

Weighted height of random trees

N. Broutin L. Devroye E. McLeish*

January 3, 2008

Abstract

We consider a model of random trees similar to the split trees of Devroye [30] in which a set of items is recursively partitioned. Our model allows for more flexibility in the choice of the partitioning procedure, and has weighted edges. We prove that for this model, the height H_n of a random tree is asymptotic to $c \log n$ in probability for a constant c that is uniquely characterized in terms of multivariate large deviations rate functions. This extension permits us to obtain the height of pebbled tries, pebbled ternary search tries, d -ary pyramids, and to study geometric properties of partitions generated by k -d trees. The model also includes all polynomial families of increasing trees recently studied by Broutin, Devroye, McLeish, and de la Salle [17].

Keywords and phrases: Height, random tree, split tree, branching process, large deviations, probabilistic analysis.

1 Introduction

In this paper, we propose a general framework for finding heights of random trees with bounded degree d using branching processes. This extends previous work on the heights of a handful of random tree models [11, 14, 17, 27, 30]. Our model makes the connection between two different ways of constructing random trees: the split trees of Devroye [30] based on the partition of a set of items into recursive bins, and the branching process techniques of Broutin and Devroye [14] and generalizes both of them. The resulting framework encompasses all the kinds of trees that were captured by other branching process techniques such as binary search trees [27], random recursive trees [28, 56], plane oriented trees [56], median-of- $(2k + 1)$ trees [66], d -ary pyramids [11] and increasing trees [17].

The split tree model is a very natural way to build a random tree: a tree shall be seen as the outcome of a recursive process partitioning a finite set of items.

*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.

The process starts from the root with a number n of items. Some items remain stored at the root, the others proceed to further levels, N_i going to the i -th subtree for $1 \leq i \leq d$. The latter subsets of items are further partitioned in a recursive way. At each level, the items are partitioned according to some rule specifying the distribution of the integers N_i , $1 \leq i \leq d$. The rules considered by Devroye [30] are essentially mixtures of multinomials: each item that goes to further levels chooses a subtree independently of the others with some probability specified by some random variable whose distribution is fixed (but the value is independent for each node). Although this model is of great interest, as shown by the number of different cases it captures, it is still too constrained for some important applications. An example is the increasing tree of Bergeron, Flajolet, and Salvy [6]. The partitioning rules we allow here are very general, and the edges of the random trees we consider are weighted to allow for more flexibility. Trees constructed using this discrete model are called *weighted split trees*.

The second important case our study covers is the bidimensional model introduced by Broutin and Devroye [14], which is related to weighted branching processes studied by Biggins [9, 10]. Here, instead of recursively splitting a set of n discrete items, the tree represents nested partitions of the unit interval $[0, 1]$. In this sense, this is a continuous counterpart of the split trees of the previous paragraph. In this model, every node is associated with two random vectors, \mathcal{Z} , describing the lengths of the edges to the children, and \mathcal{V} , describing the sizes of the next level intervals. The process is stopped when all the intervals are smaller than some specified value. The model treated by Broutin and Devroye [14] has independent \mathcal{Z} and \mathcal{V} . We generalize it and consider trees in which \mathcal{Z} and \mathcal{V} may be dependent. Trees constructed using the continuous model are called *ideal trees*.

Although less natural than the weighted split trees, the ideal trees are easier to analyze. In particular, the height of a random ideal tree stopped when the intervals are all smaller than $1/n$ is asymptotic to $c \log n$ in probability. The constant c is uniquely characterized as the only solution of an (often implicit) equation involving large deviations rate functions. Although the model of ideal trees slightly generalizes previously known results, the main contribution of this paper concerns the more difficult model of weighted split trees. In analyzing this general discrete model, ideal trees appear to be a crucial intermediate object. In particular, we make the connection between a large class of weighted split trees and the continuous *ideal trees*. We prove that under some mild conditions, the heights of a random (discrete) split tree of size n and of a suitable ideal tree stopped at $1/n$ are asymptotically comparable, i.e., are both asymptotic to $c \log n$ in probability.

The paper is organized as follows. We first introduce an *ideal* model of trees and discuss asymptotic properties of their heights (Section 2). Random weighted split trees are described and studied in Section 3. We highlight the generality of our main result (Theorem 2) concerning the heights of such trees by giving various

applications in Section 5. All the proofs are based on theory of large deviations [24, 26] and branching processes [2].

2 An ideal model of weighted random trees

Let T_∞ be an infinite rooted d -ary tree (with d^k nodes at level k), and let r be its root. Let $\pi(u)$ be the set of edges on the unique path from a node u up to the root. We assign independently to each node of T_∞ a vector

$$((Z_1, V_1), (Z_2, V_2), \dots, (Z_d, V_d)),$$

where $V_i \geq 0$, $\sum_{i=1}^d V_i = 1$ and $Z_i \in [-\infty, \infty)$. We do not assume any independence between the V_i 's and the Z_i 's. If an edge e connects u with its i -th child, then, for convenience, we define $V_e = V_i$ and $Z_e = Z_i$.

THE SHAPE OF THE TREE. With each node $u \in T_\infty$ we can associate an interval of length L_u . For the root, we set $L_r = 1$. The children of a node u have intervals of lengths $L_u \cdot V_1, \dots, L_u \cdot V_d$ so that the total length $\sum_{i=1}^d L_u V_i = L_u$ is preserved. In this model, the sums of the lengths of the intervals at each level of T_∞ remain 1. The tree thus describes a random sequence of nested partitions. The length of the interval of a node u is $L_u = \prod_{e \in \pi(u)} V_e$. The *ideal tree* with parameter n , T_n , consists of the nodes $u \in T_\infty$ for which $L_u > 1/n$:

$$T_n = \{u \in T_\infty : L_u > 1/n\}.$$

Note that here, n is a parameter and the tree T_n does not have n nodes in general.

THE WEIGHTS. The Z_i 's represent edge lengths. More specifically, the lengths of the edges connecting u to its children are Z_1, \dots, Z_d . In some applications we may have negative values, and in general, the range of each extended random variable Z_i is $[-\infty, \infty)$. We define the *weighted depth* of a node $u \in T_\infty$ by $D_u = \sum_{e \in \pi(u)} Z_e$.

Alternatively, we can see the tree as a birth process. The root is born at time 0. The random vector of interest associated with a node u is then (E_1, \dots, E_d) , with $E_i = -\log V_i$, if $V_i > 0$; if $V_i = 0$, we define $E_i = \infty$. The variables E_i , $1 \leq i \leq d$, represent differences between the time of birth of u and the times of birth of its children. In other words, the time at which u is born is $B_u = \sum_{e \in \pi(u)} E_e$. Then, T_n consists of the nodes of T_∞ that are born before time $\log n$. We are interested in the *weighted height* H_n of T_n :

$$\begin{aligned} H_n &= \max\{D_u : u \in T_n\} \\ &= \max\{D_u : L_u > 1/n, u \in T_\infty\} \\ &= \max\{D_u : B_u < \log n, u \in T_\infty\}. \end{aligned}$$

Since we deal with heights, we may assume without loss of generality that the components of all vectors are identically distributed. Indeed, randomly permuting

them does not affect the height H_n . So, in the sequel, we write V , E and Z for the typical distributions of components of $((Z_1, V_1), \dots, (Z_d, V_d))$ or

$$\mathcal{X} = ((Z_1, E_1), \dots, (Z_d, E_d)), \quad (1)$$

and define $X = (Z, E)$.

THE ACTIVE PORTION OF THE TREE. In general, it is possible that for an edge e , $V_e = 0$, $E_e = \infty$ or $Z_e = -\infty$. This implies that for any u such that $e \in \pi(u)$, $L_u = 0$, $B_u = \infty$, or $D_u = -\infty$, respectively. So, in all these cases, the node u cannot contribute to the weighted height. Thus we call a node u *active* if $L_u > 0$, $B_u < \infty$ and $D_u > -\infty$. The *active portion* of the tree is largest subtree rooted at the root of T_∞ consisting of active nodes only. One should see active nodes as *participating*.

We are only interested in the ideal trees that satisfy some specific constraints. In particular, every node u should have at least one active child, and the active portion of the tree should not be a path. Also, we enforce some constraints on \mathcal{X} defined in (1) so that depths behave nicely in the active portion of the tree. This leads us to *valid split vectors*. Only ideal trees with such vectors will be considered. The definition requires moment conditions, and we first introduce the cumulant generating function Λ of $X = (Z, E)$ by

$$\Lambda(\lambda, \mu) = \log \mathbf{E} \left[e^{\lambda Z + \mu E} \mid Z > -\infty, E < \infty \right] + \mathbf{P} \{ Z > -\infty, E < \infty \} \quad \forall \lambda, \mu \in \mathbb{R}.$$

This definition allows to deal with Z and E taking infinite values, and matches with the usual cumulant generating function when $Z, E \in \mathbb{R}$ almost surely.

Definition 1. Consider $\mathcal{X} = ((Z_1, E_1), \dots, (Z_d, E_d))$, and write (Z, E) for the distribution of a typical component. We say that \mathcal{X} is a *valid split vector* if $Z \in [-\infty, \infty)$, $E \in [0, \infty]$, $\sum_{i=1}^d e^{-E_i} = 1$, and the following conditions hold:

- (a) $0 \in \mathcal{D}_\Lambda^\circ$, the interior of the domain $\mathcal{D}_\Lambda \stackrel{\text{def}}{=} \{(\lambda, \mu) : \Lambda(\lambda, \mu) < \infty\}$.
- (b) (Z_i, E_i) , $1 \leq i \leq d$, are identically distributed.
- (c) $\mathbf{P} \{\exists i : Z_i > -\infty, E_i < \infty\} = 1$ and $d\mathbf{P} \{Z > -\infty, E < \infty\} > 1$.
- (d) $\mathbf{E}[Z \mid Z > -\infty, E < \infty] \geq 0$ and $\mathbf{E}[E \mid Z > -\infty, E < \infty] > 0$.

Remark. If $\mathbf{P} \{Z > -\infty, E < \infty\} = 1$, then T_∞ is active. Then, we only need the range conditions, $0 \in \mathcal{D}_\Lambda^\circ$, $\mathbf{E}Z \geq 0$ and $\mathbf{E}E > 0$ for \mathcal{X} to be a valid split vector.

For ideal trees with a valid split vector \mathcal{X} , the first term of the asymptotic expansion of the weighted height can be characterized by an implicit equation involving the large deviations rate function $\Lambda^* = \Lambda_X^*$ associated with $X = (Z, E)$. The function Λ^* is the convex dual of Λ [24, 58]:

$$\Lambda^*(\alpha, \rho) = \sup_{\lambda, \mu} \{ \lambda \alpha + \mu \rho - \Lambda(\lambda, \mu) \} \quad \forall \alpha, \rho \in \mathbb{R}.$$

Theorem 1. Let T_n be an ideal tree with split vector \mathcal{X} and let H_n be its weighted height. Let Λ^* be the large deviations rate function of X , a typical component of \mathcal{X} . Then

$$H_n = c \log n + o(\log n)$$

in probability, as $n \rightarrow \infty$, where $c = \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$.

Remarks. (a) Note that for a valid split vector \mathcal{X} , $\rho = 0$ is never possible in the supremum defining c (see Lemma 1). This is the case every time we write such a supremum.

(b) If Z and E are independent and do not take infinite values, then $\Lambda^*(\alpha, \rho) = \Lambda_Z^*(\alpha) + \Lambda_E^*(\rho)$, where $\Lambda_Z^*(\alpha)$ and $\Lambda_E^*(\rho)$ are defined as the usual Fenchel–Legendre transforms of $\Lambda_Z(\lambda) = \log \mathbf{E}[e^{\lambda Z}]$ and $\Lambda_E(\mu) = \log \mathbf{E}[e^{\mu E}]$, respectively. Hence Theorem 2 agrees with the result of Broutin and Devroye [14] which claims that c is the maximal value of α/ρ in $\{\Lambda_Z^*(\alpha) + \Lambda_E^*(\rho) \leq \log d\}$. Actually, under their assumptions, the optimal value is attained at a point in $\{\Lambda_Z^*(\alpha) + \Lambda_E^*(\rho) = \log d\}$.

THE DEFINITION OF THE CONSTANT c . The level sets of Λ^* play an essential role in the properties of c . Let $\Psi(\ell) = \{(\alpha, \rho) \in \mathbb{R}^2 : \Lambda^*(\alpha, \rho) \leq \ell\}$. Then, we have

$$c = \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\} = \sup\{\alpha/\rho : (\alpha, \rho) \in \Psi(\log d)\}.$$

In the sequel, we let (Z^r, E^r) be distributed as (Z, E) conditional on the components being real, i.e., $\{Z > -\infty, E < \infty\}$. We first argue about the definition of c itself.

Lemma 1. Assume $0 \in \mathcal{D}_\Lambda^o$ and $p = \mathbf{P}\{Z > -\infty, E < \infty\} \geq 1/d$.

- (a) $(\mathbf{E}Z^r, \mathbf{E}E^r) \in \Psi(\log d) \neq \emptyset$ and $c = \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$ is well defined.
- (b) If furthermore, $\mathbf{E}Z^r \geq 0$ then $c \geq 0$.
- (c) If $\mathbf{P}\{E = 0\} < 1/d$, then there exists $\delta > 0$ such that

$$c \leq \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + \delta\} < \infty.$$

Proof. Note that since $0 \in \mathcal{D}_\Lambda^o$, by Lemma 15 (c), $(\mathbf{E}Z^r, \mathbf{E}E^r) \in \mathbb{R}^2$.

- (a) For any $\lambda, \mu \in \mathbb{R}$, by Jensen's inequality,

$$\Lambda(\lambda, \mu) = \log \mathbf{E}\left[e^{\lambda Z^r + \mu E^r}\right] + \log p \geq \lambda \mathbf{E}Z^r + \mu \mathbf{E}E^r + \log p.$$

It follows that $\lambda \mathbf{E}Z^r + \mu \mathbf{E}E^r - \Lambda(\lambda, \mu) \leq -\log p$ and thus, $\Lambda^*(\mathbf{E}Z^r, \mathbf{E}E^r) \leq -\log p$. Since $p \geq 1/d$, $(\mathbf{E}Z^r, \mathbf{E}E^r) \in \{(\alpha, \rho) : \Lambda^*(\alpha, \rho) \leq \log d\} \neq \emptyset$.

- (b) If $\mathbf{E}Z^r \geq 0$ we have $c \geq \mathbf{E}Z^r/\mathbf{E}E^r \geq 0$, potentially infinite if $\mathbf{E}E^r = 0$.

(c) For all $\delta > 0$, $c \leq \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + \delta\}$, so we need only prove that the right-hand side is finite for some $\delta > 0$. Since $\mathbf{P}\{E = 0\} < 1/d$, we can pick $\delta > 0$ such that $\mathbf{P}\{E = 0\} < e^{-\delta}/d$. By Lemma 15 (c), Λ^* is a good rate function (see Appendix B), and hence the level sets $\Psi(\cdot)$ are compact. As a consequence, it suffices to prove that $\{\rho = 0\} \cap \Psi(\log d + \delta) = \emptyset$. We show that $\inf_{\alpha \in \mathbb{R}} \liminf_{\rho \downarrow 0} \Lambda^*(\alpha, \rho) \geq \log d + \delta$, which would prove the claim. For all $\alpha, \rho \in \mathbb{R}$,

$$\Lambda^*(\alpha, \rho) = \sup_{\lambda, \mu \in \mathbb{R}} \{\lambda \alpha + \mu \rho - \Lambda(\lambda, \mu)\} \geq \sup_{\mu \in \mathbb{R}} \{\mu \rho - \Lambda(0, \mu)\} = \Lambda_{E^r}^*(\rho) - \log p.$$

Let $q = \mathbf{P}\{E^r = 0\}$. Then,

$$\Lambda_{E^r}(\mu) = \log \mathbf{E} [e^{\mu E^r}] = \log (q + (1 - q) \cdot \mathbf{E} [e^{\mu E^r} \mid E^r > 0]).$$

For $\rho_n \downarrow 0$,

$$\begin{aligned} \Lambda_{E^r}^*(\rho_n) &= \sup_{\mu \in \mathbb{R}} \{ \mu \rho_n - \log (q + (1 - q) \mathbf{E} [e^{\mu E^r} \mid E^r > 0]) \} \\ &\geq -\sqrt{\rho_n} - \log \left(q + (1 - q) \mathbf{E} \left[e^{-E^r/\sqrt{\rho_n}} \mid E^r > 0 \right] \right) \rightarrow -\log q. \end{aligned}$$

So $\liminf_{n \rightarrow \infty} \Lambda_{E^r}^*(\rho_n) \geq -\log q$. Therefore, for any $\alpha \in \mathbb{R}$,

$$\liminf_{\rho \downarrow 0} \Lambda^*(\alpha, \rho) \geq -\log \mathbf{P}\{E^r = 0\} - \log p = -\log \mathbf{P}\{E = 0\} > \log d + \delta,$$

and $\inf_{\alpha} \liminf_{\rho \downarrow 0} \Lambda^*(\alpha, \rho) \geq \log d + \delta$, which completes the proof. \square

A GEOMETRIC INTERPRETATION. Observe that in a diagram of α versus ρ , α/ρ is the slope of the line connecting the origin with the point (α, ρ) . In such a diagram, $\Psi(\log d)$ is a compact convex set by Lemma 15 (see Appendix). Then, if one imagines that $\Psi(\log d)$ is embossed, c is the just the slope of the line with a joint at the origin that would be dropped from the vertical at the origin. This is illustrated by Figure 1.

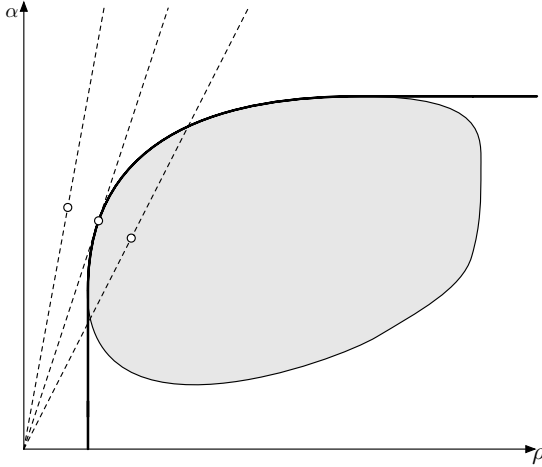


Figure 1. Typical level sets for Λ^* and I are shown. The shaded region is the set $\Psi(\log d) = \{\Lambda^*(\alpha, \rho) \leq \log d\}$. The thick line is the border of $\{I(\alpha, \rho) \leq \log d\}$. We also show three points together with the lines of interest. The steepest is used for the upper bound, the most gentle for the lower bound. The intermediate one is the optimal line.

3 Random weighted split trees

3.1 An embedding to construct random trees

In ideal trees, all nodes receive an independent copy of the *same* random vector, and the description of a finite tree is done in a very natural way by pruning the branches of T_∞ . However, ideal trees do not cover some important kinds of random trees.

Just consider trees for which the distribution of the split vector varies, like most increasing trees [6, 17]. Moreover, the size of an ideal tree is random in general. We now propose a model that is not subject to these constraints, and hence captures many more applications. As stated in the introduction, our model is related to the split trees of Devroye [30] and a tree of size n is seen as a recursive partition of a set of n items.

Random trees can be constructed using a variety of methods, also called embeddings. We propose an embedding which emphasizes an underlying structure consisting of independent random variables. Many important brands of random trees can be captured by this model. Some examples are presented in Section 5.

DIFFERENT TYPES OF NODES. Consider a family $\{\mathcal{X}^m, m \geq 0\}$ of random vectors, where $\mathcal{X}^m = ((Z_1^m, E_1^m), \dots, (Z_d^m, E_d^m))$. Assume that for all m , and $1 \leq i \leq d$, $m \exp(-E_i^m)$ is almost surely integer-valued, and $E_i^m \geq 0$. Assign independently a copy of the sequence $\{\mathcal{X}^m, m \geq 0\}$ to each one of the nodes of an infinite d -ary tree T_∞ . The different elements of the sequence $\{\mathcal{X}^m, m \geq 0\}$ provide sufficient flexibility.

BUILDING A RANDOM TREE ON n ITEMS. Given an integer n and the copies of $\{\mathcal{X}^m, m \geq 0\}$, we build a sequence $\{(D_u, B_u), u \in T_\infty\}$ of weighted depths and birth times of the nodes of T_∞ . Observe that although the dependence is not explicitly written, $\{(D_u, B_u), u \in T_\infty\}$ depends on n . The construction is made easier by using the auxiliary sequence $\{N_u, u \in T_\infty\}$, where N_u is the *cardinality* of a node u , that is the number of items in its subtree. Let $n \geq 0$ and consider $((Z_1^n, E_1^n), \dots, (Z_d^n, E_d^n))$, the copy of \mathcal{X}^n at the root of T_∞ . The children u_1, \dots, u_d of the root are assigned cardinalities $N_{u_i} = n \exp(-E_i^n) \in \mathbb{N}$, $1 \leq i \leq d$. Given the values of N_{u_1}, \dots, N_{u_d} , the sequences $\{N_v : v \in T_\infty(u_i)\}$, $1 \leq i \leq d$, describing the trees rooted at u_i , $1 \leq i \leq d$ are recursively built in the same way, unless $1 \leq N_{u_i} \leq b$ or $N_{u_i} = 0$. Here b is the number of items that a node can contain.

Using $\{N_u, u \in T_\infty\}$, and the copies of $\{\mathcal{X}^m, m \geq 0\}$, we now assign random variables (Z_e, E_e) to the edges of T_∞ . Let e be the i -th edge out of a node $u \in T_\infty$. We set

$$Z_e = Z_i^{N_u} \quad \text{and} \quad E_e = E_i^{N_u}.$$

Recall that $\pi(u)$ denotes the set of edges on the path from u up to the root in T_∞ . As for the case of ideal trees, we define the weighted depth of a node u , $D_u = \sum_{e \in \pi(u)} Z_e$ and the birth time of a node u , $B_u = \sum_{e \in \pi(u)} E_e$. This finishes the construction of $\{(D_u, B_u), u \in T_\infty\}$ which fully describes our random weighted tree. Then we have

$$T_n \stackrel{\text{def}}{=} \{u \in T_\infty : N_u > 0\} = \{u \in T_\infty : B_u < \log n\}.$$

We are interested in the weighted height $H_n = \max\{D_u : u \in T_n\}$ of the random tree T_n . Again, it is sufficient to consider the trees for which, for each n , the

components of \mathcal{X}^n are identically distributed. We say that a random tree is *valid* if it can be constructed by the above process and the following conditions hold:

- PERMUTATION INVARIANCE. For any integer n , and any permutation σ , we have

$$((Z_{\sigma(1)}^n, E_{\sigma(1)}^n), \dots, (Z_{\sigma(d)}^n, E_{\sigma(d)}^n)) = ((Z_1^n, E_1^n), \dots, (Z_d^n, E_d^n)) \quad \text{in distribution.}$$

- CONVERGENCE. There exists a valid split vector \mathcal{X}^∞ (see Definition 1) such that the cumulant generating functions of the *vectors* \mathcal{X}^n and \mathcal{X}^∞ satisfy $\Lambda_{\mathcal{X}^n} \rightarrow \Lambda_{\mathcal{X}^\infty} \leq \infty$ everywhere as $n \rightarrow \infty$ and $0 \in \mathcal{D}_{\Lambda_{\mathcal{X}^\infty}}^o$.
- BOUNDED HEIGHT. There exists a deterministic function ψ such that for all n , $H_n \leq \psi(n)$.

Remarks. (a) Observe that since $0 \in \mathcal{D}_{\Lambda^\infty}^o$, $\Lambda_{\mathcal{X}} \rightarrow \Lambda_{\mathcal{X}^\infty}$ implies that $\mathcal{X} \rightarrow \mathcal{X}^\infty$ in distribution [see 12, p. 390].

(b) We can slightly relax the constraint that the height be bounded. For instance, subexponential tails for the height would suffice: for all $M \geq 1$, there exists a function f with $f(t)/t \rightarrow \infty$ as $t \rightarrow \infty$ such that

$$\sup_{n \leq M} \left\{ e^{f(t)} \cdot \mathbf{P} \{H_n \geq t\} \right\} \leq 1.$$

Ordinary tries violate this condition, and are be treated separately [15, 16].

Lemma 2. *Let T_n be a valid random tree as defined above. Then we have:*

- CONDITIONAL INDEPENDENCE. *For any node u , the σ -algebras generated by the variables associated with edges in the subtrees rooted at the children u_1, \dots, u_d are independent, conditioned on the sizes N_{u_1}, \dots, N_{u_d} .*
- SIZE-DEPENDENT DISTRIBUTION. *Conditioning on $N_u = k$, the subtree rooted at u , $T_n(u)$, is distributed as T_k .*

As in the case of ideal trees, the height may be characterized using large deviations functions. The height of T_n and of an ideal tree with split vector \mathcal{X}^∞ are asymptotically comparable in probability. The main result of this paper is the following theorem.

Theorem 2. *Let \mathcal{X}^∞ be a valid split vector. Let T_n be a valid random tree with limit split vector \mathcal{X}^∞ and let H_n be its weighted height. Then,*

$$H_n = c \log n + o(\log n)$$

in probability, as $n \rightarrow \infty$ where $c = \sup\{\alpha/\rho : \Lambda^(\alpha, \rho) \leq \log d\}$ and Λ^* is the large deviations rate function for X^∞ , a typical component of \mathcal{X}^∞ .*

The heights of many known trees fall within the scope of Theorem 2. These include binary search trees [27], bounded degree increasing trees [6, 17], random recursive trees [28, 56], plane-oriented trees [56], scale-free trees [3, 56], pyramids [11, 50], and most models captured by the less general result of Broutin and Devroye [14]. Many applications are treated in Section 5. For more applications see [13].

3.2 Relying on ideal trees

In the proof of Theorem 2 we approximate depths in random trees by those in a suitable ideal tree that is easier to deal with. The approximation is based on a coupling of the random trees with the ideal tree with split vector \mathcal{X}^∞ .

By assumption, $\mathcal{X}^n \rightarrow \mathcal{X}^\infty$ in distribution as $n \rightarrow \infty$, hence by Skorohod's theorem [see, e.g., 12], we can find a coupling for which the convergence holds almost surely. In the following, we let \mathcal{X}^n be the copies of the random variables such that $\mathcal{X}^n \rightarrow \mathcal{X}^\infty$ almost surely. If we use copies of this coupled sequence $\{\mathcal{X}^m, m \geq 0\}$ to build the random trees, we obtain a coupled sequence $\{T_n, n \geq 0\}$. Since the convergence of \mathcal{X}^n to \mathcal{X}^∞ is almost sure, each node has a copy of \mathcal{X}^∞ as well. These copies, in turn, define a proper ideal tree with split vector \mathcal{X}^∞ . This latter tree is called the *ideal tree associated* with the coupled sequence $\{T_n, n \geq 0\}$.

Lemma 3. *Consider the coupled sequence of random trees $\{T_n, n \geq 0\}$, and the associated ideal tree. Let ℓ be a fixed positive integer. Let \mathcal{L}_ℓ be the set of nodes at level ℓ in T_∞ . Then, as $n \rightarrow \infty$,*

$$\{(D_v, B_v) : v \in \mathcal{L}_\ell\} \rightarrow \{(D_v^\infty, B_v^\infty) : v \in \mathcal{L}_\ell\}$$

a.s., where D_v^∞ and B_v^∞ are the weighted depth and birth time of v in an ideal tree with valid split vector \mathcal{X}^∞ .

Proof. Since $\mathcal{X}^n \rightarrow \mathcal{X}^\infty$ a.s., each node has an independent copy of the limit as well. These limit random variables are used to define $\{(D_u^\infty, B_u^\infty), u \in T_\infty\}$, which characterizes fully a coupled *ideal tree*. Assume for now that, for all $u \in T_\infty$, and some coupled random variables (D'_u, B'_u) is distributed as (D_u, B_u) ,

$$(D'_u, B'_u) \xrightarrow[n \rightarrow \infty]{} (D_u^\infty, B_u^\infty) \quad \text{almost surely.} \quad (2)$$

Then $\{(D'_v, B'_v) : v \in \mathcal{L}_\ell\} \rightarrow \{(D_v^\infty, B_v^\infty) : v \in \mathcal{L}_\ell\}$ a.s., as $n \rightarrow \infty$. Therefore, to prove the lemma, it suffices to show that (2) holds for all $u \in T_\infty$.

Let A be a set of probability 1 on which, for all u , $\mathcal{X}_u^n \rightarrow \mathcal{X}_u^\infty$. We prove by induction on the (unweighted) depth that

$$\forall \omega \in A, \quad (D'_u(\omega), B'_u(\omega)) \rightarrow (D_u^\infty(\omega), B_u^\infty(\omega)).$$

For the sake of simplicity, we drop the ω and simply write (D'_u, B'_u) and (D_u^∞, B_u^∞) , remembering that, in fact, these values are measurable functions of ω . If u is the root, then $(D'_u, B'_u) = (0, 0) = (D_u^\infty, B_u^\infty)$. Otherwise, u is the i -th child of some node v . The induction hypothesis tells us that $(D'_v, B'_v) \rightarrow (D_v^\infty, B_v^\infty)$ as $n \rightarrow \infty$. Let the components of \mathcal{X}^∞ be (Z_i^∞, E_i^∞) , $1 \leq i \leq d$. Assume first that $B_v^\infty = \infty$, then $B_u^\infty = B_v^\infty + E_i^\infty(v) = \infty$. As $B'_u \geq B'_v$, it follows that $B'_u \rightarrow B_u^\infty$. If $B_v^\infty < \infty$, we have $N'_v = n \exp(-B'_v) \sim n \exp(-B_v^\infty) \rightarrow \infty$ as $n \rightarrow \infty$. As a consequence,

$$\begin{aligned} D'_u &= D'_v + Z_i^{N'_v}(u) \xrightarrow[n \rightarrow \infty]{} D_v^\infty + Z_i^\infty(v) = D_u^\infty, \quad \text{and} \\ B'_u &= B'_v + E_i^{N'_v}(v) \xrightarrow[n \rightarrow \infty]{} B_v^\infty + E_i^\infty(v) = B_u^\infty. \end{aligned}$$

Therefore, $(D'_u, B'_u) \rightarrow (D_u^\infty, B_u^\infty)$, as $n \rightarrow \infty$, which completes the proof. \square

Important remark. Proving Theorem 2 amounts to showing that a property holds in probability. As a consequence, we can use the coupled sequence of trees we have just described. In the remainder of the paper, the trees we consider are always taken from this coupled sequence. In particular, there always exists a coupled ideal tree to rely on, and it does make sense to condition on events happening on this ideal tree to study random variables in T_n . We let Z^∞ , E^∞ , D^∞ , and B^∞ be the variables associated with the coupled ideal tree, so for a node $u \in T_\infty$ the variables of interest in the ideal tree are

$$D_u^\infty = \sum_{e \in \pi(u)} Z_e^\infty \quad \text{and} \quad B_u^\infty = \sum_{e \in \pi(u)} E_e^\infty.$$

3.3 Towards the proof

The proofs are based on large deviations theory (see, e.g., [26], [24], or [25]). We are interested in the case of *extended* random vectors whose components may also take one (only) of the values ∞ or $-\infty$. We now focus on this slight generalization.

Let $\{X_i, 1 \leq i \leq n\}$ be a family of independent and identically distributed (i.i.d.) extended random vectors distributed like $X = (Z, E)$. Assume $Z \in [-\infty, \infty)$ and $E \in [0, \infty]$. Set $p = \mathbf{P}\{Z > -\infty, E < \infty\}$. For α and ρ real numbers, we are interested in the tail probability

$$\mathbf{P}\left\{\sum_{i=1}^n Z_i > \alpha n, \sum_{i=1}^n E_i < \rho n\right\}. \quad (3)$$

Before we state the result, recall that the cumulant generating function Λ of an (extended) random vector X is defined by

$$\Lambda(\lambda, \mu) = \log \mathbf{E}\left[e^{\lambda Z + \mu E} \mid Z > -\infty, E < \infty\right] + \log p \quad \forall \lambda, \mu \in \mathbb{R}.$$

Observe that if Z and E are a.s. real, then $\Lambda(\lambda, \mu) = \mathbf{E}[e^{\lambda Z + \mu E}]$, which matches the usual definition. The tail probability in (3) is characterized using Λ^* , the Fenchel–Legendre dual of Λ [see 58]: for $\alpha, \rho \in \mathbb{R}$, we define

$$\Lambda^*(\alpha, \rho) = \sup_{\lambda, \mu} \{\lambda \alpha + \mu \rho - \Lambda(\lambda, \mu)\}.$$

Theorem 3 (Cramér, see [24]). *Assume that $\{(Z_i, E_i), i \geq 1\}$ are i.i.d. random vectors distributed as (Z, E) , and that $0 \in \mathcal{D}_\Lambda^\circ$. Then for any $\alpha, \rho \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}\left\{\sum_{i=1}^n Z_i > \alpha n, \sum_{i=1}^n E_i < \rho n\right\} = -I(\alpha, \rho) \stackrel{\text{def}}{=} -\inf\{\Lambda^*(x, y) : x > \alpha, y < \rho\}.$$

Remarks.(a) It is possible that $\Lambda^* = \infty$ everywhere except at a point, and consequently I may be infinite as well.

(b) Observe that the inequalities in Theorem 3 are strict. The result is false if one allows equality (see [44] or [24, Exercise 2.2.37] for a counterexample built by taking (α, ρ) on the boundary of \mathcal{D}_Λ). This technicality may be avoided if one enforces $(\alpha, \rho) \in \mathcal{D}_{\Lambda^*}^o$ (see Lemma 15).

Proof. The quadrant $(\alpha, \infty) \times (-\infty, \rho)$ is a convex open set. Hence Theorem 6.1.8 of Dembo and Zeitouni [24] applies when $\mathbf{P}\{Z = -\infty \text{ or } E = \infty\} = 0$ (thus, $p = 1$). We now show the details in the extended case. Note that

$$\begin{aligned} & \mathbf{P} \left\{ \sum_{i=1}^n Z_i > \alpha n, \sum_{i=1}^n E_i < \rho n \right\} \\ &= \mathbf{P} \left\{ \sum_{i=1}^n Z_i > \alpha n, \sum_{i=1}^n E_i < \rho n \mid \forall i, Z_i > -\infty, E_i < \infty \right\} \cdot p^n. \end{aligned}$$

The classical form of Cramér's theorem applies to the first factor, and hence, writing

$$\Lambda_c(\lambda, \mu) \mapsto \log \mathbf{E} \left[e^{\lambda Z + \mu E} \mid Z > -\infty, E < \infty \right],$$

the cumulant generating function of (Z, E) conditioned on $\{Z > -\infty, E < \infty\}$, and Λ_c^* for its dual,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P} \left\{ \sum_{i=1}^n Z_i > \alpha n, \sum_{i=1}^n E_i < \rho n \right\} = -\inf \{ \Lambda_c^*(x, y) : x > \alpha, y < \rho \} + \log p.$$

However, $\Lambda = \Lambda_c + \log p$, and therefore $\Lambda^* = \Lambda_c^* - \log p$, which finishes the proof. \square

The Gärtner–Ellis theorem is an extension of Theorem 3 for sequences of random variables that are neither independent nor identically distributed [35, 40]. We will only use the upper bound.

Theorem 4 (Gärtner–Ellis). *Let $\{(Z_n, E_n), n \geq 1\}$ be random vectors taking values in $[-\infty, \infty) \times [0, \infty]$. Let $F_n = \{Z_i > -\infty, E_i < \infty, 1 \leq i \leq n\}$. Let $A_m, m \geq 0$, be an arbitrary sequence of events. Assume that for all $(\lambda, \mu) \in \mathbb{R}^2$, and $\delta > 0$, there exists $M = M(\lambda, \mu, \delta)$ such that*

$$\sup_{n \geq 1} \frac{1}{n} \log \mathbf{E} \left[\mathbf{1}[F_n, A_M] \cdot \exp \left(\sum_{i=1}^n \lambda Z_i + \mu E_i \right) \right] \leq \Lambda(\lambda, \mu) + \delta \leq \infty. \quad (4)$$

Assume that Λ is the cumulant generating function of some extended random vector $X = (Z, E)$. Let Γ be a closed set such that $\{-\infty\} \times [0, \infty] \cup [-\infty, +\infty) \times \{\infty\} \notin \Gamma$. Assume that $0 \in \mathcal{D}_\Lambda^o$. Then, for any $\gamma > 0$, there exists M' such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n (Z_i, E_i) \in \Gamma, A_{M'} \right\} \leq -\min \left\{ 1/\gamma, \inf_{(\alpha, \rho) \in \Gamma} \Lambda^*(\alpha, \rho) - \gamma \right\},$$

where Λ^* is the convex dual of Λ .

Remark. This is just a statement of a relaxed version of the classical Gärtner–Ellis theorem [35, 40], that requires convergence to Λ instead of our asymptotic bounds. Observe that we only require *pointwise* asymptotic bounds on the moment generating functions. A formal proof can be found in Appendix B.

The Gärtner–Ellis and Cramér theorems do not give directly the constant c in terms of Λ^* but in terms of $I(\cdot, \cdot)$. The following alternate expressions for the constant c will be useful in the proofs. Until the end of the section, $\Lambda(\cdot, \cdot)$, $\Lambda^*(\cdot, \cdot)$ and $I(\cdot, \cdot)$ are the functions associated with (Z, E) .

Lemma 4. *Assume that $0 \in \mathcal{D}_\Lambda^o$, $\mathbf{P}\{Z > -\infty, E < \infty\} \geq 1/d$ and $\mathbf{P}\{E = 0\} < 1/d$. Let $c \stackrel{\text{def}}{=} \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$. Then*

- (a) $c = \inf_{\epsilon > 0} \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + \epsilon\}$.
- (b) $c = \sup\{\alpha/\rho : I(\alpha, \rho) \leq \log d\}$.
- (c) $c = \sup\{\alpha/\rho : (\alpha, \rho) \in \Psi(\log d) \cap \mathcal{D}_{\Lambda^*}^o\}$.

Proof. For $\ell \in \mathbb{R}$, let $\Psi_I(\ell) = \{(\alpha, \rho) : I(\alpha, \rho) \leq \ell\}$. Observe that Lemma 1 ensures that $\Psi(\log d) \neq \emptyset$ and that c is well-defined.

(a) Since $\{\Lambda^*(\alpha, \rho) \leq \log d\} \subseteq \{\Lambda^*(\alpha, \rho) \leq \log d + \epsilon\}$ for all $d \geq 1$ and $\epsilon > 0$, it is straightforward that $\sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\} \leq \inf_{\epsilon > 0} \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + \epsilon\}$.

For $n \geq 1$, write $c_n \stackrel{\text{def}}{=} \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + 1/n\}$. By Lemma 1 (c), there exists n_0 large enough that $c_n < \infty$ for all $n \geq n_0$. Let $\delta > 0$. For $n \geq n_0$, let $(\alpha_n, \rho_n) \in \Psi(\log d + 1/n)$ be a sequence of points such that $\alpha_n/\rho_n \geq c_n - \delta$. Clearly $(\alpha_n, \rho_n) \in \Psi(\log d + 1/n_0)$ for all $n \geq n_0$ and since $\Psi(\log d + 1/n_0)$ is compact, there exists a subsequence $\{(\alpha_n, \rho_n), n \geq n_0\}$ that converges to $(\alpha_\infty, \rho_\infty) \in \Psi(\log d + 1/n_0)$ as $n \rightarrow \infty$. For each $n \geq n_0$ in the subsequence we have that $\Lambda^*(\alpha_n, \rho_n) \leq \log d + 1/n$, and since Λ^* is continuous in $\Psi(\log d + 1/n_0)$ (since it is compact), then $\Lambda^*(\alpha_\infty, \rho_\infty) \leq \log d$. Thus $\alpha_\infty/\rho_\infty \leq \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$. Also, since $\alpha_n/\rho_n \geq c_n - \delta$ for all $n \geq n_0$, then $\alpha_\infty/\rho_\infty \geq \lim_{n \rightarrow \infty} c_n - \delta = \inf_{\epsilon > 0} \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + \epsilon\} - \delta$. Taking $\delta \rightarrow 0$ concludes the proof.

(b) Recall that the rate function I is defined by $I(\alpha, \rho) = \inf\{\Lambda^*(x, y) : x > \alpha, y < \rho\}$, for $\alpha, \rho \in \mathbb{R}$. We first show that $\sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\} \leq \sup\{\alpha/\rho : I(\alpha, \rho) \leq \log d\}$. For any $\epsilon > 0$, we can pick $\alpha_0 < \alpha$ and $\rho_0 > \rho$ such that $\alpha_0/\rho_0 > \alpha/\rho - \epsilon$. Then $(\alpha_0, \rho_0) \in \Psi_I(\log d)$, implying that $\sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\} \leq \sup\{\alpha/\rho : I(\alpha, \rho) \leq \log d\} + \epsilon$. Since ϵ is arbitrary, $\sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\} \leq \sup\{\alpha/\rho : I(\alpha, \rho) \leq \log d\}$.

Next, we show that $\sup\{\alpha/\rho : I(\alpha, \rho) \leq \log d\} \leq \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$. Assume that $I(\alpha, \rho) \leq \log d$, or, equivalently, $(\alpha, \rho) \in \Psi_I(\log d)$. Then by definition, there exist (x, y) such that $x > \alpha$, $y < \rho$ and $\Lambda^*(x, y) \leq \log d$. Clearly, $(x, y) \in \Psi(\log d)$ and $x/y > \alpha/\rho$, which proves the claim.

(c) Since Λ^* is finite on $\Psi(\log d)$, we see that $\Psi(\log d)^o \subset \Psi(\log d) \cap \mathcal{D}_{\Lambda^*}^o \subset \Psi(\log d)$. By Lemma 1, $\{\rho = 0\} \cap \Psi(\log d) = \emptyset$, and hence α/ρ is continuous on $\Psi(\log d)$.

Accordingly, $\sup\{\alpha/\rho : (\alpha, \rho) \in \Psi(\log d)^o\} = \sup\{\alpha/\rho : (\alpha, \rho) \in \Psi(\log d)\}$. The result follows. \square

Lemma 5. *Assume $0 \in \mathcal{D}_\Lambda^o$. Let $p = \mathbf{P}\{Z > -\infty, E < \infty\}$. Then $\Psi(-\log p) \subset \{(\alpha, \rho) : \alpha \leq \mathbf{E}Z^c, \rho \geq \mathbf{E}E^c\}$, where (Z^c, E^c) denotes a random vector distributed as (Z, E) conditioned on $\{Z > -\infty, E < \infty\}$.*

Proof. We have $\Lambda^* = \Lambda_{(Z^c, E^c)}^* - \log p$. So it suffices to prove the claim when $p = 1$, and hence $(Z, E) = (Z^c, E^c)$ almost surely. Assume that $\alpha > \mathbf{E}Z^c$. For any $\rho \in \mathbb{R}$, we have $\Lambda^*(\alpha, \rho) = \sup_{\lambda, \mu} \{\lambda\alpha + \mu\rho - \Lambda(\lambda, \mu)\} \geq \sup_{\lambda} \{\lambda\alpha - \Lambda(\lambda, 0)\}$. Since $0 \in \mathcal{D}_\Lambda^o$, Λ is differentiable at 0 and $\Lambda(\lambda, 0) = \lambda\mathbf{E}Z^c + o(\lambda)$, as $\lambda \rightarrow 0$. As a consequence, $\lambda\alpha - \Lambda(\lambda, 0) = \lambda(\alpha - \mathbf{E}Z^c) + o(\lambda) \sim \lambda(\alpha - \mathbf{E}Z^c)$ by assumption. It follows that there exists $\lambda > 0$ such that $\lambda\alpha - \Lambda(\lambda, 0) > 0$, and hence $\Lambda^*(\alpha, \rho) > 0$, hence proving that $(\alpha, \rho) \notin \Psi(0)$ if $\alpha > \mathbf{E}Z^c$. The case when $\rho < \mathbf{E}E^c$ is treated in a similar way. \square

AROUND THE OPTIMAL VALUE. To prove Theorem 1, we shall need to show that, for $\epsilon > 0$, $\mathbf{P}\{H_n \geq (c + \epsilon) \log n\} = o(1)$, and $\mathbf{P}\{H_n \geq (c - \epsilon) \log n\} = 1 - o(1)$. In other words, taking for granted the link between these tail probabilities and $\Lambda^*(\cdot, \cdot)$ and $I(\cdot, \cdot)$, we need some information about the behavior of the curves around $\{\alpha = c\rho\}$. This is why the next lemma is the key to proving the upper and lower bounds of Theorems 1 and 2.

Lemma 6. *Assume $0 \in \mathcal{D}_\Lambda^o$ and $p = \mathbf{P}\{Z > -\infty, E < \infty\} > 1/d$. Let $c \stackrel{\text{def}}{=} \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$.*

- (a) $\forall \epsilon > 0$, there exists $(\alpha, \rho) \in \mathbb{R}^2$ such that $I(\alpha, \rho) < \log d$, and $c - \epsilon < \alpha/\rho < c$.
- (b) If $\mathbf{P}\{E = 0\} < 1/d$, then, for any $\epsilon > 0$, $\inf\{\Lambda^*(x, y) : x/y \geq c + \epsilon\} > \log d$.

Proof. Lemma 1 ensures that $\Psi(\log d) \neq \emptyset$.

(a) Let $\epsilon > 0$. By definition, we can pick (α_0, ρ_0) such that $\Lambda^*(\alpha_0, \rho_0) \leq \log d$ and $\alpha_0/\rho_0 > c - \epsilon/2$. Consider the region $\Psi(\log d) \cap \mathcal{B} \cap \mathcal{D}_{\Lambda^*}^o$ where \mathcal{B} is a non-empty open ball centered at (α_0, ρ_0) for which all points $(x, y) \in \mathcal{B}$ satisfy $x/y > c - \epsilon$. Since Λ^* is convex and $\Lambda^*(\mathbf{E}Z^c, \mathbf{E}E^c) = -\log p < \log d$, this implies that $\Lambda^*(x, y) < \log d$ for some $(x, y) \in \Psi(\log d) \cap \mathcal{B} \cap \mathcal{D}_{\Lambda^*}^o$. Furthermore, we can pick such an (x, y) such that $x/y < c$. Next, pick $(\alpha, \rho) \in \Psi(\log d) \cap \mathcal{B}$ such that $\alpha < x$ and $\rho > y$ (and hence $\alpha/\rho < c$). Since $\Lambda^*(x, y) < \log d$, $I(\alpha, \rho) < \log d$, and $c - \epsilon < \alpha/\rho < c$.

(b) Let $\epsilon > 0$ and assume for a contradiction that $\inf\{\Lambda^*(\alpha, \rho) : \alpha/\rho \geq c + \epsilon\} \leq \log d$. Then, for any $\delta > 0$, there exist (α, ρ) such that $\Lambda^*(\alpha, \rho) \leq \log d + \delta$ and $\alpha/\rho \geq c + \epsilon$. As a consequence, $\forall \delta > 0$, $\sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + \delta\} \geq c + \epsilon$. This implies that $\inf_{\delta > 0} \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d + \delta\} \geq c + \epsilon$. By Lemma 4 (b), we have $\inf_{\delta > 0} \sup\{x/y : \Lambda^*(\alpha, \rho) \leq \log d\} = c$, and obtain a contradiction. \square

3.4 The upper bound

Let Λ^n denote be the cumulant generating function of a typical component (Z^n, E^n) of \mathcal{X}^n . Let \mathcal{L}_k be the set of nodes k levels away from the root in T_∞ . Let u_k be the

left-most node in \mathcal{L}_k . We introduce the event F_k defined by

$$F_k \stackrel{\text{def}}{=} \{Z_e > -\infty, E_e < \infty, \forall e \in \pi(u_k)\}.$$

The upper bound is based on the Gärtner–Ellis theorem (Theorem 4). The following result proves that the conditions for its application hold, with the event A_M being $N_u \geq M$.

Lemma 7. *Let $\lambda, \mu \in \mathbb{R}$. For any $\delta > 0$, there exists M large enough such that*

$$\sup_{n,k} \left\{ \frac{1}{k} \log \mathbf{E} \left[\mathbf{1}[F_k, N_{u_k} \geq M] \cdot \exp(\lambda D_{u_k} + \mu B_{u_k}) \mid N_{u_0} = n \right] \right\} \leq \Lambda(\lambda, \mu) + \delta,$$

where Λ is the cumulant generating function of (Z^∞, E^∞) .

Proof. In order to improve the readability of the equations, and in this proof only, let us reindex the random vectors (Z_e, E_e) on the left-most path to T_∞ as $\{(Z_i, E_i), i \geq 1\}$, where the indices increase with the distance from the root. In the same spirit, for $i \geq 0$, write N_i, D_i and B_i for N_{u_i}, D_{u_i} and B_{u_i} , respectively. If $n < M$, we clearly have $\mathbf{1}[N_k \geq M] = 0$ and the result holds. With our new notation, $D_k = \sum_{i=1}^k Z_i$ and $B_k = \sum_{i=1}^k E_i$, so proving the results reduces to bounding

$$\begin{aligned} C &\stackrel{\text{def}}{=} \mathbf{E} \left[\mathbf{1}[F_k, N_k \geq M] \cdot e^{\lambda D_k + \mu B_k} \mid N_0 \right] \\ &= \mathbf{E} \left[\mathbf{1}[F_k, N_k \geq M] \cdot e^{\sum_{i=1}^k (\lambda Z_i + \mu E_i)} \mid N_0 \right]. \end{aligned}$$

The random vectors (Z_i, E_i) are not independent. However, by conditioning on N_1 ,

$$C = \mathbf{E} \left[\mathbf{E} \left[\mathbf{1}[F_k, N_k \geq M] \cdot e^{\sum_{i=1}^k (\lambda Z_i + \mu E_i)} \mid N_1 \right] \mid N_0 \right].$$

Let F_k^2 be the event that $\{Z_i, E_i \in \mathbb{R}, 2 \leq i \leq k\}$. Then, given N_0 and N_1 , the random variables $\mathbf{1}[F_k^2, N_k \geq M] \exp(\sum_{i=2}^k (\lambda Z_i + \mu E_i))$ and $\mathbf{1}[F_1] \exp(\lambda Z_1 + \mu E_1)$ are independent. Hence

$$C \leq \mathbf{E} \left[\underbrace{\mathbf{E} \left[\mathbf{1}[F_k^2, N_k \geq M] \cdot e^{\sum_{i=2}^k (\lambda Z_i + \mu E_i)} \mid N_1 \right]}_I \cdot \mathbf{E} \left[\mathbf{1}[F_1] e^{\lambda Z_1 + \mu E_1} \mid N_1 \right] \mid N_0 \right],$$

where we used $\mathbf{1}[N_1 \geq M] \leq 1$ in the second factor. The first factor can be bounded by

$$I \leq \sup_{m \geq M} \mathbf{E} \left[\mathbf{1}[F_k^2, N_k \geq M] \cdot e^{\sum_{i=2}^k (\lambda Z_i + \mu E_i)} \mid N_1 = m \right],$$

which is independent of N_1 and N_0 . Let $\delta > 0$ and let M be large enough that for all $m \geq M$, $\Lambda^m(\lambda, \mu) \leq \Lambda(\lambda, \mu) + \delta$. Then

$$\begin{aligned} C &\leq \sup_{m \geq M} \mathbf{E} \left[\mathbf{1}[F_k^2, N_k \geq M] \cdot e^{\sum_{i=2}^k (\lambda Z_i + \mu E_i)} \mid N_1 = m \right] \cdot e^{\Lambda(\lambda, \mu) + \delta} \\ &= \sup_{m \geq M} \mathbf{E} \left[\mathbf{1}[F_{k-1}, N_{k-1} \geq M] \cdot e^{\sum_{i=1}^{k-1} (\lambda Z_i + \mu E_i)} \mid N_0 = m \right] \cdot e^{\Lambda(\lambda, \mu) + \delta}. \end{aligned}$$

An easy induction then shows that

$$\sup_{n \geq M} \mathbf{E} \left[\mathbf{1}[F_k, N_k \geq M] \cdot e^{\lambda D_k + \mu B_k} \mid N_0 = n \right] \leq e^{k\Lambda(\lambda, \mu) + k\delta}.$$

Since δ was arbitrary, the proof is complete. \square

Let $\epsilon > 0$. Let $c' = c + \epsilon$, where $c = \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$ is the constant defined in the statement of Theorem 2. By definition,

$$\mathbf{P} \{H_n > c' \log n\} = \mathbf{P} \{\exists v \in T_n : D_v > c' \log n\}.$$

Recall that \mathcal{L}_k denotes the set of nodes at level k in T_∞ . The union bound yields

$$\mathbf{P} \{H_n > c' \log n\} \leq \sum_{k \geq 0} \mathbf{P} \{\exists v \in \mathcal{L}_k : D_v > c' \log n, v \in T_n\}.$$

Using a second union bound over the nodes in each level,

$$\mathbf{P} \{H_n > c' \log n\} \leq \sum_{k \geq 0} d^k \cdot \mathbf{P} \{D_{u_k} > c' \log n, u_k \in T_n\}. \quad (5)$$

In order to further bound (5), we first restrict our attention to the case $N_{u_k} \geq M$. We have

$$\mathbf{P} \{D_{u_k} > c' \log n, N_{u_k} \geq M\} \leq \mathbf{P} \{(D_{u_k}, B_{u_k}) \in \Gamma, N_{u_k} \geq M\},$$

where $\Gamma = \{(x, y) \in \mathbb{R}^2 : x \geq c'y\}$. By Lemma 7, and since $0 \in \mathcal{D}_\Lambda^o$, the upper bound of Gärtner–Ellis theorem (Theorem 4) holds: for any $\gamma > 0$, there exists M_1 such that for all $M \geq M_1$,

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathbf{P} \{D_{u_k} > c' \log n, N_{u_k} \geq M\} \leq -\min \left\{ \frac{1}{\gamma}, \inf_{(x, y) \in \Gamma} \Lambda^*(x, y) - \gamma \right\}.$$

By Lemma 6, there exists $\beta > 0$ such that $\inf\{\Lambda^*(x, y) : (x, y) \in \Gamma\} \geq \log d + \beta$. Then, choosing $\gamma < \beta/2$, we have

$$\sum_{k \geq K} d^k \cdot \mathbf{P} \{D_{u_k} \geq c' \log n, N_{u_k} \geq M\} \leq \sum_{k \geq K} d^k \cdot e^{-k(\beta/2 + \log d)} \leq C_1 \cdot e^{-K\beta/2}, \quad (6)$$

for all $K \geq K_1$ large enough and some constant $C_1 = C_1(K_1)$,

As in the proof of Theorem 1, we treat the values of $k \leq K$ using Markov's inequality. Let $\lambda > 0$, such that $(\lambda, 0) \in \mathcal{D}_\Lambda^o$. There exists $C_2 \geq 0$ and $M_2 > 0$ such that $\sup\{\Lambda^n(\lambda, 0) : n \geq M_2\} \leq C_2 < \infty$. Then, for this value of λ , by Lemma 7,

$$\mathbf{P} \{D_{u_k} \geq c' \log n, N_{u_k} \geq M_2\} \leq e^{kC_2 - \lambda c' \log n}.$$

Therefore, by the union bound,

$$\sum_{k \leq K} \mathbf{P} \{ \exists v \in \mathcal{L}_k : D_v \geq c' \log n, N_v \geq M_2 \} \leq \frac{K d^K e^{K C_2}}{n^{\lambda c'}}. \quad (7)$$

Let now $M_3 = \max\{M_1, M_2\}$. We have obtained bounds on the terms of (6) for every k when $N_{u_k} \geq M_3$. It remains to deal with the nodes at the bottom of the tree for which $N < M_3$. Recall that by assumption, $\mathbf{P} \{H_n \geq \psi(n)\} = 0$.

$$\begin{aligned} \mathbf{P} \{H_n \geq (c + 2\epsilon) \log n\} &\leq \mathbf{P} \{ \exists v \in T_n : D_v \geq (c + 2\epsilon) \log n - \psi(M_3), N_v \geq M_3 \} \\ &\leq \mathbf{P} \{ \exists v \in T_n : D_v \geq (c + \epsilon) \log n, N_v \geq M_3 \}. \end{aligned}$$

Hence, putting (6) and (7) together,

$$\mathbf{P} \{H_n \geq (c + 2\epsilon) \log n\} \leq \frac{K d^K e^{K C_2}}{n^{\lambda c'}} + C_1 e^{-K \beta/2}.$$

As $\lambda c' > 0$, this can be made as small as we want by first choosing K and next letting n go to infinity. Since ϵ was arbitrary, this finishes the proof of the upper bound.

3.5 The lower bound

The aim of this section is to build a surviving Galton–Watson process that ensures that nodes with large weighted depth exist in T_n with probability $1 - o(1)$. We split the construction of this process into stages. We first show that deep nodes do occur with positive probability. Then, we proceed with proving that “deep nodes” occur in T_n with probability $1 - o(1)$ using a standard boosting argument and considering the subtrees rooted at level $t > 0$. We first have to ensure that there are enough nodes at level t that are suitable for the boosting argument. Indeed, it is possible that $D_u = -\infty$ or $B_u = \infty$ which would prevent any node in the subtree of u from having any effect on the height.

SKIMMING THE TREE. Our aim here is to find nodes of sufficiently large weighted depth in T_n . We start by finding nodes with large weighted depth in the coupled ideal tree, and then prove that the corresponding nodes in T_n are also sufficiently deep.

Lemma 8. *Let T_n be a random tree as described in Section 3. Let $c = \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$. For all $\epsilon > 0$, there exists n_0 such that*

$$\inf_{n \geq n_0} \mathbf{P} \{ \exists u \in T_n : D_u \geq (c - \epsilon) \log n \} > 0.$$

Proof. Let $\epsilon > 0$. By Lemma 6, there exists α and ρ such that $\alpha/\rho = c'$ and $I(\alpha, \rho) < \log d$, for some c' such that $c - \epsilon/2 < c' < c$. Let α and ρ be fixed. Let ℓ be an arbitrary positive integer to be chosen later. A node $v \in T_\infty$ is called *ideally*

good if either it is the root, or v lies ℓ levels below an ideally good node u and we have

$$D_v^\infty \geq D_u^\infty + \ell\alpha \quad \text{and} \quad E_v^\infty \leq E_u^\infty + \ell\rho.$$

The set of ideally good nodes forms a Galton–Watson tree. Let Y^∞ be the size of the progeny of u in this Galton–Watson process. By linearity of expectation, writing $\pi(u, v)$ for the set of edges on the unique path from u to v in the ideal tree, with v lying ℓ levels below u ,

$$\begin{aligned} \mathbf{E}Y^\infty &= d^\ell \cdot \mathbf{P} \{D_v^\infty - D_u^\infty \geq \ell\alpha, E_v^\infty - E_u^\infty \leq \ell\rho\} \\ &= d^\ell \cdot \mathbf{P} \left\{ \sum_{e \in \pi(u, v)} Z_e^\infty \geq \ell\alpha, \sum_{e \in \pi(u, v)} E_e^\infty \leq \ell\rho \right\}. \end{aligned}$$

By Cramér’s theorem (Theorem 3), and because of our choice for α and ρ , we have

$$\mathbf{E}Y^\infty = d^\ell \cdot e^{-I(\alpha, \rho)\ell + o(\ell)} = e^{\ell \log d - \ell I(\alpha, \rho) + o(\ell)} \xrightarrow{\ell \rightarrow \infty} \infty.$$

Thus, there exists ℓ large enough such that $\mathbf{E}Y^\infty > 1$. This choice makes the process supercritical. Let ℓ now be fixed.

Consider the coupled random trees T_n , with size-dependent vectors. A node $v \in T_\infty$ is called *good* if it is the root, or it lies ℓ levels below a good node u and we have

$$E_v \leq E_u + \rho\ell \quad \text{and} \quad D_v \geq D_u + \ell\alpha.$$

The set of good nodes is a branching process. However, the progeny distribution Y_u of a node u now depends on u and the process is *not* a Galton–Watson process. We deal with this minor issue using Lemma 13. By Lemma 3, we have

$$\liminf_{n \rightarrow \infty} \mathbf{P} \{Y_u \geq t \mid N_u = n\} \geq \mathbf{P} \{Y^\infty \geq t\},$$

for all $0 \leq t \leq d^\ell$. Since $\mathbf{E}Y^\infty > 1$, there exists M large enough that for all $n \geq M$,

$$\mathbf{P} \{Y_u \geq t \mid N_u = n\} \geq \mathbf{P} \{Y^\infty \geq t\} + \frac{1 - \mathbf{E}Y^\infty}{2d^\ell}.$$

Now, by Lemma 13, there exists a random variable Y' such that, for all t ,

$$\mathbf{P} \{Y' \geq t\} = \max \left(\mathbf{P} \{Y^\infty \geq t\} + \frac{1 - \mathbf{E}Y^\infty}{2d^\ell}, 0 \right).$$

Further, there exist coupled copies of Y' , $\{Y'_u, u \in T_\infty\}$ such that have $Y'_u \leq Y_u$ if

$N_u \geq M$. The Galton–Watson process with progeny distribution Y' is supercritical:

$$\begin{aligned}
\mathbf{E}[Y'] &= \sum_{t=1}^{d^\ell} \mathbf{P}\{Y' \geq t\} \\
&\geq \sum_{t=1}^{d^\ell} \left(\mathbf{P}\{Y^\infty \geq t\} + \frac{1 - \mathbf{E}Y^\infty}{2d^\ell} \right) \\
&= \mathbf{E}Y^\infty + \frac{1 - \mathbf{E}Y^\infty}{2} \\
&= \frac{1 + \mathbf{E}Y^\infty}{2} > 1.
\end{aligned}$$

Therefore, it survives with probability $1 - q > 0$. Note that the guarantee $Y'_u \leq Y_u$ only holds if $N_u \geq M$. In particular, it is *not* true that every node u in the coupled Galton–Watson process with progeny distribution Y' is also a good node.

However, in the case of survival, either (a) there is an infinite path of good nodes u with $N_u \geq M$, or (b) there is some good node w with $N_w < M$. Now, if (a) happens, for every integer k , there exists a node v such that $D_v \geq \alpha k$ and $N_v \geq M$. So in particular, with

$$k_1 = \left\lfloor \frac{\log n}{\rho \ell} \right\rfloor,$$

$D_v \geq c' \log n - \alpha \ell$, and $v \in T_n$ since $N_v \geq M \geq 1$. In case (b), consider the shallowest good node w such that $N_w < M$. Then, w is part of some generation k_2 of the process (at level $k_2 \ell$ in T_∞). Since w is good, $M > N_w \geq n \exp(-\rho k_2 \ell)$, and hence,

$$k_2 \geq \frac{\log n - \log M}{\rho \ell}.$$

It follows that $D_w \geq c' \log n - c' \log M$. As a consequence, in both cases, for n large enough, there exists a node $u \in T_n$ with $D_u \geq (c - \epsilon) \log n$, and this happens with probability at least $1 - q > 0$. \square

Lemma 8 ensures that nodes with a large weighted depth exist with positive probability. We now show that such nodes actually exist with probability $1 - o(1)$. To this aim, we intend to use the standard boosting technique: we run multiple copies of the branching process to increase the chance that one survives. Instead of using the root as a first individual, we want to use some of the d^t nodes at level t as starting individuals of independent processes. However, as for the case of ideal trees, not all such nodes are suitable as starting individuals.

THE NICE PORTION OF THE TREE. Since $\mathbf{P}\{Z = -\infty, E = \infty\}$ may be positive, we cannot expect in general that all d^t nodes at level t are good starting individuals. Indeed, some node u may not even be active, i.e., $D_u = -\infty$ or $B_u = \infty$. In spite of this fact, we claim that there are enough good starting individuals when \mathcal{X}^∞ is a

valid split vector. In order to prove this claim, we use a second branching process defined on the top t levels.

We first look at the ideal tree. Let $v \in T_\infty$ be called *ideally nice* if either it is the root, or it is linked to an ideally nice node by an edge e and we have

$$Z_e^\infty > a \quad \text{and} \quad E_e^\infty < b.$$

Let R_t^∞ be the number of ideally nice nodes in \mathcal{L}_t , the set of nodes t levels away from the root in T_∞ . Then $\{R_t^\infty, t \geq 0\}$ is a Galton–Watson process. By Definition 1 (c) of a valid split vector, $\mathbf{P}\{Z^\infty > -\infty, E^\infty < \infty\} > 1/d$, hence there exist $\delta > 0$, a_0 and b_0 such that for all $a \leq a_0$ and $b \geq b_0$, $\mathbf{P}\{Z^\infty > a, E^\infty < b\} > 1/d + \delta$. Now, by Definition 1 (c), $\mathbf{P}\{\exists i : Z_i^\infty > -\infty, E_i^\infty < \infty\} = 1$, and thus $\mathbf{P}\{\exists i : Z_i^\infty > a, E_i^\infty < b\} \rightarrow 1$, as $a \rightarrow -\infty$ and $b \rightarrow \infty$. By Theorem 12, the process survives with probability at least $1 - q'$, and $q' = q'(a, b)$ can be made as small as we want by choice of a and b . If $R_t^\infty > 0$ for all $t \geq 0$, then $R_t^\infty \rightarrow \infty$ as $t \rightarrow \infty$ with probability one [2]. As a consequence, for any integer r , there exists t_0 such that $\mathbf{P}\{R_{t_0}^\infty \leq r \mid R_t^\infty > 0, \forall t \geq 0\} \leq 1/r$.

We return to the random tree T_n . A node at level t in T_n is called *nice* if $D_u \geq at$ and $B_u \leq bt$. By Lemma 3, the number R_{t_0} of nice nodes u at level t_0 satisfies, for n large enough,

$$\mathbf{P}\{R_{t_0} \leq r \mid R_t^\infty > 0, \forall t \geq 0\} \leq 2/r. \quad (8)$$

Observe in particular that the conditioning is meaningful since we consider the coupled sequence of trees. Equation (8) gives us the handle we need on the number of nodes we can use as starting individuals in the boosting step.

BOOSTING THE SURVIVAL PROBABILITY. Let $\epsilon > 0$. Let $\{T_\infty(v_i), 1 \leq i \leq R_{t_0}\}$ be the family of subtrees of T_∞ rooted at the nice nodes $\{v_i, 1 \leq i \leq R_{t_0}\}$. The processes of good nodes described in the proof of Lemma 8 evolve independently in every $T_\infty(v_i)$. Furthermore, by Lemma 8, there is n_0 such that for all $1 \leq i \leq R_{t_0}$ and for all $m \geq n_0$,

$$\mathbf{P}\left\{\exists u \in T_\infty(v_i) : D_u - D_{v_i} \geq \left(c - \frac{\epsilon}{2}\right) \log m, B_u - B_{v_i} < \log m\right\} \geq 1 - q. \quad (9)$$

By construction, we have $D_{v_i} \geq at_0$ and $B_{v_i} \leq bt_0$, for $1 \leq i \leq R_{t_0}$. Let n be large enough, and let m be such that $\log m = \log n - bt_0$. If one can find a node u in $T_\infty(v_i)$ as described in (9), then

$$D_u \geq at_0 - \left(c - \frac{\epsilon}{2}\right) t_0 b + \left(c - \frac{\epsilon}{2}\right) \log n \geq (c - \epsilon) \log n,$$

for n large enough. Such a node u is called a *deep* node. Moreover, $B_u < \log m + B_{v_i} \leq \log n$ so $u \in T_n$ and $H_n \geq D_u \geq (c - \epsilon) \log n$.

If no deep node exists, then one of the following must occur: either $\{R_t, t \geq 0\}$ dies, or it survives but $R_{t_0} \leq r$, or we cannot find a deep node in any of the $R_{t_0} \geq r$ independent trees $T_\infty(v_i)$. As a consequence, for n large enough,

$$\begin{aligned} \mathbf{P}\{H_n \leq (c - \epsilon) \log n\} &\leq \mathbf{P}\{R_{t_0} < r\} + \mathbf{P}\{H_n \leq (c - \epsilon) \log n \mid R_{t_0} \geq r\} \\ &\leq \mathbf{P}\{\exists t \geq 0 : R_t = 0\} + \mathbf{P}\{1 \leq R_{t_0} < r\} + q^r, \end{aligned}$$

by independence of $T_\infty(v_i)$, $1 \leq i \leq R_{t_0}$. It follows that

$$\mathbf{P}\{H_n \leq (c - \epsilon) \log n\} \leq q' + \frac{2}{r} + q^r.$$

This can be made as small as we want by choice of $q' = q'(a, b)$ and r . This completes the proof of the lower bound.

4 The height of trees of effective size n

In some applications, one wants to express the height of the tree in terms of the number of significant nodes. Only the active portion of the tree is significant for the height, and we shall define the *effective size* $\#T_n$ of T_n as the size of its active portion:

$$\#T_n = |\{u \in T_n : D_u > -\infty\}|.$$

When $\mathbf{P}\{Z = -\infty\} = 0$, the effective size is just the number of nodes $|T_n|$. The only difference between the height H_n and that of a tree of effective size n is a scale factor.

Theorem 5. *Let T_n be a valid weighted random tree with limit split vector \mathcal{X} , of (random) effective size $s_n = \#T_n$. Let Λ^* be the rate function of a typical component X of \mathcal{X} . Then, the height of T_n satisfies $H_n = \frac{c}{\gamma} \log s_n + o(\log s_n)$ in probability, as $n \rightarrow \infty$, where $c = \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$ and $\gamma = -\sup\{\phi : \Lambda_Y(\phi) \leq -\log d\}$, with $Y = E + \infty \cdot \mathbf{1}[Z = -\infty]$.*

We start by proving an equivalent result for ideal trees.

Theorem 6. *Let T_n be an ideal tree with valid split vector \mathcal{X} , of (random) effective size $s_n = \#T_n$. Let Λ^* be the rate function of a typical component X of \mathcal{X} . Then, the height of T_n satisfies $H_n = \frac{c}{\gamma} \log s_n + o(\log s_n)$ in probability, as $n \rightarrow \infty$, where $c = \sup\{\alpha/\rho : \Lambda^*(\alpha, \rho) \leq \log d\}$ and $\gamma = -\sup\{\phi : \Lambda_Y(\phi) \leq -\log d\}$, with $Y = E + \infty \cdot \mathbf{1}[Z = -\infty]$.*

Theorem 6 follows easily from following lemma about the effective size of T_n . The ideas are borrowed from Nerman [55]. See also the papers of Biggins [9, 10] who uses Nerman's results in settings that are similar to ours.

Lemma 9. *Let T_n be an ideal tree with valid split vector \mathcal{X} . Let (Z, E) be a typical component of \mathcal{X} . Let $\gamma = -\sup\{\phi : \Lambda_Y(\phi) \leq -\log d\}$, where $Y = E + \infty \cdot \mathbf{1}[Z = -\infty]$. Then $\log \#T_n \sim \gamma \log n$ in probability, as $n \rightarrow \infty$.*

Proof. The effect of $Z = -\infty$ is to cut down a subtree. We introduce a modified time random variable Y producing the same effect: $Y = E + \infty \cdot \mathbf{1}[Z = -\infty]$. Because the proofs rely on the renewal theorem, Biggins [10] assumes the distributions are nonlattice. However, the theorems can be proved in the lattice case as well [10, 55]. Theorem 2.1 of Biggins [10] can be used without modification, provided we translate it to our setting. We use the cumulant generating function Λ_Y defined by

$$\Lambda_Y(\phi) = \log \mathbf{E} \left[e^{\phi Y} \right] + \log \mathbf{P} \{Y < \infty\},$$

for $\phi \in \mathbb{R}$. The *Malthusian* parameter

$$\gamma = -\sup\{\phi : \Lambda_Y(\phi) \leq -\log d\}$$

is the quantity of interest. For all ϕ ,

$$\Lambda_Y(\phi) = \log \mathbf{P} \{Z > -\infty, E < \infty\} + \log \mathbf{E} \left[e^{\phi E} \mid Z > -\infty, E < \infty \right].$$

Also, $\Lambda_Y(0) = \log \mathbf{P} \{Z > -\infty, E < \infty\} > -\log d$ by assumption. Hence, $\gamma > 0$ (which just means that the process is supercritical). Clearly, $\sup_t e^{-\gamma t} < \infty$. By Theorem 2.1 of [10], we thus conclude that $\log \#T_n \sim \gamma \log n$ on the surviving set. However, by Definition 1 the process survives with probability 1. As a consequence, we have $\log \#T_n \sim \gamma \log n$ a.s. and thus in probability. \square

Remark. Lemma 9 can also be proved using properties of recursive equations and the contraction method (see, e.g., [54, 57, 59, 60]).

Proof of Theorem 6. By Theorem 1, $H_n = c \log n + o(\log n)$ in probability as $n \rightarrow \infty$. Also, by Lemma 9, $s_n = \#T_n \sim \gamma \log n$ in probability, as $n \rightarrow \infty$. Now,

$$\frac{H_n}{\log n} \rightarrow c \quad \text{and} \quad \frac{\log n}{\log s_n} \rightarrow \frac{1}{\gamma}$$

in probability, as $n \rightarrow \infty$. Therefore, the product converges in probability. \square

We can now sandwich the effective sizes of random trees between that of two ideal trees. This provides a result similar to Lemma 9. The proof of Theorem 5 goes along the same lines as that of Theorem 6, and is omitted.

Lemma 10. *Let T_n be a weighted random tree with valid limit split vector \mathcal{X} . Let (Z, E) be a typical component of \mathcal{X} . Then, as $n \rightarrow \infty$, $\log \#T_n \sim \gamma \log n$ in probability, where $\gamma = -\sup\{\phi : \Lambda_Y(\phi) \leq -\log d\}$.*

Proof. The modified size-dependent time random variables are now $\{Y^m, m \geq 0\}$ where $Y^m = E^m + \infty \cdot \mathbf{1}[Z^m = -\infty]$. Upper and lower bounds on $\#T_n$ may be obtained by respectively lower, and upper bounding Y^n so as to have i.i.d. variables,

and then use Lemma 9 above. We describe the upper bound and omit the proof of the lower bound since it follows along the same lines. We have

$$\begin{aligned}\Lambda_{Y^m}(\phi) &= \log \mathbf{P} \{Z^m > -\infty, E^m < \infty\} + \Lambda^m(0, \phi) \\ &\rightarrow \log \mathbf{P} \{Z > -\infty, E < \infty\} + \Lambda(0, \phi),\end{aligned}$$

as $m \rightarrow \infty$. Since $0 \in \mathcal{D}_\Lambda^o$, $Y^m \rightarrow Y$ in distribution [12, p. 390]. We use a coupling argument. Let F_m and F be the distribution functions of Y^m and Y , respectively. Let $G_M(x) = \sup\{F_m(x), m \geq M\}$. The function G_M is the distribution function of a proper random variable W . By the dominated convergence theorem, we have $\Lambda_W(\gamma + \epsilon) \rightarrow \Lambda(\gamma + \epsilon) < \log d$. As a consequence, there exists M large enough that $\Lambda_W(\gamma + \epsilon) \leq \log d$.

Now, for $m \geq M$, Y^m stochastically dominates W_M . Let U be a $[0, 1]$ -uniform random variable. For each node $u \in T_\infty$, $F(Y_u)$ is a $[0, 1]$ -random variable, and $G_M^{-1} \circ F(Y_u) \leq Y_u$ is distributed as W . Let T_n^M be the subtree of T_n consisting of nodes u with $N_u \geq M$. There are at most $\#T_n^M \cdot d$ hanging subtrees with $N_u < M$, each one of effective size at most M . It follows that $\#T_n \leq \#T_n^M(1 + dM)$ and

$$\limsup_{n \rightarrow \infty} \frac{\log \#T_n}{\log n} \leq \limsup_{n \rightarrow \infty} \frac{\log \#T_n^M + \log(1 + dM)}{\log n} \leq \gamma + \epsilon,$$

where the last inequality follows from the choice of M and Lemma 9. Since ϵ was arbitrary, the proof is complete. \square

5 Applications

5.1 Variations on binary search trees

Binary search trees [47] are search trees built on a set of keys $\{1, 2, \dots, n\}$. Given a permutation $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ of the keys, the first element σ_1 is stored at the root of a binary tree. The set of keys is then partitioned according to their values into $\{\sigma_i : \sigma_i < \sigma_1\}$ and $\{\sigma_i : \sigma_i > \sigma_1\}$. Both subsets are then treated recursively to form the left and right subtrees of the root, respectively.

If the permutation is taken uniformly at random from the set of permutations of $\{1, \dots, n\}$, the tree is called a random binary search tree. This model is of great interest, particularly because of its ubiquity in computer science as, e.g., the tree emerging from the branching structure of quicksort [45]. In this model of randomness, σ_1 is an element of $\{1, \dots, n\}$ taken uniformly at random and hence the sizes of the left and right subtrees are distributed as $\text{Bin}(n-1, U)$ and $\text{Bin}(n-1, 1-U)$, respectively, where U is a $[0, 1]$ -uniform random variable. More precisely, writing (N_1, N_2) for a vector that is distributed as a multinomial $(n-1; U, 1-U)$, the vector of interest is

$$\mathcal{X}^n = \left(\left(1, -\log \left(\frac{N_1}{n} \right) \right), \left(1, -\log \left(\frac{N_2}{n} \right) \right) \right). \quad (10)$$

One can show that the conditions required to apply Theorem 2, are satisfied. In particular:

Lemma 11. *Let \mathcal{X}^n be defined by (10). Let $\mathcal{X} = ((1, -\log U), (1, -\log(1 - U)))$. Then, $\Lambda_{\mathcal{X}^n} \rightarrow \Lambda_{\mathcal{X}}$ everywhere.*

Proof. The weights are irrelevant here, and we consider

$$(E_1^n, E_2^n) = (\log n - \log N_1, \log n - \log N_2)$$

only. Observe that (N_1, N_2) is distributed as $(\lfloor nU \rfloor, \lfloor n(1 - U) \rfloor)$, where U is a $[0, 1]$ -uniform random variable. For all $\mu_1, \mu_2 \in \mathbb{R}$,

$$M_n(\mu_1, \mu_2) \stackrel{\text{def}}{=} \mathbf{E} [e^{\mu_1 E_1 + \mu_2 E_2}] = \mathbf{E} \left[\left(\frac{\lfloor nU \rfloor}{n} \right)^{-\mu_1} \cdot \left(\frac{\lfloor n(1 - U) \rfloor}{n} \right)^{-\mu_2} \right].$$

Thus,

$$\left(\frac{\lfloor nU \rfloor}{n} \right)^{-\mu_1} \cdot \left(\frac{\lfloor n(1 - U) \rfloor}{n} \right)^{-\mu_2} \xrightarrow{n \rightarrow \infty} U^{-\mu_1} \cdot (1 - U)^{-\mu_2}$$

almost surely. Therefore, if $\mu_1 < 1$ and $\mu_2 < 1$, by the bounded convergence theorem,

$$M_n(\mu_1, \mu_2) \rightarrow \mathbf{E} [U^{-\mu_1} \cdot (1 - U)^{-\mu_2}].$$

If, on the other hand, either $\mu_1 \geq 1$ or $\mu_2 \geq 1$, then by Fatou's Lemma [see, e.g., 12],

$$\liminf_{n \rightarrow \infty} M_n(\mu_1, \mu_2) \geq \mathbf{E} [U^{-\mu_1} \cdot (1 - U)^{-\mu_2}] = \infty.$$

Thus, we have convergence everywhere in $\mathbb{R} \cup \{+\infty\}$, which completes the proof. \square

Hence, for this model, $E = -\log U$ and $Z = 1$. The random variable E is then distributed as an exponential random variable with mean 1 and Theorem 2 immediately implies the following theorem of Devroye [27].

Theorem 7 (Devroye [27]). *Let T_n be a random binary search tree. Let H_n be its height. Then $H_n \sim c \log n$, in probability as $n \rightarrow \infty$, where $c = 1/\rho = 4.311\dots$ and ρ is the smallest solution of $\rho - 1 - \log \rho = \log 2$.*

Remark. In the following, we will not prove the convergence of the cumulant generating functions any more, and only refer to Lemma 11.

The value $4.311\dots \log n$ is fairly large compared to $\log_2 n$, the height of a complete binary tree with n nodes. As this value represents the worst case search time, various methods have been used to shrink it and hence obtain more efficient searching data structures. One of the easiest approaches is to ensure that the splits are more balanced towards $(1/2, 1/2)$. One way to achieve more balanced splits is to use the median of $2k + 1$ keys as a pivot [66]. When k is fixed, the split at every node is still given by (10) but now (N_1, N_2) is distributed as a multinomial $(n - 1; U_k, 1 - U_k)$ and U_k is a beta $(k + 1, k + 1)$ random variable. Again, we see that for $\mathcal{X} = ((1, -\log U_k), (1, -\log(1 - U_k)))$, $\Lambda_{\mathcal{X}^n} \rightarrow \Lambda_{\mathcal{X}}$ everywhere as $n \rightarrow \infty$. This suffices for the hypothesis of Theorem 2 to hold.

Theorem 8 (Devroye [29]). *Let T_n be a binary search tree built with the medians of $2k+1$ keys as pivots. Then the height H_n of T_n satisfies $H_n \sim c_k \log n$ in probability as $n \rightarrow \infty$, where c_k is the unique solution of*

$$\frac{s}{c_k} + \sum_{i=k+1}^{2k+1} \log \left(1 - \frac{s}{i}\right) = \log 2,$$

and s is implicitly defined by

$$\frac{1}{c_k} = \sum_{i=k+1}^{2k+1} \frac{1}{i-s}.$$

If k is fixed, we can make c_k close to $1/\log 2$. However, for each k we have $c_k > 1/\log 2$. One can improve this by taking values of k that depend on the number of keys stored in a subtree. If $k \rightarrow \infty$ as $n \rightarrow \infty$, we see that $\mathcal{X}^n \rightarrow \mathcal{X} = ((1, \log 2), (1, \log 2))$ a.s. as $n \rightarrow \infty$. Theorem 2 then implies that $H_n \sim \log_2 n$, in probability as $n \rightarrow \infty$. This strengthens the theorem of Martínez and Roura [51] which asserts that the average depth, in this case, is asymptotic to $\log_2 n$ [see also 66].

5.2 Digital search trees

This example is a weighted version of the one of Broutin et al. [17]. We consider tries on a finite alphabet $\mathcal{A} = \{1, 2, \dots, d\}$ with the Bernoulli model of randomness: each datum consists of an infinite sequence $A^i = A_1^i, A_2^i, \dots$ of i.i.d. random elements of \mathcal{A} [23, 39, 65]. A string A^i corresponds to an infinite path in a d -ary tree defined in the following way: from the root, take the A_1^i -th first child, next the A_2^i -th, and so forth. We prune the subtrees of each node that contain only one single string. The remaining tree is the trie associated with the n strings.

For tries, there is no deterministic bound on the height of a trie built from n or even two strings. Neither Theorem 1 nor Theorem 2 applies to tries. Various techniques have been used to shrink the height of tries such as PATRICIA [53] and digital search trees [22, 48]. See also the recent survey by Flajolet [37]. We focus on digital search trees, which are, strictly speaking, not *search trees*. 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. In this “pebbled” version of tries, a string, taken at random, is associated to the root. Then, the $n-1$ remaining strings are distributed to the k subtrees depending on the value of their first character. The tree is then built recursively.

In a computer, the characters are coded in binary. The *cost* of a character in terms of bit comparisons is then the length of its binary code. The model of pebbled tries has been studied by Broutin et al. [17] in the case where all k characters have

the same cost. However, if one uses an optimal code (one that minimizes the costs of the characters), the lengths of the codewords depend on the character, and hence the costs of characters vary. Also, in such a code, the length of a codeword is obviously dependent of the probability that the corresponding character occurs (prefix codes of Huffman [46]). Hence, this model of pebbled tries built with Huffman coded characters is a perfect application for Theorem 2.

Let p_i be the probability that character i occurs at some fixed position of a string. Let ℓ_i be the length of the binary codeword for character i . Then, at a node u with $N_u = n + 1$, the split \mathcal{V}^n is distributed as a multinomial(n, p_1, p_2, \dots, p_d) random vector. The weights (Z_1, \dots, Z_d) are deterministic and equal to $(\ell_1, \ell_2, \dots, \ell_d)$. Now, $\mathcal{V}^n \rightarrow (p_1, p_2, \dots, p_d)$ almost surely, and hence it is easily checked that the required conditions on the random variables are satisfied with $X = (\ell_K, p_K)$ where K is uniform in $\{1, \dots, d\}$. It follows that

$$\Lambda(\lambda, \mu) = -\log k + \log \left(\sum_{i=1}^d e^{\lambda \ell_i - \mu \log p_i} \right).$$

Also, since for all i , $\ell_i > 0$ and $\log p_i < 0$ (or there is a.s. only one character in the alphabet and the tree is degenerate), $e^{\Lambda + \log d}$ is a sum of positive convex functions whose gradient spans $(0, \infty)^2$. As a result, for $\alpha, \rho \in (0, \infty)$, there exist λ and μ for which $\sup_{\lambda', \mu'} \{\lambda' \alpha + \mu' \rho - \Lambda(\lambda', \mu')\} = \lambda \alpha + \mu \rho - \Lambda(\lambda, \mu)$ which are given implicitly by

$$\alpha = \frac{\sum_{i=1}^d \ell_i e^{\lambda \ell_i} p_i^{-\mu}}{\sum_{i=1}^d e^{\lambda \ell_i} p_i^{-\mu}} \quad \text{and} \quad \rho = \frac{-\sum_{i=1}^d \log p_i e^{\lambda \ell_i} p_i^{-\mu}}{\sum_{i=1}^d e^{\lambda \ell_i} p_i^{-\mu}}.$$

Then, by Theorem 1, the height of the pebbled trie is asymptotic to $c \log n$ in probability, where c the maximum value of α/ρ along the curve

$$\lambda \alpha + \mu \rho = \log \left(\sum_{i=1}^d e^{\lambda \ell_i - \mu \log p_i} \right).$$

Numerical values can easily be obtained for every set of parameters $\{(p_i, \ell_i), 0 \leq i \leq d\}$.

5.3 Pebbled TST

In the same vein, we can study the height of a pebbled version of ternary search trees (TST). The (non-pebbled) TST structure introduced by Bentley and Sedgewick [5] uses early ideas of Clampett [19] to improve on array-based implementations of tries. If an array is used to implement the branching structure of a node, the number of null pointers can become an issue when the alphabet is large. In TSTs, instead of the usual array, the node structure consists of a binary search tree (BST), therefore forcing small branching factors and limiting the amount of null pointers. The TST may be seen as a hybrid structure combining tries and binary search trees: the

high level structure is still that of a trie ; only the structure of a node and the way character matching are handled changes. The additive parameters of TSTs have been studied thoroughly by Clément, Flajolet, and Vallée [20, 21], but the question of the height was left open.

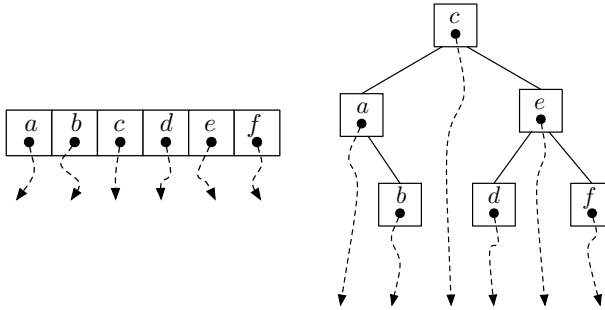


Figure 2. The structure of a node in an array-based trie (left) and TST (right) over the alphabet $\{a, b, c, d, e, f\}$. The pointers used for the high level trie structure are dashed.

We now describe the modified pebbled version. Let $\{A^1, A^2, \dots, A^n\}$ be the set of strings, where we write $A^i = A_1^i A_2^i \dots$ for the sequence of characters. We distinguish the *nodes* of the trie structure from the *slots* of the local binary search trees. As shown in Figure 2, each node contains d slots. The nodes at the same distance j from the root are said to be *at level j* . At level j , the key used for the comparisons is the j -th character of the sequences.

The tree is built by assigning the sequences to the first empty *slot* as they come along in order A^1, A^2, \dots . The first string A^1 is stored in the first slot of the root of the TST and partitions the following sequences with respect to their first characters into three groups $\{A^i : A_1^i < A_1^1\}$, $\{A^i : A_1^i = A_1^1\}$ and $\{A^i : A_1^i > A_1^1\}$. In general, a sequence A^j stored in a slot at level ℓ induces a partition of further strings into $\{A^i : A_\ell^i < A_\ell^j\}$, $\{A^i : A_\ell^i = A_\ell^j\}$ and $\{A^i : A_\ell^i > A_\ell^j\}$. Given the TST built from the first $m - 1$ sequences, we may think of A^m as an item moving down the tree: it starts at the root and moves as indicated by the sequences stored in the slots encountered, the comparisons being done on the ℓ -th character at level ℓ . It changes level only upon finding a matching character, in the other cases, it moves in the slots of the *same* node until it eventually finds either an empty slot, or a matching character. The sequence A^m is stored in the first empty slot encountered, and is used to partition the next strings according the same rules. As a consequence, the structure of a node is that of a structure of binary search tree (see Figures 2 and 3).

Assume now that the strings are independent sequences of i.i.d. characters where a character $a \in \{1, \dots, d\}$ has probability $p_a > 0$. We are interested in the height of a pebbled TST built from n of these independent sequences. Consider a node u whose subtree stores $n + 1$ strings. As in the previous section, the split vector at u , (N_1, N_2, \dots, N_d) , is clearly multinomial(n, p_1, \dots, p_d). Looking at the high level trie structure, the edges may be seen as being weighted by the number of edges in the local binary search tree structure (Figure 3). Clearly, the cost of the edge leading to a character a is the 1 plus the depth of the node labeled a in the BST of the

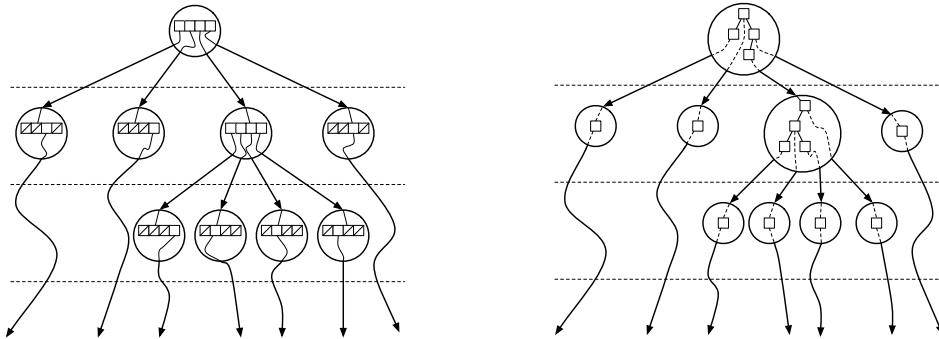


Figure 3. The outer structure of trie (left) and the expanded binary search tree structure of the nodes of a TST on an alphabet of size four. The *nodes* are shown as circles whereas the *slots* are represented by squares.

node considered. Let Z_a^n be the random variable accounting for this value. Then the vector of interest is

$$\mathcal{X}^n = \left(\left(Z_1^n, -\log \left(\frac{N_1}{n} \right) \right), \dots, \left(Z_d^n, -\log \left(\frac{N_d}{n} \right) \right) \right).$$

The random variable Z_a^n has been studied by Clément et al. [20] and Archibald and Clément [1]. In particular they studied the expected values and variances of $\{Z_a^n, 1 \leq a \leq d\}$. However, we need information about the distributions of Z_a^n and their limits as $n \rightarrow \infty$. Let τ be the smallest n for which $\{A_1^i, 0 \leq i \leq n\}$ contains a copy of each character. Then, for each $n \geq \tau$, the distribution of Z_a^n is that of $Z_a = Z_a^\tau$, independent of n . The random variable τ is a stopping time and $\mathbf{P}\{\tau \geq n\} \leq (1 - \min\{p_i, 1 \leq i \leq d\})^n$. This proves that τ is a.s. finite and that $Z_a^n \rightarrow Z_a$, in distribution. Then, with \mathcal{X} distributed as $((Z_1, -\log p_1), \dots, (Z_d, -\log p_d))$, one can show that $\Lambda_{\mathcal{X}^n} \rightarrow \Lambda_{\mathcal{X}}$ everywhere as $n \rightarrow \infty$.

This is sufficient for Theorem 2 to apply. The height of the pebbled TST is asymptotic to $c \log n$ in probability, where c is the maximum value of α/ρ in $\{(\alpha, \rho) : \Lambda^*(\alpha, \rho) \leq \log d\}$, and Λ^* is the Cramér function associated with $X = (Z_K, -\log p_K)$ where K is uniform in $\{1, \dots, d\}$. Numerical values can be obtained for examples of $\{p_i, 1 \leq i \leq d\}$.

Remark. The height of the non-pebbled version of TST requires more care and studied by Broutin and Devroye [15, 16].

Example: symmetric pebbled TST. We can obtain more concrete results for symmetric TST, i.e., when $p_1 = p_2 = \dots = p_d = 1/d$. In such a case, $E = \log d$ almost surely, and only Z matters. Also, Z is distributed as one plus the depth of a random node in a random binary search tree of size d . This distribution is known

exactly [18, 49]:

$$\mathbf{P} \{Z = k\} = \frac{2^{k-1}}{d \cdot d!} \sum_{j=k}^d \begin{bmatrix} d \\ j \end{bmatrix} \quad \forall 1 \leq k \leq d, \quad (11)$$

where $\begin{bmatrix} n \\ k \end{bmatrix}$ is the Stirling number of the first kind with parameters n and k [43, 63]. Using (11) one can compute Λ , Λ^* , and values of the constant c_d such that the height of a pebbled TST of size n is asymptotic to $c_d \log n$ in probability, as $n \rightarrow \infty$. Numerical values are found in Table 1.

| d | 2 | 3 | 4 | 5 | 10 | 20 |
|-------|------------|------------|------------|------------|------------|------------|
| c_d | 2.88539... | 2.63747... | 2.64612... | 2.69384... | 2.91785... | 3.14445... |

Table 1. Some numerical values for the height of symmetric pebbled TST.

5.4 Skinny cells in k -d trees

We consider the k -d tree introduced by Bentley [4]. This geometric structure generalizes binary search trees to multidimensional data sets. Given a set \mathcal{D} of d -dimensional data points $\{Y^1, Y^2, \dots, Y^n\}$, where $Y^i = (y_1^i, \dots, y_d^i)$ for all i , we recursively build the following binary tree structure partitioning the data set using comparisons of some of their components. The first datum Y^1 is stored at the root. The remaining data are processed as follows: $\{Y^i : i \geq 2, y_1^i \leq y_1^1\}$ and $\{Y^i : y_1^i > y_1^1\}$ are assigned respectively to the left and right subtrees, and both subtrees are recursively built using the same method. The comparisons are done in a cyclical way depending on the depth of the node at which they occur: the key used at a node at depth ℓ is the $(\ell \bmod d + 1)$ -st component of a vector. For a more complete account on k -d trees see [42] or [61, 62].

When the data points are i.i.d. $[0, 1]^d$ -uniform random variables, one can see a k -d tree as a random refining partition of $[0, 1]^d$. The root represents $[0, 1]^d$, and more generally, a node u represents the set of points $x \in [0, 1]^d$ that would be stored in its subtree if they were data points inserted after u . Therefore, each cell is split into two along a dividing line, on which lies one of the points Y^i , and whose direction changes in a cyclical way. The cells are obviously rectangular. Let C_u be the cell associated with a node u . Let $L_1(u), L_2(u), \dots, L_d(u)$ be the its lengths with respect to the d dimensions. We are interested in the worst case ratio of two dimensions of a cell. For example, if $d = 2$, this is the worst case ratio length over width. By symmetry, since d is bounded, we can always consider the worst case of the first two dimensions, L_1 and L_2 . Such a parameter is of great importance in applications. Indeed, for partial match queries, the running times of algorithms depend on the shape of the cells, and in particular on how close they are to squares [31, 33, 38, 52]. We prove the following:

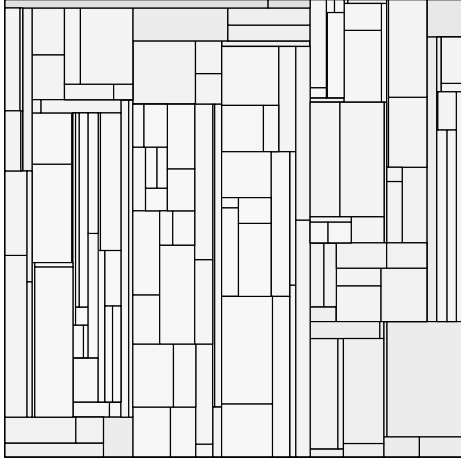


Figure 4. A (randomly generated) k -d tree on 150 points uniformly distributed in $[0, 1]^2$.

Theorem 9. Let T_n be a k -d tree built from n i.i.d. $[0, 1]^d$ -uniform random points. Let

$$R_n \stackrel{\text{def}}{=} \max \left\{ \frac{L_1(u)}{L_2(u)} : u \in T_n \right\}.$$

Then $R_n = n^{c_d + o(1)}$ in probability, as $n \rightarrow \infty$. Furthermore, c_d is the maximum value of α/ρ in the set $\{(\alpha, \rho) : \lambda\alpha + \mu\rho + \log((1 - \mu)^2 - \lambda^2) + (d - 2)\log(1 - \mu) \leq d \log 2\}$, where

$$\mu = 1 - \frac{\rho(d - 1) + \sqrt{\rho^2 + \alpha^2 d(d - 2)}}{\rho^2 - \alpha^2} \quad \text{and} \quad \lambda = -\alpha \cdot \frac{d(d - 2) - 2\rho(d - 1)(1 - \mu)}{(\rho^2 - \alpha^2)(\rho(1 - \mu) - (d - 2))}.$$

In particular, $c_d < 1$ for all $d \geq 2$.

Remark. The same result holds for the shape of cells in the quadtrees of Finkel and Bentley [36]. Indeed, the distribution of (Z, E) given by (12) also describes the shape of cells in quadtrees.

| d | 2 | 3 | 5 | 10 | 40 | 100 |
|-------|------------|------------|------------|------------|------------|------------|
| c_d | 0.86602... | 0.79047... | 0.71246... | 0.63483... | 0.54976... | 0.52442... |

Table 2. Some numerical values for the constant c_d describing the asymptotic values of R_n in a d -dimensional k -d tree.

Proof. We intend to express the maximum ratio as a weighted height. Since in k -d trees, *not* all the levels in the tree are equivalent, proceed in two stages: first consider the levels of depth $0 \bmod d$, and consider the levels $1 \bmod d$ since for the other $d - 2$ choices $\bmod d$, the ratio L_1/L_2 is not modified.

If we group the levels by bunches of d , then all bunches behave similarly. We obtain a 2^d -ary tree. In this tree, each node corresponds to a rectangular region

of $[0, 1]^d$, and its children are the result of its split into 2^d subregions. The points come uniformly at random, and hence the probability that a region is hit is its area. Figure 5 illustrates the way we turn the question about the ratio into the weighted height of some tree.

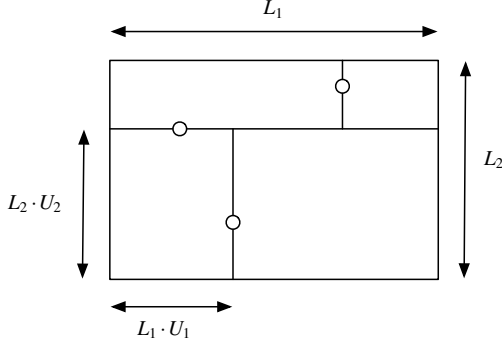


Figure 5. The way a node is interpreted when $d = 2$.

The area of a rectangle is the product of $[0, 1]$ -uniform random variables determining the splits, and the ratio L_1/L_2 is the ratio of two products of some of these random variables. More precisely, let U_i , $1 \leq i \leq d$ be i.i.d. $[0, 1]$ -uniform random variables. Taking the logarithms of the areas and of the ratios, we see that the increments are distributed like

$$E = -\sum_{i=1}^d \log U_i \quad \text{and} \quad Z = -\log U_1 + \log U_2. \quad (12)$$

To use Theorem 2, we need to compute Λ^* associated with $X = (Z, E)$. We start with the moment generating function: for any real numbers λ and μ , by independence,

$$\mathbf{E} \left[e^{\lambda Z + \mu E} \right] = \mathbf{E} \left[U_1^{-\lambda - \mu} \cdot U_2^{\lambda - \mu} \prod_{i=3}^d U_i^{-\mu} \right] = \mathbf{E} \left[U_1^{-\lambda - \mu} \right] \cdot \mathbf{E} \left[U_2^{\lambda - \mu} \right] \cdot \prod_{i=3}^d \mathbf{E} \left[U_i^{-\mu} \right].$$

As a consequence,

$$\mathbf{E} \left[e^{\lambda Z + \mu E} \right] = \begin{cases} \frac{1}{(1 - \mu)^2 - \lambda^2} \cdot \frac{1}{(1 - \mu)^{d-2}} & \text{if } \mu < 1 - |\lambda| \\ \infty & \text{otherwise.} \end{cases}$$

It follows that

$$\Lambda(\lambda, \mu) = \begin{cases} -\log((1 - \mu)^2 - \lambda^2) - (d - 2) \log(1 - \mu) & \text{if } \mu < 1 - |\lambda|, \\ \infty & \text{otherwise.} \end{cases}$$

So $\mathcal{D}_\Lambda = \{(\lambda, \mu) : \mu < 1 - |\lambda|\}$. To compute Λ^* , we find the maximum of $(\lambda, \mu) \mapsto \lambda\alpha + \mu\rho - \Lambda(\lambda, \mu)$, which is achieved for λ and μ such that

$$\begin{cases} \alpha = \frac{\partial \Lambda(\lambda, \mu)}{\partial \lambda} = \frac{2\lambda}{(1 - \mu)^2 - \lambda^2} & \text{and} \\ \rho = \frac{\partial \Lambda(\lambda, \mu)}{\partial \mu} = \frac{2(1 - \mu)}{(1 - \mu)^2 - \lambda^2} + \frac{d - 2}{1 - \mu}, \end{cases} \quad (13)$$

if such a point exists. For $d \geq 2$, such a point does exist and the system above has solution

$$\begin{cases} \mu &= 1 - \frac{\rho(d-1) \pm \sqrt{\rho^2 + \alpha^2 d(d-2)}}{\rho^2 - \alpha^2} \\ \lambda &= \frac{\alpha}{\rho^2 - \alpha^2} \cdot \frac{2\rho(\mu-1)(d-1) + d(d-2)}{\rho(\mu-1) + d-2}. \end{cases} \quad (14)$$

Observe that the mere fact that this is a solution of (13) ensures that $(\lambda, \mu) \in \mathcal{D}_\Lambda$. Hence, we have $\mathcal{D}_{\Lambda^*} = \mathbb{R}^2$, and by Theorem 2, c is the maximum of α/ρ with (α, ρ) in the set

$$\Psi(d \log 2) = \{(\alpha, \rho) : \lambda\alpha + \mu\rho + \log((1-\mu)^2 - \lambda^2) + (d-2)\log(1-\mu) \leq d \log 2\},$$

where λ and μ are defined by (14). Since this only accounts for the levels whose depths are $0 \pmod d$, this gives only a lower bound on the actual weighted depth of the tree. However, one can find a matching upper bound easily. Indeed, to account for the levels $1 \pmod d$, it suffices to group the levels starting at level 1. Doing this, the distribution for E and Z remains unchanged, but the ratio L_1/L_2 is now off by one single multiplicative factor of $1/U$. It follows immediately that the weighted height on the levels $1 \pmod d$ is also $c \log n$, which finishes the proof of Theorem 9. \square

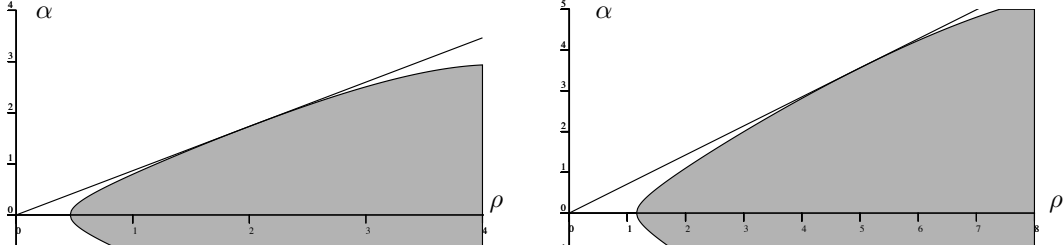


Figure 6. The sets $\{(\rho, \alpha) : \Lambda^*(\alpha, \rho) \leq d \log 2\}$ together with the lines of maximum slopes for the maximum ratio of two dimensions of a cell in a k -d tree in \mathbb{R}^2 and \mathbb{R}^5 .

Corollary 1. For $d = 2$, we have $c_2 = \sqrt{3}/2$.

Proof. If $d = 2$, μ and λ simplify and we have

$$\mu = 1 - \frac{2\rho}{\rho^2 - \alpha^2} \quad \text{and} \quad \lambda = \frac{2\alpha}{\rho^2 - \alpha^2}.$$

The condition $\mu - 1 < -|\lambda|$ is equivalent to $\rho > |\alpha|$, so the set to consider is $\{(\alpha, \rho) : \rho > |\alpha|\}$. It follows that

$$\Lambda^*(\alpha, \rho) = \frac{2\alpha^2}{\rho^2 - \alpha^2} + \rho - \frac{2\rho^2}{\rho^2 - \alpha^2} + \log\left(\frac{2\lambda}{\alpha}\right) = \rho - 2 + 2 \log 2 - \log(\rho^2 - \alpha^2).$$

Therefore, we need to find the maximum value of α/ρ subject to $\rho - 2 \leq \log(\rho^2 - \alpha^2)$. The optimum is clearly obtained on the boundary of the set, i.e., for $\rho - 2 = \log(\rho^2 - \alpha^2)$. Then, we have

$$\frac{\alpha}{\rho} = \sqrt{1 - \frac{e^{\rho-2}}{\rho^2}},$$

which is maximum when the derivative vanishes:

$$\frac{d}{d\rho} \left(\frac{\alpha}{\rho} \right) = \frac{-e^{\rho-2} \left(\frac{1}{\rho^2} - \frac{2}{\rho^3} \right)}{2\sqrt{1 - \frac{e^{\rho-2}}{\rho^2}}} = 0.$$

This happens when $\rho = 2$ and then $\alpha/\rho = \sqrt{3}/2$. (Note that $\rho > |\alpha|$.) □

Remark. We have $\lim_{d \rightarrow \infty} c_d = 1/2$. Indeed, the optimal point is at $\rho = d$. Using $\alpha \sim cd$, note that

$$1 - \mu = \frac{1}{1-c} + o(1) \quad \text{and} \quad \lambda = \frac{1}{1-c} + o(1).$$

So,

$$\Lambda^*(\alpha, \rho) = \frac{cd}{1-c} + d - \frac{d}{1-c} - d \log(1-c) + o(d).$$

Finally, if $c = 1/2$, we have $\Lambda^*(\alpha, \rho) = d \log 2 + o(d)$.

5.5 Skinny cells in relaxed k -d trees

The model of k -d trees described above is a bit constrained due to the cyclical way in which the components of a vector are used as keys. In particular, k -d trees are data structures that are mostly static: they are built *once*, and then used to perform multiple queries on the *same* data. To cope with the issue of updating k -dimensional search structures, Duch, Estivill-Castro, and Martínez [34] introduced a randomized data structure that is similar to k -d trees, but are not subject to the same constraints. The symmetry is reintroduced by choosing the index of the component used as a key at random when a node is inserted in the structure. This tree structure is naturally called relaxed k -d tree. The structure leads to update algorithms that are easy to implement, but it is not known whether the structure is indeed efficient.

Theorem 10. *Let T_n be a relaxed k -d tree built from n i.i.d. $[0, 1]^2$ -uniform random points. Let*

$$R_n = \max \left\{ \frac{L_1(u)}{L_2(u)} : u \in T_n \right\}.$$

Then $R_n = n^{1+o(1)}$ in probability, as $n \rightarrow \infty$.

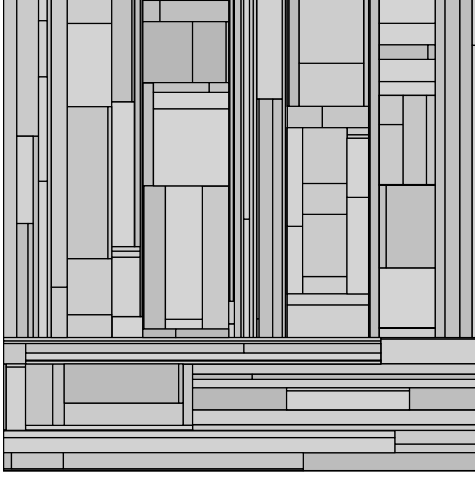


Figure 7. A (randomly generated) relaxed k -d tree on 150 uniformly distributed points in $[0, 1]^2$. One can notice in at the first glance that the cells look skinnier than those shown in Figure 7.

Remark. The cells of 2-dimensional relaxed k -d trees are skinnier than those of k -d trees. This explains why partial match queries are more costly in relaxed k -d trees [32, 33] than in k -d trees [38].

Proof of Theorem 10. Consider a cell that does not contain any data point. In the tree, it corresponds to an external node u . A new incoming point falls in this cell with probability $L_1(u) \cdot L_2(u)$. If this happens, two new cells are created. Clearly, the cell gets divided uniformly. Let U be a $[0, 1]$ -uniform random variable. Then, if the number N_u of nodes contained in the subtree rooted at u is n , the sizes of the subcells are distributed as a multinomial($n - 1, U, 1 - U$) random vector.

As in the case of k -d trees, the ratio L_1/L_2 is either multiplied or divided by U . Each of this cases happens with probability $1/2$ at every split, so with the additive formalism, the increase in $\log(L_1/L_2)$ is

$$Z(U) = \begin{cases} -\log U & \text{w.p. } 1/2 \\ \log U & \text{w.p. } 1/2. \end{cases}$$

Again, $\mathcal{X}^n \rightarrow ((Z(U), -\log U), (Z(1 - U), -\log(1 - U)))$ almost surely as $n \rightarrow \infty$. Hence we have $X = (Z(U), -\log U)$, and for $\lambda, \mu \in \mathbb{R}$,

$$\mathbf{E} \left[e^{\lambda Z + \mu E} \right] = \frac{1}{2} \mathbf{E} \left[U^{-\lambda - \mu} \right] + \frac{1}{2} \mathbf{E} \left[U^{\lambda - \mu} \right].$$

Therefore, we have

$$\mathbf{E} \left[e^{\lambda Z + \mu E} \right] = \begin{cases} \infty & \text{if } \lambda \geq 1 - \mu \text{ or } \lambda \leq \mu - 1 \\ \frac{1 - \mu}{(1 - \mu)^2 - \lambda^2} & \text{otherwise,} \end{cases}$$

and

$$\Lambda(\lambda, \mu) = \begin{cases} \infty & \text{if } \lambda \geq 1 - \mu \text{ or } \lambda \leq \mu - 1 \\ \log(1 - \mu) - \log((1 - \mu)^2 - \lambda^2) & \text{otherwise.} \end{cases}$$

The maximum of $(\lambda, \mu) \mapsto \lambda\alpha + \mu\rho - \Lambda(\lambda, \mu)$ is achieved for λ and μ , with $\mu - 1 \leq \lambda \leq 1 - \mu$, satisfying

$$\alpha = \frac{\partial\Lambda(\lambda, \mu)}{\partial\lambda} = \frac{2\lambda}{(1-\mu)^2 - \lambda^2} \quad \text{and} \quad \rho = \frac{\partial\Lambda(\lambda, \mu)}{\partial\mu} = \frac{2(1-\mu)}{(1-\mu)^2 - \lambda^2} - \frac{1}{1-\mu},$$

if such a point exists. This implies in particular that

$$\begin{aligned} \rho^2 - \alpha^2 &= \frac{4(1-\mu)^2}{((1-\mu)^2 - \lambda^2)^2} + \frac{1}{(1-\mu)^2} \\ &\quad - \frac{4}{(1-\mu)^2 - \lambda^2} - \frac{4\lambda^2}{((1-\mu)^2 - \lambda^2)^2} \\ &= \frac{1}{(1-\mu)^2} \geq 0, \end{aligned}$$

for $\mu - 1 \leq \lambda \leq 1 - \mu$. Then, provided $|\alpha| < |\rho|$, the solution is given by

$$\mu = 1 - \frac{1}{\sqrt{\rho^2 - \alpha^2}} \quad \text{and} \quad \lambda = \frac{\alpha}{\rho^2 - \alpha^2} \cdot \frac{1}{1 + (1-\mu)\rho}. \quad (15)$$

If $|\alpha| \geq |\rho|$, then $\Lambda^*(\alpha, \rho) = \infty$. Indeed, assume that $\alpha = \rho + \delta$, for some $\delta > 0$ (a symmetric argument holds when $\alpha = -\rho - \delta$). Let $\epsilon > 0$, and write $\lambda_0 = 1 - \mu - \epsilon \leq 1 - \mu$. Then,

$$\begin{aligned} \Lambda^*(\alpha, \rho) &\geq \lambda_0\alpha + \mu_0\rho - \Lambda(\lambda_0, \mu_0) \\ &= \lambda_0\delta + \rho - \rho\epsilon + \log 2 + \log \epsilon + O(1/\lambda_0) \rightarrow \infty \end{aligned}$$

as $\lambda_0 \rightarrow \infty$. It follows by Theorem 2 that c is the maximum value of α/ρ in the set

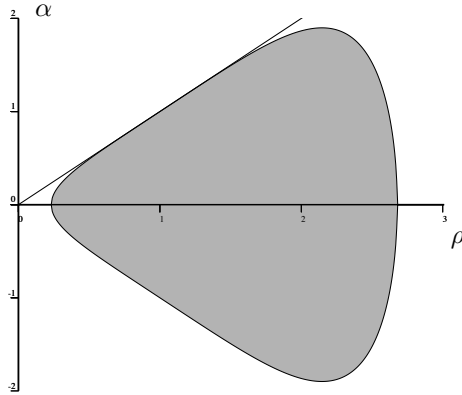


Figure 8. The set $\{(\rho, \alpha) : \Lambda^*(\alpha, \rho) \leq \log 2\}$ and the line of maximum slope for the maximum ratio of two dimensions of a cell in a relaxed k -d tree in \mathbb{R}^2 .

$$\Psi(\log 2) = \{\lambda\alpha + \mu\rho - \log(1-\mu) + \log((1-\mu)^2 - \lambda^2) \leq \log 2\},$$

where λ and μ are defined in (15).

Since $\mathcal{D}_{\Lambda^*} \subset \{|\alpha| \leq |\rho|\}$, it is clear that $c \leq 1$, so we only need to prove that $c \geq 1$. In particular, it suffices to find points $(\alpha, \rho) \in \Psi(\log 2)$ with α/ρ arbitrarily

close to 1. Because μ and hence Λ^* is not properly defined for $\alpha = \rho$, we consider $\Lambda^*(1 - \epsilon, 1)$ for $\epsilon \in (0, 1)$. One can verify that, as $\epsilon \rightarrow 0$,

$$\Lambda^*(1 - \epsilon) = \log 2 - \sqrt{2\epsilon} + o(\sqrt{\epsilon}),$$

and therefore, $(1 - \epsilon, 1) \in \Psi(\log 2)$ for ϵ small enough. This proves that $c \geq 1 - \epsilon$ for any small enough $\epsilon > 0$ and hence, by Theorem 2, that $\log R_n \sim \log n$ in probability as $n \rightarrow \infty$. \square

5.6 d -ary pyramids

Allowing $Z = -\infty$ can be useful when one needs to exclude some tree paths in the definition of the height. Pyramids [7, 41] are built incrementally as follows. A d -ary pyramid of size 1 is a single node. Given a d -ary pyramid of size n , pick a node u uniformly at random among those that have degree at most $d - 1$. The next node becomes a child of u . The height of a 2-ary pyramid has been studied by Mahmoud [50]. Biggins and Grey [11] obtained it for $d \geq 2$.

Theorem 11 ([11, 50]). *The height H_n of a d -ary pyramid of size n is $H_n \sim \log n / (\gamma \rho_0)$ in probability, as $n \rightarrow \infty$, where γ is given by*

$$\sum_{i=1}^d \frac{1}{(1 + \gamma)^i} = 1, \quad (16)$$

and ρ_0 is defined as the smallest root of

$$\rho \cdot \mu = \log \left(\sum_{i=1}^d \frac{1}{(1 - \mu)^i} \right) \quad \text{where} \quad \rho = \frac{\sum_{i=1}^d i(1 - \mu)^{-i-1}}{\sum_{i=1}^d (1 - \mu)^{-i}}, \quad (17)$$

where $\mu < 1$. Numerical values are given in Table 3.

| d | 2 | 3 | 5 | 10 |
|------------------------|-----------------|-----------------|-----------------|-----------------|
| γ | 0.6180339880... | 0.8392867552... | 0.9659482366... | 0.9990186327... |
| ρ_0 | 0.4056580492... | 0.3759749401... | 0.3684055189... | 0.3678801695... |
| $(\gamma \rho_0)^{-1}$ | 3.988664818... | 3.169061969... | 2.810088635... | 2.720946695... |

Table 3. Some numerical values for the height of d -ary pyramids of effective size n . It is not surprising to observe that $(\gamma \rho_0) \rightarrow 1/e$ as $d \rightarrow \infty$, since the height of the random recursive tree is asymptotic to $e \log n$.

We derive Theorem 11 using our framework. Random recursive trees [64] are ∞ -ary pyramids. A random recursive tree of size one consists of a single node. A random recursive tree of size $n + 1$ is built from one of size n by picking a uniform random node u , and adding a new node as a child of u . Clearly, conditioning on the new node being a child of an unsaturated node u , u is still uniform among the

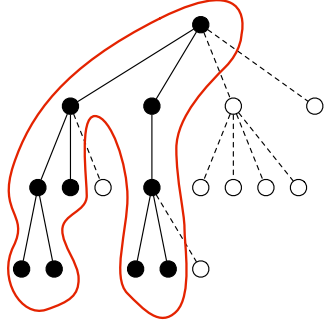


Figure 9. A 2-ary pyramid seen as the subtree of a random recursive tree consisting of the first two children of any node. The black vertices are part of the 2-ary pyramid.

unsaturated nodes. Hence, one can see a d -ary pyramid as the subtree of a random recursive tree consisting only of the first d children of any node (Figure 9).

This gives a simple way to obtain the height of d -ary pyramids: build a random recursive tree in which the first d children of any node have an edge of weight 1 leading to their parent, and the others a weight of $-\infty$: $(Z_1, \dots, Z_d, Z_{d+1}, \dots) = (1, \dots, 1, -\infty, \dots)$. One can verify [see, e.g., 14] that the (infinite) split vector $(V_1, V_2, \dots, V_i, \dots)$ for a random recursive tree is distributed like

$$\left(U_1, (1 - U_1)U_2, \dots, \prod_{j=1}^{i-1} (1 - U_j)U_i, \dots \right),$$

where $\{U_i, i \geq 1\}$ is a family of i.i.d. $[0, 1]$ -uniform random variables. Since our result only holds for trees of bounded degree, we can rewrite the split vector by collecting the children with index greater than $d + 1$ in a single “bin”:

$$(V_1, V_2, \dots, V_{d+1}) = \left(U_1, (1 - U_1)U_2, \dots, \prod_{j=1}^{d-1} (1 - U_j)U_d, \prod_{j=1}^d (1 - U_j) \right), \quad (18)$$

and $(Z_1, Z_2, \dots, Z_{d+1}) = (1, \dots, 1, -\infty)$. Write $E_k = -\log V_k$. The height is not affected by a random permutation of the children, so the random variable of interest is $X = (Z, E) = (Z_K, E_K)$, where K be taken uniformly at random in $\{1, \dots, d + 1\}$. Then, according to the definition of Λ , we have that for all λ and μ real numbers,

$$\Lambda(\lambda, \mu) = \log \mathbf{E} \left[e^{\lambda + \mu E_K} \mid K \leq d \right] + \log d - \log(d + 1).$$

Using the definition (18) for the split vector (V_1, \dots, V_{d+1}) , we find that,

$$\Lambda(\lambda, \mu) = \lambda + \log \left(\sum_{i=1}^d \mathbf{E} \left[U_1^{-\mu} \right]^i \right) - \log(d + 1),$$

and therefore,

$$\Lambda(\lambda, \mu) = \begin{cases} \infty & \text{if } \mu \geq 1, \\ \lambda + \log \left(\sum_{i=1}^d (1 - \mu)^{-i} \right) - \log(d + 1) & \text{otherwise.} \end{cases}$$

We have $\mathcal{D}_\Lambda = \{(\lambda, \mu) : \mu < 1\}$. It follows that the optimum value for $\mu = \mu(\rho)$ is obtained for

$$\alpha = \frac{\partial \Lambda(\lambda, \mu)}{\partial \lambda} = 1 \quad \text{and} \quad \rho = \frac{\partial \Lambda(\lambda, \mu)}{\partial \mu} = \frac{\sum_{i=1}^d i(1-\mu)^{-i-1}}{\sum_{i=1}^d (1-\mu)^{-i}}. \quad (19)$$

The function f defined on $(0, \infty)$ by

$$x \mapsto \sum_{i=1}^d i x^{-i-1} / \sum_{i=1}^d x^{-i}$$

is continuous, $\lim_{x \rightarrow \infty} f(x) = 0$ and $\lim_{x \downarrow 0} f(x) = \infty$. Therefore, (19) admits a well-defined solution $\mu(\rho)$ for all $\rho > 0$. By Theorem 5, one needs a rescaling factor γ to express the height of d -ary pyramids of size n . The constant γ is given by

$$\sum_{i=1}^d \frac{1}{(1+\gamma)^i} = 1.$$

Then, the height of a d -ary pyramid is asymptotic to $\log n / (\gamma \rho_0)$, where ρ_0 satisfies

$$\lambda + \mu(\rho_0)\rho_0 - \Lambda(\lambda, \mu(\rho_0)) = \log(d+1),$$

which proves (17) since λ cancels.

6 Concluding remarks

The model we have introduced in this paper allowed us to unify number of results concerning heights of random trees. As a payback, we were able to obtain new results about geometric structures such as the shape of the skinniest cell in random k -d trees. We forced Z to take values in \mathbb{R} . However, our results apply equally to some, but not all, parameters for multivariate Z . We could ask about the maximum value of a function of the components of Z . For example, Biggins [8] treats the question of the maximum of projections of the positions of particles in a branching random walk.

The model developed here does not encompass tries, a species of trees for which the bounded height condition does not hold (see p. 8). The height of a random trie under the standard model of randomness is the sum of two contributions: the first one is explained by the bulk of the tree and is similar to the present analysis; the second is a corrective term accounting for the fringe of the tree. See the related study of digital structures by Broutin and Devroye [15, 16].

References

- [1] M. Archibald and J. Clément. Average depth in binary search tree with repeated keys. In *Fourth Colloquium on Mathematics and Computer Science*, volume AG of *DMTCS Proceedings*, pages 309–320. Discrete Mathematics and Theoretical Computer Science, 2006.
- [2] K. B. Athreya and P. E. Ney. *Branching Processes*. Springer, Berlin, 1972.
- [3] A.L. Barabási and R. Albert. Emergence of scaling in random network. *Science*, 286: 509–512, 1999.
- [4] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communication of the ACM*, 18:509–517, 1975.
- [5] J. L. Bentley and R. Sedgewick. Fast algorithm for sorting and searching strings. In *Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 360–369. SIAM, 1997.
- [6] F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP*, volume 581 of *Lecture Notes in Computer Science*, pages 24–48. Springer, 1992.
- [7] P.K. Bhattacharya and J.L. Gastwirth. A nonhomogeneous markov model of a chain-letter scheme. In M.H. Rizvi, J.S. Rustagi, and D. Siegmund, editors, *Recent Advances in Statistics: Papers in Honor of Herman Chernoff*. Academic Press, New York, 1983.
- [8] J. D. Biggins. The asymptotic shape of the branching random walk. *Advanced Applied Probability*, 10:62–84, 1978.
- [9] J. D. Biggins. The growth and spread of the general branching random walk. *The Annals of Applied Probability*, 5:1008–1024, 1995.
- [10] J. D. Biggins. How fast does a general branching random walk spread. In K. B. Athreya and P. Jagers, editors, *Classical and modern branching processes*, New York, 1996. Springer-Verlag.
- [11] J. D. Biggins and D. R. Grey. A note on the growth of random trees. *Statistics and Probability letters*, 32:339–342, 1997.
- [12] P. Billingsley. *Probability and Measure*. Wiley, New York, 3rd edition, 1995.
- [13] N. Broutin. *Shedding New Light on Random Trees*. Phd thesis, McGill University, Montreal, 2007.
- [14] N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. *Algorithmica*, 46:271–297, 2006.
- [15] N. Broutin and L. Devroye. An analysis of the height of tries with random weights on the edges. *Combinatorics, Probability and Computing*, 2007. (42 pages), in press.
- [16] N. Broutin and L. Devroye. The height of list tries and TST. In *International Conference on Analysis of Algorithms*, DMTCS Proceedings. Discrete Mathematics and Theoretical Computer Science, 2007. (13 pages). To appear.

- [17] N. Broutin, L. Devroye, E. McLeish, and M. de la Salle. The height of increasing trees. *Random Structures and Algorithms*, 2007. (25 pages), in press.
- [18] G.G. Brown and B.O. Shubert. On random binary trees. *Mathematics of Operations Research*, 9:43–65, 1984.
- [19] H. A. Clampett. Randomized binary searching with tree structures. *Communications of the ACM*, 7(3):163–165, 1964.
- [20] J. Clément, P. Flajolet, and B. Vallée. The analysis of hybrid trie structures. In *9th annual ACM-SIAM Symposium on Discrete Algorithms*, pages 531–539, Philadelphia, PA, 1998. SIAM Press.
- [21] J. Clément, P. Flajolet, and B. Vallée. Dynamical source in information theory: a general analysis of trie structures. *Algorithmica*, 29:307–369, 2001.
- [22] E. G. Coffman and J. Eve. File structures using hashing functions. *Communications of the ACM*, 13:427–436, 1970.
- [23] R. de la Briandais. File searching using variable length keys. In *Proceedings of the Western Joint Computer Conference, Montvale, NJ, USA*. AFIPS Press, 1959.
- [24] A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*. Springer Verlag, second edition, 1998.
- [25] F. den Hollander. *Large Deviations*. American Mathematical Society, Providence, RI, 2000.
- [26] J.-D. Deuschel and D.W. Stroock. *Large Deviations*. American Mathematical Society, Providence, RI, 1989.
- [27] L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33: 489–498, 1986.
- [28] L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24:277–298, 1987.
- [29] L. Devroye. On the expected height of fringe balanced trees. *Acta Informatica*, 30: 459–466, 1993.
- [30] L. Devroye. Universal limit laws for depth in random trees. *SIAM Journal on Computing*, 28(2):409–432, 1998.
- [31] L. Devroye, J. Jabbour, and C. Zamora-Cura. Squarish k -d trees. *SIAM Journal on Computing*, 30:1678–1700, 2001.
- [32] A. Duch. *Design and Analysis of Multidimensional Data Structures*. PhD thesis, UPC, Barcelona, 2004.
- [33] A. Duch and C. Martínez. On the average performance of orthogonal range search in multidimensional data structures. *Journal of Algorithms*, 44(1):226–245, 2002.

- [34] A. Duch, V. Estivill-Castro, and C. Martínez. Randomized k -dimensional binary search trees. In K.-Y. Chwa and O.H. Ibarra, editors, *Proc. of the 9th International Symposium on Algorithms and Computation (ISAAC'98)*, volume 1533 of *Lecture Notes in Computer Science*, pages 199–208. Springer Verlag, 1998.
- [35] R. S. Ellis. Large deviations for a general class of random vectors. *The Annals of Probability*, 12:1–12, 1984.
- [36] R. A. Finkel and J. L. Bentley. Quad trees, a data structure for retrieval on composite keys. *Acta Informatica*, 4:1–19, 1974.
- [37] P. Flajolet. The ubiquitous digital tree. In B. Durand and W. Thomas, editors, *STACS 2006, Annual Symposium on Theoretical Aspects of Computer Science*, volume 3884 of *Lecture Notes in Computer Science*, pages 1–22. Springer, 2006.
- [38] P. Flajolet and C. Puech. Partial match retrieval of multidimensional data. *Journal of the ACM*, 33(2):371–407, 1986.
- [39] E. Fredkin. Trie memory. *Communications of the ACM*, 3(9):490–499, 1960.
- [40] J. Gärtner. On large deviations from the invariant measure. *Theory of Probability and its Applications*, 22:24–39, 1977.
- [41] J.L. Gastwirth and P.K. Bhattacharya. Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Operations Research*, 32(3):527–536, 1984.
- [42] G. H. Gonnet and R. Baeza-Yates. *Handbook of Algorithms and Data Structures*. Addison-Wesley, Workingham, second edition, 1991.
- [43] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, Reading, MA, second edition, 1994.
- [44] P. Groeneboom, J. Oosterhoff, and F.H. Ruymgaart. Large deviation theorems for empirical probability measures. *The Annals of Probability*, 7:553–586, 1979.
- [45] C. A. R. Hoare. Quicksort. *The Computer Journal*, 5:10–15, 1962.
- [46] D.A. Huffman. A method for constructing minimum redundancy codes. *Proceedings of the Institute of Radio Engineers*, pages 1098–1102, 1952.
- [47] D. E. Knuth. *The Art of Computer Programming: Sorting and Searching*, volume 3. Addison-Wesley, Reading, MA, 1973.
- [48] A.G. Konheim and D.J. Newman. A note on growing binary trees. *Discrete Mathematics*, 4:57–63, 1973.
- [49] W.C. Lynch. More combinatorial properties of certain trees. *Computing J.*, 7:299–302, 1965.
- [50] H. M. Mahmoud. A strong law for the height of random binary pyramids. *The Annals of Applied Probability*, 4:923–932, 1994.
- [51] C. Martínez and S. Roura. Optimal sampling strategies in quicksort and quickselect. *SIAM Journal on Computing*, 31:683–705, 2001.

- [52] C. Martínez, A. Panholzer, and H. Prodinger. Partial match in relaxed multidimensional search trees. *Algorithmica*, 29(1–2):181–204, 2001.
- [53] D. R. Morrison. PATRICIA — Practical Algorithm To Retrieve Information Coded in Alphanumeric. *Journal of the ACM*, 15:514–534, 1968.
- [54] R. Neininger and L. Rüschemdorf. A survey of multivariate aspects of the contraction method. *Discrete Mathematics and Theoretical Computer Science*, 8:31–56, 2006.
- [55] O. Nerman. On the convergence of a supercritical general (C-M-J) branching process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57:365–395, 1981.
- [56] B. Pittel. Note on the height of random recursive trees and m -ary search trees. *Random Structures and Algorithms*, 5:337–347, 1994.
- [57] S.T. Rachev and L. Rüschemdorf. Probability metrics and recursive algorithms. *Advances in Applied Probability*, 27:770–799, 1995.
- [58] R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [59] U. Rösler. A fixed point theorem for distributions. *Stochastic Processes and their Applications*, 37:195–214, 1992.
- [60] U. Rösler and L. Rüschemdorf. The contraction method for recursive algorithms. *Algorithmica*, 29:3–33, 2001.
- [61] H. Samet. *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*. Addison-Wesley, Reading, MA, 1990.
- [62] H. Samet. *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, Reading, MA, 1990.
- [63] R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, Reading, MA, 1996.
- [64] R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Theoretical Probability and Mathematical Statistics*, 51:1–27, 1995.
- [65] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley, New York, 2001.
- [66] M. Van Emden. Increasing the efficiency of quicksort. *Communications of the ACM*, 13:563–567, 1970.

A Galton–Watson processes and beyond

We bound the extinction probabilities of some Galton–Watson processes.

Theorem 12. *Let $d \geq 1$ be a fixed integer. Consider a sequence of Galton–Watson processes with progeny distributions $Z^{(x)}$ on $\{0, 1, \dots, d\}$, $\mathbf{E}Z^{(x)} = \mu_x$, and extinction probabilities q_x , $x \in \mathbb{R}$. Assume that there exists x_0 and $\delta > 0$ such that $\inf_{x \geq x_0} \mathbf{E}Z^{(x)} \geq 1 + \delta$. If $\mathbf{P}\{Z^{(x)} = 0\} \rightarrow 0$, as $x \rightarrow \infty$, then $q_x \rightarrow 0$.*

The proof of Theorem 12 is based on the following Lemma providing an explicit bound on the extinction probability.

Lemma 12. *Let $d \geq 1$ be a fixed integer. Consider a Galton–Watson process with progeny distribution Z on $\{0, 1, \dots, d\}$ and extinction probability q . Let $\mu = \mathbf{E}Z > 1$ and $p_i = \mathbf{P}\{Z = i\}$, $0 \leq i \leq d$. Then*

$$q \leq \frac{2p_0}{p_0 + \frac{\mu-1}{2}} \quad \text{if} \quad \frac{2p_0}{p_0 + \frac{\mu-1}{2}} < 1 - \mu^{-\frac{1}{d-1}}.$$

Proof. The proof is based on an analysis of the probability generating function $f(s) = \sum_{i \geq 1} p_i s^i$. We know that q satisfies $f(q) = q$. Observe that

$$\mu = p_1 + \sum_{i \geq 2} i p_i \geq p_1 + 2 \sum_{i \geq 2} p_i.$$

Note that $p_0 + p_1 < 1$. Define the auxiliary generating function

$$g(s) = \frac{f(s) - p_0 - p_1 s}{1 - p_0 - p_1},$$

and note that $g(s) \leq s^2$ for all $s \in [0, 1]$, and $g(0) = 0$. Now,

$$g(q) = \frac{f(q) - p_0 - p_1 q}{1 - p_0 - p_1} = \frac{q(1 - p_1) - p_0}{1 - p_0 - p_1} \leq q^2.$$

We rewrite the above equation in order to obtain

$$q \leq q^2 \cdot \frac{1 - p_1 - p_0}{1 - p_1} + \frac{p_0}{1 - p_1} \leq q^2 + \frac{p_0}{p_0 + \sum_{i > 1} p_i} \leq q^2 + \frac{p_0}{p_0 + \frac{\mu - p_1}{2}}.$$

Finally, we see that

$$q \leq q^2 + \alpha \quad \text{where} \quad \alpha = \frac{p_0}{p_0 + \frac{\mu-1}{2}}.$$

If $4\alpha \geq 1$, then clearly $q \leq 1 \leq 4\alpha$. Otherwise, $4\alpha < 1$ and this implies either

$$q \leq \frac{1 - \sqrt{1 - 4\alpha}}{2} \quad \text{or} \quad q \geq \frac{1 + \sqrt{1 - 4\alpha}}{2}.$$

For all $x \in [0, 1]$, we have $\sqrt{1 - x} \geq 1 - x$ and thus, we can conclude that, when $4\alpha < 1$, either

$$q \leq 2\alpha \quad \text{or} \quad q \geq 1 - 2\alpha. \tag{20}$$

We now assume that $q \neq 0$, for otherwise, the result trivially holds. Note that, in this case, since $q = f(q)$, $q \leq \sum_{i=1}^d i p_i q^{i-1}$. By monotonicity, for the solution r of $1 = \sum_{i=1}^d i p_i r^{i-1}$, we have $q \leq r$. Observe also that $\sum_{i=1}^d i p_i = \mu$, which we have assumed greater than 1. As a consequence, $r \leq 1$ and

$$1 = \sum_{i=1}^d i p_i r^{i-1} \geq r^{d-1} \sum_{i=1}^d i p_i = \mu r^{d-1},$$

so $q \leq r \leq \mu^{-\frac{1}{d-1}}$. Recalling (20), if $1 - 2\alpha > \mu^{-\frac{1}{d-1}}$, then we must have $q \leq 2\alpha$. This proves the lemma. \square

We are now ready to prove Theorem 12 which is, in fact, an easy corollary of Lemma 12.

Proof of Theorem 12. If $\mathbf{P}\{Z^{(x)} = 0\} = 0$, the result is clear. Assuming $\mathbf{P}\{Z^{(x)} = 0\} > 0$, recall Lemma 12. We have, for $x \geq x_0$,

$$1 - \mu_x^{-\frac{1}{d-1}} \geq 1 - (1 + \delta)^{-\frac{1}{d-1}} \stackrel{\text{def}}{=} \xi > 0.$$

As a consequence, since $p_0 = \mathbf{P}\{Z^{(x)} = 0\} \rightarrow 0$, for x large enough,

$$\frac{2p_0}{p_2 + \frac{\mu_x - 1}{2}} = \frac{2p_0}{p_0 + \delta/2} \leq \frac{4p_0}{\delta} < \xi.$$

Therefore, for $x \rightarrow \infty$, we have $q_x = O(p_0) = o(1)$, which completes the proof. \square

In the course of the proofs, we will need the following technical lemma. One should see it as a tool to deal with branching processes for which the progeny distribution may depend on the node. It asserts that if there is a deterministic lower bound for the reproduction distribution function, then one can find a subprocess that is a proper Galton–Watson process, that is, for which every node has the same progeny distribution.

Lemma 13. *Let Z be a random variable and let $N \geq 0$ be a random integer such that*

$$\inf_n \mathbf{P}\{Z \geq k \mid N = n\} \geq t_k,$$

where $t_k \downarrow 0$ as $k \rightarrow \infty$. Then there exists a random variable Y with $Y \leq Z$ such that $\mathbf{P}\{Y \geq k\} = t_k$ for all k .

Proof. For all k ,

$$\begin{aligned} \mathbf{P}\{Z \geq k\} &= \sum_{n \geq 0} \mathbf{P}\{Z \geq k, N = n\} \\ &= \sum_{n \geq 0} \mathbf{P}\{N = n\} \cdot \mathbf{P}\{Z \geq k \mid N = n\} \\ &\geq t_k \\ &= \mathbf{P}\{W \geq k\}, \end{aligned}$$

where W is chosen such that $\mathbf{P}\{W \geq k\} = t_k$ for all k . Then, $Z \succ W$ and, by the inverse transform, there exists a coupled pair (Y, Z) , Y distributed as W , with $Y \leq Z$. \square

B Large deviations and Cramér functions

B.1 Proof of Theorem 4

The proof follows roughly the lines of that presented by [24]. Let $\gamma > 0$. Observe first that, since $\{-\infty\} \times [0, \infty] \cup [-\infty, +\infty) \times \{\infty\} \notin \Gamma$, we have, for all M ,

$$\mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n (Z_i, E_i) \in \Gamma, A_M \right\} = \mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n (Z_i, E_i) \in \Gamma, A_M, F_n \right\}. \quad (21)$$

REDUCING Γ TO A COMPACT SET. The first step consist in bounding (21) to a similar probability involving a compact set. Since $0 \in \mathcal{D}_\Lambda^0$, there exists λ and $\mu > 0$ and $A > 0$ such that $\Lambda(\lambda, \mu) < A$. For any $r > 0$, we have

$$\mathbf{P} \left\{ \sum_{i=1}^n \lambda Z_i + \mu E_i > rn, A_M, F_n \right\} \leq \mathbf{E} \left[\mathbf{1}_{[A_M, F_n]} \cdot \exp \left(\sum_{i=1}^n \lambda Z_i + \mu E_i \right) \right] e^{-rn}.$$

Applying assumption (4) for this λ and μ , for all $M \geq M_1$ large enough, since $\Lambda(\lambda, \mu) \leq A$,

$$\mathbf{P} \left\{ \sum_{i=1}^n \lambda Z_i + \mu E_i > rn, A_M, F_n \right\} \leq e^{(A+\delta-r)n}.$$

Therefore, for $r = A + \delta + 1/\gamma$, writing $C = \{(x, y) : \lambda x + \mu y \leq r\}$, and $\Gamma' = \Gamma \cap C$ we see that, for $M \geq M_1$,

$$\begin{aligned} \mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in \Gamma, A_M, F_n \right\} &\leq \mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in \Gamma', A_M, F_n \right\} \\ &\quad + \mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in C^c, A_M, F_n \right\}, \end{aligned}$$

and hence, for $M \geq M_1$,

$$\mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in \Gamma, A_M, F_n \right\} = \mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in \Gamma', A_M, F_n \right\} + e^{-n/\gamma}. \quad (22)$$

COVERING Γ' WITH SMALL SETS. We now proceed by covering Γ' with a finite set of balls. For any $\omega = (x_\omega, y_\omega) \in \mathbb{R}^2$, there exists $(\lambda_\omega, \mu_\omega)$ such that

$$\lambda_\omega x_\omega + \mu_\omega y_\omega - \Lambda(\lambda_\omega, \mu_\omega) > \min \left\{ \frac{1}{\gamma} + \frac{2\gamma}{3}, \Lambda^*(x_\omega, y_\omega) - \frac{\gamma}{3} \right\}.$$

For all $\omega \in \mathbb{R}^2$, there exists an open ball \mathcal{B}_ω such that for all $(x, y) \in \mathcal{B}_\omega$, $|\lambda_\omega(x - x_\omega) + \mu_\omega(y - y_\omega)| \leq \gamma/3$. Hence we have

$$\inf_{(x,y) \in \mathcal{B}_\omega} \{\lambda_\omega x + \mu_\omega y\} \geq \Lambda(\lambda_\omega, \mu_\omega) - \frac{\gamma}{3} + \min \left\{ \frac{1}{\gamma} + \frac{2\gamma}{3}, \Lambda^*(x_\omega, y_\omega) - \frac{\gamma}{3} \right\}. \quad (23)$$

The set $\{\mathcal{B}_\omega, \omega \in \mathbb{R}^2\}$ covers \mathbb{R}^2 but is uncountable. However Γ' is contained in a compact set, and it can be covered by $\{\mathcal{B}_\omega, \omega \in \mathcal{C}\}$, where \mathcal{C} is finite. Thus, by the union bound,

$$\begin{aligned} \mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in n\Gamma', A_M, F_n \right\} &\leq \mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n (Z_i, E_i) \in \bigcup_{\omega \in \mathcal{C}} \mathcal{B}_\omega, A_M, F_n \right\} \\ &\leq \sum_{\omega \in \mathcal{C}} \mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n (Z_i, E_i) \in \mathcal{B}_\omega, A_M, F_n \right\}. \end{aligned}$$

Consider one term in the sum above. Note that

$$\sum_{i=1}^n (Z_i, E_i) \in \mathcal{B}_\omega \quad \Rightarrow \quad \sum_{i=1}^n \lambda_\omega Z_i + \mu_\omega E_i \geq \inf_{(x,y) \in \mathcal{B}_\omega} \{\lambda_\omega x + \mu_\omega y\}.$$

Then, using assumption (4) with λ_ω and μ_ω and $\delta = \gamma/3$, there exists $M_2 = M_2(\omega)$ such that for all $M \geq M_2$,

$$\mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in \mathcal{B}_\omega, A_M \right\} \leq \exp \left(n\Lambda(\lambda_\omega, \mu_\omega) + n\frac{\gamma}{3} - n \inf_{(x,y) \in \mathcal{B}_\omega} \{\lambda_\omega x + \mu_\omega y\} \right).$$

Then, recalling the bound (23), we obtain for all $M \geq M_2(\omega)$,

$$\mathbf{P} \left\{ \sum_{i=1}^n (Z_i, E_i) \in \mathcal{B}_\omega, A_M \right\} \leq \exp \left(n\frac{2\gamma}{3} - n \cdot \min \left\{ \frac{1}{\gamma} + \frac{2\gamma}{3}, \Lambda^*(x_\omega, y_\omega) - \frac{\gamma}{3} \right\} \right).$$

Finally, plugging the bound above in (22), and observing that $\inf\{\Lambda^*(x, y) : (x, y) \in \Gamma\} = I(\alpha, \rho)$, for all $M \geq \max\{M_1, M_2(\omega) : \omega \in \mathcal{C}\}$

$$\mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n (Z_i, E_i) \in \Gamma, A_M \right\} \leq (1 + |\mathcal{C}|) \cdot \exp \left(-n \cdot \min \left\{ \frac{1}{\gamma}, I(\alpha, \rho) - \gamma \right\} \right).$$

Taking logarithms completes the proof.

B.2 Properties of Λ , Λ^* and I

The functions Λ , Λ^* and I are well understood [24]. They are the corner stone of the characterization of first order asymptotic properties of the height of random trees. Consider a mapping $f : \mathbb{R}^2 \rightarrow (-\infty, \infty]$. Let \mathcal{D}_f be its domain: $\mathcal{D}_f = \{(\alpha, \rho) : f(\alpha, \rho) < \infty\}$, and let \mathcal{D}_f^o be the interior of \mathcal{D}_f .

The mapping f is said to be *convex* if, for all $x_1, x_2 \in \mathbb{R}^2$, and $\theta \in [0, 1]$, we have $f(\theta x_1 + (1 - \theta)x_2) \leq \theta f(x_1) + (1 - \theta)f(x_2)$, where it is understood that if the left-hand side is ∞ , then either $f(x_1) = \infty$ or $f(x_2) = \infty$. If the *level sets* $\Psi_f(\ell) = \{x : f(x) \leq \ell\}$ are closed for all $\ell \in \mathbb{R}$, we say that f is *lower*

semicontinuous, and call f a *rate function*. The mapping f is said to be a *good rate function* if its level sets are compact.

THE FUNCTION $\Lambda(\cdot, \cdot)$. The cumulant generating function is the link between the random variables and the rate functions, and its properties imply those of Λ^* and I .

Lemma 14. *The function $\Lambda(\cdot, \cdot)$*

- (a) *takes values in $(-\infty, \infty]$ if $p = \mathbf{P}\{Z > -\infty, E < \infty\} > 0$;*
- (b) *is convex on \mathbb{R}^2 , and continuous in \mathcal{D}_Λ^o .*

Proof. (a) By definition, $\forall \lambda, \mu \in \mathbb{R}$, we have

$$\Lambda(\lambda, \mu) = \log p + \log \mathbf{E} \left[e^{\lambda Z + \mu E} \mid Z > -\infty, E < \infty \right].$$

Both Z and E are real on $\{Z > -\infty, E < \infty\}$, and hence $\mathbf{E} [e^{\lambda Z + \mu E}] > 0$. Since $p > 0$, this yields $\Lambda(\lambda, \mu) > -\infty$.

(b) The convexity follows from Hölder's inequality. The continuity in \mathcal{D}_Λ^o is a straightforward consequence of the convexity. For details see [24]. \square

THE FUNCTION $\Lambda^*(\cdot, \cdot)$. The level sets of Λ^* are of particular interest, and we write $\Psi = \Psi_{\Lambda^*}$. Indeed, as we will see later, the heights will be characterized using optimizations of some objective functions on the level sets of Λ^* .

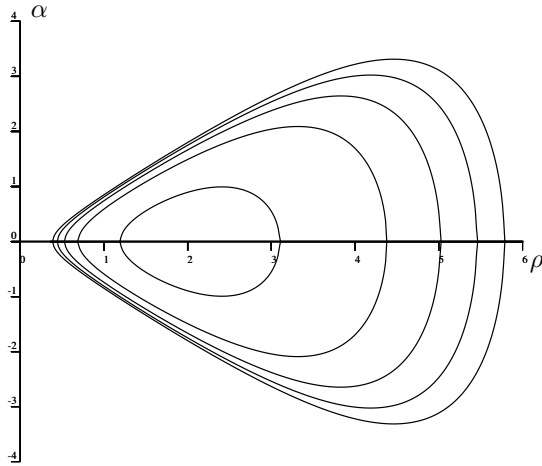


Figure 10. An increasing family of level sets $\Psi(\ell)$ for the function Λ^* corresponding to the study of k -d trees (Section 5.4)

Lemma 15. *The function $\Lambda^*(\cdot, \cdot)$ is*

- (a) *convex on \mathbb{R}^2 ;*
- (b) *continuous on $\mathcal{D}_{\Lambda^*}^o$;*
- (c) *a good rate function if $0 \in \mathcal{D}_\Lambda^o$.*

Proof. (a) The convexity of Λ^* is a direct consequence of its definition: for $t_1, t_2, \lambda \in \mathbb{R}^2$ and $\theta \in [0, 1]$, using \cdot to denote the standard inner product,

$$\begin{aligned}\Lambda^*(\theta t_1 + (1 - \theta)t_2) &= \sup_{\lambda \in \mathbb{R}^2} \{\lambda \cdot (\theta t_1 + (1 - \theta)t_2) - \Lambda(\lambda)\} \\ &\leq \sup_{\lambda \in \mathbb{R}^2} \{\theta \lambda \cdot t_1 - \theta \Lambda(\lambda)\} + \sup_{\lambda \in \mathbb{R}^2} \{(1 - \theta) \lambda \cdot t_2 - (1 - \theta) \Lambda(\lambda)\} \\ &= \theta \Lambda^*(t_1) + (1 - \theta) \Lambda^*(t_2).\end{aligned}$$

(b) Since Λ^* is convex, it is continuous on $\mathcal{D}_{\Lambda^*}^o$.

(c) Let $\ell \geq 0$. For $r \geq 0$, let $\mathcal{C}_r = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} \geq r\}$. Since $0 \in \mathcal{D}_{\Lambda}^o$, there exists a ball \mathcal{B} , centered at the origin with radius $\delta > 0$, and $A < \infty$ such that for all $(\lambda, \mu) \in \mathcal{B}$, $\Lambda(\lambda, \mu) \leq A$. For any $r \geq 0$ and $(\alpha, \rho) \in \mathcal{C}_r$,

$$\Lambda^*(\alpha, \rho) = \sup_{\lambda, \mu} \{\lambda \alpha + \mu \rho - \Lambda(\lambda, \mu)\} \geq \sup_{(\lambda, \mu) \in \mathcal{B}} \{\lambda \alpha + \mu \rho - \Lambda(\lambda, \mu)\} \geq \delta \cdot r - A.$$

As a consequence, for R large enough, \mathcal{C}_R does not intersect $\Psi(\ell)$, proving that $\Psi(\ell)$ is bounded.

We now show that $\Psi(\ell)$ is closed (Λ^* is lower semicontinuous). It suffices to prove that $\Psi(\ell)$ contains all its accumulation points: for any $(\alpha, \rho) \in \mathbb{R}^2$ such that there exists $(\alpha_n, \rho_n) \in \Psi(\ell)$ with $(\alpha_n, \rho_n) \rightarrow (\alpha, \rho)$, we should have $(\alpha, \rho) \in \Psi(\ell)$. For any $\lambda, \mu \in \mathbb{R}$,

$$\liminf_{n \rightarrow \infty} \Lambda^*(\alpha_n, \rho_n) \geq \liminf_{n \rightarrow \infty} \{\lambda \alpha_n + \mu \rho_n - \Lambda(\lambda, \mu)\} = \lambda \alpha + \mu \rho - \Lambda(\lambda, \mu).$$

As a result,

$$\liminf_{n \rightarrow \infty} \Lambda^*(\alpha_n, \rho_n) \geq \sup_{\lambda, \mu} \{\lambda \alpha + \mu \rho - \Lambda(\lambda, \mu)\} = \Lambda^*(\alpha, \rho).$$

Hence, $\Lambda^*(\alpha, \rho) \leq \ell$ and $(\alpha, \rho) \in \Psi(\ell)$, which proves that $\Psi(\ell)$ is closed. \square

THE FUNCTION $I(\cdot, \cdot)$. The function that appears in the Cramér and Gärtner–Ellis theorems is $I(\cdot, \cdot)$.

Lemma 16. For $\alpha, \rho \in \mathbb{R}^2$, let $I(\alpha, \rho) \stackrel{\text{def}}{=} \inf\{\Lambda^*(x, y) : x > \alpha, y < \rho\}$. Then

- (a) $(\alpha, \rho) \mapsto I(\alpha, \rho)$ is non-decreasing in α , and non-increasing in ρ ;
- (b) for $(\alpha, \rho) \in \mathcal{D}_{\Lambda^*}^o$, $I(\alpha, \rho) = \inf\{\Lambda^*(x, y) : x \geq \alpha, y \leq \rho\}$.

Proof. (a) This is clear from the definition as an infimum.

(b) Clearly $\inf\{\Lambda^*(x, y) : x \geq \alpha, y \leq \rho\} \leq I(\alpha, \rho)$. So we prove that $I(\alpha, \rho) \leq \inf\{\Lambda^*(x, y) : x \geq \alpha, y \leq \rho\}$. Consider a sequence $(x_n, y_n) \in \mathcal{D}_{\Lambda^*}$ such that

$$\lim_{n \rightarrow \infty} \Lambda^*(x_n, y_n) = \inf\{\Lambda^*(x, y) : x \geq \alpha, y \leq \rho\}.$$

We build an auxiliary sequence (x'_n, y'_n) , $n \geq 1$. Let $k \geq 1$. If $x_k \neq \alpha$ and $y_k \neq \rho$, then $(x'_k, y'_k) = (x_k, y_k)$. Assume now that $x_k = \alpha$ or $y_k = \rho$. Then we construct

a new point (x'_k, y'_k) such that $\Lambda^*(x'_k, y'_k) < \Lambda^*(x_k, y_k) + 1/k$ where $x'_k > \alpha$ and $y'_k < \rho$. This construction is done in the following way. Assume at first that for small enough $\epsilon > 0$ there exists a ball \mathcal{B}_ϵ centered at (x_k, y_k) with radius ϵ contained within $\mathcal{D}_{\Lambda^*}^o$. In this case, by the continuity of Λ^* , we find a point $(x'_k, y'_k) \in \mathcal{B}_\epsilon$ with $x_k > \alpha, y_k < \rho$ such that $\Lambda^*(x'_k, y'_k) < \Lambda^*(x_k, y_k) + 1/k$. In the second case, no such ball exists for any ϵ , which means in particular (x_k, y_k) lies on the boundary of \mathcal{D}_{Λ^*} . Consider the region $R_\epsilon = \mathcal{B}_\epsilon \cap \mathcal{D}_{\Lambda^*}^o \cap \{(\alpha, \infty) \times (-\infty, \rho)\}$. Since $(\alpha, \rho) \in \mathcal{D}_{\Lambda^*}^o$, an open convex set, this region is non-empty. Let $\beta = \inf_{\epsilon \downarrow 0} \sup\{\Lambda^*(x, y) : (x, y) \in R_\epsilon \setminus (x_k, y_k)\}$. Assume for a contradiction that $\beta > \Lambda^*(x_k, y_k)$. Then there exist (x, y) such that the line joining (x, y) to (x_k, y_k) lies below Λ^* , contradicting the convexity of Λ^* . Hence $\beta \geq \Lambda^*(x_k, y_k)$ and, for ϵ small enough, there exist (x'_k, y'_k) in R_ϵ such that $\Lambda^*(x'_k, y'_k) \leq \Lambda^*(x_k, y_k)$. Therefore, using the auxiliary sequence, we see that $\inf\{\Lambda^*(x, y) : x > \alpha, y < \rho\} \leq \lim_{n \rightarrow \infty} \Lambda^*(x'_n, y'_n) = \inf\{\Lambda^*(x, y) : x \geq \alpha, y \leq \rho\}$. This finishes the proof. \square