Random Trees, Heights, and Large Deviations

Nicolas Broutin

1 Introduction

Random trees are of prime importance for studying the average case behavior of algorithms and data structures. The canonical examples are Hoare's quicksort algorithm and the binary search tree. The most natural parameter to study is the average running time of the algorithm, and the average time for answering a query in a data structure. This does not provide sufficient information to properly dimension the systems, and one usually wants to quantify the extreme values that should occur (in an average sense). In the case of trees, one of the extreme values of interest is the height.

We present here a general framework to devise a law of large numbers for the height of random trees. Our model unifies the treatment of many examples of the literature such as binary search trees, median-of-$(2k+1)$ trees, random recursive trees, plane oriented trees, digital search trees, scale-free trees, and all polynomial families of increasing trees, among others.

The approach is based on the branching processes techniques used by Devroye (1986, 1987) in his analysis of the height of random binary search trees, in which he exhibited a deep connection between the height of a binary search tree and extremes in branching random walks. He proved that for a binary search tree of size n, the height is asymptotic to $c \log n$ in probability, where $c = 4.311\ldots$ is the unique solution greater than 1 of $c \log(2e/c) = 1$. The branching random walk is much too constrained to capture many of examples of interest. We introduce two generalizations of Devroye's techniques that permit to broaden the scope of the model.

It is convenient to see a tree of size n as a recursive partition of set of n balls, or *items*. A random tree may then be described by the way in which each node partitions the items contained in its subtree among its children. This characteristic of a node is called the *split*.

Split relaxation. The models captured so far were such that all the nodes had to split the items according to the *same* distribution, independently of the number of items to be partitioned. Some important kinds of trees like general families of increasing trees (Bergeron et al., 1992) exhibit splits that vary with the number of items. Accordingly, in our model, a node may partition the n items contained in its subtree using a distribution depending on n. However, there should be a limit distribution as $n \to \infty$ so that the distributions of the splits are controlled far away from the fringe. Then, the nodes storing many items in their subtrees behave somewhat similarly, and the others should not contribute much to the height (for they contain few nodes, and the heights of the corresponding subtrees should be small, see below).

Weighted trees. We consider trees with weighted edges to allow for more flexibility. In particular, the weights are useful to transform kinds of trees where the degree is not bounded into binary trees, or to account for the running times to branch to children that may not be constant. The latter situation occurs for instance when one prefers a variable-size structure like a linked-list to store the pointers to subtrees. Considering such weighted

trees is close in spirit to the continuous-time branching random walks of Biggins (1996) or the multidimensional trees of Broutin and Devroye (2006).

For the general model sketched above, and under mild conditions, we prove that the height H_n of a tree on n items is asymptotic to $c \log n$ in probability as $n \to \infty$. The constant c is computable. It is uniquely characterized geometrically using large deviations rate functions related to the distributions of the splits and weights, though the equation it satisfies is often implicit.

The remaining of the document is organized as follows. In Section 2 we present the model in detail and state the main result, a law of large numbers for the height of random trees. In Section 3, we give an overview of the possible applications of Theorem 1. This abstract is based on the long version by Broutin et al. (2007) and we refer the reader to this document for proofs and many more applications. See also Broutin et al. (2007+) for an account of the unweighted version.

2 A general model of random trees

Consider a family of random vectors $\{((Z_1^n, V_1^n), (Z_2^n, V_2^n)), n \geq 0\}$. The joint distribution of $((Z_1^n, V_1^n), (Z_2^n, V_2^n))$ will characterize the behavior of nodes having n items in their subtree. The components V_i^n, $i = 1, 2$, describe the way the data is partitioned among the subtrees, and Z_i^n, $i = 1, 2$, the costs of the edges. The numbers nV_1^n and nV_2^n represent the numbers of items be stored in the left and right subtrees, respectively, and they shall only take integer values such that $V_1^n + V_2^n \leq 1$.

Let T_∞ be the infinite complete binary tree, and assign an independent copy of the *entire* family $\{((Z_1^n, V_1^n), (Z_2^n, V_2^n)), n \geq 0\}$ to each node $u \in T_\infty$. Given an integer number n, the construction of a weighted tree T_n on n items is done in two stages: a *shape* is first given to the random tree; then the shape being fixed, some *weights* are assigned to the edges.

The shape of the tree. The shape is simply a subtree of T_∞ described by the number of items N_u stored in every subtree rooted at $u \in T_\infty$. The values N_u, $u \in T_\infty$ are built recursively from the root. Given $N_u = n$, we define the number of items of the two subtrees to be $nV_1^n(u)$ and $nV_2^n(u)$. Both values are natural numbers summing to at most n, so that the process actually distributes atomic items in the subtrees. The tree of interest is $T_n = \{u \in T_\infty : N_u \geq 1\}$.

The weights. We now assume $\{N_u, u \in T_\infty\}$ fixed. For a node u, and given $N_u = n$, we define the weights of the two edges e_1, e_2 incident at u to be $Z_{e_1} = Z_1^n(u)$ and $Z_{e_2} = Z_2^n(u)$. The *weighted depth* of a node $u \in T_\infty$ is then defined to be $D_u = \sum_{u \in \pi(u)} Z_e$, where $\pi(u)$ denotes the set of edges on the shortest path between u and the root.

We are interested in the height H_n of T_n, or the maximum weighted depth of a node in the tree, defined by $H_n = \max\{D_u : u \in T_n\}$. The height is easier characterized using the alternative distributions $\mathcal{X}^n = ((Z_1^n, E_1^n), (Z_2^n, E_2^n)) = ((Z_1^n, -\log V_1^n), (Z_2^n, -\log V_2^n))$. We have the following conditions:

- Permutation invariance. Randomly permuting the children of each node does not change $\{D_u, u \in T_\infty\}$. Thus, it suffices to consider vectors such that $(Z_1^n, E_1^n) = (Z_2^n, E_2^n)$ in distribution, for all $n \geq 0$. Let (Z^n, E^n) a typical component of $\mathcal{X}^n$.
- Convergence. We assume that there exists a random vector $X = (Z, E)$ such that, for all $\lambda, \mu \in \mathbb{R}$,

$$\Lambda_n(\lambda, \mu) \stackrel{\text{def}}{=} \log \mathbf{E}\left[e^{\lambda Z^n + \mu E^n}\right] \xrightarrow[n \to \infty]{} \Lambda(\lambda, \mu) \stackrel{\text{def}}{=} \log \mathbf{E}\left[e^{\lambda Z + \mu E}\right] \leq \infty,$$

and the function $\Lambda(\cdot,\cdot)$ is finite in a neighborhood of the origin. This is our notion of limit behavior when the number of items tend to infinity. (Strictly speaking, we need to require this for the *vectors* $\mathcal{X}^n$ and $\mathcal{X}$, but this exact condition is only technical.)

- FINITENESS. It is not clear that either T_n or H_n are finite. However, it is the case if $0 \leq Z$, $\mathbf{E}Z < \infty$, $\mathbf{P}\{V > 0\} > 1/d$, and $\mathbf{P}\{\exists i : V_i > 0\} = 1$.
- BOUNDED HEIGHT. There exists a deterministic function $\psi(\cdot)$ such that for all $n \in \mathbb{N}$, $H_n \leq \psi(n)$ almost surely.

The characterization of the height relies on the theory of large deviations for sums of random vectors,

$$\sum_{e \in \pi(u)} (Z_e, E_e).$$

The summands are neither independent (the variables associated with an edge depend on those along the path to the root), nor identically distributed (the distribution depends on the number of items in the corresponding subtree). However, the dependence is weak and the Gärtner–Ellis theorem applies (Gärtner, 1977; Ellis, 1984). In particular, despite the dependence of elementary steps, the collections of depths $\{D_u, u \in T_n\}$ is close enough to a branching random walk (i.e., in which all the steps are independent and identically distributed with distribution $\mathcal{X}$) truncated when $\sum_{e \in \pi(u)} E_e \geq \log n$ so that the first order asymptotic terms of the heights are comparable. What matters are the tail probabilities related to the truncated branching random walk mentioned above: for a node u lying k levels away from the root, we have

$$\begin{aligned} \mathbf{P}\left\{D_u > \frac{\alpha}{\rho}\log n, u \in T_n\right\} &= \mathbf{P}\left\{\sum_{e\in\pi(u)} Z_e > \frac{\alpha}{\rho}\log n, \sum_{e\in\pi(u)} E_e < \log n\right\} \\ &= \exp\left(-I(\alpha,\rho)k + o(k)\right), \end{aligned}$$

where $I(\alpha,\rho) = \inf\{\Lambda^\star(\alpha',\rho') : \alpha' > \alpha, \rho' < \rho\}$ and $\Lambda^\star(\alpha,\rho) = \sup\{\lambda\alpha + \mu\rho - \Lambda(\lambda,\mu) : \lambda,\mu \in \mathbb{R}\}$.

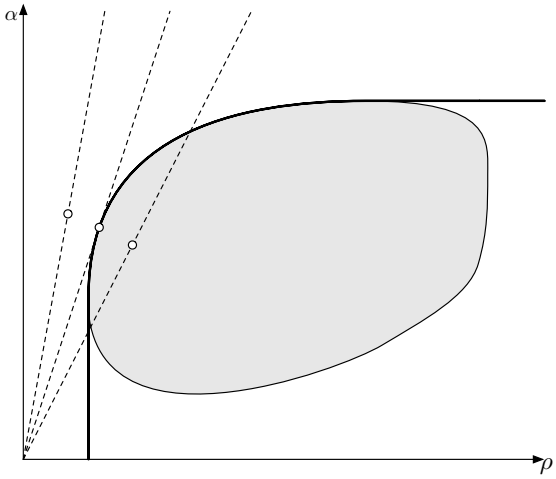

Figure 1. A typical *level set* $\Psi \stackrel{\text{def}}{=} \{\Lambda^\star(\alpha,\rho) \leq \log 2\}$, and the line whose slope gives the constant $c = \sup\{\alpha/\rho : \Lambda^\star(\alpha,\rho) \leq \log 2\}$. A couple (α,ρ) corresponds to nodes $u \in T_\infty$ lying $\lceil \rho^{-1}\log n \rceil$ levels away from the root and $D_u \geq \alpha \log n$. If the point (α,ρ) lies inside the level set, then such nodes exist with high probability, whereas no such exist if it lies outside. This is why the weighted height is characterized by the line of maximum slope with one end at the origin and the other in Ψ.

Theorem 1 (Broutin, Devroye, and McLeish, 2007)**.** *Let T_n be a random tree and let H_n be its height. Let $c = \sup\{\alpha/\rho : \Lambda^\star(\alpha,\rho) \leq \log 2\}$, where $\Lambda^\star(\cdot,\cdot)$ is the Cramér function associated with the vector (Z,E). Then,*

$$H_n = c\log n + o(\log n) \qquad \textit{in probability,}$$

as $n \to \infty$.

The techniques used in the proof of Theorem 1 are very close to an analysis of the expected weighted profile of the tree, i.e., a count of the number of nodes per level and weighted depth. Indeed, the height corresponds the the maximal value of $\frac{\alpha}{\rho}\log n$ such that the expected number of nodes $u \in T_n$ lying k levels away from the root,

$$2^k \mathbf{P}\left\{D_u > \frac{\alpha}{\rho}\log n, u \in T_n\right\} = \exp\left(k(\log 2 - I(\alpha,\rho) + o(1))\right)$$

tends to infinity as $n \to \infty$. The constant c appearing in Theorem 1 may be interpreted geometrically. See Figure 1.

3 Applications

The general model presented in Section 2 is applicable to a large class of examples. Some of the classical ones are collected with the descriptions of the corresponding distributions (Z, E) in Table 1. These examples are treated in detail by Broutin et al. (2007).

Model	$\mathbf{X} = (\mathbf{Z}, \mathbf{E})$	**c**	**References**
Binary search tree	$(1, -\log U)$,	4.331...	Devroye (1986, 1987)
Random recursive ree	$(\mathrm{Be}(1/2), -\log U)$	e	Pittel (1994)
Optimal BST	$(1, \log 2)$	$\frac{1}{\log 2}$	Martínez and Roura (2001)
			Broutin et al. (2007)
d-ary increasing tree	$(1, \mathrm{beta}(\frac{1}{d-1}, 1))$	$(d-1)c\log(\frac{de}{(d-1)c})$	Bergeron et al. (1992)
		$c \geq \frac{1}{d-1}$	Broutin et al. (2007+)

Table 1. Some of the possible applications of Theorem 1. In the entire table, U denotes a $[0,1]$-uniform random variable, $\mathrm{Be}(p)$ a Bernoulli random variable with parameter p, and $\mathrm{beta}(\alpha,\beta)$ a beta random variable with parameters α and β.

Some other examples are less directly related to the problem of the height of a tree. For instance, the shape of skinny cells in geometric structures may (surprisingly?) be studied using Theorem 1. Consider the recursive partitions given by k-d or relaxed k-d trees. We are given a set of n uniform random points in $[0,1]^2$ and we recursively cut the square into smaller cells using hyperplanes parallel to the axis, and going through a random point. In the k-d tree, the direction of the hyperplanes change in a cyclical way at each level of the partition, whereas the direction is always random and independent for relaxed k-d trees. For precise definitions and references, see Broutin, Devroye, and McLeish (2007).

Let R_n and $R_n^\star$ be the worst-case ratios of the lengths of two sides of a cell in a random k-d and relaxed k-d tree, respectively. Then, we have $R_n = n^{\sqrt{3}/2+o(1)}$ in probability while $R_n^\star = n^{1+o(1)}$ in probability. This structural discrepancy accounts for the different behaviors of the two structures with respect to partial match queries for example (Duch and Martínez, 2002).

4 Concluding remarks

The model we have described here is very general, as shown by the sample examples given in Section 3. Many more applications are presented by Broutin and Devroye (2006) and Broutin et al. (2007). There is an important family of random trees that are very similar but do not fit in the model: the tries (see, e.g., Szpankowski, 2001), do not satisfy the "bounded height" condition and are not covered. It is possible to derive a modified version

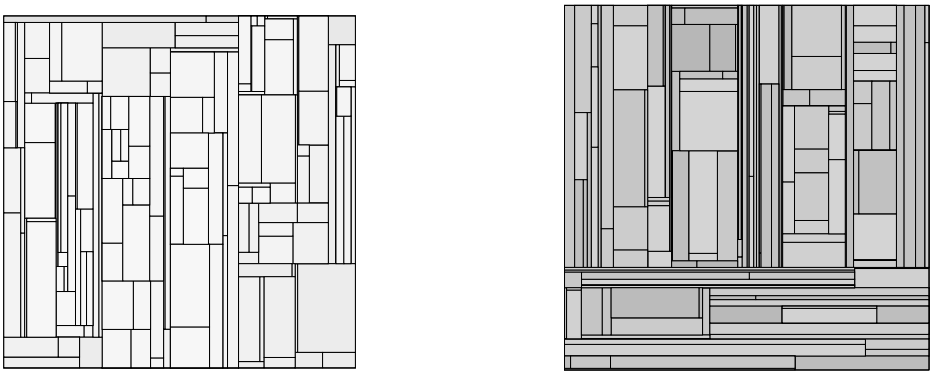

Figure 2. Two randomly generated partitions: on the left a k-d tree and on the right a relaxed k-d tree, both on 150 points. One can see at first glance that cells are skinnier on the right.

of Theorem 1 involving a corrective term that accounts for the height not being bounded. For more on this, see Broutin and Devroye (2007) and Broutin and Devroye (2007+).

References

F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP*, volume 581 of *Lecture Notes in Computer Science*, pages 24–48, 1992.

J. D. Biggins. 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.

N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. *Algorithmica*, 46:271–297, 2006.

N. Broutin and L. Devroye. An analysis of the height of tries with random weights on the edges. *Combinatorics, Probability and Computing*, 2007+. (39 pages). To appear.

N. Broutin and L. Devroye. The height of list tries and TST. In *International Conference on Analysis of Algorithms*, 2007. (13 pages). To appear.

N. Broutin, L. Devroye, and E. McLeish. Weighted height of random trees. Manuscript (46 pages), 2007.

N. Broutin, L. Devroye, E. McLeish, and M. de la Salle. The height of increasing trees. *Random Structures and Algorithms*, 2007+. (23 pages), to appear.

L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33:489–498, 1986.

L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24:277–298, 1987.

A. Duch and C. Martínez. On the average performance of orthogonal range search in multidimensional data structures. *Journal of the Algorithms*, 44(1):226–245, 2002.

R. S. Ellis. Large deviations for a general class of random vectors. *The Annals of Probability*, 12:1–12, 1984.

J. Gärtner. On large deviations from the invariant measure. *Theory of Probabability and its Applications*, 22:24–39, 1977.

C. Martínez and S. Roura. Optimal sampling strategies in quicksort and quickselect. *SIAM Journal on Computing*, 31:683–705, 2001.

B. Pittel. Note on the height of random recursive trees and m-ary search trees. *Random Structures and Algorithms*, 5:337–347, 1994.

W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley, New York, 2001.