

# The height of increasing trees

N. Broutin

L. Devroye

E. McLeish

M. de la Salle\*

May 21, 2007

## Abstract

We extend results about heights of random trees (Devroye, 1986, 1987, 1998b). In this paper, a general split tree model is considered in which the normalized subtree sizes of nodes converge in distribution. The height of these trees is shown to be in probability asymptotic to  $c \log n$  for some constant  $c$ . We apply our results to obtain a law of large numbers for the height of all polynomial varieties of increasing trees (Bergeron et al., 1992).

**Keywords and phrases:** Height, random tree, branching process, probabilistic analysis, increasing tree.

## 1 Introduction

The present paper gives a general result on the heights of random trees. It applies in particular to general  $d$ -ary increasing trees (Bergeron, Flajolet, and Salvy, 1992), a random tree model whose height was not known until now, although the variance of the height is known to be bounded (Drmota, 2006). The general approach we adopt is based on branching processes (Athreya and Ney, 1972; Harris, 1963). The hunt for the height of binary search trees has motivated the development of these branching processes techniques. Pittel (1984) was the first to introduce a crucial continuous embedding and to show that the height of a tree of  $n$  nodes is asymptotic to  $c \log n$  for some positive constant  $c$ . Using earlier work by Biggins (1976, 1977), Devroye (1986) proved that  $c \geq 2$  is given by the solution of  $c \log(2e/c) = 1$ , hence showing that the height of random binary search trees of size  $n$  is asymptotic to  $4.311\dots \log n$  in probability. Using branching random walks Biggins and Grey (1997) were able to generalize these theorems and extend the class of random trees that can be handled using this single method. More recently, Broutin and Devroye (2006) gave further results based on 2-dimensional branching processes that seem to encompass many extremal problems related to paths in trees.

In our tree model, we distinguish the *nodes* of the tree from the *items* they may contain. A node may contain zero, one, or more items. In the sequel, the *size* is the number of items in a tree. For a given node  $u$ , we use  $N_u$  to denote the size of the subtree rooted at  $u$ . We introduce  $d$ -ary *ideal* trees, in which each node receives an independent copy of the vector  $\mathcal{V} = (V_1, V_2, \dots, V_d)$ , with  $V_i \geq 0$ , for  $i = 1, \dots, d$ , and  $\sum_{i=1}^d V_i = 1$ .  $\mathcal{V}$  is called the split vector. It turns out that one can relate ideal trees to a larger family of random trees in which, for a node  $u$  and its  $i$ -th child  $u_i$ ,  $N_{u_i}$  is close to  $N_u \cdot V_i$  in distribution. We will be more precise in the next section. This model is close to the split tree model of Devroye (1998b). Note that in the ideal tree model, the split vectors are independent and identically

---

\*Research of the authors was supported by NSERC Grant A3456 and a James McGill fellowship. Address: School of Computer Science, McGill University, Montreal H3A2K6 Canada. Email: [nbrout,luc,mcleish@cs.mcgill.ca](mailto:{nbrout,luc,mcleish}@cs.mcgill.ca), [mikael.de-la-salle@normalesup.org](mailto:mikael.de-la-salle@normalesup.org).

distributed (i.i.d.). In particular, the copy at a node  $u$  does not depend on the value of  $N_u$ . However, there are kinds of trees in which there is such a dependence. These include the increasing trees, for instance. With a bit more machinery, one can relate the model to an ideal tree, and obtain results concerning the height (see Section 3).

We now describe the increasing tree model of Bergeron et al. (1992). This tree model relates to ordered trees. Thus, all trees in the next few paragraphs are to be considered ordered. A *labeled* tree of size  $n$  is a tree on  $n$  vertices that are given distinct elements of  $\{1, 2, \dots, n\}$ . A labeled tree is called *increasing* if the labels are increasing along any path starting at the root. Let  $\{s_r, r \geq 0\}$  be a sequence of nonnegative integers with  $s_0 \neq 0$ . One can describe a variety of trees by its *characteristic sequence*  $\{s_r\}$  or its *degree function*, the power series  $\phi(w) = \sum_{i \geq 0} s_i w^i$ . Varieties for which  $s_d \neq 0$  and  $s_r = 0$  for all  $r > d$  are called *polynomial*. The variety with characteristic sequence  $\{s_r\}$  consists of a collection of trees in which a specific tree  $T$ , with fixed shape and labeling, occurs exactly

$$\prod_{r \geq 0} s_r^{D_r}$$

times, where  $D_r$  is the number of nodes of outdegree  $r$  in  $T$ , and we adopt the convention  $0^0 = 1$ . Then, a *random increasing tree* of size  $n$  is an element of size  $n$  taken uniformly at random from the collection. As a consequence, a specific tree is drawn with probability proportional to the number of its occurrences in the collection. For counting purposes, it is helpful to think of a node of outdegree  $r$  as being colored in one of  $s_r$  colors. In that case, we see that each labeled colored tree  $T$  of size  $n$  occurs exactly once in the collection. The model embraces many important tree families like random recursive trees, plane oriented trees, and binary search trees. We now describe these three examples to help the reader become more familiar with the increasing tree model.

**EXAMPLE 1: UNIFORM RECURSIVE TREES.** Uniform recursive trees are described using an incremental process (Meir and Moon, 1978). A recursive tree of size 1 is a single node, labeled 1. At any time  $n$ , a new node labeled  $n$  comes along, and becomes a child of a *uniformly* chosen random vertex. The process produces any properly (increasingly) labeled tree of size  $n$  with the same probability  $1/(n-1)!$ . The order of the children of a node is irrelevant, and without loss of generality, one can consider the children as ordered by increasing value of their labels. As a consequence, a random recursive tree of size  $n$  is distributed as a random (ordered) increasing tree of the same size with associated sequence  $\{1/r!, r \geq 0\}$  or, equivalently, degree function  $\phi(w) = \exp w$ . The factorial compensates the fact that a node of degree  $d$  can label its children with  $d$  labels in  $d!$  ways. Put differently, in the ordered tree labeling, we only keep the one labeling in which the child labels are ordered. The fact that  $s_r$  is not integer-valued does not matter here, as we can still build a collection of validly labeled trees.

**EXAMPLE 2: PLANE ORIENTED TREES.** Plane oriented trees (or plane recursive trees) are an ordered version of recursive trees due to Szymański (1987). Plane oriented trees are also discussed and enumerated by Prodinger and Urbanek (1983). They may be obtained using successive insertions as well. The difference lies in that a parent is no longer chosen uniformly, but rather with probability proportional to one plus its outdegree. This is also the preferential attachment model of Barabási and Albert (1999), and a particular case of the more general linear recursive trees of Pittel (1994). Because the trees are now ordered, an increasing tree with characteristic sequence  $\{1, 1, \dots\}$ , and degree function  $\phi(w) = 1/(1-w)$ , is distributed as a random plane oriented tree. This equivalence is best seen by observing that in an increasing labeling, the maximal label attaches itself as a child of a node of degree  $d$  with probability proportional to  $d+1$ , as it can be inserted anywhere among the  $d$  existing children. These trees are the main example of random ordered trees. For more information see Mahmoud (1992), Mahmoud, Smythe, and Szymanski (1993) or the survey by Smythe and Mahmoud (1995).

**EXAMPLE 3: RANDOM BINARY SEARCH TREE.** The random binary search tree is a particularly important data structure. It also occurs in the analysis of quicksort (Knuth, 1973). One is given  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ , a random permutation of  $\{1, 2, \dots, n\}$ . A random binary search tree is built by first storing  $\sigma_1$  in the root. The left and right subtrees are then built similarly from the permutations  $\{\sigma_i : \sigma_i < \sigma_1\}$  and  $\{\sigma_j : \sigma_j > \sigma_1\}$ , respectively. Note that the values  $\sigma_i$  should not be seen as the labels of the increasing tree model. Observe that a tree with the same distribution may be obtained using a sequence of insertions. Let  $k \geq 1$  and let  $T_k$  be the binary search tree corresponding to  $\{\sigma_i, 1 \leq i \leq k\}$ . Then, since  $\{\sigma_i, 1 \leq i \leq k+1\}$  is a uniform random permutation,  $\sigma_{k+1}$  is uniform on  $\{1, \dots, k+1\}$ . A position corresponds to an external node of  $T_k$ , and hence  $\sigma_{k+1}$  is assigned to a uniform random external node of  $T_k$  to make  $T_{k+1}$ . The tree  $T_n$  whose nodes are labeled with the indices of the elements of the random permutation  $\sigma$  is clearly increasing. The probability that  $T_n$  has a given shape  $t$  is proportional to the number of proper labelings of  $t$ . Now, binary search trees are *position* trees, i.e., left and right children are distinguished. Hence, a random binary search tree is distributed as a random increasing tree taken from the sequence  $\{s_0 = 1, s_1 = 2, s_2 = 1\}$ .

Although true for our three examples, it is *not true* that all increasing tree families may be obtained using an incremental process. When there is a hole in the sequence  $\{s_r\}$ , as in  $\{s_0 = 1, s_1 = 0, s_2 = 1\}$ , for example, no incremental process exists that can model a random increasing tree. We refer the reader to Panholzer and Prodinger (2006) who characterize the varieties of increasing trees that admit a construction using successive insertions. From now on, we consider only polynomial varieties. See also the related papers of Łuczak and Winkler (2004) and Janson (2006).

Asymptotics of the heights of particular instances of increasing trees are known, but a general theorem was still missing. In this paper, we prove that the height of any random increasing tree of size  $n$  with bounded maximum degree  $d$  is asymptotic to  $c_d \log n$  in probability. The constant  $c_d$  depends only on  $d$  and is characterized as the solution of an implicit equation (Theorem 3).

In Section 3, we introduce a general model of random trees. We prove the following result (Theorem 2): for a node  $u$  with children  $u_1, \dots, u_d$ , if the distribution of  $(N_{u_1}/N_u, \dots, N_{u_d}/N_u)$ , conditioned on  $N_u = n$  converges in distribution to the split vector  $\mathcal{V}$  as  $n \rightarrow \infty$ , then the height is asymptotic to  $c_d \log n$  in probability. The theorem covers a broad spectrum of applications, including polynomial families of increasing trees discussed above. We start by describing the simpler model of ideal trees.

## 2 Ideal trees

Let  $\mathcal{V} = (V_1, \dots, V_d)$  be a random vector with nonnegative components such that  $\sum_{i=1}^d V_i = 1$  almost surely (a.s.). Let  $T_\infty$  be an infinite rooted  $d$ -ary tree (with  $d^k$  nodes at every level  $k \geq 0$ ). Assign an independent copy of  $\mathcal{V}$  to each node of  $T_\infty$ . Consider any node  $u \in T_\infty$  and its copy of the vector,  $(V_1, \dots, V_d)$ . If  $e$  is the edge between  $u$  and its  $i$ -th child, we define  $V_e = V_i$ . It should always be clear whether the subscript refers to an index or an edge. Let  $\pi(u)$  be the set of edges on the unique path from  $u$  to the root in  $T_\infty$ . We define  $L_u = \prod_{e \in \pi(u)} V_e$ . The subtree  $T_n$  of  $T_\infty$  consisting of the nodes  $u$  with  $L_u \geq 1/n$  is called the *ideal tree* built from the *split vector*  $\mathcal{V}$ . Its height, i.e., the maximum distance from a node to the root, is denoted by  $H_n$ .

Alternatively, we can describe  $T_n$  using additive settings. For each edge  $e$  of  $T_\infty$ , let  $E_e = -\log V_e \geq 0$ . Then,  $B_u = -\log L_u = \sum_{e \in \pi(u)} E_e$  is similar to the birth time of  $u$ , and  $T_n$  is the tree of nodes  $u$  born before time  $\log n$ , i.e., for which  $B_u \leq \log n$ . Then,

studying the height  $H_n$  reduces to analyzing the time at which the first node is born at level  $k$  (Biggins, 1976, 1977). We use the notation  $E_e$  because in many applications,  $E_e$  is distributed as an exponential random variable.

Without loss of generality, we assume that the components of  $\mathcal{V} = (V_1, \dots, V_d)$  are identically distributed. Indeed, randomly permuting the components does not affect the height and yields i.i.d. components. Let now  $V$  denote a typical component, and write  $E = -\log V \geq 0$ , agreeing that  $\log 0 = -\infty$ . Observe that  $\mathbf{EV} = 1/d$ , and that  $\mathbf{P}\{E = 0\} = \mathbf{P}\{V = 1\} \leq \mathbf{EV} = 1/d$ .

Ideal trees and their heights have already been studied under the name of branching random walk by Biggins (1976, 1977) and Biggins and Grey (1997). The split trees of Devroye (1998b) are close to ideal trees as well. The theory of large deviations (see, e.g., Dembo and Zeitouni, 1998) is at the heart of the analysis. We now introduce the necessary background. Let  $\Lambda = \Lambda_E$  be the cumulant generating function associated with the random variable  $E$ : for  $\lambda < 0$ , we have

$$\Lambda(\lambda) = \log \mathbf{E} e^{\lambda E}.$$

Then the left-tail Cramér function  $\Lambda^*$  is the Fenchel-Legendre dual of  $\Lambda$  (see Rockafellar, 1970), defined for  $\rho \leq \mathbf{EE}$  by

$$\Lambda^*(\rho) = \Lambda_E^*(\rho) = \sup\{\lambda\rho - \Lambda(\lambda) : \lambda \leq 0\}. \quad (1)$$

The function  $\Lambda^*$  characterizes large deviation tail probabilities for sums of i.i.d. variables distributed as  $E$ : Let  $\{E_i, i \geq 1\}$  be a family of i.i.d. random variables distributed as  $E$ , then Cramér's theorem (Cramér, 1938; Chernoff, 1952) states that for  $\rho \leq \mathbf{EE}$  (possibly infinite)

$$\mathbf{P}\left\{\sum_{i=1}^n E_i \leq \rho n\right\} = \exp(-n\Lambda^*(\rho) + o(n)), \quad (2)$$

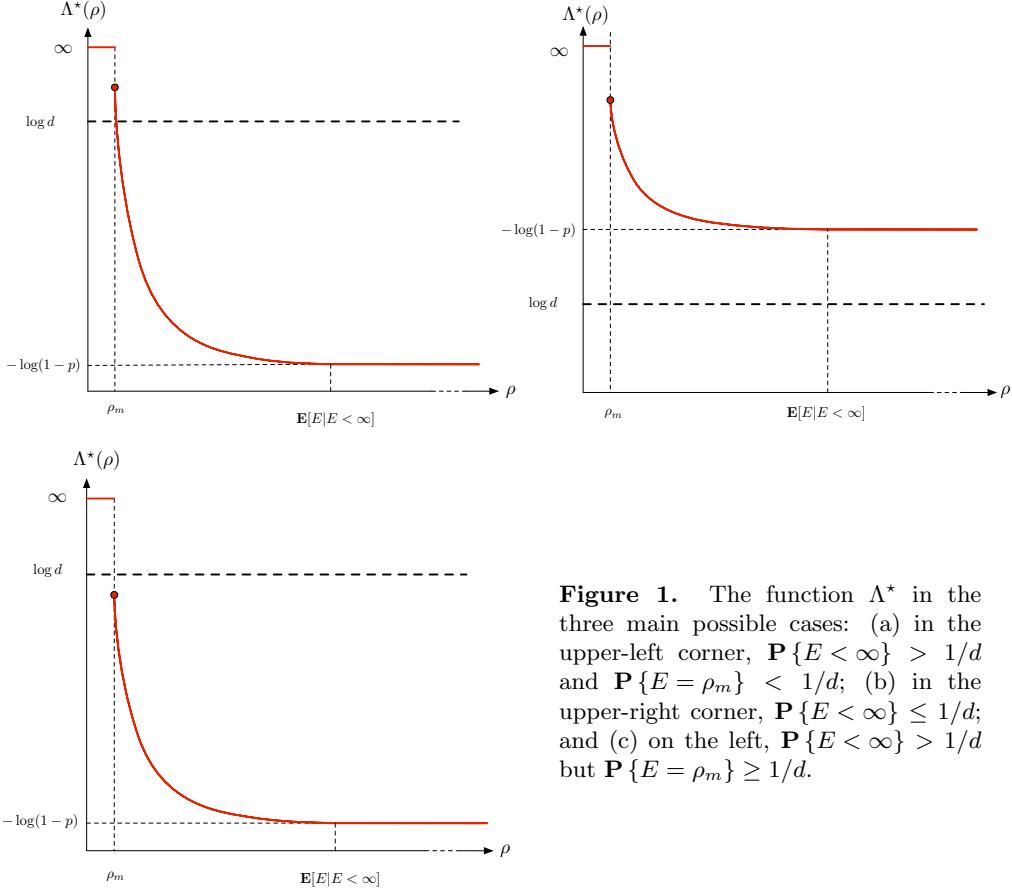
as  $n$  tends to infinity. For more information about the Cramér functions and large deviations, we refer the reader to Dembo and Zeitouni (1998) or Deuschel and Stroock (1989). In particular,  $\Lambda^*(\rho) \in [0, \infty]$  and is non-increasing in  $\rho$ . Trivially, since  $E \geq 0$ , we have  $\Lambda^*(\rho) = \infty$  for  $\rho < 0$ . Broutin and Devroye (2006) studied this very model. In particular, they showed that if  $\mathbf{EE} < \infty$ ,  $E$  takes values down to 0 (its essential infimum  $\inf\{\rho : \mathbf{P}\{E > \rho\} > 0\} = 0$ ), and  $\mathbf{P}\{E = 0\} = 0$ , then

$$H_n \sim c \log n \quad \text{in probability}, \quad (3)$$

as  $n \rightarrow \infty$ , where  $c$  is the unique solution of  $\Lambda^*(1/c) = \log d$  such that  $1/c \leq \mathbf{EE}$ .

The result of Broutin and Devroye has conditions that we would like to relax here. Neither of the two extensions is straightforward. The main issue is that, when relaxing the conditions, the equation  $\Lambda^*(1/c) = \log d$  may no longer have a solution (Figure 1). This happens in particular if either  $E$  has a fairly large mass at the left end of its support, or  $\mathbf{P}\{E = \infty\}$  is too large. Indeed, assume first that  $\rho_m = \inf\{\rho : \mathbf{P}\{E > \rho\} > 0\}$  and  $\mathbf{P}\{E = \rho_m\} = p > 1/d$ . Then, for all  $t \geq \rho_m$ ,  $\Lambda^*(t) \leq -\log p < \log d$  and clearly  $\Lambda^*(t) = \infty$  when  $t < \rho_m$ . On the other hand, if  $\mathbf{P}\{E < \infty\} = p < 1/d$  then, for all  $\rho$ ,  $\Lambda^*(\rho) \geq -\log p > \log d$  (see Figure 1). Therefore, going beyond such apparently pathological cases, requires to modify the definition of the constant  $c$ . We have

**Theorem 1.** *Let  $T_n$  be an ideal tree built from the split vector  $\mathcal{V}$ . Let  $H_n$  be its height. Assume that  $\mathbf{P}\{V > 0\} > 1/d$  and  $\mathbf{P}\{V = 1\} < 1/d$ . Then, for  $c = \sup\{1/\rho : \Lambda^*(\rho) \leq \log d\} < \infty$ , we have  $H_n = c \log n + o(\log n)$  in probability, as  $n \rightarrow \infty$ .*



**Figure 1.** The function  $\Lambda^*$  in the three main possible cases: (a) in the upper-left corner,  $\mathbf{P}\{E < \infty\} > 1/d$  and  $\mathbf{P}\{E = \rho_m\} < 1/d$ ; (b) in the upper-right corner,  $\mathbf{P}\{E < \infty\} \leq 1/d$ ; and (c) on the left,  $\mathbf{P}\{E < \infty\} > 1/d$  but  $\mathbf{P}\{E = \rho_m\} \geq 1/d$ .

Theorem 1 has been proved by Biggins (1977) in a slightly different form based on Laplace transforms. Note that enforcing  $\mathbf{P}\{V = 1\} < 1/d$  only excludes the trivial case in which the split vector  $\mathcal{V}$  is a uniform random permutation of  $(1, 0, \dots, 0)$ . In such a case,  $H_n = \infty$  for all  $n$ , and it happens that  $c = \infty$ . If  $\mathbf{P}\{V > 0\} \leq 1/d$ , then the set  $\{\rho \in \mathbb{R} : \Lambda^*(\rho) \leq \log d\} = \emptyset$  but  $\Lambda^*(\infty) = 0$ . So, at an intuitive level, an *informal* equivalent of Theorem 1 holds in those two extreme cases as well. By definition, if  $c$  is well-defined, it is clearly unique.

**Lemma 1.** Assume that  $\mathbf{P}\{E = 0\} < 1/d$  and  $\mathbf{P}\{E < \infty\} > 1/d$ . Then  $c = \sup\{1/\rho : \Lambda^*(\rho) \leq \log d\}$  is well-defined and  $c < \infty$ .

*Proof.* Set  $p = \mathbf{P}\{E = \infty\} = \mathbf{P}\{V = 0\}$ . Note that  $1-p = \mathbf{P}\{V > 0\} > 1/d$ . Introduce the random variable  $S$ , distributed as  $E$  conditional on  $E < \infty$ . It is possible that  $\mathbf{P}\{S = 0\} > 0$  and  $\mathbf{E}S = \infty$ . For  $\lambda < 0$ , we have  $\Lambda(\lambda) = \log \mathbf{E}e^{\lambda E} = \log(1-p) + \log \mathbf{E}e^{\lambda S}$ , and therefore,

$$\Lambda^*(\rho) = \Lambda_S^*(\rho) - \log(1-p). \quad (4)$$

Since  $S \in [0, \infty)$ , we have  $\inf_\rho \Lambda_S^*(\rho) = 0$  (Dembo and Zeitouni, 1998) and  $\inf_\rho \Lambda^*(\rho) = -\log(1-p) < \log d$ . As a consequence,  $c$  is well-defined. In other words, the case (b) in Figure 1 is forbidden.

We now show that there exists  $\delta > 0$  such that for all  $\rho \leq \delta$ ,  $\Lambda^*(\rho) > \log d$ . Since  $\Lambda^*$  is nonincreasing, this proves that  $c \leq 1/\delta < \infty$ . Let  $q = \mathbf{P}\{S = 0\}$ . We claim that

$$\liminf_{\rho \downarrow 0} \Lambda_S^*(\rho) \geq \begin{cases} -\log q & \text{if } q > 0; \\ \infty & \text{if } q = 0. \end{cases} \quad (5)$$

Clearly, taking (5) for granted, and using (4), either  $q = 0$  and  $\liminf_{\rho \downarrow 0} \Lambda^*(\rho) = \infty$ , or we have

$$\begin{aligned}\liminf_{\rho \downarrow 0} \Lambda^*(\rho) &\geq \liminf_{\rho \downarrow 0} \Lambda_S^*(\rho) - \log(1-p) \\ &\geq -\log \mathbf{P}\{S=0\} - \log \mathbf{P}\{E < \infty\} \\ &= -\log \mathbf{P}\{E=0\} > \log d,\end{aligned}$$

and the result holds as well. So it remains only to show (5). Let  $X$  be distributed as  $S$  conditional on  $S > 0$ . Then, for any  $\rho$ ,

$$\begin{aligned}\Lambda_S^*(\rho) &= \sup_{\lambda \leq 0} \{\rho\lambda - \log(q + (1-q)\mathbf{E}e^{\lambda X})\} \\ &\geq -\sqrt{\rho} - \log\left(q + (1-q)\mathbf{E}e^{-X/\sqrt{\rho}}\right),\end{aligned}$$

by taking  $\lambda = -1/\sqrt{\rho}$ . If  $q = 0$ , then by the dominated convergence theorem,  $\Lambda_S^*(\rho) \rightarrow \infty$ , as  $\rho \rightarrow 0$ . Otherwise,  $q > 0$  and, as  $\rho \rightarrow 0$ ,

$$\Lambda_S^*(\rho) \geq o(1) - \log q - \log\left(1 + \frac{1-q}{q} \cdot \mathbf{E}e^{-X/\sqrt{\rho}}\right) = o(1) - \log q,$$

by the dominated convergence theorem, as  $\mathbf{P}\{X > 0\} = 1$ .  $\square$

**EXAMPLE.** Consider a simple example to illustrate the computation of the constant  $c$ . Suppose  $d = 2$  and let the split vector  $\mathcal{V}$  be defined by

$$\mathcal{V} = \begin{cases} (0, 1) & \text{w.p. } p/2 \\ (1, 0) & \text{w.p. } p/2 \\ (1/2, 1/2) & \text{w.p. } 1-p, \end{cases}$$

where  $p \in (0, 1)$ . Observe that  $\mathbf{E}E = \infty$ . Also,  $\mu \stackrel{\text{def}}{=} \mathbf{E}[E | E < \infty] = \frac{1-p}{2-p} \cdot \log 4$ . We have

$$\Lambda^*(\rho) = \begin{cases} -\log(1-p/2) & \text{when } \mu < \rho < \infty, \\ \sup_{\lambda \leq 0} \{\lambda\rho - \log(2^\lambda(1-p) + \frac{p}{2})\} & \text{when } \rho \leq \mu. \end{cases}$$

The supremum above is achieved at

$$\lambda = \log_2\left(\frac{p}{2(1-p)(\rho^{-1}\log 2 - 1)}\right),$$

which is nonpositive for the given range of  $\rho$ .  $\Lambda^*(\rho), \rho > 0$  spans  $[-\log(1-p/2), \log(2/p)]$  and hence there exists a value  $c$  such that  $\Lambda^*(1/c) = \log 2$ . For  $\lambda$  defined above, we get the following implicit equation for  $c$ :

$$\frac{1}{c} \log_2\left(\frac{p}{2(1-p)(c\log 2 - 1)}\right) - \log\left(\frac{p}{c\log 2 - 1} + p\right) = 0.$$

For all values of  $p \in (0, 1)$ , the ideal tree with split vector  $\mathcal{V}$  has height  $H_n \sim c \log n$  in probability. In particular, one can show that  $c \rightarrow 1/\log 2$  as  $p \rightarrow 0$ . So, the shape of the tree approaches that of a complete binary tree on  $n$  nodes. Numerical values can be found in Table 1.

Note that along any path from leaf to root, we have  $\lfloor \log_2 n \rfloor$  internal nodes with two children corresponding to the  $(1/2, 1/2)$  split. All other nodes have  $(0, 1)$  type splits, and thus, by eliminating the child edge corresponding to zero, they can be viewed as one-child nodes. The number of consecutive one child nodes on a path is geometrically distributed with parameter  $1-p$ . Eliminating the one-child nodes, the height could thus equivalently have been studied by looking at binary trees with perfect splits in which the edges have independent geometrically distributed weights in the manner of Broutin and Devroye (2006).

$p$	1/3	1/5	1/10	1/20
$c$	4.328085123...	3.251524856...	2.579634951...	2.237874866...

Table 1.

### 3 A general model of random trees

Let  $T_\infty$  be a copy of the infinite complete  $d$ -ary tree. We place  $n$  items in the tree in some way and associate each item with one node of the tree. It is possible that several items are associated with the same node, although none of our examples cover this. The item distribution can equivalently be captured by the collection of cardinalities  $\{N_u, u \in T_\infty\}$ , where  $N_u$  is the number of items contained in the subtree rooted at  $u$ . We define the finite tree  $T_n$  consisting of all nodes with  $N_u > 0$ . We are interested in asymptotic properties of  $H_n$ , the height of  $T_n$ . The  $d$  children of  $u$  are denoted by  $u_1, u_2, \dots, u_d$ . Given  $N_u, N_u > 0$ , the *split* at  $u$  in  $T_\infty$  is the random vector  $(N_{u_1}/N_u, \dots, N_{u_d}/N_u)$ . When  $N_u = 0$ , we say that the split vector consists of all zeros. Our model is as follows:

- CONDITIONAL INDEPENDENCE. For every node  $u$ , the subtrees  $T_n(u_1), \dots, T_n(u_d)$  rooted at  $u_1, \dots, u_d$  are independent conditional on the sizes  $N_{u_1}, \dots, N_{u_d}$ .
- PERMUTATION INDEPENDENT DISTRIBUTION. If  $u_1, u_2, \dots, u_d$  are the children of  $u \in T_\infty$ , and  $\sigma$  is any permutation of  $\{1, \dots, d\}$ , then given  $N_u$ ,  $(N_{u_{\sigma(1)}}, \dots, N_{u_{\sigma(d)}})$  is distributed as  $(N_{u_1}, \dots, N_{u_d})$ .
- SIZE-DEPENDENT DISTRIBUTIONS. There exist random vectors  $\{\mathcal{V}^m, m \geq 0\}$  such that for all  $u$ , conditioned on  $N_u = m > 0$ ,  $(N_{u_1}/m, \dots, N_{u_d}/m)$  is distributed as  $\mathcal{V}^m = (V_1^n, \dots, V_d^n)$ . This implies that for all  $u \in T_\infty$ , and conditioned on  $N_u = m$ , the subtree  $T_n(u)$  rooted at  $u$  is distributed as  $T_m$ .
- CONVERGENCE IN DISTRIBUTION. We have  $\mathcal{V}^n \rightarrow \mathcal{V}$  in distribution as  $n \rightarrow \infty$ .
- BOUNDED HEIGHT. There exists a *deterministic* function  $\psi$  such that  $H_n \leq \psi(n)$  a.s. for all  $n \geq 0$ . For example, in most tree models, we have  $H_n \leq n$ . However, tries (Fredkin, 1960; Knuth, 1973) are an exception to this rule, since such a function does not exist.

The following theorem is our main result.

**Theorem 2.** *Let  $T_n$  be a random tree satisfying the above conditions with  $\mathcal{V}$  such that  $\mathbf{P}\{V > 0\} > 1/d$  and  $\mathbf{P}\{V = 1\} < 1/d$ . Let  $H_n$  be the height of  $T_n$ . Let  $c = \sup\{1/\rho : \Lambda^*(\rho) \leq \log d\}$ . Then  $H_n = c \log n + o(\log n)$  in probability, as  $n \rightarrow \infty$ .*

The lower and upper bounds are proved in Sections 4 and 5, respectively. Taken together, they prove Theorem 2.

### 4 Proof of the lower bound

As in the proofs for branching random walks or ideal trees, the argument is based on branching processes: One aims at building a surviving Galton–Watson tree (Devroye, 1987, 1998a). The key ideas can be traced back to Biggins (1977).

#### 4.1 Relying on ideal trees

Depths in  $T_n$  should not be far from those in ideal trees for large  $n$ , if all split vectors in  $T_n$  are replaced by their asymptotic counterparts. This is formalized in Lemma 2 below.

**Lemma 2.** Let  $\{T_n, n \geq 0\}$  be a family of random trees satisfying the conditions of the previous section, and let the limiting split vector be denoted by  $\mathcal{V}$ . Let  $\ell$  be a fixed integer. Let  $v_1, v_2, \dots, v_k$  be the  $k = d^\ell$  nodes at level  $\ell$  in  $T_\infty$ . Then

$$(N_{v_1}/n, N_{v_2}/n, \dots, N_{v_k}/n) \xrightarrow[n \rightarrow \infty]{} (L_{v_1}, L_{v_2}, \dots, L_{v_k})$$

in distribution, where the random variables  $L_u, u \in T_\infty$ , are as defined for the ideal trees.

*Proof.* The proof is based on a coupling argument: we find a coupled copy of the family  $\{T_n, n \geq 0\}$  such that  $(N_{v_1}/n, \dots, N_{v_k}/n) \rightarrow (L_{v_1}, \dots, L_{v_k})$  almost surely as  $n \rightarrow \infty$ . Recall that, by assumption,  $\mathcal{V}^n \rightarrow \mathcal{V}$  in distribution. Accordingly, by Skorohod's representation theorem (see, e.g., Billingsley, 1995), there exist copies of these random vectors such that the convergence holds almost surely. In the following, we let  $\mathcal{V}^n$  and  $\mathcal{V}$  denote these copies.

We now construct our coupled sequence of random trees  $T_n$ . To do so, assign independently to each node  $u \in T_\infty$  a copy of the sequence of vectors  $\{\mathcal{V}^m, m \geq 0\}$ . Given these sequences, we *deterministically* build a sequence  $\{T'_n, n \geq 0\}$  distributed as  $\{T_n, n \geq 0\}$ . More precisely, for all  $n$  and  $u \in T_\infty$ , we construct a new random variable  $N'_u$  distributed as the size  $N_u$  of the subtree  $T_n(u)$ . Let now  $n \geq 0$ . We proceed inductively on the distance from the root. If  $u$  is the root of  $T_\infty$ , then  $N'_u = n$ . Otherwise,  $u$  is the  $i$ -th child of some node  $v$ . Let  $\{V_i^m(v), m \geq 0\}$  be the sequence of the  $i$ -th components of the given vectors  $\{\mathcal{V}^m(v), m \geq 0\}$  at  $v$ . Assume that  $N'_v$  has already been defined. Then we set

$$N'_u = N'_v \cdot V_i^{N'_v}(v).$$

Observe that, although  $v$  has an infinite sequence  $\{\mathcal{V}^m(v), m \geq 0\}$ , the split vector to be used at  $v$ , namely  $\mathcal{V}^{N'_v}(v)$ , is selected at the time that  $N'_v$  is defined. Note also that  $N'_u$  is integer. Equivalently, we can write

$$N'_v = \sum_{m \geq 0} m \cdot V_i^m(u) \cdot \mathbf{1}[N'_u = m].$$

This finishes the construction of the tree  $T'_n$  characterized by  $\{N'_u, u \in T_\infty\}$ .

Since  $\mathcal{V}^n \rightarrow \mathcal{V}$  almost surely as  $n \rightarrow \infty$ , each node  $u \in T_\infty$  has actually a coupled copy of  $\mathcal{V}$  as well. These are used to obtain the coupled sequence  $\{L_u, u \in T_\infty\}$  as in the ideal tree. We now proceed to show that for all  $u \in T_\infty$ ,  $N'_u/n$  converges almost surely to  $L_u$ . Note that this implies that the random vector  $(N'_{v_1}/n, \dots, N'_{v_k}/n)$  converges a.s. to  $(L_{v_1}, \dots, L_{v_k})$ . Since, by construction,  $(N'_{v_1}/n, \dots, N'_{v_k}/n)$  is distributed as  $(N_{v_1}/n, \dots, N_{v_k}/n)$ , this is sufficient to prove the lemma.

Consider a set  $A$  of probability 1 on which  $\mathcal{V}^n \rightarrow \mathcal{V}$ . We prove by induction on the depth of  $u \in T_\infty$  that, for all  $\omega \in A$ ,  $N'_u(\omega)/n \rightarrow L_u(\omega)$  as  $n \rightarrow \infty$ . Fix  $\omega \in A$ . To simplify the notation, we drop the  $\omega$ 's but remember that, in fact, all values are *deterministic* and taken at  $\omega \in A$ . If  $u$  is the root, then  $N'_u/n = 1 = L_u$ . Let now  $u \in T_\infty$ , and consider  $u_i$  its  $i$ -th child. We assume that  $N'_u/n \rightarrow L_u$ , and we want to prove that  $N'_{u_i}/n \rightarrow L_{u_i}$ . There are two cases: if  $L_u = 0$ , then  $L_{u_i} = L_u \cdot V_i(u) = 0$ . Since  $N'_{u_i} \leq N'_u$ , it is clear that  $N'_{u_i}/n \rightarrow L_{u_i}$ . Otherwise,  $L_u > 0$ . Then by the induction hypothesis,  $N'_u \sim L_u n \rightarrow \infty$ , so

$$\frac{N'_{u_i}}{n} = \frac{N'_{u_i}}{N'_u} \cdot \frac{N'_u}{n} = V_i^{N'_u} \cdot \frac{N'_u}{n} \xrightarrow{n \rightarrow \infty} V_i(u) \cdot L_u = L_{u_i}.$$

This concludes the proof.  $\square$

## 4.2 Finding Galton–Watson processes

In the trees  $T_n$ , the splits used at different nodes do not have the same distribution. As a consequence, the construction of *proper* Galton–Watson processes based on  $\{N_u, u \in T_\infty\}$

is more intricate than in the models considered by Devroye (1998b) or Broutin and Devroye (2006). Lemma 3 provides us with the result we need to find Galton–Watson processes inside other branching processes.

**Lemma 3.** *Let  $N$  be a random positive integer. Let  $X$  be a random variable. Assume that, conditioning on  $N = n$ ,  $X$  is distributed like  $X_n$ , where, for all  $k$ ,*

$$\inf_n \mathbf{P} \{X_n \geq k\} \geq t_k$$

and  $t_k \downarrow 0$  as  $k \rightarrow \infty$ . Then one can find a random variable  $Y$  such that  $Y \leq X$  and, for all  $k$ ,

$$\mathbf{P} \{Y \geq k\} = t_k.$$

*Proof.* Let  $W$  be a random variable with tail distribution  $t_k$ :  $\mathbf{P} \{W \geq k\} = t_k$ . Let  $F_n$  be the distribution function of  $X_n$  and  $G$  be the distribution function of  $W$ . Let  $U$  be uniformly distributed on  $[0, 1]$ , then we couple  $W$  and all  $X_n$ 's in the Skorohod manner:

$$X_n = F_n^{-1}(U), \quad \text{and} \quad W = G^{-1}(U).$$

It is easy to see that with probability one,  $W \leq X_n$  for all  $n$ , and thus  $W \leq X$ .  $\square$

Lemma 4 below allows us to tune the branching processes arguments that we are about to use in the proof of Lemma 5.

**Lemma 4.** *Assume that  $\mathbf{P} \{V > 0\} > 1/d$  and  $\mathbf{P} \{V = 1\} < 1/d$ . Let  $c = \sup\{1/\rho : \Lambda^*(\rho) \leq \log d\}$ . For any  $\epsilon > 0$ , there exists  $\rho$  such that  $1/\rho > c - \epsilon$  and  $\Lambda^*(\rho) < \log d$ .*

*Proof.* Let  $\epsilon > 0$ . By definition of  $c$ , we can find  $\rho_0$  such that  $c - \epsilon/2 \leq 1/\rho_0 \leq c$  and  $\Lambda^*(\rho_0) \leq \log d$ . There exists an open ball  $\mathcal{B}$  centered at  $\rho_0$  such that, for all  $\rho \in \mathcal{B}$ ,  $1/\rho \geq c - \epsilon$ . Now,  $\Lambda^*$  is convex (Dembo and Zeitouni, 1998), and strictly convex except possibly when  $\Lambda^*(\rho) = \infty$  or its minimum value,  $-\log \mathbf{P} \{E \mid E < \infty\} < 1/d$ . As a consequence, there exists  $\rho \in \mathcal{B}$  such that  $\Lambda^*(\rho) < \log d$ .  $\square$

The essence of the lower bound argument is the following lemma.

**Lemma 5.** *Let  $H_n$  be the height of  $T_n$  defined as in Theorem 2. Assume that  $\mathbf{P} \{V > 0\} > 1/d$  and  $\mathbf{P} \{V = 1\} < 1/d$ . Let  $c = \sup\{1/\rho : \Lambda^*(\rho) \leq \log d\}$ . Then, for any  $\epsilon > 0$ , there exists  $n_0 \geq 0$  such that*

$$\inf_{n \geq n_0} \mathbf{P} \{H_n \geq (c - \epsilon) \log n\} > 0.$$

*Proof.* Let  $\epsilon > 0$  be arbitrary. We want to prove that we can find deep nodes in  $T_n$ . The hunt for such nodes is simplified by first finding nodes with similar properties in the corresponding ideal tree. So consider first ideal trees. Let  $V^\infty$  and  $E^\infty$  be the distributions of the variables associated with the ideal tree. For example, we will write  $E_e^\infty$  for the value of a random sample drawn from  $E^\infty$  and associated with edge  $e$ .

By Lemma 4, there exists  $\rho > 0$  such that  $c - \epsilon < 1/\rho$  and  $\Lambda^*(\rho) < \log d$ . Let such  $\rho$  now be fixed. A node  $u$  is called *ideally good* if either it is the root, or if lies  $\ell$  levels below an ideally good node  $v$  and  $L_u \geq L_v \cdot e^{-\rho\ell}$  (in the notation introduced in the section on ideal trees). The constant  $\ell$  is to be determined later. The collection of ideally good nodes form a Galton–Watson tree. Let  $W$  denote the progeny distribution of this process. While its distribution may be a bit complicated to describe, we know that

$$\mathbf{E}W = d^\ell \cdot \mathbf{P} \{L_u \geq L_v \cdot e^{-\rho\ell}\} = d^\ell \cdot \mathbf{P} \left\{ \sum_{e \in \pi(u,v)} E_e^\infty \leq \rho\ell \right\},$$

where  $\pi(u,v)$  is the set of edges on the path from  $u$  to  $v$ . By Cramér's theorem, we have

$$\mathbf{E}W = d^\ell \cdot e^{-\ell\Lambda^*(\rho)+o(\ell)} \rightarrow \infty$$

as  $\ell \rightarrow \infty$ . Accordingly, there exists  $\ell$  large enough that  $\mathbf{E}W > 1$ . Take such an  $\ell$  and fix it until the end of the proof.

We now move on to finding deep nodes in  $T_n$ . A node  $u \in T_\infty$  is called *good* if either  $u$  is the root, or  $u$  lies  $\ell$  levels below a good node  $v$  and  $N_u \geq N_v \cdot e^{-\rho\ell}$ . Let  $X_u$  be the number of good descendants of  $u$ ,  $\ell$  levels below  $u$ . Observe that unlike  $W$ , the distribution of  $X_u$  depends on  $u$ , and hence, the process consisting of good nodes is *not* a Galton–Watson process. However, Lemma 3 allows us to show that there exists a subset of the good nodes that is distributed as a supercritical Galton–Watson tree with progeny distribution  $Y$ . In the rest of the proof, we construct the random variable  $Y$  and the necessary coupled Galton–Watson process.

Consider  $\{v_i, 1 \leq i \leq d^\ell\}$ , the  $d^\ell$  nodes  $\ell$  levels below  $u$  in  $T_n$ . By Lemma 2, conditioning on  $N_u = n$ ,  $\{N_{v_i}/n, 1 \leq i \leq d^\ell\}$  converges to  $\{L_{v_i}/L_u, 1 \leq i \leq d^\ell\}$  in distribution, as  $n \rightarrow \infty$ . This implies in particular that

$$\liminf_{n \rightarrow \infty} \mathbf{P}\{X_u \geq t \mid N_u = n\} \geq \mathbf{P}\{W \geq t\},$$

for all  $0 \leq t \leq d^\ell$ . As a consequence, since  $\mathbf{E}W > 1$ , we can find  $M$  large enough such that for all  $n \geq M$ ,

$$\mathbf{P}\{X_u \geq t \mid N_u = n\} - \mathbf{P}\{W \geq t\} > \frac{1 - \mathbf{E}W}{2d^\ell}. \quad (6)$$

Consider the random variable  $Y$  such that for all  $t$ ,

$$\mathbf{P}\{Y \geq t\} = \max \left\{ \mathbf{P}\{W \geq t\} + \frac{1 - \mathbf{E}W}{2d^\ell}, 0 \right\}.$$

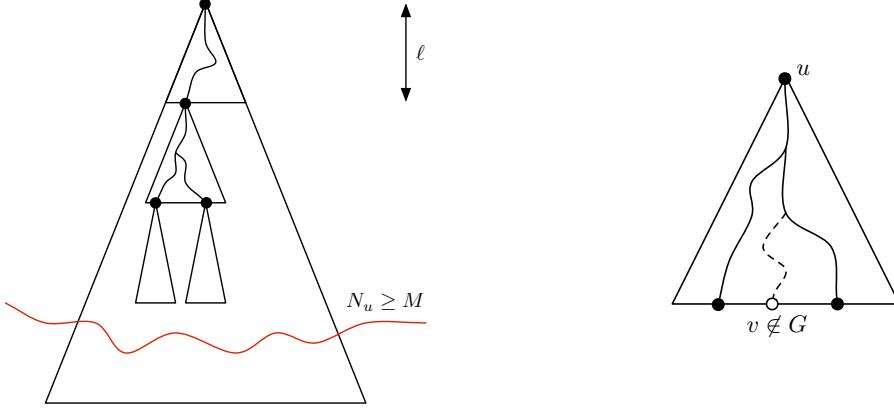
Then, we have

$$\begin{aligned} \mathbf{E}Y &= \sum_{t=1}^{d^\ell} \mathbf{P}\{Y_u \geq t\} \\ &\geq \sum_{t=1}^{d^\ell} \left( \mathbf{P}\{W \geq t\} + \frac{1 - \mathbf{E}W}{2d^\ell} \right) \\ &= \mathbf{E}W + \frac{1 - \mathbf{E}W}{2} \\ &= \frac{1 + \mathbf{E}W}{2} > 1. \end{aligned} \quad (7)$$

Consider a Galton–Watson tree with progeny distribution  $Y \leq d^\ell$ . Assume that this process finds its individuals in the layers  $k\ell$ ,  $k \geq 0$  of  $T_\infty$ . Let  $Y_u$  denote the copy of  $Y$  at a node  $u$ . By Lemma 3 and (6), there exists a coupling such that for all  $u \in \mathcal{L}_{k\ell}$ ,  $k \geq 0$ ,  $Y_u \leq X_u$  as long as  $N_u \geq M$ . Let  $G$  collect all nodes  $u$  from  $T_\infty$  that are also nodes at which the Galton–Watson process is alive. Recalling that  $\{X_u\}$  counts the good nodes, this implies that, in the coupled process, every node  $u \in G$  with  $N_u \geq M$  is guaranteed to be good (see Figure 2). By (7), the Galton–Watson process  $G$  is supercritical, and survives, therefore, with probability  $1 - q > 0$  (Harris, 1963; Athreya and Ney, 1972). This probability depends upon constants  $M, d$  and  $\ell$  which were already chosen earlier. If  $G$  survives, then look only at the nodes of  $G$  that also belong to  $\{u \in T_\infty : N_u \geq M\}$ . Either there is an infinite path of nodes  $u \in G$  satisfying  $N_u \geq M$ , or  $G$  contains a node  $u$  such that  $N_u < M$ . In the first case,  $H_n = \infty$  and the result holds. In the latter situation, let  $u \in G$  be of minimum depth  $k$  (as measured in  $T_\infty$ , not  $G$ ) such that  $N_u < M$ . Since  $u \in G$ , we have  $M > N_u \geq n \cdot e^{-\rho k}$ . It follows that

$$H_n \geq k \geq \frac{\log n - \log M}{\rho} \geq (c - \epsilon) \log n,$$

for  $n \geq n_0$  large enough. This proves that  $\inf_{n \geq n_0} \mathbf{P}\{H_n \geq (c - \epsilon) \log n\} \geq 1 - q > 0$ .  $\square$



**Figure 2.** On the left, the coupled Galton–Watson process  $G$ . On the right, an individual  $u$  such that  $N_u \geq M$  and its progeny (black dots). The white node  $v$  is good, but  $x \notin G$ . This illustrates the fact that  $X_u \geq Y_u$ , and that on the set  $\{u : N_u \geq M\}$ , the Galton–Watson process  $G$  provides a lower bound on the number of good nodes.

### 4.3 Boosting the probability up to $1 - o(1)$

Finally, we show the lower bound:

**Lemma 6.** *Consider random trees  $T_n$  as described in Section 3. Assume that  $\mathbf{P}\{V > 0\} > 1/d$  and  $\mathbf{P}\{V = 1\} < 1/d$ . Let  $c = \sup\{1/\rho : \Lambda^*(\rho) \leq \log d\}$ . For any  $\epsilon > 0$ , we have*

$$\mathbf{P}\{H_n \geq (c - \epsilon) \log n\} \xrightarrow[n \rightarrow \infty]{} 1.$$

*Proof.* We run multiple independent copies of the process used in the proof of Lemma 5, therefore multiplying the chances that one survives (Devroye, 1998a). This is achieved by starting the processes at level  $t$ , and using the conditional independence of disjoint subtrees. Since we may have  $V = 0$  with some positive probability  $p$ , we cannot ensure that all  $d^t$  nodes at level  $t$  are suitable as starting individuals. However, a fair number of them are.

To see this, we use a second branching process argument in the top  $t$  levels of  $T_\infty$ . More precisely, let  $Z_t$  denote the number of nodes  $u$  at level  $t$  in  $T_n$  such that  $N_u > ne^{-at}$ , where the value of  $a > 0$  is chosen such that  $\mathbf{P}\{V > e^{-a}\} > 1/d$  and  $e^{-a} < 1/d$ . Such a choice is possible since  $\mathbf{P}\{V > 0\} > 1/d$ . We claim that for any integer  $r$ , there exists a natural number  $t$  such that, for all  $n$  large enough,  $\mathbf{P}\{Z_t \leq r\} < 1/r$ .

To prove this fact consider  $T_\infty$  as an ideal tree and let  $Z_t^\infty$  denote the number of nodes  $u$  at level  $t$  such that  $L_u > e^{-at}$ . Define an auxiliary Galton–Watson process starting at the root in the following way: a node  $v$  is called *ideally nice* if it is the root, or if its parent is ideally nice and  $V_e > e^{-a}$ . The set of ideally nice nodes is clearly a Galton–Watson process. Since  $\sum_{i=1}^d V_i = 1$ , each individual has at least one child and the process survives with probability 1. Also, we have ... and so the process is not trivial. In particular, the number of elements of the process at level  $t$  tends to infinity in probability as  $t \rightarrow \infty$  (see, e.g., Athreya and Ney, 1972). Accordingly, there exists  $t_0$  such that the number of ideally nice nodes at level  $t_0$  is less than or equal to  $r$  with probability less than  $1/r$ . Since for any ideally nice node  $u$  at level  $t_0$ ,  $L_u > e^{-at}$ , this implies  $\mathbf{P}\{Z_{t_0}^\infty \leq r\} < 1/r$ .

But now, by Lemma 2,

$$\liminf_{n \rightarrow \infty} \mathbf{P}\{Z_{t_0} \leq r\} \leq \mathbf{P}\{Z_{t_0}^\infty \leq r\} < 1/r.$$

So, there exists  $n_1 = n_1(t_0, r)$  large enough such that for all  $n \geq n_1$ ,  $\mathbf{P}\{Z_{t_0} \leq r\} \leq 2/r$ . Assume now that  $Z_{t_0} > r$ . Let  $\{v_i, 1 \leq i \leq r\}$  be the first  $r$  nice individuals at level  $t_0$ . For

$1 \leq i \leq r$ , we have  $N_{v_i} \geq n \cdot e^{-at_0}$ . Conditioning on the sizes at level  $t$ , we can run the processes described in the proof of Lemma 5 independently in  $T_\infty(v_i)$ , the subtrees of  $T_\infty$  rooted at  $v_i$ ,  $1 \leq i \leq r$ . By Lemma 5, for all  $n$  large enough, in any of the  $r$  subtrees  $T_\infty(v_i)$ , with probability at least  $1 - q$ , there exists a node at depth at least  $(c - \epsilon/2)(\log n - at_0)$ . In such a case,  $H_n \geq (c - \epsilon/2)(\log n - at_0) + t_0 \geq (c - \epsilon)\log n$ , for  $n$  large enough.

Finally, still for  $n$  large enough,

$$\begin{aligned}\mathbf{P}\{H_n < (c - \epsilon)\log n\} &\leq \mathbf{P}\{Z_{t_0} \leq r\} + \mathbf{P}\{H_n < (c - \epsilon)\log n, Z_{t_0} > r\} \\ &\leq \frac{2}{r} + q^r.\end{aligned}$$

This can be made arbitrarily close to 0 by the choice of  $r$ . This concludes the proof of the lower bound.  $\square$

## 5 Proof of the upper bound

Let  $\epsilon > 0$  and pick  $\rho_0$  such that  $1/\rho_0 \in (c, c + \epsilon)$ . Since  $c = \sup\{1/\rho : \Lambda^*(\rho) \leq \log d\}$ , we have  $\Lambda^*(\rho_0) > \log d$ . Now by definition,  $\Lambda^*(\rho_0) = \sup\{\lambda\rho_0 - \Lambda(\lambda) : \lambda \leq 0\}$ , and hence there exists  $\lambda \leq 0$  such that  $\lambda\rho_0 - \Lambda(\lambda) > \log d$ . Let this particular  $\lambda$  be fixed in rest of the proof.

Let  $\mathcal{L}_t$  be the set of nodes at level  $t$  in  $T_\infty$ . Let  $u_t$  be the left-most node in  $\mathcal{L}_t$ . Let  $M > 0$  be any integer. Since  $H_M \leq \psi(M)$ , for  $H_n$  to be larger than  $(c + \epsilon)\log n$  there must be a node  $u$  in level  $\mathcal{L}_k$ , with  $k = \lfloor (c + \epsilon)\log n - \psi(M) \rfloor$ :

$$\mathbf{P}\{H_n \geq (c + \epsilon)\log n\} \leq \mathbf{P}\{\exists u \in \mathcal{L}_k : N_u \geq M\}.$$

Then, by the union bound over nodes in  $\mathcal{L}_k$ , and the symmetry condition on the split vectors,

$$\mathbf{P}\{H_n \geq (c + \epsilon)\log n\} \leq d^k \cdot \mathbf{P}\{N_{u_k} \geq M\}. \quad (8)$$

If  $N_{u_0} = n$ , then the random variable  $N_{u_k}$  is a product of random variables. However, they are neither independent nor identically distributed. We aim at bounding  $N_{u_k}$  by a product of i.i.d. random factors distributed as  $X_M$ , say.

We now construct the random variable  $X_M$ . Let  $F_n$  and  $F$  be the distribution functions of  $V^n$  and  $V$ , the typical components of  $\mathcal{V}^n$  and  $\mathcal{V}$ , respectively. By assumption,  $V^n \rightarrow V$  in distribution, so that for every point of continuity  $x$  of  $F$ , we have  $F_n(x) \rightarrow F(x)$  (see, for example, Billingsley, 1995). Then, for all such  $x$ , and  $M > 0$ ,

$$G_M(x) \stackrel{\text{def}}{=} \inf\{F_n(x) : n \geq M\} \xrightarrow[M \rightarrow \infty]{} F(x). \quad (9)$$

Since  $G_M$  is increasing, right-continuous, and bounded between 0 and 1, it is indeed the distribution function of some random variable  $X_M$ . Also, by (9),  $X_M$  converges in distribution to  $V$ . In particular, since  $0 \leq X_M \leq 1$ , and  $\lambda \leq 0$ , by the dominated convergence theorem,

$$\log \mathbf{E}X_M^{-\lambda} \rightarrow \log \mathbf{E}V^{-\lambda} = \Lambda(\lambda)$$

as  $M \rightarrow \infty$ . As a consequence, by our choice of  $\lambda$ , there exists  $M$  large enough such that  $\lambda\rho_0 - \log \mathbf{E}X_M^{-\lambda} - \log d > 0$ . Let this  $M$  be fixed until the end of the proof.

With the random variable  $X_M$  in hand, let us go back to bounding (8). With the value of  $M$  described above, we claim that

$$\mathbf{P}\{N_{u_k} \geq M\} \leq \mathbf{P}\left\{\prod_{i=1}^k X_M^{(i)} \geq \frac{M}{n}\right\}, \quad (10)$$

where  $X_M^{(i)}$ ,  $1 \leq i \leq k$ , are i.i.d. copies of  $X_M$ . To see this, we couple the random variables appearing on both sides of (10). Let  $U_1, \dots, U_k$  be i.i.d.  $[0, 1]$ -uniform random variables. Then, by the inverse transform technique,

$$X_M^{(i)} = G_M^{-1}(U_i), \quad 1 \leq i \leq k,$$

are i.i.d. copies of  $X_M$  (Grimmett and Stirzaker, 2001). We now construct the coupled copies of  $N_{u_0}, N_{u_1}, \dots, N_{u_k}$  inductively. We set  $N_{u_0} = n$  and define

$$N_{u_{i+1}} = N_{u_i} \cdot F_{N_{u_i}}^{-1}(U_{i+1}), \quad 0 \leq i \leq k-1.$$

By construction, conditional upon  $N_{u_i} = m$ ,  $N_{u_{i+1}}/m$  is distributed as  $V^m$  and  $N_{u_{i+1}} \in \mathbb{N}$  almost surely, as desired. Recall that, by (9), for  $m \geq M$ , we have  $F_m \geq G_M$ . This implies that, for all  $i$  such that  $0 \leq i \leq k-1$ ,

$$\frac{N_{u_{i+1}}}{N_{u_i}} = F_{N_{u_i}}^{-1}(U_{i+1}) \leq G_M^{-1}(U_{i+1}) = X_M^{(i)}.$$

Finally, on the set  $\{N_{u_k} \geq M\}$ ,

$$N_{u_k} = n \cdot \prod_{i=0}^{k-1} \frac{N_{u_{i+1}}}{N_{u_i}} \leq n \cdot \prod_{i=0}^{k-1} X_M^{(i)},$$

which proves (10).

Let  $Y = -\log X_M$ . Observe that for  $n$  large enough,  $\rho_0 k = \rho_0 \lfloor (c + \epsilon) \log n - \psi(M) \rfloor \geq \log n$ . Now, recalling (8), and using (10), we see that

$$\begin{aligned} \mathbf{P}\{H_n \geq (c + \epsilon) \log n\} &\leq d^k \cdot \mathbf{P}\left\{\prod_{i=1}^k X_M^{(i)} \geq \frac{M}{n}\right\} \\ &= d^k \cdot \mathbf{P}\left\{\sum_{i=1}^k -\log X_M^{(i)} \leq \log n - \log M\right\} \\ &\leq d^k \cdot \mathbf{P}\left\{\sum_{i=1}^k -\log X_M^{(i)} \leq \rho_0 k\right\}, \end{aligned}$$

for  $n$  large enough. Finally, by Chernoff's bound (Chernoff, 1952), and because  $\Lambda_Y^*(\rho_0) \geq \rho_0 \lambda - \log \mathbf{E} X_M^{-\lambda} > \log d$ ,

$$\mathbf{P}\{H_n \geq (c + \epsilon) \log n\} \leq e^{-k(\Lambda_Y^*(\rho) - \log d)} \xrightarrow[n \rightarrow \infty]{} 0.$$

This completes the proof of the upper bound.

## 6 Simple examples

RANDOMLY BALANCED BINARY SEARCH TREES. Suppose we are given distinct real numbers  $x_1, \dots, x_n$ . We build a random tree in the following way. Select a random sample of size  $m_n$ , where  $m_n \rightarrow \infty$  slowly enough that  $m_n$  is  $o(n)$ , and find the median  $M_n$  of this random sample. Split the nodes into two subtrees with  $M_n$  as the root, as one does in a binary search tree:  $\{x_i : x_i < M_n\}$  and  $\{x_i : x_i > M_n\}$  are sent to the left and right subtrees, respectively. Repeat recursively at all nodes. It is easy to show that the split for this random tree,  $(N_1/n, N_2/n)$ , where  $N_1$  and  $N_2$  are the sizes of the first and second subtrees, converges in probability to  $(1/2, 1/2)$  as  $n \rightarrow \infty$ . Thus, without any further work, we can conclude

from Theorem 2 that  $H_n \sim \log_2 n$  in probability. This yields a simple randomized way of asymptotically balancing a binary search tree, strengthening a result of Martínez and Roura (2001) who proved that the average depth is asymptotically  $\log_2 n$ .

**ALMOST PERFECT SPLITS.** As a further example, we consider the situation where the split vector of the random tree is close to  $(1/2, 1/2)$  with high probability, but can sometimes have a very unbalanced split. Let  $(X_n, Y_n, Z_n)$ ,  $n \geq 1$ , be a random sequence where  $0 \leq X_n, Y_n \leq 1/2$  and  $0 \leq Z_n \leq 1$  have a quite arbitrary dependence structure. We only assume that  $X_n \rightarrow 0$  and  $Y_n \rightarrow 0$  in probability, as  $n \rightarrow \infty$ . Suppose the split vector at a node  $u$  of size  $n$  is:

$$\begin{cases} (\lfloor nZ_n \rfloor, \lfloor n(1 - Z_n) \rfloor) & \text{w.p. } X_n \\ (\lfloor n(\frac{1}{2} + Y_n) \rfloor, \lfloor n(\frac{1}{2} - Y_n) \rfloor) & \text{w.p. } 1 - X_n. \end{cases}$$

We flip a perfect coin and exchange the components of the split vector to satisfy the permutation invariance constraint. The split vector converges in distribution to  $(1/2, 1/2)$ . This holds whatever the distribution of  $Z_n$ , even in the extreme case where  $Z_n = 1$  almost surely. Thus, as in the previous example,  $H_n \sim \log_2 n$  in probability.

## 7 Increasing trees

The varieties of increasing trees discussed in Section 1 have been introduced and thoroughly studied by Bergeron, Flajolet, and Salvy (1992), especially the polynomial families, with bounded maximum degree. We start by recalling the model of increasing trees and some of the important results. First of all, we must verify that the increasing trees of bounded maximum degree are indeed covered by our conditions. An increasing tree on  $n$  nodes has  $n$  items, one per node. Clearly, the bounded height condition holds. Furthermore, we can make a  $d$ -ary tree out of the ordered tree by adding, if necessary, empty subtrees to make all degrees  $d$ , and randomly permuting all children of all nodes. This creates the required symmetry. Given subtree sizes of siblings, the subtrees are indeed independent, and their distributions depend upon the subtree size only. Finally, as we will show in the proof of our main result, the crucial split vector convergence condition holds with a limiting Dirichlet split.

Bergeron, Flajolet, and Salvy (1992) showed that, in a random increasing tree sampled from any polynomial family of maximum degree  $d$ , the depth  $D_n$  of a typical node has mean  $\mu_n$  asymptotic to  $\frac{d}{d-1} \log n$  and variance  $\sigma_n^2 \sim \frac{d}{d-1} \log n$ . Furthermore, they showed that

$$\frac{D_n - \mu_n}{\sigma_n} \rightarrow \mathcal{N}(0; 1),$$

the standard Gaussian, in distribution, as  $n \rightarrow \infty$ . Depths in increasing tree families have also been studied by Panholzer and Prodinger (2004, 2006). Drmota (2006) showed that the expected height is  $O(\log n)$  and that, under some extra conditions, the variance of the height is  $O(1)$ , just as in the more usual model of binary search trees (Reed, 2000, 2003; Drmota, 2003). The fact that the first order asymptotics for the depths are identical for all polynomial varieties of outdegree  $d$  motivated this paper. We show that similarly, the first order limit behaviour for the height depends upon  $d$  only.

**Theorem 3.** *Let  $T_n$  be a random increasing tree of size  $n$  from the family described by  $\{s_r : 0 \leq r \leq d\}$ ,  $s_0 \neq 0$ ,  $s_d \neq 0$ . Let  $p$  be the period of the variety (the greatest common divisor of all indices  $i > 0$  for which  $s_i > 0$ ). Let  $H_n$  be the height of  $T_n$ . Then  $H_n \sim c_d \log n$  in probability, as  $n \rightarrow \infty$  along integers  $n$  that are  $1 \bmod p$ . The constant  $c_d$  is the unique solution in  $[1/(d-1), \infty)$  of*

$$(d-1)c_d \log \left( \frac{de}{(d-1)c_d} \right) = 1. \quad (11)$$

$d$	2	3	4	5	10	50	100
$c_d$	4.311...	3.540...	3.272...	3.137...	2.907...	2.753...	2.735...

**Table 2.** Some numerical values for  $c_d$ . Observe that  $\lim_{d \rightarrow \infty} c_d = e$ .

*Proof.* The main result of interest to us is the asymptotic enumeration of polynomial families of increasing trees provided by Bergeron et al. (1992). Let  $\delta = 1/(d-1)$ , and let  $Y_n$  be the number of increasing trees with  $n$  nodes in the variety described by  $\{s_r\}$ . Define

$$R_n = \frac{Y_n}{n! \rho^{-n} n^{\delta-1}} \quad \text{with} \quad \rho = \int_0^\infty \frac{dw}{\phi(w)}.$$

Then, Bergeron et al. (1992) proved that there exists  $m$  large enough such that if  $n \geq m$ , and if the limit is taken along values  $n = 1 \bmod p$ , we have

$$\lim_{n \rightarrow \infty} R_n = pC, \quad \text{and} \quad a \leq R_n \leq b, \quad (12)$$

for some constants  $0 < a \leq b < \infty$  and  $C > 0$ . If  $n$  does not satisfy  $n = 1 \bmod p$ , then  $R_n = Y_n = 0$ . The limit for  $R_n$  is the key to obtaining the height of random increasing trees.

From Theorem 2, we only need to compute the limit distribution of the split. We are particularly interested in integer-valued vectors  $(n_1, \dots, n_d)$  satisfying

$$n_i = 1 \bmod p, \quad \text{and} \quad \sum_{i=1}^d n_i = n - 1. \quad (13)$$

Equation (13) implies  $n = 1 + d \bmod p$ , so that, since  $d$  is a multiple of  $p$ , we have  $n = 1 \bmod p$ . The number of increasing trees such that the children of the root have (positive) sizes  $n_1, n_2, \dots, n_d$  is

$$s_d \cdot \binom{n-1}{n_1, n_2, \dots, n_d} \cdot \prod_{i=1}^d Y_{n_i} = s_d \cdot f(n_1, \dots, n_d) \cdot (n-1)! \cdot \rho^{1-n} \prod_{i=1}^d n_i^{\delta-1}. \quad (14)$$

Here, by (12), we have  $a^d \leq f(n_1, \dots, n_d) \leq b^d$  and  $f(n_1, \dots, n_d) \rightarrow p^d C^d$ , provided that (13) holds and  $\min\{n_i : 1 \leq i \leq d\} \rightarrow \infty$ .

Let  $X_1, X_2, \dots, X_d$  be the sizes of the  $d$  subtrees of the root of a random increasing tree on  $n$  nodes. Then, normalizing (14) by  $Y_n$  we obtain that

$$\mathbf{P}\{(X_1, \dots, X_d) = (n_1, \dots, n_d)\} = \rho \cdot s_d \cdot g(n_1, \dots, n_d) \cdot n^{1-d} \cdot \prod_{i=1}^d \left(\frac{n_i}{n}\right)^{\delta-1}, \quad (15)$$

where  $a^d/b \leq g(n_1, \dots, n_d) \leq b^d/a$ , for  $\min\{n_i : 1 \leq i \leq d\} \geq m$ , and  $g(n_1, \dots, n_d) \rightarrow p^{d-1} C^{d-1}$  if (13) holds and the  $\min\{n_i : 1 \leq i \leq d\} \rightarrow \infty$ . The upper bound on  $g$  holds if we merely have  $n \geq m$ . Observe that, for  $z = (z_1, \dots, z_d) \in (0, 1]^d$  and

$$S(z) = \left\{ (n_1, \dots, n_d) : n_i \geq nz_i, \quad n_i = 1 \bmod p, \quad \sum_{i=1}^d n_i = n - 1 \right\},$$

we have

$$\mathbf{P}\left\{\bigcap_{i=1}^d \left[\frac{X_i}{n} \geq z_i\right]\right\} = \rho \cdot s_d \cdot n^{1-d} \cdot \sum_{(n_1, \dots, n_d) \in S(z)} g(n_1, \dots, n_d) \cdot \prod_{i=1}^d \left(\frac{n_i}{n}\right)^{\delta-1}. \quad (16)$$

Equation (16) describes the distribution of the split whose asymptotics we are after. Let  $n = 1 \bmod p$ . Consider the  $(d-1)$ -dimensional grid of integers that are  $1 \bmod p$ : we place  $(n_1, \dots, n_{d-1})$  on this grid, and set  $n_d = n - 1 - \sum_{i=1}^{d-1} n_i$  (so that  $n_d = 1 \bmod p$  as well). We now bound the sum in the right-hand side of (16) from above and below by an integral in the standard manner. For the upper bound, consider the  $(d-1)$ -dimensional cube with side  $p/n$  having as its topmost point  $(n_1/n, \dots, n_{d-1}/n)$ . Then, setting

$$S'(z) = \left\{ (w_1, \dots, w_{d-1}) : w_i \geq z_i, \sum_{i=1}^{d-1} w_i \leq 1 \right\}, \quad (17)$$

for a continuous version of  $S(z)$ , and writing  $w_d = 1 - \sum_{i=1}^{d-1} w_i$  to save space, we have

$$\begin{aligned} & \sum_{S(z)} g(n_1, \dots, n_d) \cdot \prod_{i=1}^d \left( \frac{n_i}{n} \right)^{\delta-1} \\ & \leq \int_{S'(z-p/n)} g(n_1, \dots, n_d) \cdot \left( \frac{n}{p} \right)^{d-1} \cdot \left( w_d - \frac{(d-1)p}{n} \right)^{\delta-1} \cdot \prod_{i=1}^{d-1} w_i^{\delta-1} \cdot dw_i. \end{aligned}$$

Recall that  $g(n_1, \dots, n_d)$  is bounded, and  $g(n_1, \dots, n_d) \rightarrow (pC)^{d-1}$  if  $\min\{n_i : 1 \leq i \leq d\} \rightarrow \infty$  in such a way that (13) holds. Therefore,

$$\mathbf{P} \left\{ \bigcap_{i=1}^d \left[ \frac{X_i}{n} \geq z_i \right] \right\} \xrightarrow{n \rightarrow \infty} \rho \cdot s_d \cdot C^{d-1} \cdot \int_{S'(z)} w_d^{\delta-1} \cdot \prod_{i=1}^{d-1} w_i^{\delta-1} \cdot dw_i,$$

by the dominated convergence theorem. A matching lower bound is easily obtained in a similar way. We are now ready to prove the following lemma, giving the limit distribution we are looking for.

**Lemma 7.** *Let  $\mathcal{V} = (V_1, \dots, V_d)$  have the symmetric Dirichlet distribution with parameter  $\delta$ . Then,  $(X_1, \dots, X_d) \rightarrow \mathcal{V}$  in distribution as  $n \rightarrow \infty$ .*

*Proof.* Recall that by definition (Evans et al., 2000), for all  $z = (z_1, \dots, z_d) \in [0, 1]$ , and  $u = (w_1, \dots, w_{d-1})$ ,

$$\mathbf{P} \left\{ \bigcap_{i=1}^d [V_i \geq z_i] \right\} = D \cdot \int_{S'(z)} \left( 1 - \sum_{i=1}^{d-1} w_i \right)^{\delta-1} \cdot \prod_{i=1}^{d-1} w_i^{\delta-1} \cdot dw_i,$$

where  $D$  is a normalization constant depending upon  $\delta$  and  $d$  only, and  $S'(z)$  is given by (17).

It seems that we have convergence in distribution of  $(X_1/n, \dots, X_d/n)$  to  $(V_1, V_2, \dots, V_d)$ . However, we only know that the distribution functions are asymptotically proportional, as we did not bother with the constants—we did not show that  $D = \rho s_d C^{d-1}$ . To complete the proof, it suffices to show that

$$\lim_{\epsilon \downarrow 0} \mathbf{P} \left\{ \bigcap_{i=1}^d \left[ \frac{X_i}{n} \geq \epsilon \right] \right\} = 1.$$

As a side-product, this also shows that the probability that the root has fewer than  $d$  children is  $o(1)$ , and this is the reason why we did not bother with calculations for those cases. By symmetry, to show that  $\mathbf{P}\{\bigcup_{i=1}^d [X_i \leq \epsilon n]\} = o(1)$  it suffices to prove that  $\mathbf{P}\{X_1 \leq \epsilon n\} = o(1)$ . Let

$$S_1(\epsilon) \stackrel{\text{def}}{=} \left\{ (n_1, \dots, n_d) : n_1 \leq \epsilon n, n_i = 1 \bmod p, \sum_{i=1}^d n_i = n - 1 \right\}.$$

For  $n \geq m$ , we have

$$\begin{aligned} \mathbf{P}\{X_1 \leq \epsilon n\} &= \rho \cdot s_d \cdot n^{1-d} \cdot \sum_{(n_1, \dots, n_d) \in S_1(\epsilon)} g(n_1, \dots, n_d) \cdot \prod_{i=1}^d \left(\frac{n_i}{n}\right)^{\delta-1} \\ &\leq \rho \cdot s_d \cdot n^{1-d} \cdot \sum_{(n_1, \dots, n_d) \in S_1(\epsilon)} \frac{b^d}{a} \cdot \prod_{i=1}^d \left(\frac{n_i}{n}\right)^{\delta-1} \\ &\xrightarrow[n \rightarrow \infty]{} \rho \cdot s_d \cdot \frac{b^d}{a} \cdot p^{1-d} \cdot \int_{w_1 \leq \epsilon} \left(1 - \sum_{i=1}^{d-1} w_i\right)^{\delta-1} \cdot \prod_{i=1}^{d-1} w_i^{\delta-1} \cdot dw_i, \end{aligned}$$

by the integral comparison method used earlier. The last integral is  $o(1)$  as  $\epsilon \downarrow 0$ . This finishes the proof.  $\square$

With Lemma 7 in hand, it remains only to compute  $\Lambda^*$  and characterize  $c_d$ . Let  $V$  be a typical component of  $\mathcal{V}$  with the symmetric Dirichlet distribution with parameter  $\delta$ . It is well-known that  $V$  is beta  $(\delta, 1)$  (Evans et al., 2000). In fact,  $V$  is distributed as  $U^{1/\delta}$ , where  $U$  is uniform on  $[0, 1]$ . With  $E = -\log V$ , we have  $\mathbf{E}E = 1/\delta$ . For  $\lambda \leq 0$ , we have

$$\Lambda(\lambda) \stackrel{\text{def}}{=} \log \mathbf{E}[e^{\lambda E}] = \log \mathbf{E}[V^{-\lambda}] = \log \mathbf{E}[U^{-\lambda/\delta}] = \log\left(\frac{\delta}{\delta - \lambda}\right).$$

The function  $\lambda \mapsto \lambda t - \Lambda(\lambda)$  is maximized for  $\lambda = \delta - 1/t$ , and thus  $\Lambda^*(\rho) = \delta\rho - 1 - \log \delta - \log \rho$ . Since  $\mathbf{P}\{V = 0\} = \mathbf{P}\{V = 1\} = 0$ , there exists  $\rho$  such that  $\Lambda^*(\rho) = \log d$ . As a consequence, by Theorem 2, we have  $H_n = c_d \log n + o(\log n)$  in probability, as  $n \rightarrow \infty$  where  $c_d$  is the solution greater than  $\delta = 1/(d-1)$  of

$$\frac{1}{(d-1)c_d} - 1 + \log(d-1) + \log c_d = \log d,$$

which is equivalent to (11). This completes the proof.  $\square$

**RANDOM BINARY SEARCH TREE.** As we have seen in the introduction, this model is equivalent to a random increasing tree with characteristic sequence  $\{s_0 = 1, s_1 = 2, s_2 = 1\}$ . The height given by Theorem 3 matches the result of Devroye (1986), that is,  $H_n \sim 4.311\dots \log n$  in probability, as  $n \rightarrow \infty$ . Observe that in this case, the distribution of the split is exact and  $V$  is uniformly distributed on  $[0, 1]$ .

**INCREASING  $d$ -ARY POSITION TREE.** Consider the following growth process for  $d$ -ary position trees. An *external* node is a node with no children, and an *internal* node is a node with  $d$  children (of which some may be internal and some external). Furthermore, each node contains a time-stamp value that indicates the time at which it was added to the tree. We start the growth process with a single internal node (with time-stamp 1) which has  $d$  external children. At each time step, an external node is picked uniformly at random and replaced by an internal node with  $d$  external nodes—its potential children. The time-stamp value of the  $i$ -th internal node is exactly  $i$ . A random tree of size  $n$  built in this way is distributed as a random increasing tree taken from the variety associated with the sequence  $\{\binom{d}{i} : 0 \leq i \leq d\}$ . To see this, note first that the number of proper increasing labelings of a tree  $T$  with a fixed shape is exactly the number of ways  $T$  can be obtained by our process. Also, the growth process produces every properly labeled position tree with equal probability. It suffices now to observe that the number of trees  $T$  of a given shape and labeling in the increasing family is

$$\prod_{i=0}^d \binom{d}{i}^{D_i(T)},$$

where  $D_i(T)$  is the number of nodes of degree  $i$  in  $T$ . This turns out to be the number of embeddings of  $T$  in a  $d$ -ary position tree, and therefore we can consider the family of increasing trees as containing each embedding once. Since the number of labelings also matches up, a tree produced by the incremental growth process is a random increasing tree. Binary search trees correspond to the case  $d = 2$ . In some sense, the increasing trees associated with  $\phi(w) = (1 + w)^d$  are the typical increasing trees of degree  $d$  because they have a natural incremental construction. Since the children in this model are identified by their position between 1 and  $d$ , we have in effect created a random  $d$ -ary position tree (of internal nodes).

## 8 Increasing conditional Galton–Watson trees

There is an analogy between Galton–Watson trees conditional on their size and simply generated trees (see, e.g., Moon (1970), Kennedy (1975), or Kolchin (1980, 1986)). If  $T$  is a Galton–Watson tree with reproduction law  $\mu = \{\mu_n, n \geq 0\}$ , then for all integers  $n$ , the probability that, conditional on  $|T| = n$ ,  $T$  is equal to an ordered tree  $t$  of size  $n$  is proportional to

$$\prod_{r \geq 0} \mu_r^{D_r(t)}, \quad (18)$$

where  $D_r(t)$  is the number of nodes in  $t$  of outdegree  $r$ . This formula is the same as the one defining a random simply generated tree with defining sequence  $\sum_{r \leq 0} \mu_r w^r$ : for a certain unlabeled tree of shape  $t$ , the probability of selection is proportional to (18).

In the random increasing trees of the polynomial variety, we made the selection probability of an ordered tree of shape  $t$  equal to (18) times the number of possible increasing labelings of nodes of  $t$  using the integers  $1, \dots, |t|$ . Let us call the latter number  $\Psi(t)$ . While it will not matter in the discussion, it is easy to see that

$$\Psi(t) = \frac{n!}{\prod_{u \in t} |N_u|}.$$

We can formally extend this definition to size-conditioned Galton–Watson trees: select a tree of shape  $t$  with probability proportional to

$$\Psi(t) \times \prod_{r \geq 0} \mu_r^{D_r(t)}.$$

For reference below, we will call this the *increasing conditional Galton–Watson tree*. Here, of course, the  $\mu_r$  have no restrictions except that they define a probability distribution on  $\{0, 1, \dots, d\}$ . Equivalently, the probability of a given increasingly-labeled tree  $t$  is proportional to (18). Another equivalent way of thinking of these trees is via simulation: generate an unconditional random Galton–Watson tree  $T$  and label the nodes by a random equiprobable permutation of  $1, \dots, |T|$ . Repeat this until for the first time, a tree is obtained that has precisely size  $n$  and whose labeling is increasing. That tree has the desired distribution.

A curiosity of the increasing conditional Galton–Watson tree is that if  $\mu_r$  is replaced by  $\alpha \mu_r \theta^r$  for arbitrary constants  $\alpha > 0$  and  $\theta > 0$ , then (18) is rewritten as

$$\prod_{r \geq 0} \left( \alpha^{D_r(t)} \mu_r^{D_r(t)} \theta^{r D_r(t)} \right) = \alpha^n \theta^{n-1} \prod_{r \geq 0} \mu_r^{D_r(t)}.$$

In other words, increasing conditional Galton–Watson trees (as well as their special case, the increasing trees of Bergeron et al. (1992)) have distributions that do not depend upon

the actual values of  $\alpha$  and  $\theta$ . This fact was already known for standard conditional Galton–Watson trees (Kennedy, 1975). It carries over to the increasing labels model.

Our main theorem still applies to increasing conditional Galton–Watson trees, and, interestingly, the first-order asymptotics of the height do not depend upon the particular values of  $\mu_r$ , but only on the index  $d$  of the largest strictly positive value  $\mu_r$ . There is an analog asymptotic result on  $Y_n$  (see above Theorem 3). The difference is that then one no longer has the combinatorial interpretation of  $Y_n$  as the *number* of trees of size  $n$  in the variety, as it is replaced by a *weighting* of the tree of size  $n$ . To check its validity, follow the proof of Bergeron et al. (1992) step by step, and verify that everything goes through.

## 9 Digital search trees

Consider random tries over a finite alphabet  $\mathcal{A} = \{1, \dots, k\}$ , where each datum consists of an infinite string of i.i.d. symbols drawn from a fixed distribution on  $\mathcal{A}$  (Fredkin, 1960; Szpankowski, 2001). A string corresponds to an infinite path in the tree, with the  $i$ -th symbol having the index of the  $i$ -th child on the path. On each path, we mark the highest node visited by only one string and associate it with that string. The trie consists of these  $n$  marked nodes (leaves) and their paths to the root.

The probability of the  $i$ -th symbol is denoted by  $p_i$ . When there are  $n$  independent strings, we obtain the standard random trie. Its properties are well-known (Szpankowski, 2001): for example, if  $D_n$  is the depth of a random leaf, then

$$\frac{D_n}{\log n} \rightarrow \frac{1}{\mathcal{E}} \quad \text{in probability}$$

as  $n \rightarrow \infty$ , where

$$\mathcal{E} = \sum_{i=1}^k p_i \log \left( \frac{1}{p_i} \right)$$

is the entropy of the distribution  $\{p_i, 1 \leq i \leq k\}$  (Pittel, 1985). Also, if  $H_n$  denotes the height of the trie, then

$$\frac{H_n}{\log n} \rightarrow \frac{2}{Q} \quad \text{in probability}$$

as  $n \rightarrow \infty$ , where

$$Q = \log \left( \frac{1}{\sum_{i=1}^k p_i^2} \right)$$

(Pittel, 1985). From Jensen's inequality and  $(\max_{1 \leq i \leq k} p_i)^2 \leq \sum_{i=1}^k p_i^2 \leq \max_{1 \leq i \leq k} p_i$ , we have

$$\frac{1}{H} \leq \frac{1}{Q} \leq \frac{1}{\log \left( \frac{1}{\max_{1 \leq i \leq k} p_i} \right)} \leq \frac{2}{Q},$$

so that the height is always at least twice as big as the typical depth of a node.

In some applications, it is important to reduce the height. Attempts in this direction include PATRICIA trees (Morrison, 1968), and digital search trees (Coffman and Eve, 1970; Konheim and Newman, 1973). In both cases, we have

$$\frac{H_n}{\log n} \rightarrow \frac{1}{\log \left( \frac{1}{\max_{1 \leq i \leq k} p_i} \right)} \quad \text{in probability}$$

(Pittel, 1985). In other words, the height improves over that of the random trie, but not by more than 50%. Also, both PATRICIA trees and digital search trees introduce slight inconveniences: inorder traversal of digital search trees does not visit the nodes in sorted order, and internal edges of PATRICIA trees have cumbersome labels.

The term of *digital search tree* seems misleading to us, since digital search trees are not *search trees*, where a search query is carried over using the values stored in the nodes. We have kept the name digital search tree since it is standard. However, we prefer the term *pebbled tries*, to emphasize the trie structure: a string (a “pebble”) is assigned to each node in the tree instead of to each leaf. We may also assign up to  $b$  pebbles to each node. In the *pebbled trie*, we recursively split the collection of strings up as follows: at the root, where  $n$  strings have to split up, select one string uniformly at random and associate it with the root. Assign the  $n - 1$  remaining strings according to the values of their first symbol to the subtrees of the root. Recursively apply this rule for each node, but use the  $\ell$ -th symbol for a node at level  $\ell$  in the tree. In this manner, the height  $H_n$  of the pebbled trie is at most  $n - 1$ . The fact that we can bound the height by a deterministic function distinguishes pebbled tries from ordinary tries. It permits us to apply our results to pebbled tries. The vector of subtree sizes  $(N_1, \dots, N_k)$  at the root is multinomial  $(n - 1, p_1, p_2, \dots, p_k)$ .

Since  $H_n \leq n - 1$  and  $(N_1/n, \dots, N_k/n)$  tends in distribution to  $(p_1, \dots, p_k)$ , we can apply Theorem 1, after randomly permuting the split vector. Thus, the random variable  $V$  in Theorem 1 is distributed as  $p_Z$ , where  $Z$  is uniform in  $\{1, \dots, k\}$ . From trivial calculations, we deduce that

$$\frac{H_n}{\log n} \rightarrow \frac{1}{\log \left( \frac{1}{\max_{1 \leq i \leq k} p_i} \right)} \quad \text{in probability,}$$

just as for the more intricate PATRICIA trees. Note that Theorem 1 does not apply to the standard trie because  $H_n$  cannot be bounded in a deterministic manner by a function of  $n$ .

While Theorem 1 is only valid for trees with a finite branch factor, it is easy to verify that the above result remains valid for infinite branch factors, that is, pebbled tries in which the symbols take any positive integer value.

## 10 Concluding remarks

Using Theorem 2, we have obtained the height of increasing trees with bounded degree  $d$ . It is interesting to observe that the first order asymptotics do not depend on the sequence of values  $\{s_0, s_1, \dots, s_d\}$ . In particular, having ordered or unordered trees does not change the height. The case of  $d = \infty$ , that is, families with unbounded degree is fundamentally different. Indeed, uniform recursive trees and plane-oriented trees (Pittel, 1994) are special cases of this model, and their heights are  $e \log n$  and  $1.7956\dots \log n$ , respectively. Ordering nodes clearly modifies the height in this case.

Another model worthy of study is that of Prodinger and Urbanek (1983) in which increasing trees pick their labels from  $\{1, 2, \dots, \varphi(n)\}$ , with  $\varphi(n) < n$ , and ties between parent-child pairs are allowed.

## 11 Acknowledgment

We thank Michael Drmota for discussions regarding the random increasing tree model, and two anonymous referees for their careful reading.

## References

- K. B. Athreya and P. E. Ney. *Branching Processes*. Springer, Berlin, 1972.
- A.L. Barabási and R. Albert. Emergence of scaling in random network. *Science*, 286: 509–512, 1999.
- F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP*, volume 581 of *Lecture Notes in Computer Science*, pages 24–48, 1992.
- J. D. Biggins. The first and last-birth problems in a multitype age-dependent branching process. *Advances in Applied Probability*, 8:446–459, 1976.
- J. D. Biggins. Chernoff’s theorem in the branching random walk. *Journal of Applied Probability*, 14:630–636, 1977.
- J. D. Biggins and D. R. Grey. A note on the growth of random trees. *Statistics and Probability letters*, 32:339–342, 1997.
- P. Billingsley. *Probability and Measure*. Wiley, New York, 3rd edition, 1995.
- N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. *Algorithmica*, 46:271–297, 2006.
- H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–507, 1952.
- E. G. Coffman and J. Eve. File structures using hashing functions. *Communications of the ACM*, 13:427–436, 1970.
- H. Cramér. Sur un nouveau théorème-limite de la théorie des probabilités. In *Colloque Consacré à la Théorie des Probabilités*, volume 736, pages 5–23. Hermann, Paris, 1938.
- A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*. Springer Verlag, second edition, 1998.
- J.-D. Deuschel and D.W. Stroock. *Large Deviations*. American Mathematical Society, Providence, RI, 1989.
- L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33:489–498, 1986.
- L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24:277–298, 1987.
- L. Devroye. Branching processes and their application in the analysis of tree structures and tree algorithms. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Springer Series on Algorithms and Combinatorics*, pages 249–314, Berlin, 1998a. Springer.
- L. Devroye. Universal limit laws for depth in random trees. *SIAM Journal on Computing*, 28(2):409–432, 1998b.
- M. Drmota. An analytic approach to the height of binary search trees II. *Journal of the ACM*, 50:333–374, 2003.
- M. Drmota. The height of increasing trees. *Annals of Combinatorics*, 2006. Submitted.
- M. Evans, N.A.J. Hastings, and J.B. Peacock. *Statistical Distributions*. Wiley, New York, NY, third edition, 2000.

- E. Fredkin. Trie memory. *Communications of the ACM*, 3(9):490–499, 1960.
- G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, second edition, 2001.
- T. E. Harris. *The Theory of Branching Processes*. Springer, Berlin, 1963.
- S. Janson. Conditioned Galton–Watson trees do not grow. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees Combinatorics and Probability*, pages 331–334, 2006.
- D.P. Kennedy. The Galton–Watson process conditioned on the total progeny. *Journal of Applied Probability*, 12:800–806, 1975.
- D. E. Knuth. *The Art of Computer Programming: Sorting and Searching*, volume 3. Addison-Wesley, Reading, MA, 1973.
- V. F. Kolchin. Branching processes and random trees. In *Problems in Cybernetics, Combinatorial Analysis and Graph Theory (in Russian)*, pages 85–97, Moscow, Nauka, 1980.
- V. F. Kolchin. *Random Mappings*. Optimization Software, New York, 1986.
- A.G. Konheim and D.J. Newman. A note on growing binary trees. *Discrete Mathematics*, 4:57–63, 1973.
- M. Luczak and P. Winkler. Building uniformly random subtrees. *Random Structures and Algorithms*, 24:420–443, 2004.
- H. Mahmoud. Distances in plane-oriented recursive trees. *Journal of Computational and Applied Mathematics*, 41:237–245, 1992.
- H. Mahmoud, R. T. Smythe, and J. Szymanski. On the structure of random plane-oriented recursive trees and their branches. *Random Structures and Algorithms*, 4:151–176, 1993.
- C. Martínez and S. Roura. Optimal sampling strategies in quicksort and quickselect. *SIAM Journal on Computing*, 31:683–705, 2001.
- A. Meir and J.W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.
- J.W. Moon. *Counting Labelled Trees*. Number 1 in Canadian Mathematical Monographs. Canadian Mathematical Congress, Montreal, 1970.
- D. R. Morrison. PATRICIA — Practical Algorithm To Retrieve Information Coded in Alphanumeric. *Journal of the ACM*, 15:514–534, 1968.
- A. Panholzer and H. Prodinger. Analysis of some statistics for increasing tree families. *Discrete Mathematics and Theoretical Computer Science*, 6:437–460, 2004.
- A. Panholzer and H. Prodinger. The level of nodes in increasing trees revisited. *Random Structures and Algorithms*, 2006. To appear.
- B. Pittel. On growing random binary trees. *Journal of Mathematical Analysis and its applications*, 103:461–480, 1984.
- B. Pittel. Asymptotic growth of a class of random trees. *The Annals of Probability*, 13: 414–427, 1985.
- B. Pittel. Note on the height of random recursive trees and  $m$ -ary search trees. *Random Structures and Algorithms*, 5:337–347, 1994.

- H. Prodinger and F.J. Urbanek. On monotone functions of tree structures. *Discrete Applied Mathematics*, 5:222–239, 1983.
- B. Reed. How tall is a tree? In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 479–483, New York, NY, USA, 2000. ACM Press.
- B. Reed. The height of a random binary search tree. *Journal of the ACM*, 50:306–332, 2003.
- R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Theoretical Probability and Mathematical Statistics*, 51:1–27, 1995.
- W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley, New York, 2001.
- J. Szymański. On a non-uniform random recursive tree. In M. Karoński and Z. Palka, editors, *Random Graphs '85*, volume 33 of *Annals of Discrete Mathematics*, pages 297–306, Amsterdam, 1987. North Holland.