The sequence $(B_n/n)_{n\geq 1}$ is tight.
(c) For any sequence $(\omega(n))_{n\geq 0}$ such that $n = o(\omega(n))$, $B_n/\omega(n) \to 0$ in probability as $n \uparrow \infty$.
(d) For any $\eta > 0$, $B_n/n^{1+\eta} \to 0$ almost surely as $n \uparrow \infty$.

Remark 1.2. This result comes close to, but does not quite reach, establishing weak convergence of the sequence $(B_n/n)_{n\geq 1}$. Indeed, if we knew that there was a *unique* probability measure μ on $[0, \infty)$ whose moments are the sequence $(m_k)_{k\geq 1}$, then weak convergence would follow from standard compactness arguments. The most famous condition for uniqueness is *Carleman's criterion*, which states that μ is unique provided $\sum_{k\geq 1} m_k^{-1/2k} = \infty$. In our case, however, this condition is not satisfied.

Our second main result says that the degree distribution is the same as for random recursive trees (without doubling):

Theorem 1.3 (*Degree Distribution*). *For all $i, n \geq 0$, let $U_i(n)$ be the number of nodes in τ_n that have exactly i children. Almost surely as $n \uparrow \infty$,*

$$\frac{U_i(n)}{|\tau_n|} \to \frac{1}{2^{i+1}}.$$

The equivalent result for the random recursive tree is due to Mahmoud and Smythe [1] (see also [2]), who also prove that the fluctuations are Gaussian. The fact that the asymptotic degree distribution is the same as in the random recursive trees can be expected from the observation that, except at doubling times, the tree does grow like a random recursive tree, while at doubling times the degree distribution stays roughly unchanged. In the case of the random recursive tree, one can use standard results for Pólya urns since, for all $m \geq 1$, the vector $(U_1(n), \dots, U_{m-1}(n), \sum_{i\geq m} U_i(n))$ is a Pólya urn. In our case, because of the doubling steps, we no longer have a Pólya urn. Instead we use stochastic approximation methods, which are also classical in the context of urns; see [3] for a survey on stochastic approximation, and [4,5] for books on the topic.

Our third main result concerns what is called the (height) profile, i.e. the joint distribution of the heights of uniformly random nodes:

Theorem 1.4 (*Height Profile*). *For all $k, n \geq 1$, given τ_n, let $u_n^{(1)}, \dots, u_n^{(k)}$ be k nodes taken uniformly, independently at random in τ_n. Then, in distribution as $n \uparrow \infty$,*

$$\left(\frac{|u_n^{(1)}| - \frac{2\log n}{1+\log 2}}{\sqrt{\frac{\log n}{1+\log 2}}}, \dots, \frac{|u_n^{(k)}| - \frac{2\log n}{1+\log 2}}{\sqrt{\frac{\log n}{1+\log 2}}} \right) \Rightarrow (V + W_1, \dots, V + W_k),$$

where V is an almost surely finite random variable, and $W_1, \dots, W_k$ are i.i.d. standard Gaussian, independent of V.

It is interesting to compare this result to its equivalent for the random recursive tree: in the case of the random recursive tree, it is known that

$$\left(\frac{|u_n^{(1)}| - \log n}{\sqrt{\log n}}, \dots, \frac{|u_n^{(k)}| - \log n}{\sqrt{\log n}} \right) \Rightarrow (W_1, \dots, W_k).$$

(See Devroye [6] and Dobrow [7] for convergence of the marginals, and [8] for the joint convergence.) Perhaps as expected, the height of a typical node in the doubling tree is larger than in the random recursive tree ($\frac{2}{1+\log 2} \log n > \log n$). Interestingly, the doubling events add some dependencies between the height of i.i.d. nodes (these dependencies are expressed in the random variable V in the limit).

Note that the height profile of random trees is the object of interest of a large amount of literature: see, e.g., Drmota and Gittenberger [9] for the Catalan tree, Chauvin, Drmota and Jabbour-Hattab [10] and Chauvin, Klein, Marckert and Rouault [11] for the binary search tree, Schopp [12] for the m-ary increasing tree, Katona [13] and Sulzbach [14] for the preferential attachment tree, and the very recent universal result of Kabluchko, Marynych, and Sulzbach [15]. All of these papers use a martingale method that dates back to Biggins [16] in the context of branching random walks; as far as we know, this method does not apply to our setting because the doubling events remove the branching property that is crucial to this approach. Also, our result is, as far as we know, the only one to show some dependence between the marginals in the limit: we will see in the proof that the dependent term V does come directly from the doubling events, which dramatically impact the shape, and thus the height profile, of the whole tree.

To supplement our result on the height profile, we also prove the following lower bound on the height H_n of the tree τ_n itself, i.e. the maximal height of a node:

Proposition 1.5 (*Lower Bound on the Height*). *Let H_n denote the height of τ_n. Almost surely as $n \to +\infty$,*

$$H_n \geq \frac{1+\mathrm{e}}{1+\log 2} \cdot \log n + o(\log n).$$

Note that this lower bound is strictly larger than $\frac{2}{1+\log 2} \log n$, the order of the height of a typical node as given in Theorem 1.4, which is as expected. Again, it is interesting to compare this result to the equivalent in the case of the random recursive tree, which is due to Pittel [17]: in the case of the random recursive tree, $H_n / \log n \to \mathrm{e}$ almost surely as $n \uparrow \infty$. Because $(1+\mathrm{e})/(1+\log 2) < \mathrm{e}$, our lower bound does not allow for any definite comparison between the height of the doubling tree and that of the random recursive tree. We leave this as an open problem.

The rest of the paper is organised as follows: in Section 2, we prove Proposition 1.1 as well as some asymptotic results on the times at which doubling events happen, which are used in the rest of the paper. We prove convergence of the degree distribution (i.e. Theorem 1.3) in Section 3, convergence of the height profile (Theorem 1.4) and the lower bound on the height of the tree in Section 4. Finally, Section 5, we look at the original model of Bodini in which doubling events happen at all nodes and not only at the root, and prove that, in expectation, the size of the tree is superlinear.

2. Asymptotic analysis of the number of nodes and the doubling times

In this section, we prove Proposition 1.1, and state and prove a number of preliminary results which will subsequently be used in the proofs of our other main results.

2.1. Asymptotics of the number of nodes

The aim of this section is to prove Proposition 1.1. We start with the following lemma:

Lemma 2.1. *Almost surely, $\sum_{n\geq 0} \frac{1}{B_n} = \infty$.*

Proof. This follows from Lévy's extension of the Borel–Cantelli lemma (see, e.g. [18, 12.15]). Let $\mathcal{F}_n$ denote the σ-algebra generated by $\tau_0, \dots, \tau_n$ and let $\mathcal{D}_n$ be the event that $v_n = \varnothing$, i.e. at time n we pick the root of the tree. By definition of the model, for all $n \geq 0$,

$$\mathbb{P}(\mathcal{D}_n | \mathcal{F}_n) = \frac{1}{|\tau_n|} = \frac{1}{B_n + 1}.$$

On the event that $\sum_{n\geq 0} \frac{1}{B_n} < \infty$, we have that, almost surely,

$$\sum_{n\geq 0} \mathbb{P}(\mathcal{D}_n | \mathcal{F}_n) = \sum_{n\geq 0} \frac{1}{B_n + 1} \leq \sum_{n\geq 0} \frac{1}{B_n} < \infty,$$

which implies that, almost surely, there exists n_0 such that, for all $n \geq n_0$, $\mathcal{D}_n$ does not occur. Then

$$\sum_{n\geq 0} \mathbb{P}(\mathcal{D}_n|\mathcal{F}_n) \geq \sum_{n\geq n_0} \frac{1}{B_n+1} = \sum_{n\geq n_0} \frac{1}{B_{n_0}+n-n_0} = \infty.$$

This contradiction means that $\mathbb{P}(\sum_{n\geq 0} \frac{1}{B_n} < \infty) = 0$, as required. $\square$

Remark 2.2. In fact, we will prove later that $\frac{1}{\log n} \sum_{i=0}^{n-1} \frac{1}{B_i} \to \frac{1}{1+\log 2}$, in probability as $n \to \infty$ (see (4.10)) and in fact almost surely (see Remark 4.5).

Proof of Proposition 1.1. We start with (1.1); as we explain below, the remaining claims are simple consequences of the convergence in (1.1). We proceed by induction on $k \geq 1$. First note that, if we let $\mathcal{F}_n = \sigma(B_0, \dots, B_n)$, then

$$\mathbb{E}[B_{n+1}|\mathcal{F}_n] = \frac{1}{B_n+1} \cdot (2B_n+2) + \frac{B_n}{B_n+1} \cdot (B_n+1) = B_n+2.$$

Indeed, with probability $1/(B_n+1)$ a doubling occurs, in which case $B_{n+1} = 2B_n + 2$, while with probability $1/(B_n+1)$, we just add one non-root node, i.e. $B_{n+1} = B_n + 1$. Thus $\mathbb{E}[B_{n+1}] = \mathbb{E}[B_n] + 2$, which implies $\mathbb{E}[B_n] = 2n$ for all $n \geq 0$. This concludes the proof of (1.1) in the base case $k = 1$.

For the induction step, we assume that (1.1) holds for all $\ell < k$. Now note that, for all $n \geq 0$,

$$\begin{aligned}\mathbb{E}[B_{n+1}^k \mid \mathcal{F}_n] &= \frac{1}{B_n+1} \cdot (2B_n+2)^k + \frac{B_n}{B_n+1} \cdot (B_n+1)^k = \big(B_n+2^k\big)(B_n+1)^{k-1} \\ &= \big(B_n+2^k\big) \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} B_n^\ell = B_n^k + \sum_{\ell=1}^{k-1} \left[\binom{k-1}{\ell-1} + 2^k \binom{k-1}{\ell}\right] B_n^\ell + 2^k,\end{aligned}$$

which implies

$$\mathbb{E}[B_{n+1}^k] = \mathbb{E}[B_n^k] + \sum_{\ell=1}^{k-1} \left[\binom{k-1}{\ell-1} + 2^k \binom{k-1}{\ell}\right] \mathbb{E}[B_n^\ell] + 2^k.$$

This implies that, for all $n \geq 0$,

$$\mathbb{E}[B_n^k] = 1 + \sum_{\ell=1}^{k-1} \left[\binom{k-1}{\ell-1} + 2^k \binom{k-1}{\ell}\right] \sum_{i=0}^{n} \mathbb{E}[B_i^\ell] + (n+1)2^k.$$

By the induction hypothesis, for all $\ell < k$, $\mathbb{E}[B_n^\ell] = (m_\ell + o(1))n^\ell$ as $n \uparrow \infty$. Thus, for all $\ell < k$, we can write $\mathbb{E}[B_i^\ell] = (m_\ell + \varepsilon(i,\ell))i^\ell$ where $\varepsilon(i,\ell) \to 0$ as $i \to \infty$, and obtain

$$\sum_{i=0}^{n} \mathbb{E}[B_i^\ell] = \sum_{i=0}^{n} i^\ell (m_\ell + \varepsilon(i,\ell)) = m_\ell \frac{n^{\ell+1}(1+o(1))}{\ell+1} + n^{\ell+1} \frac{1}{n} \sum_{i=0}^{n} \big(\tfrac{i}{n}\big)^\ell \varepsilon(i,\ell) = \frac{m_\ell + o(1)}{\ell+1} \cdot n^{\ell+1},$$

since $\frac{1}{n}\sum_{i=0}^{n}\big(\frac{i}{n}\big)^\ell \varepsilon(i,\ell) \to 0$. In total, we thus get that

$$\mathbb{E}[B_n^k] = \left(\frac{(k-1+2^k)m_{k-1}}{k} + o(1)\right) n^k.$$

Using the expression $m_{k-1} = 2\prod_{i=2}^{k-1}\big(1 - \frac{1}{i} + \frac{2^i}{i}\big)$, we get

$$\mathbb{E}[B_n^k] \sim 2n^k \prod_{i=2}^{k} \left(1 - \frac{1}{i} + \frac{2^i}{i}\right),$$

as claimed, which concludes the proof of (1.1).

Tightness follows from the fact that $\mathbb{E}[B_n/n] \to 2$: indeed, by Markov's inequality, for any $K > 0$,

$$\sup_{n\geq 0} \mathbb{P}(B_n/n \geq K) \leq \frac{\sup_{n\geq 0} \mathbb{E}[B_n/n]}{K}.$$

Because $\mathbb{E}[B_n/n] \to 2$, we have that $\sup_{n\geq 0} \mathbb{E}[B_n/n] < \infty$, and thus, for all $\varepsilon > 0$, there exists $K = K(\varepsilon)$ such that $\sup_{n\geq 0} \mathbb{P}(B_n/n \geq K) \leq \varepsilon$, as desired.

Now let $(\omega(n))_{n\geq 0}$ be a sequence such that $n = o(\omega(n))$ as $n \uparrow \infty$; for all $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{B_n}{\omega(n)} > \varepsilon\right) = \mathbb{P}\left(\frac{B_n}{n} > \frac{\varepsilon\omega(n)}{n}\right) \leq \frac{\mathbb{E}[B_n/n]}{\varepsilon\omega(n)/n} \to 0,$$

as $n \uparrow \infty$, as claimed. Finally, fix $\eta > 0$ and choose an integer k such that $k\eta > 1$. For all $\varepsilon > 0$, by Markov's inequality,

$$\mathbb{P}\left(\frac{B_n}{n^{1+\eta}} > \varepsilon\right) \leq \frac{\mathbb{E}[(B_n/n)^k]}{\varepsilon^k n^{k\eta}},$$

which is summable because $(\mathbb{E}[(B_n/n)^k])_{n\geq 0}$ is convergent and thus bounded. By the first Borel–Cantelli lemma, this implies that $B_n/n^{1+\eta}$ converges almost surely to 0, as claimed. $\square$

2.2. Asymptotics of doubling times

We now consider the number of doubling events before time n, i.e. the random variable

$$\kappa(n) := \sum_{i=1}^{n} \mathbf{1}_{\nu_{i-1}=\varnothing}. \tag{2.1}$$

The following results will be useful in the proof of Theorem 1.4.

Proposition 2.3. *As $n \uparrow \infty$,*

$$\frac{\kappa(n) - \frac{\log n}{1+\log 2}}{\frac{\sqrt{\log n}}{(1+\log 2)^{3/2}}} \Rightarrow \mathcal{N}(0,1).$$

Remark 2.4. Using the continuous-time embedding in Section 4.3 (see Remark 4.5) one can also show that

$$\frac{\kappa(n)}{\log n} \to \frac{1}{1+\log 2}, \qquad \text{almost surely as } n \to \infty.$$

We define the sequence $(s_n)_{n\geq 0}$ of doubling times as follows: $s_0 = 0$ and, for all $n \geq 0$,

$$s_{n+1} = \min\{k > s_n : \nu_{k-1} = \varnothing\}.$$

To prove Proposition 2.3, we start by looking at the sequence $(s_n, B_{s_n})_{n\geq 0}$. To simplify notation, we set $C_n = B_{s_n}$ and we set $\Delta s_{n+1} = s_{n+1} - s_n$, for all $n \geq 0$. Note that, by definition, for all $n \geq 0$, for all $x \in \{0, 1, 2, \dots\}$,

$$\mathbb{P}(\Delta s_{n+1} > x | s_n, C_n) = \frac{C_n}{C_n+1} \cdot \frac{C_n+1}{C_n+2} \cdots \frac{C_n+x-1}{C_n+x} = \frac{C_n}{C_n+x}.$$

Equivalently,

$$\mathbb{P}(\Delta s_{n+1} \leq x | s_n, C_n) = \frac{x}{C_n+x}.$$

Thus, if $(U_n)_{n\geq 1}$ is a sequence of i.i.d. uniform random variables on $[0,1]$, then for all $n \geq 0$,

$$\Delta s_{n+1} \stackrel{\mathrm{d}}{=} \left\lceil \frac{U_{n+1} C_n}{1-U_{n+1}} \right\rceil, \tag{2.2}$$

where $\stackrel{\mathrm{d}}{=}$ means equality in distribution. Furthermore, because, at time $s_{n+1}-1$, the number of non-root nodes in the tree is $C_n + \Delta s_{n+1} - 1$, we have

$$C_{n+1} = 2(C_n + \Delta s_{n+1} - 1) + 2 = 2(C_n + \Delta s_{n+1}) \stackrel{\mathrm{d}}{=} 2\left(C_n + \left\lceil \frac{U_{n+1} C_n}{1-U_{n+1}} \right\rceil\right). \tag{2.3}$$

In what follows, we treat the distributional equalities in (2.2) and (2.3) as actual equalities, in other words we replace the random variables s_n and C_n with distributional copies satisfying these equalities.

Lemma 2.5. *In distribution as $n \uparrow \infty$,*

$$\left(\frac{\log C_n - (1+\log 2)n}{\sqrt{n}}, \frac{\log s_n - (1+\log 2)n}{\sqrt{n}}\right) \Rightarrow (N, N),$$

where $N \sim \mathcal{N}(0,1)$ is a standard normal random variable.

Proof. First note that, by (2.3), for all $n \geq 0$,

$$C_{n+1} \geq 2\left(C_n + \frac{U_{n+1} C_n}{1-U_{n+1}}\right) = \frac{2C_n}{1-U_{n+1}}.$$

By induction, we thus get

$$C_n \geq 2^{n-1} C_1 \prod_{i=2}^{n} \frac{1}{1-U_i} = 2^n \prod_{i=2}^{n} \frac{1}{1-U_i}, \tag{2.4}$$

because, by definition, $s_1 = 1$ and $C_1 = B_{s_1} = 2$. On the other hand, (2.3) also implies that, for all $n \geq 0$

$$C_{n+1} \leq 2\left(C_n + \frac{U_{n+1} C_n}{1-U_{n+1}} + 1\right) = 2\left(\frac{C_n}{1-U_{n+1}} + 1\right).$$

By induction, we thus get

$$
\begin{aligned}
C_n &\le 2^{n-1} C_1 \prod_{i=2}^{n} \frac{1}{1-U_i} + \sum_{k=3}^{n+1} 2^{n+2-k} \prod_{i=k}^{n} \frac{1}{1-U_i} = 2^n \prod_{i=2}^{n} \frac{1}{1-U_i} + \sum_{k=3}^{n+1} 2^{n+2-k} \prod_{i=k}^{n} \frac{1}{1-U_i} \\
&\le \left(2^n + \sum_{k=3}^{n+1} 2^{n+2-k} \right) \prod_{i=2}^{n} \frac{1}{1-U_i} = 2^n \left(1 + \sum_{k=1}^{n-1} (1/2)^k \right) \prod_{i=2}^{n} \frac{1}{1-U_i} \le 2^n \left(1 + \sum_{k\ge1} (1/2)^k \right) \prod_{i=2}^{n} \frac{1}{1-U_i} \\
&= 2^{n+1} \prod_{i=2}^{n} \frac{1}{1-U_i}.
\end{aligned} \tag{2.5}
$$

In total, we have thus proved that

$$
2^n \prod_{i=2}^{n} \frac{1}{1-U_i} \le C_n \le 2^{n+1} \prod_{i=2}^{n} \frac{1}{1-U_i},
$$

which implies that, as $n \uparrow \infty$,

$$
\log C_n = n \log 2 + \sum_{i=2}^{n} \log\left(\frac{1}{1-U_i} \right) + \mathcal{O}(1). \tag{2.6}
$$

Next, by (2.2), for all $n \ge 0$,

$$
s_{n+1} \ge s_n + \frac{U_{n+1} C_n}{1-U_{n+1}},
$$

which implies, by induction,

$$
s_n \ge s_1 + \sum_{k=2}^{n} \frac{U_k C_{k-1}}{1-U_k} \ge \frac{U_n C_{n-1}}{1-U_n} \ge 2^{n-1} U_n \prod_{i=2}^{n} \frac{1}{1-U_i},
$$

where we have used (2.4). On the other hand, using (2.2) again we get that, for all $n \ge 0$,

$$
s_{n+1} \le s_n + \frac{U_{n+1} C_n}{1-U_{n+1}} + 1,
$$

which implies, using induction and (2.5), that

$$
s_n \le n + \sum_{k=2}^{n} \frac{U_k C_{k-1}}{1-U_k} \le n + \sum_{k=2}^{n} 2^k U_k \prod_{i=2}^{k} \frac{1}{1-U_i}.
$$

Now, because $U_i \in (0,1)$ almost surely for all $i \ge 1$, we get

$$
s_n \le n + 2^n \left(\prod_{i=2}^{n} \frac{1}{1-U_i} \right) \sum_{k=0}^{n-2} (1/2)^k \le n + 2^n \left(\prod_{i=2}^{n} \frac{1}{1-U_i} \right) \sum_{k\ge0} (1/2)^k = n + 2^{n+1} \prod_{i=2}^{n} \frac{1}{1-U_i}.
$$

In total, we have thus proved that for all $n \ge 1$,

$$
2^{n-1} U_n \prod_{i=2}^{n} \frac{1}{1-U_i} \le s_n \le n + 2^{n+1} \prod_{i=2}^{n} \frac{1}{1-U_i},
$$

which implies that

$$
\log s_n = n \log 2 + \sum_{i=2}^{n} \log\left(\frac{1}{1-U_i} \right) + \mathcal{O}(\log n). \tag{2.7}
$$

Applying the central limit theorem applied to the sequence of i.i.d. random variables $(\log(1/(1-U_i))_{i\ge1})$, which have expectation and variance both equal to 1, the result follows from (2.6) and (2.7). □

Proof of Proposition 2.3. The argument is inspired by standard arguments in renewal theory, with s_k playing the role of the time of the k'th renewal and $\kappa(n)$ the number of renewals up to time n. By definition, for any $k \in \mathbb{N}$, we have that $\kappa(n) \ge k$ if and only if $s_k \le n$. Now, for any $x \in \mathbb{R}$,

$$
\mathbb{P}\left(\frac{\kappa(n) - \frac{\log n}{1+\log 2}}{\frac{\sqrt{\log n}}{(1+\log 2)^{3/2}}} \ge -x \right) = \mathbb{P}(\kappa(n) \ge k_n(x)) = \mathbb{P}(s_{k_n(x)} \le n)
$$

where

$$
k_n(x) = \left\lceil \frac{\log n}{1+\log 2} - \frac{x\sqrt{\log n}}{(1+\log 2)^{3/2}} \right\rceil.
$$

But

$$
\mathbb{P}(s_{k_n(x)} \le n) = \mathbb{P}\left(\frac{\log s_{k_n(x)} - (1+\log 2) k_n(x)}{\sqrt{k_n(x)}} \le \frac{\log n - (1+\log 2) k_n(x)}{\sqrt{k_n(x)}} \right)
$$

and as $n \to \infty$ we have $k_n(x) \to \infty$ and

$$\frac{\log n - (1+\log 2)k_n(x)}{\sqrt{k_n(x)}} \to x.$$

Thus by Lemma 2.5, with Φ the cumulative density function of the standard normal distribution,

$$\mathbb{P}(s_{k_n(x)} \le n) \to \Phi(x) = 1 - \Phi(-x),$$

as required. $\square$

3. The degree distribution: proof of Theorem 1.3

The proof of Theorem 1.3 is based on stochastic approximation, specifically the following result, attributed to Robbins and Siegmund [19]:

Theorem 3.1 (*e.g.* [5, *Theorem 1.3.12*]). *Suppose that $(V(n))_{n\ge 0}$, $(\alpha_n)_{n\ge 0}$, $(\beta_n)_{n\ge 0}$, and $(\gamma_n)_{n\ge 0}$ are four non-negative sequences adapted to a filtration $(\mathcal{F}_n)_{n\ge 0}$ and satisfying, for all $n \ge 0$,*

$$\mathbb{E}[V(n+1)|\mathcal{F}_n] \le (1+\alpha_n)V(n) - \beta_n + \gamma_n.$$

Then, on the event that $\sum_{n\ge 0}\alpha_n < \infty$ and $\sum_{n\ge 0}\gamma_n < \infty$, we have that, almost surely, $(V(n))_{n\ge 0}$ converges to a finite random variable, and $\sum_{n\ge 0}\beta_n < \infty$.

Proof of Theorem 1.3. Recall that $U_i(n)$ denotes the number of nodes in τ_n that have exactly i children. We fix $m \ge 1$ and let, for all $0 \le i \le m$,

$$X_i(n) = \begin{cases} U_i(n) & \text{if } 0 \le i \le m-1 \\ \sum_{j\ge m} U_j(n) & \text{if } i = m, \end{cases}$$

and set

$$\hat{X}_i(n) = \frac{X_i(n)}{B_n + 1}.$$

By definition, for all $0 \le i \le m-1$, $\hat{X}_i(n)$ is the proportion of nodes having i children in τ_n. Let us set $\Delta X_i(n+1) = X_i(n+1) - X_i(n)$ and $\Delta B_{n+1} = B_{n+1} - B_n$. For all $0 \le i \le m$, for all $n \ge 0$,

$$\begin{aligned}\hat{X}_i(n+1) &= \frac{X_i(n) + \Delta X_i(n+1)}{B_{n+1}+1} = \hat{X}_i(n)\cdot\frac{B_n+1}{B_{n+1}+1} + \frac{\Delta X_i(n+1)}{B_{n+1}+1} \\ &= \hat{X}_i(n) + \frac{1}{B_{n+1}+1}\big(\Delta X_i(n+1) - \Delta B_{n+1}\hat{X}_i(n)\big).\end{aligned}$$

Note that, by definition of the model, with probability $1/(B_n+1)$, we pick the root and double the number of nodes with i children (for all $i \ge 1$) and add one node with two children, whilst, with probability $B_n/(B_n+1)$, we pick a non-root node uniformly at random and increase its number of children by one (in this case, with probability $\hat{X}_i(n)$, the number of nodes with i children decreases by one, and with probability $\hat{X}_{i-1}(n)$, it increases by one). Hence, for all $1 \le i \le m-1$,

$$\begin{aligned}\mathbb{E}\big[\Delta X_i(n+1)|\mathcal{F}_n\big] &= \frac{1}{B_n+1}\cdot(X_i(n) + \mathbf{1}_{i=2}) + \frac{B_n}{B_n+1}\cdot(\hat{X}_{i-1}(n) - \hat{X}_i(n)) \\ &= \hat{X}_{i-1}(n) + \frac{\mathbf{1}_{i=2} + \hat{X}_i(n) - \hat{X}_{i-1}(n)}{B_n+1}\end{aligned}$$

Similarly, in the case when the tree does not double, the number of leaves (nodes with 0 children) always increases by one, except if the node we have picked was itself a leaf. Hence

$$\mathbb{E}\big[\Delta X_0(n+1)|\mathcal{F}_n\big] = \frac{1}{B_n+1}\cdot X_0(n) + \frac{B_n}{B_n+1}\cdot(1-\hat{X}_0(n)) = 1 - \frac{1-\hat{X}_0(n)}{B_n+1}.$$

Because $\sum_{i=0}^m \hat{X}_i(n) = 1$, we can write

$$\mathbb{E}\big[\Delta X_0(n+1)|\mathcal{F}_n\big] = \sum_{i=0}^m \hat{X}_i(n) - \frac{1-\hat{X}_0(n)}{B_n+1}.$$

Finally,

$$\mathbb{E}\big[\Delta X_m(n+1)|\mathcal{F}_n\big] = \frac{1}{B_n+1}\cdot X_m(n) + \frac{B_n}{B_n+1}\cdot\hat{X}_{m-1}(n) = \hat{X}_m(n) + \hat{X}_{m-1}(n) - \frac{\hat{X}_{m-1}(n)}{B_n+1}.$$

Note that, also,

$$\mathbb{E}\big[\Delta B_{n+1}|\mathcal{F}_n\big] = \frac{1}{B_n+1}\cdot(B_n+2) + \frac{B_n}{B_n+1} = 2.$$

Introduce, for all $0 \le i \le m$,

$$\Delta M_i(n+1) = \Delta X_i(n+1) - \Delta B_{n+1}\hat{X}_i(n) - \mathbb{E}\big[\Delta X_i(n+1) - \Delta B_{n+1}\hat{X}_i(n)|\mathcal{F}_n\big],$$

and set, for all $x = (x_0, \dots, x_m) \in \mathbb{R}^{m+1}$,

$$F_i(x) = \begin{cases} \sum_{i=1}^m x_i - x_0 & \text{if } i = 0,\\ x_{i-1} - 2x_i & \text{if } 1 \le i \le m-1,\\ x_{m-1} - x_m & \text{if } i = m. \end{cases}$$

Also let

$$\varepsilon_i(n+1) = \frac{1}{B_n+1}\begin{cases} -(1-\hat{X}_0(n)) & \text{if } i = 0,\\ \mathbf{1}_{i=2} + \hat{X}_i(n) - \hat{X}_{i-1}(n) & \text{if } 1 \le i \le m-1,\\ -\hat{X}_{m-1}(n) & \text{if } i = m. \end{cases} \tag{3.1}$$

Using the above, we can write

$$\hat{X}_i(n+1) = \hat{X}_i(n) + \frac{1}{B_{n+1}+1}\big(F_i(\hat{X}(n)) + \Delta M_i(n+1) + \varepsilon_i(n+1)\big), \tag{3.2}$$

We write (3.2) as an identity on vectors:

$$\hat{X}(n+1) = \hat{X}(n) + \frac{1}{B_{n+1}+1}\big(F(\hat{X}(n)) + \Delta M(n+1) + \varepsilon(n+1)\big). \tag{3.3}$$

Now, because B_{n+1} is not $\mathcal{F}_n$-measurable, we re-write this as

$$\begin{aligned}\hat{X}(n+1) = \hat{X}(n) &+ \frac{1}{B_n+1}\big(F(\hat{X}(n)) + \Delta M(n+1) + \varepsilon(n+1)\big)\\ &- \frac{B_{n+1}-B_n}{(B_n+1)(B_{n+1}+1)}\cdot Y(n+1),\end{aligned}$$

where we have set

$$Y(n+1) = F(\hat{X}(n)) + \Delta M(n+1) + \varepsilon(n+1). \tag{3.4}$$

In total, this gives

$$\hat{X}(n+1) = \hat{X}(n) + \frac{1}{B_n+1}\big(F(\hat{X}(n)) + \Delta M(n+1) + \eta(n+1)\big), \tag{3.5}$$

where

$$\eta(n+1) = \varepsilon(n+1) - \frac{\Delta B_{n+1}}{B_{n+1}+1}\cdot Y(n+1). \tag{3.6}$$

This recursion is of the form of a stochastic approximation. However, the step sizes $(1/(B_n+1))_{n\ge 0}$ are random, and we have a random error term $(\eta(n+1))_{n\ge 0}$. Because of these two reasons, we cannot apply a theorem directly from the literature, but need instead to write a specific argument. We now let $v_i = 2^{-i-1}$ for all $0 \le i \le m-1$ and $v_m = 2^{-m}$. One can check that $F(v) = 0$; in fact, for all $x \in \mathbb{R}^d$, $F(x) = Ax$, with

$$A = \begin{pmatrix} -1 & 1 & 1 & \dots & \dots & 1\\ 1 & -2 & 0 & \dots & \dots & 0\\ 0 & 1 & \ddots & \ddots & & \vdots\\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & 1 & -2 & 0\\ 0 & \dots & \dots & 0 & 1 & -1 \end{pmatrix}$$

and one can check that the largest eigenvalue of A is 0, it is a simple eigenvalue with eigenvector v, the unique one with non-negative coefficients and satisfying $\sum_{i=0}^m v_i = 1$. We thus have (write $\|\cdot\|$ for the L^2 norm on $\mathbb{Z}^{m+1}$), for all $n \ge 0$,

$$\begin{aligned}\|\hat{X}(n+1) - v\|^2 = \|\hat{X}(n) - v\|^2 &+ \frac{2}{B_n+1}\langle \hat{X}(n) - v, A\hat{X}(n) + \Delta M(n+1) + \eta(n+1)\rangle\\ &+ \frac{1}{(B_n+1)^2}\|A\hat{X}(n) + \Delta M(n+1) + \eta(n+1)\|^2.\end{aligned} \tag{3.7}$$

We use the triangle inequality and the fact that $(x+y)^2 \le 2x^2 + 2y^2$ for all $x, y \in \mathbb{R}$, to get

$$\begin{aligned}\|A\hat{X}(n) + \Delta M(n+1) + \eta(n+1)\|^2 &\le 2\|A\hat{X}(n)\|^2 + 2\|\Delta M(n+1) + \eta(n+1)\|^2\\ &\le 2\|A(\hat{X}(n) - v)\|^2 + 4\|\Delta M(n+1)\| + 4\|\eta(n+1)\|^2\\ &\le 2|||A|||^2\|\hat{X}(n) - v\|^2 + 4\|\Delta M(n+1)\|^2 + 4\|\eta(n+1)\|^2,\end{aligned} \tag{3.8}$$

where $\|\!|A|\!\|$ is the operator norm of A. Before proceeding, we show that there exists a constant $C > 0$ such that

$$\sup_{n\geq 0} \|\Delta M(n+1)\| \leq C \quad \text{and} \quad \sup_{n\geq 0} \|\eta(n+1)\| \leq C. \tag{3.9}$$

Indeed, first recall that

$$\Delta M(n+1) = \Delta X(n+1) - \Delta B_{n+1}\hat{X}(n) - \mathbb{E}[\Delta X(n+1) - \Delta B_{n+1}\hat{X}(n)|\mathcal{F}_n].$$

On the event that the tree doubles at time $n+1$, we have (with $e_2 = (0,0,1,0,\dots,0)^t$; to be consistent with $X(n) = (X_0(n),\dots,X_m(n))^t$, we let $e_0 = (1,0,0,\dots,0)^t$, $e_1 = (0,1,0,\dots,0)^t$, etc)

$$\begin{aligned}\Delta X(n+1) - \Delta B_{n+1}\hat{X}(n) &= X(n) + 2e_2 - (B_n+2)\hat{X}(n)\\ &= X(n) - (B_n+1)\hat{X}(n) + 2e_2 - \hat{X}(n) = 2e_2 - \hat{X}(n),\end{aligned}$$

which implies $\|\Delta X(n+1) - \Delta B_{n+1}\hat{X}(n)\| \leq 3$ (because $\|\hat{X}(n)\|^2 \leq \sum_{i=0}^m \hat{X}_i(n) = 1$, as $\hat{X}(n)$ has non-negative coefficients that sum to 1). On the event that the tree does not double at time $n+1$,

$$\|\Delta X(n+1) - \Delta B_{n+1}\hat{X}(n)\| = \|\Delta X(n+1) - \hat{X}(n)\| \leq \|\Delta X(n+1)\| + 1,$$

and $\|\Delta X(n+1)\|$ is bounded by the maximum of the norms of the columns of $A+I$, which is a constant, which we let K denote. In total, we thus get that

$$\|\Delta X(n+1) - \Delta B_{n+1}\hat{X}(n)\| \leq K+3,$$

for all $n \geq 0$, which implies that $\sup_{n\geq 0} \|\Delta M(n+1)\| \leq 2(K+3)$. We now prove that $\sup_{n\geq 0} \|\eta(n+1)\| < \infty$ (see (3.6) for the definition of $\eta(n+1)$). First note that, by definition (see (3.1) for the definition of $\varepsilon(n+1)$),

$$\|\varepsilon(n+1)\| \leq \frac{4}{B_{n+1}+1} \leq 4.$$

Thus, by the triangle inequality (see (3.4) for the definition of $Y(n+1)$),

$$\|Y(n+1)\| \leq \|A\hat{X}(n)\| + \|\Delta M(n+1)\| + \|\varepsilon(n+1)\| \leq \|\!|A|\!\| + (K+3) + 4,$$

implying that $\sup_{n\geq 0} \|Y(n+1)\| \leq \|\!|A|\!\| + K + 7$. Thus, for all $n \geq 0$,

$$\|\eta(n+1)\| \leq \|\varepsilon(n+1)\| + \frac{\Delta B_{n+1}}{B_{n+1}+1} \cdot \|Y(n+1)\| \leq \|\varepsilon(n+1)\| + \|Y(n+1)\| \leq \|\!|A|\!\| + K + 11, \tag{3.10}$$

which concludes the proof of (3.9) (we choose $C \geq \|\!|A|\!\| + K + 11$). We now let $V(n) = \|\hat{X}(n) - v\|^2$ for all $n \geq 0$; with this notation, and using the triangle inequality, we get from (3.7) and (3.8) that

$$\begin{aligned}\mathbb{E}[V(n+1)|\mathcal{F}_n] \leq& V(n) + \frac{2}{B_n+1}\langle \hat{X}(n) - v, A\hat{X}(n)\rangle + \frac{2}{B_n+1}\langle \hat{X}(n) - v, \mathbb{E}[\eta(n+1)|\mathcal{F}_n]\rangle\\ &+ \frac{1}{(B_n+1)^2}\big(\|\!|A|\!\|V(n) + 2C\big),\end{aligned} \tag{3.11}$$

where we have chosen C larger than $\sup_{n\geq 0} \Delta M(n+1)$ and $\sup_{n\geq 0} \eta(n+1)$. Now, by the Cauchy–Schwarz and Jensen inequalities,

$$\langle \hat{X}(n) - v, \mathbb{E}[\eta(n+1)|\mathcal{F}_n]\rangle \leq \|\hat{X}(n) - v\|\mathbb{E}[\|\eta(n+1)\||\mathcal{F}_n] \leq 2\mathbb{E}[\|\eta(n+1)\||\mathcal{F}_n], \tag{3.12}$$

because $\|\hat{X}(n) - v\| \leq \|\hat{X}(n)\| + \|v\| \leq 2$. Now, by (3.10), and the fact that $\varepsilon(n+1) \leq 4 \leq C$ and $\|Y(n+1)\| \leq C$ for all $n \geq 0$, we have

$$\mathbb{E}[\|\eta(n+1)\||\mathcal{F}_n] \leq C\mathbb{E}\left[1 + \frac{\Delta B_{n+1}}{B_{n+1}+1}\middle|\mathcal{F}_n\right] = C\left(\frac{1}{B_n+1}\cdot\frac{B_n+2}{2B_n+3} + \frac{B_n}{B_n+1}\cdot\frac{1}{B_n+2}\right),$$

by definition of the model. Thus,

$$\mathbb{E}[\|\eta(n+1)\||\mathcal{F}_n] \leq \frac{C}{B_n+1}\left(\frac{B_n+2}{2B_n+3} + \frac{B_n}{B_n+2}\right) \leq \frac{2C}{B_n+1}.$$

Thus, by (3.12) and (3.11), for all $n \geq 0$,

$$\mathbb{E}[V(n+1)|\mathcal{F}_n] \leq \left(1 + \frac{\|\!|A|\!\|}{(B_n+1)^2}\right)V(n) + \frac{2}{B_n+1}\langle \hat{X}(n) - v, A\hat{X}(n)\rangle + \frac{10C}{(B_n+1)^2}.$$

We want to apply Theorem 3.1 with $\alpha_n = \frac{\|\!|A|\!\|}{(B_n+1)^2}$, $\beta_n = -\frac{2}{B_n+1}\langle \hat{X}(n) - v, A\hat{X}(n)\rangle$, and $\gamma_n = \frac{10C}{(B_n+1)^2}$, so we need to check the conditions on these sequences. Because, by definition, $B_n \geq n$ for all $n \geq 0$, we have that, almost surely, $\sum_{n\geq 0}\alpha_n < \infty$ and $\sum_{n\geq 0}\gamma_n < \infty$. To check that $\beta_n \geq 0$ first note that $\langle \hat{X}(n) - v, A\hat{X}(n)\rangle = \langle \hat{X}(n) - v, A(\hat{X}(n) - v)\rangle$. While the eigenvalues of A are all non-positive, A is not Hermitian and hence not negative semidefinite; however, one may check explicitly that $\langle x, Ax\rangle \leq 0$ for all $x = (x_i)_{i=0}^m$ that satisfy $\sum_{i=0}^m x_i = 0$. Indeed, for such x,

$$\begin{aligned}\langle x, Ax\rangle &= x_0\left(-x_0+\sum_{i=1}^m x_i\right)+\sum_{i=1}^{m-1} x_i(x_{i-1}-2x_i)+x_m(x_{m-1}-x_m)\\ &= -2\left(\sum_{i=1}^m x_i\right)^2+\sum_{i=1}^m x_ix_{i-1}-2\sum_{i=1}^{m-1}x_i^2-x_m^2 \le -2\sum_{i=1}^m x_i^2-2\sum_{i=1}^m\sum_{j=1,j\neq i}^m x_ix_j\\ &= -2\sum_{i=1}^m x_i^2-2\sum_{i=1}^m x_i(-x_0-x_i)=2x_0\sum_{i=1}^m x_i=-2x_0^2\le 0.\end{aligned}$$

Applying this observation to $x=\hat X(n)-v$ gives $\beta_n\ge 0$. Therefore, by Theorem 3.1, almost surely, $W := \lim_{n\uparrow\infty} V(n)$ exists and is finite, and $\sum_{n\ge 0}\beta_n<\infty$. On the event that $W\neq 0$, there exists $\varepsilon>0$ such that, for all n large enough, $V(n)\ge\varepsilon$. Now note that, on the set $\{x\in[0,1]^{m+1} : \sum_{i=0}^m x_i=1\}$ to which $\hat X(n)$ belongs for all $n\ge 0$, $x\mapsto\langle x-v, A(x-v)\rangle$ is continuous, non negative, and its unique zero is v. Thus, on $\{x\in[0,1]^{m+1} : \sum_{i=0}^m x_i=1\}\cap\{\|x-v\|\ge\varepsilon\}$, the maximum of $x\mapsto\langle x-v, A(x-v)\rangle$ is negative; we let $-c$ denote this maximum. We thus get that, for all n large enough, $\beta_n\ge 2c/(B_n+1)$. By Lemma 2.1, this implies that $\sum_{n\ge 0}\beta_n=\infty$, which is an event of probability zero. Thus, $W=0$ almost surely, i.e. $\lim_{n\uparrow\infty}\hat X(n)=v$ almost surely as $n\uparrow\infty$. In other words, for all $0\le i\le m-1$,

$$\frac{U_i(n)}{B_n+1}=\frac{X_i(n)}{B_n+1}\to\frac{1}{2^{i+1}}.$$

Because m can be chosen arbitrarily large, this concludes the proof. □

Remark 3.2. Because of the step-sizes in (3.5) being random, we were unable to prove a central limit theorem for $\hat X(n)$. We leave this as an open problem.

4. The distribution of heights

We now turn to the height profile, Theorem 1.4, as well as the lower bound on the height of the tree, Proposition 1.5. We will give full details for the case $k=1$ of Theorem 1.4 (the height of a typical node) in Section 4.1. In Section 4.2 we describe the necessary modifications for the case $k=2$, and give an outline of the case $k\ge 3$. Proposition 1.5 is proved in Section 4.3

4.1. The height of a typical node

Let us reformulate the case $k=1$ of Theorem 1.4:

Proposition 4.1. *For all $n\ge 0$, let u_n be a uniformly random node in τ_n. There is an a.s. finite random variable Λ such that, as $n\uparrow\infty$,*

$$\frac{|u_n|-\frac{2\log n}{1+\log 2}}{\sqrt{\frac{\log n}{1+\log 2}}}\Rightarrow\Lambda.$$

For the proof of Proposition 4.1, we define a process $(\tilde\tau_n,\tilde u_n)_{n\ge 0}$ such that, for all $n\ge 0$, $(\tau_n,u_n)\overset{\mathrm{d}}{=}(\tilde\tau_n,\tilde u_n)$. The process $(\tilde\tau_n,\tilde u_n)_{n\ge 0}$ will have the properties: (i) at non-doubling times, the height of $\tilde u_n$ is either unchanged or increases by 1, and (ii) at doubling-times, the height of $\tilde u_n$ either increases by 1, or is reset to 0. The process is an adaptation of a standard construction for the random recursive tree (see [8] for a description of it in a more general case). In the latter case there are no doubling-times, and the height of the uniform node is monotonically increasing. In our case we need the possibility to reset, since at doubling steps there is a new node at height 0; however, we prove that there are only finitely many reset-times (almost surely), so they can effectively be ignored. In addition, we control the increase of the height of $\tilde u_n$ due to doublings by using Proposition 2.3.

Let us now define the process $(\tilde\tau_n,\tilde u_n)_{n\ge 0}$. We let $\tilde\tau_0=\{\varnothing\}$ and $\tilde u_0=\varnothing$. Then, for all $n\ge 0$, given $(\tilde\tau_n,\tilde u_n)$, we sample

- $K(n+1)$ a Bernoulli-distributed random variable of parameter $1/(2|\tilde\tau_n|+1)$, and
- $L(n+1)$, a Bernoulli-distributed random variable of parameter $1/(|\tilde\tau_n|+1)$.

Then $(\tilde\tau_{n+1},\tilde u_{n+1})$ is constructed as follows:

(1) We let $\tilde v_n$ be a node taken uniformly at random among the nodes of $\tilde\tau_n$;
(2) If $\tilde v_n=\varnothing$, then we define

$$\tilde\tau_{n+1}=\{\varnothing\}\cup\{1w : w\in\tilde\tau_n\}\cup\{2w : w\in\tilde\tau_n\}.$$

Furthermore,

- if $K(n+1)=1$, then we set $\tilde u_{n+1}=\varnothing$, and
- if $K(n+1)=0$, then we set $\tilde u_{n+1}=1\tilde u_n$ or $\tilde u_{n+1}=2\tilde u_n$ with probability $1/2$ each.

(3) If $\tilde v_n\neq\varnothing$, then

- if $L(n+1)=1$, then we set $\tilde\tau_{n+1}=\tilde\tau_n\cup\{\tilde u_n i\}$ and $\tilde u_{n+1}=\tilde u_n i$, where $i=\min\{j\geq 1:\tilde u_n j\notin\tilde\tau_n\}$;
- if $L(n+1)=0$, then we set $\tilde\tau_{n+1}=\tilde\tau_n\cup\{\tilde v_n i\}$, where $i=\min\{j\geq 1:\tilde v_n j\notin\tilde\tau_n\}$, and $\tilde u_{n+1}=\tilde u_n$.

Lemma 4.2. *For all $n\geq 0$, $(\tilde\tau_n,\tilde u_n)\stackrel{\mathrm{d}}{=}(\tau_n,u_n)$.*

Proof. By induction. □

Proof of Proposition 4.1. Throughout the proof, we identify (τ_n,u_n) with the distributional copy $(\tilde\tau_n,\tilde u_n)$ and omit the tilde from the notation. Then, for each $n\geq 0$ there are three cases:

- either $|u_{n+1}|=|u_n|$ (if $v_n\neq\varnothing$ and $L(n+1)=0$),
- or $|u_{n+1}|=|u_n|+1$ (if $v_n=\varnothing$ and $K(n+1)=0$, or if $v_n\neq\varnothing$ and $L(n+1)=1$),
- or $|u_{n+1}|=0$ (if $v_n=\varnothing$ and $K(n+1)=1$).

Let us write $R(n)=\max\{k\leq n: v_k=\varnothing \text{ and } K(k+1)=1\}$ for the last time before n when $v_k=\varnothing$ and $K(k+1)=1$ (we set $R(n)=0$ if there are no such times). Then by the above,

$$
\begin{aligned}
|u_n| &= \sum_{i=R(n)+1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}+\sum_{i=R(n)+1}^{n}\mathbf{1}_{v_{i-1}\neq\varnothing}L(i)\\
&= \sum_{i=R(n)+1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}+\sum_{i=R(n)+1}^{n}L(i)-\sum_{i=R(n)+1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}L(i).
\end{aligned}
$$

Let $\mathcal{F}_n=\sigma(\tau_0,\tau_1,\dots,\tau_n)$. Then, since $|\tau_k|\geq k+1$ almost surely, for each $k\geq 0$

$$\mathbb{P}(v_k=\varnothing,K(k+1)=1)=\mathbb{E}[\mathbb{P}(v_k=\varnothing,K(k+1)=1\mid\mathcal{F}_k)]=\mathbb{E}\Big[\frac{1}{|\tau_k|(2|\tau_k|+1)}\Big]\leq\frac{1}{(k+1)(2k+3)}.$$

It follows, by the Borel–Cantelli lemma, that there is an a.s. finite random variable R such that $R(n)\to R$ almost surely as $n\uparrow\infty$. Similarly,

$$\mathbb{P}(\mathbf{1}_{v_k=\varnothing}L(k+1)=1\mid\mathcal{F}_k)=\mathbb{E}[\mathbb{P}(\mathbf{1}_{v_k=\varnothing}L(k+1)=1\mid\mathcal{F}_k)]=\mathbb{E}\Big[\frac{1}{|\tau_k|(|\tau_k|+1)}\Big]\leq\frac{1}{(k+1)(k+2)},$$

and thus, by the Borel–Cantelli lemma,

$$\sum_{i=R(n)+1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}L(i)\leq\sum_{i=1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}L(i)=\mathcal{O}(1),$$

almost surely as $n\uparrow\infty$. Thus, almost surely as $n\uparrow\infty$,

$$|u_n|=\sum_{i=1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}+\sum_{i=1}^{n}L(i)+\mathcal{O}(1). \tag{4.1}$$

We write (4.1) in the following form:

$$|u_n|=2\sum_{i=1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}-\sum_{i=1}^{n}\Big(\mathbf{1}_{v_{i-1}=\varnothing}-\frac{1}{B_{i-1}+1}\Big)+\sum_{i=1}^{n}\Big(L(i)-\frac{1}{B_{i-1}+2}\Big)+\mathcal{O}(1). \tag{4.2}$$

To do this, we have used the fact that

$$\sum_{i=1}^{n}\Big(\frac{1}{B_{i-1}+1}-\frac{1}{B_{i-1}+2}\Big)=\sum_{i=1}^{n}\frac{1}{(B_{i-1}+1)(B_{i-1}+2)}\leq\sum_{i=1}^{n}\frac{1}{i^2}=\mathcal{O}(1). \tag{4.3}$$

The first summand in (4.2) is taken care of by Proposition 2.3: as $n\uparrow\infty$,

$$G_0(n):=\frac{\sum_{i=1}^{n}\mathbf{1}_{v_{i-1}=\varnothing}-\frac{\log n}{1+\log 2}}{\frac{\sqrt{\log n}}{(1+\log 2)^{3/2}}}\Rightarrow G_0\sim\mathcal{N}(0,1). \tag{4.4}$$

We claim that, as $n\uparrow\infty$,

$$G_1(n):=\frac{\sum_{i=1}^{n}\big(\mathbf{1}_{v_{i-1}=\varnothing}-\frac{1}{B_{i-1}+1}\big)}{\sqrt{\frac{\log n}{1+\log 2}}}\Rightarrow G_1\sim\mathcal{N}(0,1) \tag{4.5}$$

and

$$W(n):=\frac{\sum_{i=1}^{n}\big(L(i)-\frac{1}{B_{i-1}+2}\big)}{\sqrt{\frac{\log n}{1+\log 2}}}\Rightarrow W\sim\mathcal{N}(0,1), \tag{4.6}$$

where W is independent of (G_0, G_1). Taken together, (4.4), (4.5) and (4.6) give the claim, with $\Lambda = G + W$, where

$$G := \frac{2G_0}{1+\log 2} - G_1, \tag{4.7}$$

is independent of W. We now proceed with the proofs of (4.5) and (4.6).

For (4.5), recalling that $\mathbb{E}\big[\mathbf{1}_{v_{i-1}=\emptyset}|\mathcal{F}_{i-1}\big] = \frac{1}{|\tau_{i-1}|} = \frac{1}{B_{i-1}+1}$, we can write

$$\begin{aligned}\sum_{i=1}^n \mathbf{1}_{v_{i-1}=\emptyset} &= \sum_{i=1}^n \mathbb{E}\big[\mathbf{1}_{v_{i-1}=\emptyset}|\mathcal{F}_{i-1}\big] + \sum_{i=1}^n \big(\mathbf{1}_{v_{i-1}=\emptyset} - \mathbb{E}\big[\mathbf{1}_{v_{i-1}=\emptyset}|\mathcal{F}_{i-1}\big]\big)\\ &= \sum_{i=1}^n \frac{1}{B_{i-1}+1} + M_n,\end{aligned} \tag{4.8}$$

where $M_n := \sum_{i=1}^n \big(\mathbf{1}_{v_{i-1}=\emptyset} - \mathbb{E}\big[\mathbf{1}_{v_{i-1}=\emptyset}|\mathcal{F}_{i-1}\big]\big)$ defines a martingale. We need to prove that $M_n/\sqrt{\frac{\log n}{1+\log 2}} \Rightarrow \mathcal{N}(0,1)$. The quadratic variation of $(M_n)_{n\geq 0}$ is given by

$$\begin{aligned}\langle M\rangle_n &= \sum_{i=1}^n \mathbb{E}\big[\big(\mathbf{1}_{v_{i-1}=\emptyset} - \mathbb{E}\big[\mathbf{1}_{v_{i-1}=\emptyset}|\mathcal{F}_{i-1}\big]\big)^2|\mathcal{F}_{i-1}\big] = \sum_{i=1}^n \mathbb{E}\big[\mathbf{1}_{v_{i-1}=\emptyset}|\mathcal{F}_{i-1}\big]\big(1 - \mathbb{E}\big[\mathbf{1}_{v_{i-1}=\emptyset}|\mathcal{F}_{i-1}\big]\big)\\ &= \sum_{i=1}^n \frac{1}{B_{i-1}+1}\left(1 - \frac{1}{B_{i-1}+1}\right) = \sum_{i=1}^n \frac{1}{B_{i-1}+1} + \mathcal{O}(1),\end{aligned} \tag{4.9}$$

where we used that $B_n \geq n$ almost surely for all $n \geq 0$. Furthermore, by Lemma 2.1, $\langle M\rangle_n \to +\infty$ almost surely as $n \uparrow \infty$. Thus, by the martingale law of large numbers [18, 12.14], $M_n = o(\langle M\rangle_n)$ almost surely as $n \uparrow \infty$. Using again the fact that $B_n \geq n$, we get that $\langle M\rangle_n = \mathcal{O}(\log n)$ almost surely as $n \uparrow \infty$ and hence $M_n = o(\log n)$ almost surely as $n \uparrow \infty$. Now note that, by Proposition 2.3,

$$\sum_{i=1}^n \mathbf{1}_{v_{i-1}=\emptyset} \sim \frac{\log n}{1+\log 2}, \qquad \text{in probability.}$$

Thus, by (4.8),

$$\begin{aligned}\sum_{i=1}^n \frac{1}{B_{i-1}+1} = \sum_{i=1}^n \mathbf{1}_{v_{i-1}=\emptyset} - M_n &= \sum_{i=1}^n \mathbf{1}_{v_{i-1}=\emptyset} + o(\log n) \qquad \text{almost surely,}\\ &\sim \frac{\log n}{1+\log 2} \qquad \text{in probability.}\end{aligned} \tag{4.10}$$

By (4.9), this implies that $\langle M\rangle_n \sim \log n/(1+\log 2)$ in probability. Thus, by the martingale central limit theorem [20, Thm 8.2.8],

$$\frac{M_n}{\sqrt{\frac{\log n}{1+\log 2}}} \Rightarrow \mathcal{N}(0,1), \tag{4.11}$$

as required.

For (4.6), we first reason conditionally on $\mathbf{v} = (v_k)_{k\geq 0}$ and thus on $(B_k)_{k\geq 0}$: conditionally on $\mathbf{v}$, $(L(i))_{i\geq 0}$ is a sequence of independent Bernoulli random variables of respective parameters $1/(B_{i-1}+2)$, $i \geq 0$. By (4.3) and (4.10),

$$\sum_{i=1}^n \frac{1}{B_{i-1}+2} = \sum_{i=1}^n \frac{1}{B_{i-1}+1} + \mathcal{O}(1) = \frac{\log n}{1+\log 2} + \mathcal{O}(1), \qquad \text{in probability.}$$

Thus, by the Lindeberg central theorem [21, Theorem 7.2.1] (whose conditions are easily checked since the $L(i)$ are bounded), we get that, conditionally on $\mathbf{v}$,

$$W(n) = \frac{\sum_{i=1}^n L(i) - \frac{1}{B_{i-1}+2}}{\sqrt{\frac{\log n}{1+\log 2}}} \Rightarrow W, \tag{4.12}$$

where W is a standard Gaussian. Explicitly, this means that for all continuous and bounded functions $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[\varphi(W(n)) \mid \mathbf{v}] \to \mathbb{E}[\varphi(W)].$$

Because the limit does not depend on $\mathbf{v}$, by dominated convergence, we can take expectations on both sides of the limit, which gives (4.6).

It only remains to show that W is independent of (G_0, G_1). First, for all continuous and bounded functions $\varphi, \psi : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}[\varphi(W(n))\psi(G_0(n))] &= \mathbb{E}\big[\mathbb{E}[\varphi(W(n))\psi(G_0(n)) \mid \mathbf{v}]\big] = \mathbb{E}\big[\psi(G_0(n))\mathbb{E}[\varphi(W(n)) \mid \mathbf{v}]\big]\\ &= \mathbb{E}\big[\psi(G_0(n))\mathbb{E}[\varphi(W)]\big] + \mathbb{E}\big[\psi(G_0(n))\big(\mathbb{E}[\varphi(W(n)) \mid \mathbf{v}] - \mathbb{E}[\varphi(W)]\big)\big].\end{aligned} \tag{4.13}$$

On the one hand, by linearity (because $\mathbb{E}[\varphi(W)]$ is a constant), and by (4.4),

$$\mathbb{E}\big[\psi(G_0(n))\mathbb{E}[\varphi(W)]\big] = \mathbb{E}[\varphi(W)]\mathbb{E}\big[\psi(G_0(n))\big] \to \mathbb{E}[\varphi(W)]\mathbb{E}[\psi(G_0)].$$

On the other hand,

$$\Big|\mathbb{E}\big[\psi(G_0(n))\big(\mathbb{E}[\varphi(W(n)) \mid \boldsymbol{v}] - \mathbb{E}[\varphi(W)]\big)\big]\Big| \le \mathbb{E}\Big[\psi(G_0(n))\Big|\mathbb{E}[\varphi(W(n)) \mid \boldsymbol{v}] - \mathbb{E}[\varphi(W)]\Big|\Big] \to 0,$$

by dominated convergence, because ψ and φ are bounded, and by (4.6). Thus, (4.13) implies

$$\mathbb{E}[\varphi(W(n))\psi(G_0(n))] \to \mathbb{E}[\varphi(W)]\mathbb{E}[\psi(G_0)].$$

Similarly, one can show that, for all continuous and bounded functions $\varphi : \mathbb{R} \to \mathbb{R}$ and $\psi : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{E}[\varphi(W(n))\psi(G_0(n), G_1(n))] \to \mathbb{E}[\varphi(W)]\mathbb{E}[\psi(G_0, G_1)].$$

This implies that (4.4), (4.5), and (4.6) hold jointly with W independent of (G_0, G_1), as desired. □

Remark. Note that we cannot say much about the distribution of $\Lambda = G + W$ since the two Gaussians G_0 and G_1 (see (4.7)) might be correlated.

4.2. The height profile: proof of Theorem 1.4

We now turn to the case $k = 2$ of Theorem 1.4, and we write (u_n, v_n) for the two uniformly random vertices in τ_n rather than $(u_n^{(1)}, u_n^{(2)})$. We follow a strategy similar to that of Section 4.1, defining a sequence $(\tilde{\tau}_n, \tilde{u}_n, \tilde{v}_n)_{n\ge 0}$ such that for each $n \ge 0$, the triple $(\tilde{\tau}_n, \tilde{u}_n, \tilde{v}_n)$ is a distributional copy of (τ_n, u_n, v_n).

First, let $\tilde{\tau}_0 = \{\varnothing\}$ and $\tilde{u}_0 = \tilde{v}_0 = \varnothing$. Then, for all $n \ge 0$, given $(\tilde{\tau}_n, \tilde{u}_n, \tilde{v}_n)$, we first sample $K_1(n+1)$ and $K_2(n+1)$, two independent Bernoulli-distributed random variables of parameter $1/(2|\tilde{\tau}_n|+1)$, and $L_1(n+1)$ and $L_2(n+1)$, two independent Bernoulli-distributed random variables of parameter $1/(|\tilde{\tau}_n|+1)$. Finally, we sample $(\alpha_n)_{n\ge1}$ and $(\beta_n)_{n\ge1}$, two independent sequences of random variables, uniformly distributed on $\{1, 2\}$.

(1) We let $\tilde{v}_n$ be a node taken uniformly at random among the nodes of $\tilde{\tau}_n$;
(2) If $\tilde{v}_n = \varnothing$, then we define

$$\tilde{\tau}_{n+1} = \{\varnothing\} \cup \{1w : w \in \tilde{\tau}_n\} \cup \{2w : w \in \tilde{\tau}_n\}.$$

Furthermore,

- if $K_1(n+1) = 1$, then we set $\tilde{u}_{n+1} = \varnothing$, and
- if $K_1(n+1) = 0$, then we set $\tilde{u}_{n+1} = \alpha_{n+1}\tilde{u}_n$.
- if $K_2(n+1) = 1$, then we set $\tilde{v}_{n+1} = \varnothing$, and
- if $K_2(n+1) = 0$, then we set $\tilde{v}_{n+1} = \beta_{n+1}\tilde{v}_n$.

(3) If $\tilde{v}_n \ne \varnothing$, then

- if $L_1(n+1) = L_2(n+1) = 1$, then we set $\tilde{\tau}_{n+1} = \tilde{\tau}_n \cup \{\tilde{u}_n i\}$ and $\tilde{u}_{n+1} = \tilde{v}_{n+1} = \tilde{u}_n i$, where $i = \min\{j \ge 1 : \tilde{u}_n j \notin \tilde{\tau}_n\}$;
- if $L_1(n+1) = 1$ and $L_2(n+1) = 0$, then we set $\tilde{\tau}_{n+1} = \tilde{\tau}_n \cup \{\tilde{u}_n i\}$ and $\tilde{u}_{n+1} = \tilde{u}_n i$, where $i = \min\{j \ge 1 : \tilde{u}_n j \notin \tilde{\tau}_n\}$; we also set $\tilde{v}_{n+1} = \tilde{v}_n$.
- if $L_1(n+1) = 0$ and $L_2(n+1) = 1$, then we set $\tilde{\tau}_{n+1} = \tilde{\tau}_n \cup \{\tilde{v}_n i\}$ and $\tilde{v}_{n+1} = \tilde{v}_n i$, where $i = \min\{j \ge 1 : \tilde{v}_n j \notin \tilde{\tau}_n\}$; we also set $\tilde{u}_{n+1} = \tilde{u}_n$.
- if $L_1(n+1) = L_2(n+1) = 0$, then we set $\tilde{\tau}_{n+1} = \tilde{\tau}_n \cup \{\tilde{v}_n i\}$, where $i = \min\{j \ge 1 : \tilde{v}_n j \notin \tilde{\tau}_n\}$, $\tilde{u}_{n+1} = \tilde{u}_n$, and $\tilde{v}_{n+1} = \tilde{v}_n$.

Note that, with this definition, $(\tilde{\tau}_n, \tilde{u}_n)_{n\ge0}$ is the same process as in Section 4.1. Recall that, in that process, we see some "resets" at the root at doubling-times when also $K_1(n+1) = 1$, while otherwise $\tilde{u}_{n+1}$ is either $\tilde{u}_n$ or a child of $\tilde{u}_n$. The evolution of $\tilde{v}_n$ is a bit more complex as it can reset at the root (a doubling-times when $K_2(n+1) = 1$), it can "jump" to $\tilde{u}_{n+1}$ (at non-doubling times when $L_1(n+1) = L_2(n+1) = 1$), and otherwise, $\tilde{v}_{n+1}$ is either $\tilde{v}_n$ or a child of $\tilde{v}_n$.

Lemma 4.3. *For all $n \ge 0$, $(\tilde{\tau}_n, \tilde{u}_n, \tilde{v}_n) \stackrel{\mathrm{d}}{=} (\tau_n, u_n, v_n)$.*

Proof. By induction. □

Proof of Theorem 1.4 for $k = 2$. Again, we identify (τ_n, u_n, v_n) with its distributional copy $(\tilde{\tau}_n, \tilde{u}_n, \tilde{v}_n)$ and drop the tilde from the notation. We let $R_1(n)$ (resp. $R_2(n)$) be the last time before (or at) time n when $v_{i-1} = \varnothing$ and $L_1(i) = 1$ (resp. $L_2(i) = 1$). With this definition, we have

$$|u_n| = \sum_{i=R_1(n)+1}^{n} \mathbf{1}_{v_{i-1}=\varnothing} + \sum_{i=R_1(n)+1}^{n} \mathbf{1}_{v_{i-1}\ne\varnothing} L_1(i) = \sum_{i=1}^{n} \mathbf{1}_{v_{i-1}=\varnothing} + \sum_{i=1}^{n} \mathbf{1}_{v_{i-1}\ne\varnothing} L_1(i) + \mathcal{O}(1), \tag{4.14}$$

as in Section 4.1. Now, we let $S(n)$ be the last time before (or at) time n when $v_{i-1} \neq \varnothing$ and $L_1(i) = L_2(i) = 1$. This is the last time before time n when v_n jumped to join u_n at a non-doubling time. If $S(n) < R_2(n)$, then v_n has reset to $\varnothing$ since last jumping to join u_n, so

$$|v_n| = \sum_{i=R_2(n)+1}^{n} \mathbf{1}_{v_{i-1}=\varnothing} + \sum_{i=R_2(n)+1}^{n} L_2(i).$$

If $R_2(n) < S(n)$ (note that they cannot be equal, by definition), then

$$|v_n| = |u_{S(n)}| + \sum_{i=S(n)+1} \mathbf{1}_{v_{i-1}=\varnothing} + \sum_{i=S(n)+1} \mathbf{1}_{v_{i-1}\neq\varnothing} L_2(i).$$

To summarise, if we let $S_2(n) = R_2(n) \vee S(n)$, then

$$|v_n| = |u_{S(n)}| \mathbf{1}_{R_2(n)<S(n)} + \sum_{i=S_2(n)+1}^{n} \mathbf{1}_{v_{i-1}=\varnothing} + \sum_{i=S_2(n)+1}^{n} \mathbf{1}_{v_{i-1}\neq\varnothing} L_2(i). \tag{4.15}$$

Note that

$$\mathbb{P}(v_{i-1} \neq \varnothing \text{ and } L_1(i) = L_2(i) = 1 \mid \mathcal{F}_{i-1}) = \frac{B_{i-1}}{B_{i-1}+1} \cdot \left(\frac{1}{B_{i-1}+2}\right)^2 \le \left(\frac{1}{B_{i-1}+2}\right)^2 \le \frac{1}{i^2},$$

and thus $\mathbb{P}(v_{i-1} \neq \varnothing \text{ and } L_1(i) = L_2(i) = 1) \le \frac{1}{i^2}$. By the Borel–Cantelli lemma, almost surely as $n \uparrow \infty$, $S(n) \to S$, where S is an almost surely finite random variable. Similarly, as proved in Section 4.1, $R_2(n) \to R_2$ almost surely as $n \uparrow \infty$, where R_2 is an almost-surely finite random variable. Thus, $S_2(n) \to S_2 = R_2 \vee S$ almost surely as $n \uparrow \infty$, and therefore

$$|v_n| = \sum_{i=1}^{n} \mathbf{1}_{v_{i-1}=\varnothing} + \sum_{i=1}^{n} \mathbf{1}_{v_{i-1}\neq\varnothing} L_2(i) + \mathcal{O}(1). \tag{4.16}$$

From this point, the rest of the argument is as in the proof of Proposition 4.1: we write (4.14) and (4.16) as

$$\begin{aligned}
|u_n| &= 2\sum_{i=1}^{n} \mathbf{1}_{v_{i-1}=\varnothing} - \sum_{i=1}^{n}\left(\mathbf{1}_{v_{i-1}=\varnothing} - \frac{1}{B_{i-1}+1}\right) + \sum_{i=1}^{n}\left(\mathbf{1}_{v_{i-1}\neq\varnothing} L_1(i) - \frac{1}{B_{i-1}+1}\right) + \mathcal{O}(1),\\
|v_n| &= 2\sum_{i=1}^{n} \mathbf{1}_{v_{i-1}=\varnothing} - \sum_{i=1}^{n}\left(\mathbf{1}_{v_{i-1}=\varnothing} - \frac{1}{B_{i-1}+1}\right) + \sum_{i=1}^{n}\left(\mathbf{1}_{v_{i-1}\neq\varnothing} L_2(i) - \frac{1}{B_{i-1}+1}\right) + \mathcal{O}(1),
\end{aligned} \tag{4.17}$$

where in each expression, the first sum is handled using Proposition 2.3 and the others using the maringale central limit theorem. Note that this gives

$$\left(\frac{|u_n| - \frac{2\log n}{1+\log 2}}{\sqrt{\frac{\log n}{1+\log 2}}}, \frac{|v_n| - \frac{2\log n}{1+\log 2}}{\sqrt{\frac{\log n}{1+\log 2}}}\right) \Rightarrow (V + W_1, V + W_2), \tag{4.18}$$

where $V = \frac{2G_0}{1+\log 2} - G_1$ as in (4.7) and W_1, W_2 are defined as in (4.6) using L_1 and L_2, respectively.

We now briefly comment on the modifications needed for the case $k \ge 3$. As before, we define a process $(\tilde{\tau}_n, \tilde{u}_n^{(1)}, \tilde{u}_n^{(2)}, \dots, \tilde{u}_n^{(k)})$ such that for each $n \ge 0$, $(\tilde{\tau}_n, \tilde{u}_n^{(1)}, \dots, \tilde{u}_n^{(k)})$ has the same distribution as $(\tau_n, u_n^{(1)}, \dots, u_n^{(k)})$. In this process, the triple $(\tilde{\tau}_n, \tilde{u}_n^{(1)}, \tilde{u}_n^{(2)})$ will be the same as the process used for $k = 2$ above. Each $\tilde{u}_n^{(j)}$ can be "reset" to $\varnothing$ (in the case $\tilde{v}_n = \varnothing$), it can "jump" to any of $\tilde{u}_n^{(1)}, \dots, \tilde{u}_n^{(j-1)}$ (in the case $\tilde{v}_n \neq \varnothing$), it can be replaced by a "new" child, or it can remain unchanged, these choices being determined by suitable random variables $K_1(n+1), \dots, K_k(n+1)$ and $L_1(n+1), \dots, L_k(n+1)$. In writing expressions such as (4.15), there are many cases to consider, but the analogs of (4.14) and (4.16) hold for each of $\tilde{u}_n^{(1)}, \dots, \tilde{u}_n^{(k)}$, and the rest of the argument is as before.

4.3. Lower bound on the height: proof of Proposition 1.5

Let $\mathfrak{w}_n$ denote the leftmost child of the root at height $\kappa(n)$, where we recall (2.1) that $\kappa(n)$ is the number of doubling events before time n. Since, by definition, the new node added at non-doubling times is always added to the right of already existing siblings, $\mathfrak{w}_n$ is a 'copy' of the original root of the tree τ_0. Let $\mathfrak{S}_n$ denote the subtree rooted at $\mathfrak{w}_n$, let $\mathfrak{s}(n)$ denote the number of nodes in $\mathfrak{S}_n$, and let $\mathfrak{h}(n)$ denote the height of $\mathfrak{S}_n$. Note that $\mathfrak{S}_n$ is a distributional copy of the random recursive tree at time $\mathfrak{s}(n)$. Clearly, the height H_n of τ_n satisfies $H_n \ge \kappa(n) + \mathfrak{h}(n)$. Moreover, we have the following estimate on the height of the random recursive tree, which gives $\mathfrak{h}(n) \sim \mathrm{e}\log \mathfrak{s}(n)$:

Theorem 4.4 (*see Pittel [17]*). *Let $h(n)$ be the height of the n-node random recursive tree. Almost surely as $n \uparrow \infty$, $h(n) \sim \mathrm{e}\log n$.*

However, rather than applying Theorem 4.4 directly, which would require finding estimates on $\mathfrak{s}(n)$, we use an embedding of our process $(\tau_n)_{n\ge 0}$ into continuous time. We define a continuous-time process $(\mathcal{T}(t))_{t\ge 0}$ of growing trees by first setting $\mathcal{T}(0) = \{\varnothing\}$ and then assigning to every node of the tree a clock that rings at exponential rate of parameter 1, the clocks for different nodes being independent. When a clock rings, if it is the clock associated to $\varnothing$, then, at that time, we double the tree as done at doubling events in the discrete time tree; otherwise, we add one child to the node whose clock rang. As before, our convention is to add the new node to the right of already existing siblings.

If we let t_n be the time of the nth ring of a clock, then $(\mathcal{T}(t_n))_{n\geq 0} \stackrel{\mathrm{d}}{=} (\tau_n)_{n\geq 0}$. Also note that the times at which the tree doubles define a Poisson point process of intensity 1; in particular, the number $D(t)$ of doubling events before time t is distributed as a Poisson of parameter t. We now let $\mathcal{S}(t)$ denote the subtree rooted at the leftmost node at height $D(t)$. Thus $\mathcal{S}(t)$ is the continuous-time version of $\mathfrak{S}_n$, and it is now simply a Yule process of parameter 1.

Note that a random recursive tree can be coupled to a Yule process so that the random recursive tree equals the Yule process taken at its successive jump times. Thus, if $H(t)$ is the height of the Yule process $\mathcal{S}(t)$ at time t, then $H(t) = h(|\mathcal{S}(t)|)$, where $|\mathcal{S}(t)|$ is the number of nodes in the Yule process at time t. Thus, by Theorem 4.4, almost surely as $t \uparrow \infty$,

$$\frac{H(t)}{\log|\mathcal{S}(t)|} \to \mathrm{e}.$$

It is well-known that $\mathrm{e}^{-t}|\mathcal{S}(t)|$ converges almost surely to a standard exponential random variable (see, e.g. [22, Section III.5]), which implies $\log|\mathcal{S}(t)| \sim t$ almost surely as $t \uparrow \infty$. Thus, almost surely as $t \uparrow \infty$,

$$H(t) \sim \mathrm{e}\, t.$$

It follows that the height of $\mathcal{T}(t)$ is at least $D(t) + H(t) \sim (1+\mathrm{e})t$, almost surely as $t \uparrow \infty$.

Proof of Proposition 1.5. It only remains to translate this lower bound into discrete time. For that, we need to understand the asymptotic behaviour of t_n, the times at which $\mathcal{T}(t)$ grows. Let $N(t) = |\mathcal{T}(t)|$ be the number of notes in the tree $\mathcal{T}(t)$ at time t. At time t, the rate at which the next clock rings is $N(t)$, so we need to understand $N(t)$.

To do this, we will couple $(N(t))_{t\geq 0}$ to a process $(Y(t))_{t\geq 0}$ which is the size of a standard Yule process. Indeed, intuitively $(N(t))_{t\geq 0}$ is a Yule process with jumps at the doubling-times; the process $(Y(t))_{t\geq 0}$ will be defined to "fill in" the instantaneous doubling events of $N(t)$ with a Yule process run for the amount of time it takes to double in size.

To express this more precisely (and we refer to Fig. 2 for this discussion), write $d_1, d_2, \dots$ for the doubling times of $N(t)$, i.e. the jump times of the Poisson process $(D(t))_{t\geq 0}$. For $0 \leq t < d_1$, we set $Y(t) = N(t)$. Then, we let $(Y(t))_{d_1 \leq t < d_1+\ell_1}$ be the size of a Yule process started at $Y(d_1^-)$ and stopped at time ℓ_1, defined as the first time it reaches $2Y(d_1^-)$+1. Then, for all $n \geq 1$, given $(Y(t))_{t<d_n+\ell_n}$, we let $Y(t) = N(t-\ell_n)$ for all $d_n + \ell_n \leq t < d_{n+1} + \ell_n$. Also, we define $(Y(t))_{d_{n+1}+\ell_n \leq t < d_{n+1}+\ell_{n+1}}$ as a Yule process started at $N(d_{n+1}^-)$ and stopped at time $\Delta\ell_{n+1} = \ell_{n+1} - \ell_n$, defined as the first time it hits $2N(d_{n+1}^-)$+1. By the strong Markov property, $(Y(t))_{t\geq 0}$ is a Yule process. Furthermore, by definition, almost surely for all $t \geq 0$, $Y(t + \ell_{D(t)}) = N(t)$.

Now $\mathrm{e}^{-t}Y(t) \to \xi$ almost surely as $t \to \infty$, where ξ is exponentially distributed. It follows that $\log N(t) \sim t + \ell_{D(t)}$ almost surely. We now show that $\ell_{D(t)} \sim t\log 2$ almost surely as $t \uparrow \infty$. First note that at doubling times we have $N(d_i) = Y(d_i + \ell_i)$, thus $\mathrm{e}^{-(d_i+\ell_i)}N(d_i) \to \xi$ almost surely as $i \to \infty$. But also $N(d_i) = 2N(d_i^-) + 1 = 2Y(d_i + \ell_{i-1}) + 1$, so that

$$\mathrm{e}^{-(d_i+\ell_i)}N(d_i) = 2\mathrm{e}^{-(d_i+\ell_{i-1}+\Delta\ell_i)}Y(d_i+\ell_{i-1}) + \mathrm{e}^{-(d_i+\ell_i)} \to \xi, \qquad \text{a.s. as } i \to \infty.$$

Since also $\mathrm{e}^{-(d_i+\ell_{i-1})}Y(d_i+\ell_{i-1}) \to \xi$, it follows that $\Delta\ell_i \to \log 2$ almost surely. Since $D(t) \sim t$ as $t \to \infty$,

$$\ell_{D(t)} = \sum_{i=1}^{D(t)} \Delta\ell_i \sim t\log 2, \qquad \text{almost surely,}$$

as claimed.

We thus have that

$$\log N(t) \sim (1+\log 2)t, \qquad \text{almost surely as } t \uparrow \infty. \tag{4.19}$$

We claim that

$$\liminf_{n\to\infty} \frac{t_n}{\log n} \geq \frac{1}{1+\log 2}, \qquad \text{almost surely as } n \uparrow \infty. \tag{4.20}$$

For this, we first note that $t_n \to \infty$ almost surely as $n \to \infty$; indeed, conditionally on all the B_i, we have that t_n is a sum of independent exponential random variables of rates $B_0+1, B_1+1, \dots, B_{n-1}+1$, thus the conditional mean of t_n diverges almost surely by Lemma 2.1, while the conditional variance is bounded since $B_n + 1 \geq n$ for all $n \geq 0$. Then, using (4.19) we have

$$\log n \leq \log(B_n+1) = \log N(t_n) \sim (1+\log 2)t_n, \qquad \text{almost surely as } n \to \infty.$$

From (4.20) and the fact that the height of the continuous-time tree $\mathcal{T}(t)$ is asymptotically at least $(1+\mathrm{e}+o(1))t$ (where the $o(1)$-term goes to 0 almost surely as $t \uparrow \infty$), we thus get that, almost surely as $n \uparrow \infty$,

$$H_n \geq (1+\mathrm{e}+o(1))t_n \geq \frac{1+\mathrm{e}+o(1)}{1+\log 2}\cdot \log n,$$

as claimed. $\square$

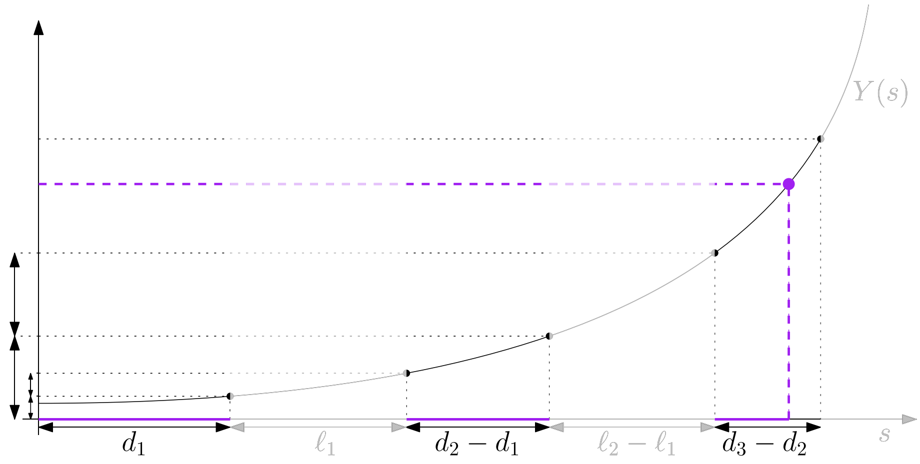

Fig. 2. One can see how the process $(N(t))_{t\geq 0}$ can be coupled with a Yule process $(Y(t))_{t\geq 0}$ so that, for all $t \geq 0$, $Y(t+\ell_{D(t)}) = N(t)$, where $D(t)$ is the number of doubling events before time t. The Yule process is the concatenation of the black and grey parts of the curve, whilst $(N(t))_{t\geq 0}$ is the curve obtained by only keeping the black parts and gluing them as if time-warps made us skip the intervals of time in grey. The two pairs of distances highlighted on the left-hand side are such that the two arrows in one pair have the same length: this means that the grey intervals are intervals during which the Yule process doubles in size. Note that, although both Y and N are jump processes that take value in $\mathbb{N}$, we have here represented them as continuous curves (with jumps for N when it doubles); this is just for ease of representation. One can see that, if the total length of all the purple intervals is t, then, indeed, $D(t) = 2$ and $N(t)$, which is the value highlighted by large a purple dot, equals $Y(t+\ell_2)$, as claimed in the proof of Proposition 1.5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Remark 4.5. Using Proposition 1.1(d), in fact $\frac{t_n}{\log n} \to \frac{1}{1+\log 2}$ almost surely as $n \to \infty$. Since $\kappa(n)$ from (2.1) satisfies $\kappa(n) = D(t_n) \sim t_n$ almost surely, it follows that $\frac{\kappa(n)}{\log n} \to \frac{1}{1+\log 2}$ almost surely.

Moreover, conditionally on $(B_i)_{i\geq 0}$, the expression $t_n - \sum_{i=0}^{n-1} \frac{1}{B_i+1}$ is a sum of independent random variables with mean zero and summable variances. Hence, it is a martingale bounded in L^2 which therefore converges almost surely. Using that $t_n \sim \log n/(1+\log 2)$ almost surely as $n \to \infty$, it follows that (4.10) can be improved to an almost sure equivalence: explicitly,

$$\sum_{i=0}^{n-1} \frac{1}{B_i+1} \sim \frac{\log n}{1+\log 2}, \qquad \text{almost surely as } n \to \infty.$$

5. A lower bound on the size of the tree doubling everywhere

As mentioned in the introduction, the model of random recursive tree that doubles at the root is a simplification of a tree that would "double everywhere". We define the random tree $(\tau_n^\infty)_{n\geq 0}$ recursively as follows: $\tau_0^\infty = \{\varnothing\}$ and, for all $n \geq 0$, given τ_n^∞, we pick a node v_n uniformly at random in τ_n^∞, let t_n be the set of (non-strict) descendants of v_n in τ_n^∞, and set

$$\tau_{n+1}^\infty = \big(\tau_n \setminus t_n\big) \cup \{v_n\} \cup \{v_n 1w : v_n w \in \tau_n^\infty\} \cup \{v_n 2w : v_n w \in \tau_n^\infty\}.$$

In other words, at every time step, we pick a node uniformly at random in τ_n^∞, remove its subtree (the node and all its descendants) from τ_n^∞ and replace it by two copies of itself as the two subtrees of a new node. Note that, by definition, for all $n \geq 0$, τ_n^∞ is binary, i.e. all its words are made on the alphabet $\{1, 2\}$.

We only make the following simple observation about this model:

Proposition 5.1. *For all $n \geq 1$ we have $\mathbb{E}[|\tau_n^\infty|] \geq \frac{n-1}{2}\log_2(\frac{n-1}{e})$.*

Proof. For all $n \geq 0$ and $u \in \tau_n^\infty$, we let $s_n(u)$ be the number of (strict) descendants of nodes u in τ_n^∞. Also recall that $|u|$ is the height of node u (i.e. the number of strict ancestors of u). Note that, if at step $n+1$, we select node $u \in \tau_n^\infty$, then $|\tau_{n+1}^\infty| = |\tau_n^\infty| + 2 + s_n(u)$. Indeed, the tree rooted at u (which contains $1 + s_n(u)$ nodes) is replaced by a node to which are attached two copies of u and its subtree (which contains $1 + 2(1 + s_n(u))$ nodes in total). Thus, for all $n \geq 0$,

$$\mathbb{E}[|\tau_{n+1}^\infty| \mid \tau_n^\infty] = |\tau_n^\infty| + \frac{1}{|\tau_n^\infty|}\sum_{u\in\tau_n^\infty}\big(2 + s_n(u)\big) = |\tau_n^\infty| + 2 + \frac{1}{|\tau_n^\infty|}\sum_{u\in\tau_n^\infty}\sum_{v\prec u} 1 = |\tau_n^\infty| + 2 + \frac{1}{|\tau_n^\infty|}\sum_{v\in\tau_n^\infty}|v|.$$

The last term is the expected height of a node chosen uniformly at random in τ_n^∞. Since τ_n^∞ is a binary tree, for any $k \geq 0$ the number of nodes at height at most k is at most 2^{k+1}. Thus, at least half the nodes of τ_n^∞ have height at least $\log_2|\tau_n^\infty| - 2 \geq \log_2 n - 2$. This implies that, almost surely,

$$\mathbb{E}[|\tau_{n+1}^{\infty}|] \geq \mathbb{E}[|\tau_n^{\infty}|] + 2 + \tfrac{1}{2}(\log_2 n - 2) \geq \mathbb{E}[|\tau_n^{\infty}|] + \tfrac{1}{2}\log_2 n.$$

By induction,

$$\mathbb{E}[|\tau_{n+1}^{\infty}|] \geq \frac{1}{2}\sum_{j=1}^{n} \log_2 j \geq \tfrac{n}{2}\log_2(\tfrac{n}{e}). \quad \square$$

Remark 5.2. Proposition 5.1 says that, in expectation, the size of the tree that "doubles everywhere" is superlinear in the number of steps. By definition, $|\tau_n^{\infty}| \leq 2^n$, where the upper-bound is attained on the event that all doubling events happen at the root. What is the exact order of $\mathbb{E}[|\tau_n^{\infty}|]$ as $n \uparrow \infty$? Can we find asymptotic equivalents for $|\tau_n^{\infty}|$ itself, either in probability, or almost surely? We leave these as open problems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research of JEB was supported by Vetenskapsrådet, grant 2023-05002, and by *Ruth och Nils Erik Stenbäcks stiftelse*. CM would like to thank Jason Schweinsberg for preliminary discussions on the asymptotic behaviour of B_n as $n \uparrow \infty$. We would like to thank Lazare Le Borgne for spotting and correcting a mistake in the proof of Theorem 1.3, and the anonymous referee for reading our paper carefully and making several very helpful comments and suggestions.

References

[1] Hosam M. Mahmoud, Robert T. Smythe, Asymptotic joint normality of outdegrees of nodes in random recursive trees, Random Structures Algorithms 3 (3) (1992) 255–266.
[2] Svante Janson, Asymptotic degree distribution in random recursive trees, Random Structures Algorithms 26 (1–2) (2005) 69–83.
[3] Robin Pemantle, A survey of random processes with reinforcement., Probab. Surv. 4 (2007) 1–79.
[4] Michel Benaïm, Dynamics of stochastic approximation algorithms, in: Seminaire de probabilites XXXIII, Springer, 2006, pp. 1–68.
[5] Marie Duflo, Random Iterative Models, Springer-Verlag, 1997.
[6] Luc Devroye, Applications of the theory of records in the study of random trees, Acta Inform. 26 (1) (1988) 123–130.
[7] Robert P. Dobrow, On the distribution of distances in recursive trees, J. Appl. Probab. 33 (3) (1996) 749–757.
[8] Cécile Mailler, Gerónimo Uribe Bravo, Random walks with preferential relocations and fading memory: a study through random recursive trees, J. Stat. Mech. Theory Exp. 2019 (9) (2019) 093206.
[9] Michael Drmota, Bernhard Gittenberger, On the profile of random trees, Random Structures Algorithms 10 (4) (1997) 421–451.
[10] Brigitte Chauvin, Michael Drmota, Jean Jabbour-Hattab, The profile of binary search trees, Ann. Appl. Probab. (2001) 1042–1062.
[11] Brigitte Chauvin, Thierry Klein, Jean-Francois Marckert, Alain Rouault, Martingales and profile of binary search trees, Electron. J. Probab. 10 (2005) 420–435.
[12] Eva-Maria Schopp, A functional limit theorem for the profile of b-ary trees, Ann. Appl. Probab. 20 (3) (2010) 907–950.
[13] Zsolt Katona, Width of a scale-free tree, J. Appl. Probab. 42 (3) (2005) 839–850.
[14] Henning Sulzbach, A functional limit law for the profile of plane-oriented recursive trees, in: DMTCS Proceedings, (1) 2008.
[15] Zakhar Kabluchko, Alexander Marynych, Henning Sulzbach, General edgeworth expansions with applications to profiles of random trees, Ann. Appl. Probab. 27 (6) (2017) 3478–3524.
[16] John D. Biggins, Martingale convergence in the branching random walk, J. Appl. Probab. 14 (1) (1977) 25–37.
[17] Boris Pittel, Note on the heights of random recursive trees and random m-ary search trees, Random Structures Algorithms 5 (2) (1994) 337–347.
[18] David Williams, Probability with Martingales, Cambridge University Press, 1991.
[19] Herbert Robbins, David Siegmund, A convergence theorem for non negative almost supermartingales and some applications, in: Optimizing Methods in Statistics, Elsevier, 1971, pp. 233–257.
[20] Rick Durrett, Probability: Theory and Examples, fifth ed., Cambridge University Press, 2019.
[21] Allan Gut, Probability: a graduate course, Springer-Verlag, 2013.
[22] Krishna B. Athreya, Peter E. Ney, Branching Processes, Springer, 1972.