Large Deviations for the Weighted Height of an Extended Class of Trees

Nicolas Broutin Luc Devroye *

March 31, 2005

Abstract

We use large deviations to prove a general theorem on the asymptotic edge-weighted height $H_n^\star$ of a large class of random trees for which $H_n^\star \sim c \log n$ for some positive constant c. A graphical interpretation is also given for the limit constant c. This unifies what was already known for binary search trees (Devroye, 1986, 1988), random recursive trees (Devroye, 1987) and plane oriented trees (Pittel, 1994) for instance. New applications include the heights of some random lopsided trees (Kapoor and Reingold, 1989) and of the intersection of random trees.

Keywords and phrases: Random binary search trees, random recursive trees, plane oriented trees, lopsided trees, probabilistic analysis, large deviations.

1 Introduction

This paper gives general laws of large numbers for the height of a class of edge-weighted random trees, which includes as special cases random binary search trees (Devroye, 1986), random recursive trees, random plane oriented trees (Pittel, 1994), and random split trees (Devroye, 1998b). But it also covers random k-ary trees not analyzed until now. The paper extends the earlier theorems of Devroye (1986, 1987, 1998a) where the theory of branching processes was used for this purpose. A special kind of branching random walk permitted Biggins and Grey (1997) to obtain the asymptotic height of various random trees including random binary search trees and random recursive trees. We propose in this paper a method based on large deviations for sums of independent random variables. The closest approach was the one of Biggins (1978) which used multidimensional branching processes. Our method makes intensive use of Cramér's Theorem for large deviations (Dembo and Zeitouni, 1992; Grimmett and Stirzaker, 2001) and some properties of the rate functions it defines. The height is characterized as the solution of a 2-dimensional optimization problem involving Cramér's functions. We apply our method in some cases where these functions can be expressed in a closed form. In particular, we are able to obtain the height for random binary search trees, random recursive trees, random median-of-$(2k+1)$ trees and random lopsided trees, thus extending the class of trees covered by a single theorem.

We first present the main result and its proof, taking for granted some results about large deviations. The proofs for these have been put in Appendix. We next make the link

*Research of the authors was supported by NSERC Grant A3456 and by a James McGill Fellowship. Address: School of Computer Science, McGill University, Montreal H3A2K6, Canada. Email: {luc,nbrout}@cs.mcgill.ca

between trees of random variables and random trees of size n, leaving the most interesting part on applications as a concluding section.

2 Main results

Consider an infinite rooted b-ary tree T_∞. Let $\pi(u)$ denote the set of edges in T_∞ on the path from the root to node u. Assign independently to each node u a random b-vector $\{(Z_1, E_1), (Z_2, E_2), \ldots, (Z_b, E_b)\}$, where each couple (Z_i, E_i) is distributed as (Z, E) for non-negative and independent Z and E, both having finite means. Also, E is not mono-atomic and has no atom at 0. Note that every Z is independent of any E, but the couples (Z_i, E_i) inside a single b-vector may be dependent. These random variables are mapped to the edges, so that each edge e receives a couple (Z_e, E_e). For each node u, let its age be $G_u = \sum_{e\in\pi(u)} E_e$ and define the weighted depth $D_u = \sum_{e\in\pi(u)} Z_e$. We now have a complete b-ary tree with nodes augmented with two independent random variables G_u and D_u. Define T_n to be the tree of nodes $u \in T_\infty$ for which $G_u \leq n$. We are interested in the weighted height $H_n = \max\{D_u : u \in T_n\}$. The following theorem characterizes H_n, whatever the distributions of Z and E. In the sequel, we let $\Lambda_Z^\star$ denote the right-tail Cramér function for a random variable Z (Dembo and Zeitouni, 1992; Rockafellar, 1970) (see below). Also, $\Lambda_E^\star$ denotes the left-tail Cramér function for E. More about large deviations can be found in Appendix.

Theorem 1. *Let $\{E_e, e \in T_\infty\}$ and $\{Z_e, e \in T_\infty\}$ be families of random variables as in the previous setting. Then*

$$\frac{H_n}{n} \xrightarrow[n\to\infty]{} c$$

in probability, where c is the unique maximum value of α/ρ along the curve $\mathscr{C}_{Z,E}$ and

$$\mathscr{C}_{Z,E} = \{(\rho, \alpha) : \Lambda_Z^\star(\alpha) + \Lambda_E^\star(\rho) = \log b, \rho \leq \mathbf{E}E, \alpha \geq \mathbf{E}Z\}. \tag{1}$$

We first argue about the existence of a solution in (1). We need to show that $\mathscr{C}_{Z,E} \neq \varnothing$. But from fact 3 of Lemma 5, $\Lambda_Z^\star(\mathbf{E}Z) = 0$. Since $\Lambda_E^\star$ is continuous where it is not infinite (Lemma 10), $\Lambda_E^\star(\mathbf{E}E) = 0$ and $\lim_{\rho\to 0} \Lambda_E^\star(\rho) = \infty$, there must be a value ρ_0 for which $\Lambda_E^\star(\rho_0) = \log b$. Thus $(\mathbf{E}Z, \rho_0) \in \mathscr{C}_{Z,E} \neq \varnothing$. The uniqueness of the constant c defined above follows from the geometry of $\mathscr{C}_{Z,E}$:

Lemma 1. *The curve $\mathscr{C}_{Z,E}$ defined in Theorem 1 is increasing and concave.*

Note that ρ/α is the slope of a line with one endpoint at the origin and the other one on $\mathscr{C}_{Z,E}$. If Z has a single atom at $\mathbf{E}Z$, then $\mathscr{C}_{Z,E}$ consists of a single point, and there is nothing to show. So assume that Z too is not mono-atomic. As we will see in Proposition 8, $\Lambda_Z^\star$ has a derivative at $\alpha = \mathbf{E}Z$ and

$$\frac{d}{d\alpha}\Lambda_Z^\star(\alpha)\bigg|_{\alpha=\mathbf{E}Z} = 0.$$

This means that the graph of $\mathscr{C}_{Z,E}$ has a horizontal tangent at $\mathbf{E}Z$, because if $\alpha = \mathbf{E}Z$, then $\rho \neq \mathbf{E}E$. Similarly, the slope is infinite at the other end of the domain, at a point $(\mathbf{E}E, \alpha^\star)$. Using Lemma 1 above, this shows that the optimal point occurs for $\alpha \in (\mathbf{E}Z, \alpha^\star)$ (see Figure 1). We conclude:

Lemma 2. *If Z is not mono-atomic, then the maximal value $c = \alpha_0/\rho_0$ on $\mathscr{C}_{Z,E}$ occurs in the interior of $\mathscr{C}_{Z,E}$. If Z is mono-atomic, then $\mathscr{C}_{Z,E}$ consists of a single point $(\rho^\star, \mathbf{E}Z)$, where $\Lambda_E^\star(\rho^\star) = \log b$.*

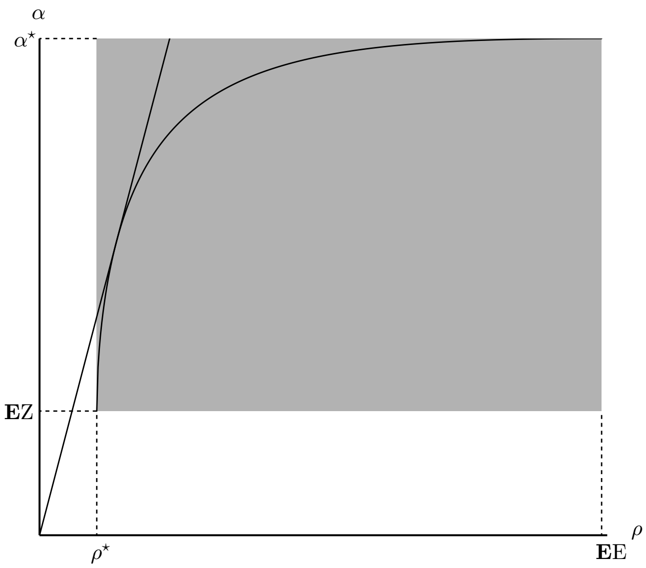

Figure 1: Geometric interpretation of the constant c in the non-degenerate case. We draw α versus ρ along the curve $\mathscr{C}_{Z,E}$. Note that $\mathbf{E}Z < \alpha < \alpha^\star$ and that $\rho^\star < \rho < \mathbf{E}E$. The slope of the tangent line is c. The grey box is the domain to be considered.

Proof of Lemma 1. Assume that none of Z and E are single masses. Since $\Lambda_Z^\star$ and $\Lambda_E^\star$ are the rate functions for the right tail of Z and left tail of E, $\Lambda_Z^\star$ increases and $\Lambda_E^\star$ decreases on its support. Thus, $\mathscr{C}_{Z,E}$ is increasing. Consider now $\alpha_1, \alpha_2 \geq \mathbf{E}Z$ and $x \in (0,1)$. Let ρ_1, ρ_2 and ρ be the points of $\mathscr{C}_{Z,E}$ corresponding to α_1, α_2 and $\alpha = x\alpha_1 + (1-x)\alpha_2$, respectively. We have to show that $\rho \leq x\rho_1 + (1-x)\rho_2$ to obtain the concavity. By convexity of both rate functions,

$$\begin{aligned}
\Lambda_E^\star(x\rho_1 + (1-x)\rho_2) &\leq x\Lambda_E^\star(\rho_1) + (1-x)\Lambda_E^\star(\rho_2) \\
&= \log 2 - (x\Lambda_Z^\star(\alpha_1) + (1-x)\Lambda_Z^\star(\alpha_2)) \\
&\leq \log 2 - \Lambda_Z^\star(x\alpha_1 + (1-x)\alpha_2) \\
&= \Lambda_E^\star(\rho).
\end{aligned}$$

Since $\Lambda_E^\star$ decreases, the concavity holds. $\square$

Knowing that c is uniquely defined, we can address the issue of the main part of the proof of Theorem 1. We split it into two lemmas, following the analysis of Devroye (1998a). Collecting the results to prove Theorem 1 is then straightforward.

Lemma 3 (Upper bound)**.** *With the notations of Theorem 1, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{H_n \geq c(1+\epsilon)n\right\} = 0.$$

Proof. Let c be defined as in Theorem 1 and let $\epsilon > 0$ be arbitrary. Let $L_{n,k}$ be the set of nodes u that are k levels away from the root in T_n. For any $\alpha \geq \mathbf{E}Z$, the union bound on the nodes at level k gives

$$
\begin{aligned}
\mathbf{P}\left\{\exists u \in L_{n,k} : D_u \geq \alpha k\right\} &\leq b^k \mathbf{P}\left\{\sum_{i=1}^k Z_i \geq \alpha k, \sum_{i=1}^k E_i \leq n\right\} \\
&= b^k \mathbf{P}\left\{\sum_{i=1}^k Z_i \geq \alpha k\right\} \cdot \mathbf{P}\left\{\sum_{i=1}^k E_i \leq n\right\},
\end{aligned}
$$

by independence. For each tail probability, we use Chernoff's bound, which is a one-sided explicit version of Crámer's theorem (Dembo and Zeitouni, 1992; Grimmett and Stirzaker, 2001). Recalling that $\Lambda_E^\star(\rho) = 0$ for $\rho \geq \mathbf{E}E$ and $\Lambda_Z^\star(\alpha) = 0$ for $\alpha \leq \mathbf{E}Z$, and setting $\rho = n/k$, we have

$$
\begin{aligned}
\mathbf{P}\left\{\exists u \in L_{n,k} : D_u \geq \alpha n/\rho\right\} &\leq b^k \mathbf{P}\left\{\sum_{i=1}^k Z_i \geq \alpha k\right\} \cdot \mathbf{P}\left\{\sum_{i=1}^k E_i \leq \rho k\right\} \\
&\leq b^k \exp\left(-k \cdot \Lambda_Z^\star(\alpha)\right) \cdot \exp\left(-k \cdot \Lambda_E^\star(\rho)\right) \\
&= \exp\left(k\left[\log b - \Lambda_Z^\star\left(\alpha\right) - \Lambda_E^\star\left(\rho\right)\right]\right) \\
&\stackrel{\text{def}}{=} \exp\left(-k\gamma\left(\alpha,\rho\right)\right).
\end{aligned}
$$

With $\alpha = (1+\epsilon)c\rho$, we have $\gamma(\alpha,\rho) = \gamma((1+\epsilon)c\rho,\rho) \geq \beta \stackrel{\text{def}}{=} \inf_\rho \gamma((1+\epsilon)c\rho,\rho) > \inf_\rho \gamma(c\rho,\rho) = 0$. The last statement follows from the continuity of γ where it is finite, and the monotonicity in both of its arguments. Using the union bound we obtain, for any integer K,

$$
\begin{aligned}
\mathbf{P}\left\{H_n \geq c(1+\epsilon)n\right\} &= \mathbf{P}\left\{\exists k, \exists u \in L_{n,k} : D_u \geq c(1+\epsilon)n\right\} \\
&\leq \sum_{k=0}^{K-1} \mathbf{P}\left\{\exists u \in L_{n,k} : D_u \geq c(1+\epsilon)n\right\} \\
&\qquad + \sum_{k\geq K} \mathbf{P}\left\{\exists u \in L_{n,k} : D_u \geq c(1+\epsilon)n\right\}.
\end{aligned}
$$

Consider first the second term:

$$
\sum_{k\geq K} \mathbf{P}\left\{\exists u \in L_{n,k} : D_u \geq c(1+\epsilon)n\right\} \leq \sum_{k\geq K} e^{-\beta k} = O\left(e^{-\beta K}\right).
$$

Furthermore, by Markov's inequality,

$$
\begin{aligned}
\sum_{k=0}^{K-1} \mathbf{P}\left\{\exists u \in L_{n,k} : D_u \geq c(1+\epsilon)n\right\} &\leq \sum_{k=0}^{K-1} \mathbf{P}\left\{\exists u \in L_{\infty,k} : D_u \geq c(1+\epsilon)n\right\} \\
&\leq K \cdot b^K \cdot \sup_{k<K, u \in L_{\infty,k}} \mathbf{P}\left\{D_u \geq c(1+\epsilon)n\right\} \\
&\leq K \cdot b^k \cdot \frac{K\mathbf{E}Z}{(1+\epsilon)n} \xrightarrow[n\to\infty]{} 0,
\end{aligned}
$$

which shows that $\mathbf{P}\left\{H_n \geq c(1+\epsilon)n\right\} \leq O\left(e^{-\beta K}\right) + o(1)$. Since K is arbitrary, the proof is complete. $\square$

Lemma 4 (Lower Bound). *For any $\epsilon > 0$, there exists some node u for which $D_u \geq c(1-\epsilon)n$ with probability tending to 1 as $n \to \infty$.*

Proof. This proof uses ideas of Biggins (1977). Let $\epsilon > 0$ be arbitrary. We exhibit a path in T_n to a node u with $D_u \geq c(1-2\epsilon)\log n$. For this purpose, we will build a surviving Galton-Watson process. We start with the root. Consider in T_∞ the nodes L levels away. A given node u is called *good* if $D_u \geq \alpha L$ and $G_u \leq \rho L$, for some α and ρ to be chosen later. We define the Galton-Watson children to be the good nodes. Each child reproduces independently according to the same reproduction distribution, that is, a node v lying L levels below a node u is a good child of u if $D_v - D_u \geq \alpha L$ and $G_v - G_u \leq \rho L$. The process of good nodes survives with positive probability if the expected number of children is more than one. Write N_L for the number of good nodes L levels away from the root.

$$\begin{aligned} \mathbf{E}N_L &= b^L \mathbf{P}\left\{D_u \geq \alpha L, G_u \leq \rho L\right\} \\ &= \exp(-\gamma(\alpha,\rho)L + o(L)), \end{aligned}$$

according to the proof of Lemma 3. Choosing α and ρ such that $\gamma(\alpha,\rho) < 0$ makes $\mathbf{E}N_L > 1$ for L large enough. Picking $\alpha = \alpha_0$ and $\rho = \rho_0/\sqrt{1-\epsilon}$ suffices, where α_0, ρ_0 are as in Lemma 2. Therefore, writing $q < 1$ for the probability of extinction, the process survives with probability $1-q > 0$. We now have to boost this probability up to $1-o(1)$. We do this by starting the Galton-Watson process at level tL instead, giving more chance that at least one of the b^{tL} processes survives. We need now to consider the joint distribution of the E random variables $\{E_1, E_2, \ldots, E_b\}$ down the same vertex: for any $\beta > 0$, we can pick a such that $\mathbf{P}\left\{E_1 \leq a, E_2 \leq a, \ldots, E_b \leq a\right\} \geq 1-\beta$. Let A be the event that all the E_e random variables in the top tL levels take values less than a. Then,

$$\mathbf{P}\left\{A^c\right\} \leq \beta\, b^{tL}.$$

Thus A occurs with probability arbitrary close to one, controlled by our choice for β. And if A is true, then all nodes v at level tL are such that $G_v \leq atL$. Let now B be the event that one of the b^{tL} Galton-Watson processes survives. Then

$$\mathbf{P}\left\{B^c\right\} = q^{b^{tL}},$$

by independence. If both A and B occur, then there is a node u at level $tL + kL$ in T_∞ such that $G_u \leq atL + \rho kL$ and $D_u \geq \alpha kL$. Taking

$$k = \left\lfloor \frac{n(1-\epsilon)}{\rho_0 L} \right\rfloor,$$

$\rho = \rho_0/\sqrt{1-\epsilon}$ and $\alpha = \alpha_0$ gives $G_u \leq n\sqrt{1-\epsilon} + atL < n$, for n large enough and $D_u \geq c(1-\epsilon)n - \alpha_0 L \geq c(1-2\epsilon)n$ for n large enough. Thus

$$\mathbf{P}\left\{H_n \geq c(1-2\epsilon)n\right\} \geq \mathbf{P}\left\{A \cup B\right\} \geq 1 - \mathbf{P}\left\{A^c\right\} - \mathbf{P}\left\{B^c\right\},$$

and we can control the lower bound and make it as close to 1 as we want by choice of β and t independently of n. Therefore, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left\{H_n \geq c(1-2\epsilon)n\right\} = 1. \qquad \square$$

3 Towards applications

Our primary aim is to obtain results about the height of some incrementally built random trees. The main problem is the number of nodes. Indeed, the tree T_m has a random number of nodes N. We would like to pick the right m for N to be almost n. In most examples below, we need to pick a node uniformly at random in order to grow the tree. This can be achieved in the following way.

We need a particular kind of Crump-Mode-Jagers process (Crump and Mode, 1968), namely a Bellman-Harris process (Athreya and Ney, 1972; Harris, 1963). Let X be a random variable of mean μ which takes non-negative integer values. Consider a branching process that starts with a single individual. It dies at a random time M_1 and gives birth to $X+1$ new independent individuals that behave similarly. Call these events *replacements*. Assume that the lifetime of each individual is exponentially distributed with mean one so that, in particular, $M_1 \stackrel{\mathcal{L}}{=}$ exponential(1). Let M_k be the random time of the kth birth, and N_k the size of the population just before M_k. Because of the memoryless property of the exponential distribution, we just start a new process at M_k with $N_{k+1} = N_k + X_k$ brand new individuals, where X_k is an independent copy of X. Symmetry shows that each individual is equally likely to be the next one to die. Let $\{E_i, i \geq 1\}$, be a family of independent exponential(1) random variables. Since

$$\min\{E_1, E_2, \ldots, E_m\} \stackrel{\mathcal{L}}{=} \frac{E_1}{m},$$

and $X_0 = 1$, we have

$$M_k \stackrel{\mathcal{L}}{=} \sum_{i=1}^{k} \frac{E_i}{\sum_{j=0}^{i-1} X_j}.$$

Estimating the number of nodes when the process is stopped at time m_n is made easier by first considering M_k.

Proposition 1. *The time M_k of the k-th birth satisfies $\mu M_k \sim \log k$, almost surely.*

Proof. We have that

$$\frac{1}{i} \sum_{j=0}^{i-1} X_j \xrightarrow[i\to\infty]{} \mu \quad a.s.,$$

by the strong law of large numbers. From a generalized law of large numbers, (Chow and Teicher, 1978, Theorem 2, p. 331),

$$\frac{1}{\log k} \sum_{i=1}^{k} \frac{E_i}{i} \xrightarrow[k\to\infty]{} 1 \quad a.s.,$$

which together yield

$$\frac{M_k}{\log k} \xrightarrow[k\to\infty]{} \frac{1}{\mu} \quad a.s. \qquad \square$$

With this in hand, we can now consider the number of nodes in the process. Let

$$m_n = \frac{1}{\mu} \cdot \log n, \tag{2}$$

and write $N(t)$ for the number of individuals when the process is stopped at the deterministic time t. It happens that $N(m_n)$ is close enough to n for us to make use of Theorem 1 in the sequel.

Proposition 2. *The number of nodes $N(m_n)$ in the process stopped at time m_n defined in (2) is such that $\log N(m_n) \sim \log n$, almost surely.*

Proof. Again, from the law of large numbers

$$\frac{N_k}{k} \xrightarrow[k\to\infty]{} \mu \quad a.s.,$$

so

$$\frac{\log N_k}{\log k} \xrightarrow[k\to\infty]{} 1 \quad a.s.$$

Using Proposition 1, this yields

$$\frac{\log N_k}{\mu M_k} \xrightarrow[k\to\infty]{} 1 \quad a.s.$$

So, stopping the process at time t,

$$\frac{\log N(t)}{\mu t} \xrightarrow[t\to\infty]{} 1 \quad a.s..$$

Taking $t = m_n$ finishes the proof. □

We now make the link between branching random walks stopped at a deterministic time m_n and random trees over n nodes. Let $H_n^\star$ be the height of a random tree with n nodes, grown by appending, at each step, $1 + X$ children to a uniform random leaf. The following proposition is of interest for most applications.

Proposition 3. *Let c be the constant defined in Theorem 1 with E having exponential(1) distribution. Then*

$$\frac{H_n^\star}{\log n} \xrightarrow[n\to\infty]{} \frac{c}{\mu}$$

in probability.

Proof. Let $\epsilon > 0$. Then, with probability tending to 1 as $n \to \infty$,

$$n^{1-\epsilon} \leq N_n \leq n^{1+\epsilon}.$$

Therefore,

$$\mu(1-\epsilon) \cdot \frac{H^\star_{n^{1-\epsilon}}}{(1-\epsilon)\log n} = \frac{H^\star_{n^{1-\epsilon}}}{m_n} \leq \frac{H^\star_{N_n}}{m_n} = \frac{H_{m_n}}{m_n} \leq \frac{H^\star_{n^{1+\epsilon}}}{m_n} = \frac{H^\star_{n^{1+\epsilon}}}{(1+\epsilon)\log n} \cdot \mu(1+\epsilon),$$

and Theorem 1 yields the conclusion. □

Remark: Since the exponential distribution is the only memoryless distribution, any other choice for the lifetimes leads to non-uniform sampling for the next individual that die. This is sometimes required for the applications such as the median-of-$(2k+1)$ trees (see section 4.3).

4 Applications

We will now present a few applications of Theorem 1, using our unifying view. We present in particular random binary search trees, random recursive trees, median-of-(2k+1) trees, and other models of random trees. The rate functions $\Lambda_Z^\star$ and $\Lambda_E^\star$ are Crámer functions which are often hard to express in a closed form. This makes it difficult to derive useful properties of the optimal point of (1). Also, the equations we obtain are often implicit.

4.1 Random Binary Search Tree

Let us test Theorem 1 on the height of the random binary search tree, which, following Knuth (1973) is defined as follows: take a random permutation $Y_1, Y_2, \ldots Y_n$ of $\{1, 2, \ldots, n\}$; insert the elements Y_i, $i = 1, 2, \ldots, n$ one after an other as nodes in an initially empty search tree. We define the partial rank R_i of Y_i to be the rank of Y_i in the sequence $\{Y_1, Y_2, \ldots, Y_i\}$. We make Y_1 the root and send all Y_i's such that $Y_i < Y_1$ to the left subtree and the others to the right subtree. We then process the elements falling into each subtree in a recursive way. Interesting functionals of this random tree are the depth of Y_n (the time to insert Y_n) and the height $H_n^\star$ (the maximal time to insert an element). Knuth (1973) and Mahmoud (1992b) summarize the known properties. Regarding the height, we have

Theorem 2 (Devroye 1986, 1987, 1998a)**.** *For a random binary search tree,*

$$\frac{H_n^\star}{\log n} \xrightarrow[n\to\infty]{} c$$

in probability, where c is the unique solution greater than 1 of $2e/c = e^{1/c}$.

Usually, Theorem 2 is proved by considering the tree in which we associate with each node the size of the subtree rooted at that node. For instance, at the root, the size of the left subtree is $\lfloor nU \rfloor$, with $U \stackrel{\mathcal{L}}{=} [0,1]$-uniform. This approach implies dealing with some tedious truncations (Devroye, 1986). Instead, we use a property of the partial ranks R_i (Mahmoud, 1992b). Indeed, at time step i, the next element Y_i is equally likely to fall into any of the i slots defined by the partition generated by the previous elements $Y_1, Y_2, \ldots, Y_{i-1}$: R_i is uniform in $\{1, 2, \ldots, i\}$ and $R_1, R_2, \ldots$ are independent (Devroye, 1988). Therefore each external node is equally likely to be picked and replaced by Y_i. This makes the link with Bellman-Harris process of the previous section when we set $1+X = 2$. All edges account for one step down the tree, so that we let $Z = 1$ with probability 1. The depth D_u induced by the values of Z_e's along the path to node u corresponds to the usual depth. For the paper the be self contained, we show the details of the computation of one of the tail function, namely $\Lambda_E^\star$, which will be used many times. We first note that $E \stackrel{\mathcal{L}}{=} -\log U$, where U is $[0,1]$-uniform. Then, the moment generating function

$$M_E(x) = \mathbf{E}\left\{e^{xE}\right\} = \mathbf{E}\left\{U^{-x}\right\} = \frac{1}{1-x}.$$

Then, it follows that the cumulant generating function (see Appendix) is

$$\Lambda_E(x) = -\log(1-x)$$

and the associated rate function

$$\Lambda_E^\star(\rho) = \sup_x\{x\rho - \Lambda_E(\rho)\} = \sup_x\{x\rho + \log(1-x)\}.$$

The maximum is achieved when taking $x = 1 - 1/\rho$ and therefore

$$\Lambda_E^\star(\rho) = \rho - 1 - \log\rho.$$

Also, with these settings,

$$\Lambda_Z^\star(\alpha) = \begin{cases} 0 & \text{if } \alpha = 1 \\ \infty & \text{if } \alpha > 1, \end{cases} .$$

The maximum of α/ρ is achieved when $\alpha_0 = 1$ and $\Lambda^\star_Z(\alpha_0) = 0$ and ρ_0 is the unique solution smaller than 1 of $\rho - 1 - \log\rho = \log 2$, or $2e\rho = e^\rho$. Proposition 3 applies, and therefore, for a random binary search tree,

$$\frac{H^\star_n}{\log n} \xrightarrow[n\to\infty]{} \frac{1}{\rho_0} \tag{3}$$

in probability. Remark that as claimed earlier, the independence is not needed for the random variables appearing in disjoint paths in the tree. For example, in the usual binary search tree, the two edges emanating from the same node have exponentials associated with them that are distributed as $(-\log U, -\log(1-U))$, where U is uniform on $[0,1]$ (Devroye, 1986). Stopping the process at $m_n = \log n$ leads to a tree of size $N_n = n$ a.s. (Devroye, 1998a). Dropping this dependence makes N_n random but the result still holds. As an example, we can generate a tree giving to the two edges out of any node u the same value. The height is not affected and still satisfies (3).

4.2 Random Recursive Tree

The random recursive tree is one of the simplest random trees (Smythe and Mahmoud, 1995). It is built incrementally: when starting, the tree T_1 consists of a single node v_1. At each step i a new vertex v_i is added to the tree and appended as a child to a node chosen uniformly in $\{v_1, v_2, \dots, v_{i-1}\}$. This is sometimes called a Yule process. Various functionals of this tree have been studied in the literature. We are particularly interested in its height $H^\star_n$ when n goes to infinity.

Theorem 3 (Devroye 1987; Pittel 1994)**.** *The height $H^\star_n$ of a random recursive tree with n nodes is $e\log n$ in probability as n goes to infinity.*

To use Theorem 1, we look at the uniform random recursive tree as a binary tree (Figure 2). Recall that a random binary search tree can be build alternatively by choosing at each step an external node uniformly at random and replacing it with an internal one. This is because the partial rank R_i of the element Y_i inserted at time i is uniformly distributed in $\{1, 2, \dots, i\}$ (Mahmoud, 1992b). Therefore, building a binary tree in which the external nodes represent the nodes of our random recursive tree solves the issue of the uniform choice. Thus we want to map the nodes of a rooted tree to the external nodes of a binary tree, in such a way that we keep the information about the distances to the root. Consider a rooted tree $\mathcal{T}_n$ on n vertices. Let $S = \{d_1, d_2, \dots, d_n\}$ be a multiset of numbers that represent the distances from the nodes to the root tree. To make the mapping more visual, we will also describe the construction of a binary tree with labeled edges $\mathcal{T}^b_n$ on n external vertices together with S^b_n, the sequence of distances in $\mathcal{T}^b_n$ (see Figure 2).

- $\mathcal{T}_1$ consists of a single node and $S_1 = \{0\}$. Appending a node yields a tree on two nodes and $S_2 = \{0, 1\}$. Let $\mathcal{T}^b_2$ be the binary tree with two external nodes. Let e and f be its edges. Label them with $z_e = 1$ and $z_f = 0$. Consider the labels as distances. Then $\mathcal{T}^b_2$ has distance sequence $S^b_2 = \{0, 1\} = S_2$.

- Suppose now we are given $\mathcal{T}_n$ and the corresponding $\mathcal{T}^b_n$. They match the distance sequence $S_n = \{d_1, d_2, \dots, d_n\}$. Appending v to node u means that we make $S_{n+1} = S_n \cup \{d+1\}$, where $d \in S$ is the distance from u to the root in both $\mathcal{T}_n$ and $\mathcal{T}^b_n$. In terms of trees, we replace the external node u in $\mathcal{T}^b_n$ by an internal node x. There are two new external nodes associated with x, and the edges e and f out of x are labeled $z_e = 1$ and $z_f = 0$. We may as well label the new external vertices v (such that $e = (x, v)$) and u (with $f = (x, u)$). Then we clearly have $S^b_{n+1} = S^b_n \cup \{d+1\}$, and the sequences S_{n+1} and S^b_{n+1} match, as required.

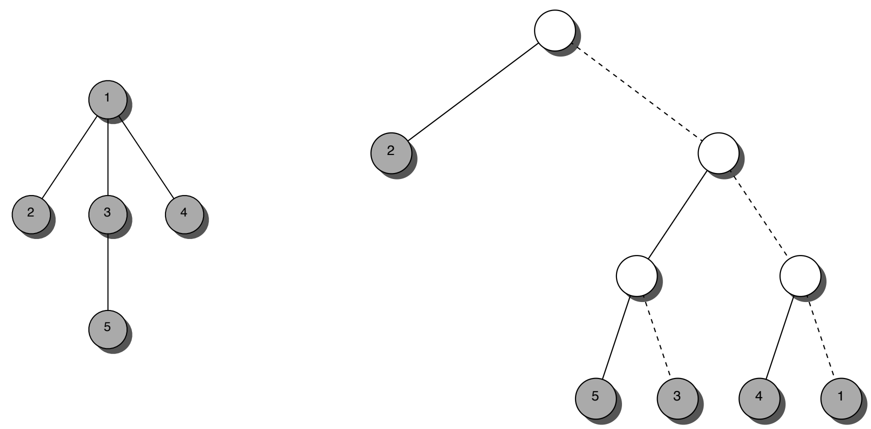

Figure 2: A rooted tree and the corresponding binary tree. The white nodes have been added for the sake of the construction. Solid lines correspond to edges with $Z = 1$ and dashed ones to those with $Z = 0$. Therefore, 1 is equivalent to the root (as the root distance is zero), 2 to the first child of the root (distance one), and so on.

Replacing deterministic labels by random variables makes this model fit for our framework. For the same reason as in binary search trees, $E \stackrel{\mathcal{L}}{=}$ exponential(1). Since on any path π from the root in T_∞, each edge e is as likely to be labeled with 0 as with 1, we have $Z \stackrel{\mathcal{L}}{=}$ Bernoulli(1/2). From Theorem 1, we have to maximize α/ρ on the curve

$$\mathscr{C}_{Z,E} = \{(\alpha, \rho) : \Lambda^\star_Z(\alpha) + \Lambda^\star_E(\rho) = \log 2\}.$$

But we have that $\Lambda^\star_Z = \alpha \log \alpha + (1-\alpha)\log(1-\alpha) + \log 2$ and $\Lambda^\star_E(\rho) = \rho - 1 - \log \rho$ (Dembo and Zeitouni, 1992), which yields

$$\mathscr{C}_{Z,E} = \{(\alpha, \rho) : \alpha \log \alpha + (1-\alpha)\log(1-\alpha) + \rho - \log \rho = 1\}. \tag{4}$$

The slope of the curve $\rho(\alpha)$ is

$$\frac{d\rho}{d\alpha} = \frac{\log \alpha - \log(1-\alpha)}{1/\rho - 1}. \tag{5}$$

Recalling the geometric interpretation shows that the optimal α satisfies

$$\frac{d\rho}{d\alpha} \cdot \alpha = \rho.$$

Straightforward manipulations using (5) give $\alpha \log \alpha - \alpha \log(1-\alpha) = 1 - \rho$. Taking the value for $1-\rho$ in the equation (4) for $\mathscr{C}_{Z,E}$ yields $\rho = 1 - \alpha$. Using this value again in (4) gives the desired result, that is, $\alpha/\rho = e$.

4.3 Median-of-$(2k+1)$ trees and split trees

A well-known improvement of Quicksort (Knuth, 1973; Sedgewick, 1975) samples $2k+1$ elements at random and uses the median as a pivot instead of splitting the data at a

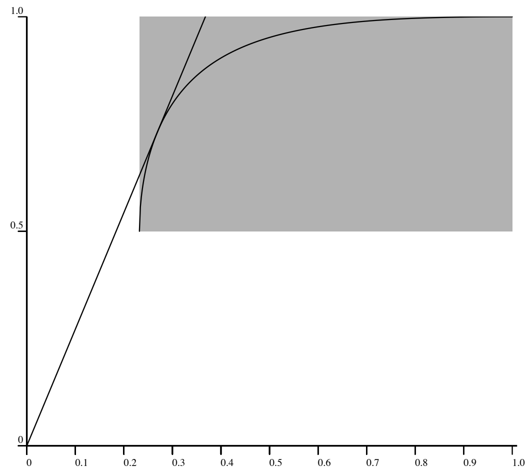

Figure 3: The curve corresponding to uniform recursive trees $\mathscr{C}_{Z,E} = \{(\rho, \alpha) : \alpha \log \alpha + (1-\alpha)\log(1-\alpha) + \rho - \log \rho = 1, \rho \leq 1, \alpha \geq 1/2\}$.

uniform random point. Such a scheme makes the splits more balanced, and therefore the tree less high. The trees produced are called median-of-$(2k+1)$ trees. They can equivalently be viewed as trees produced by a balancing heuristic applied to the fringe of the tree (Poblete and Munro, 1985; Walker and Wood, 1976). It is clear that the external nodes are no longer uniformly picked, and as a consequence Proposition 3 does not apply. However, since our goal is to demonstrate the generality of our framework, we will avoid wandering around and refer for the details to Devroye (1993).

Consider again the associated tree of subtree sizes, with the root having value n. The sizes of the subtrees of the children of the root are both distributed as multinomial $(n-1, W, 1-W)$ random variables, where W and $1-W$ are in turn distributed as the median of $2k+1$ uniform $[0,1]$ random variables, which is known to be beta$(k+1, k+1)$. The multinomial is really concentrated about its mean, and thus behaves roughly as $(nW, n(1-W))$. So we get same first order behavior for the height and other parameters if we were to associate with the edges out of the root random variables W and $1-W$, and let the tree consist of all nodes u for which the product of edge values on $\pi(u)$ is at least $1/n$. Equivalently, taking logarithms, and associating with sibling edges the values $-\log(W)$ and $-\log(1-W)$, and independently so for all other sibling pairs, we may let the tree consist of all nodes u for which the sum of edge values on $\pi(u)$ is at most $\log n$. Thus, $-\log W$ now plays the role of the lifetime of a particle. We are able to rediscover with little work the following theorem.

Theorem 4 (Devroye 1993)**.** *The height $H_n^\star$ of a median-of-$(2k+1)$ tree satisfies*

$$\frac{H_n^\star}{\log n} \xrightarrow[n\to\infty]{} c(k)$$

in probability, where $c = c(k)$ is the unique solution of the equation

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

and s is implicitly defined by the equation

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

Proof. (Outline only.) Let X be beta$(k+1, k+1)$ and $E = -\log X$. Let $Z = 1$ almost surely. Then

$$M(s) = \mathbf{E}\left\{e^{-s\log X}\right\} = \frac{\Gamma(2k+2)}{\Gamma(k+1)^2}\int_0^1 x^{-s}x^k(1-x)^k dx = \frac{(2k+1)!}{k!}\prod_{i=k+1}^{2k+1}\frac{1}{i-s}.$$

Then we have that

$$\Lambda_E^\star(\rho) = \sup_t \left\{\rho t + \log M(t)\right\}.$$

The optimization corresponds to equation

$$\frac{d}{dt}\left(\rho t + \log M(t)\right) = 0,$$

that is

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

Using Theorem 1, since $\mathscr{C}_{Z,E}$ is restricted to a single point at $\alpha = 1$, we get the second equation $\Lambda_E^\star(\rho) = \log 2$, in other words

$$\rho s + \sum_{k+1}^{2k+1}\left(\log(i-s) - \log(i)\right) = \log 2. \qquad \square$$

For example, in the case $k = 1$ which corresponds to the much studied median-of-3 quicksort, we have that

$$M(s) = \mathbf{E}\left\{e^{-s\log B}\right\} = \frac{6}{(3-s)(2-s)}.$$

Therefore,

$$\Lambda_E^\star(\rho) = \sup_s\{\rho s + \log(3-s) + \log(2-s)\}.$$

Optimizing with respect to s yields

$$s = \frac{5\rho - 2 - \sqrt{\rho^2+4}}{2\rho}.$$

Using this optimal value for s leads to an equation we can solve numerically: $\rho = .31322\ldots$, $c = 3.19258\ldots$ and

$$H_n^\star \sim 3.19258\ldots\log n.$$

We note here that the split trees of Devroye (1998b) can be dealt with similarly. If a subtree of size n is roughly split into subtrees of sizes close to $(nV_1, nV_2, \ldots, nV_b)$, where $V_i \geq 0$, and $\sum_i V_i = 1$, then we can study the height by setting $Z_e = 1$ for all edges, and $(E_1, \ldots, E_b) = (-\log V_1, \ldots, -\log V_b)$.

4.4 Random lopsided trees

In information theory, researchers are interested in building codes that are optimal with respect to various measures. Prefix-free codes are particularly interesting because they can be decoded directly by following a path in a tree, and output a character correponding to the codeword when reaching a leaf. In such trees, a node u represents a prefix p, and its children the words that can be built by appending a symbol to p. In digital applications, characters are usually encoded with bits and therefore processing each symbol has the same cost. We can think of costs by assigning lengths to the edges in the tree; in this case, they would all have equal length. But some codes have encoding length depending on the symbols. These are called Varn codes (Varn, 1971). Such encodings lead to trees whose edges have non-equal length, *lopsided trees* (Choi and Golin, 2001; Kapoor and Reingold, 1989). Let $c_1 \leq c_2 \leq \cdots \leq c_b$ be fixed positive real numbers, then a tree is said to be lopsided if it is b-ary rooted, and for each node, the edge to the i-th child has length c_i. We define a random lopsided tree, and show how its height can be studied using Theorem 1.

Define a random lopsided tree by the following inductive process: start with a single internal node, L_1. To build a random tree L_{n+1} with $n+1$ (internal) nodes, take an instance of L_n, pick an external node uniformly at random and replace it with an internal node. The weights of the b child edges of that internal node are $c_1, c_2, \ldots, c_b$, a set of nonnegative real numbers that are not all equal. In our model, we should take E exponential and $Z = c_W$, where W is uniform on $\{1, \ldots, b\}$.

Theorem 5. *The height $H_n^\star$ of a random b-ary lopsided tree having n nodes built with costs $\{c_1, c_2, \ldots, c_b\}$ satifies*

$$H_n^\star \sim \frac{c}{b-1} \cdot \log n$$

in probability, where c is the unique maximal value of α/ρ under the constraint that

$$\alpha t(\alpha) + \log \alpha - \log \left(\sum_i c_i e^{t c_i} \right) + \rho - 1 - \log \rho = 0, \tag{6}$$

where $t(\alpha)$ is uniquely defined by

$$\sum_i (\alpha - c_i) e^{t c_i} = 0. \tag{7}$$

Remark: Theorem 5 does not formally apply to the case of equal c_i's. It is easy to verify though that when $c_1 = c_2 = \cdots = c_b = 1$, we are led to

$$H_n^\star \sim \frac{c}{b-1} \log n$$

in probability, where $c = 1/\rho$, and ρ is the unique solution greater than 1 of $\Lambda_E^\star(\rho) = \rho - 1 - \log \rho = \log b$.

Our random lopsided trees may also be used when we replace a random node by a fixed deterministic tree. The growing process is as follows. Start with a grey node. Each step sees the replacement of uniformly selected random grey node by a deterministic tree consisting of k nodes (see, e.g., Figure 4). In this replacement tree, all leaves, as well as none, some or all of the internal nodes are painted grey (if the root is grey, then the node just replaced may be selected again), for a total of $\ell \leq k$ grey nodes. If we are interested in standard distances to the root, and in the classical definition of the height, then we can imagine another tree in which the replaced node receives a number ℓ of children, with edge weights equal to the distances to the root in the replacement tree. The original tree has sizes given by $1+s(k-1)$ for s integer, and the new imagined tree has sizes given by $1+s\ell$ for s integer: they are linearly related. The weighted height in the new tree corresponds to the standard height in the original tree. We work out two examples.

In Figure 4, we replace a randomly picked grey node by a subtree of five nodes, of which two grey nodes, at distances 1 and 3 from their roots. This corresponds to a random lopsided tree (modulo a proportionality constant in the size of the tree) with weights $(1, 3)$, and fanout $b = 2$. The slope of the tangent going through the origin is $9.3389\ldots$, implying

$$H_n^\star \sim 9.3389\ldots \log n$$

in probability.

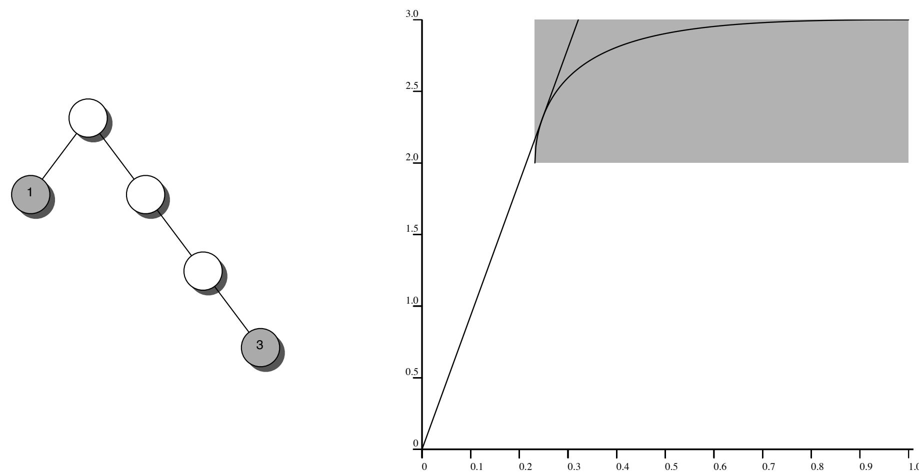

Figure 4: The pattern that replaces a grey node and the curve $\mathscr{C}_{Z,E}$ together with the optimal tangent when the set of costs is $\{1, 3\}$. The nodes are labeled with their depth.

In Figure 5, we have the same replacement, but paint all five nodes grey. This yields the random lopsided tree with fanout $b = 5$ and cost vector $(0, 1, 1, 2, 3)$. The slope of the optimal tangent is $20.966\ldots$, which gives the height after renormalization:

$$H_n^\star \sim 5.241\ldots \log n$$

in probability.

Proof of Theorem 5. In this model, external nodes are picked uniformly at random and Proposition 3 applies, with $X = b - 1$ almost surely. Therefore, $\Lambda_E^\star(\rho) = \rho - 1 - \log \rho$. Since on a path to the root, each edge is equally likely to have any cost,

$$\Lambda_Z(t) = \log\left(\mathbf{E}\left\{e^{tZ}\right\}\right) = \log\left(\sum_i e^{tc_i}\right) - \log b.$$

Using the definition $\Lambda_Z^\star(\alpha) = \sup_t\{\alpha t - \Lambda_Z(t)\}$, we see that the optimal value is obtained for

$$\alpha = \frac{\sum_i c_i e^{tc_i}}{\sum_i e^{tc_i}},$$

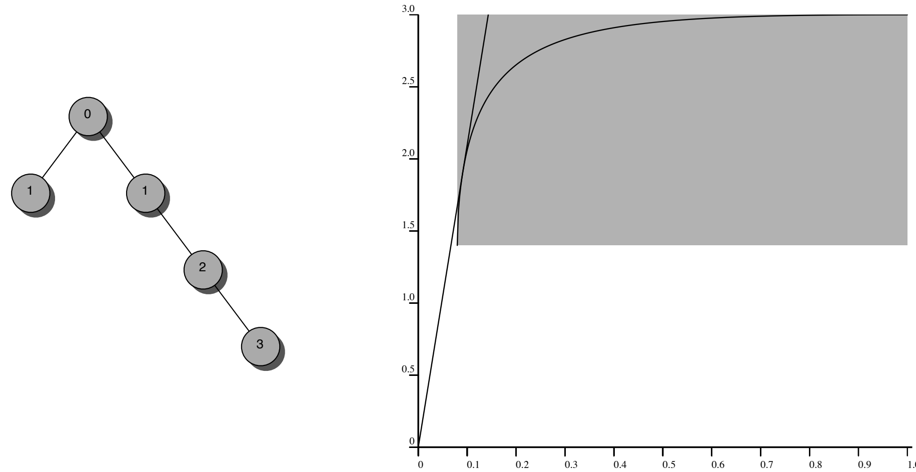

Figure 5: With the set of costs $\{0, 1, 1, 2, 3\}$, one can think of a uniform grey node being replaced by the tree pattern on the left.

which is equivalent to (7). The value $t(\alpha)$ is unique as long as at least two of the c_i's are distinct. (6) follows immediately from Proposition 3. $\square$

4.5 Plane oriented trees and linear recursive trees

Plane oriented trees (PORTs) are rooted trees in which the children of every node are oriented. A random PORT with n nodes is defined as a tree taken uniformly at random from the set of $(n-1)!$ plane oriented trees with n nodes. The depths of nodes in random PORTs have been studied by Mahmoud (1992a) and their height by Pittel (1994). An interesting property of PORTs is their recursive description: one can view a random PORT with n nodes as a random PORT with $n-1$ nodes, to which we add a node uniformly at random in the set of slots available. Nodes have labels 1 through n in order of addition, and therefore, the label numbers are increasing on paths down from the root. The slots are the positions in the tree that lead to different new trees. Because of the order, each node with k children has $k+1$ slots (external nodes) attached to it as described in Figure 6.

We may consider them as *linear recursive trees*, a more general model of Pittel (1994), which has also been dealt with by Biggins and Grey (1997). For this kind of tree, each node u has a weight w_u, and when growing a random linear recursive tree, a new node is added as a child to u at random with probability proportional to w_u. For linear recursive trees, we have $w_u = 1 + \beta \mathrm{deg}_u$, where deg_u denotes the number of children of u and $\beta \geq 0$ is called the parameter. We can obtain the same distribution on trees by taking external nodes uniformly at random and with a suitable number of external nodes for each vertex, at least when β is integer (see below).

Assume that β is integer-valued. It is easily seen that when we pick a uniform external node at depth d, and replace it by $\beta+2$ new external nodes, $\beta+1$ at depth d and one at $d+1$, then this may be seen as replacing a uniform external node by the fixed tree pattern of Figure 6. The Z values of the $\beta+2$ child edges of a node consist of one 1 and $\beta+1$ 0's. A typical Z value therefore is Bernoulli $(1/(\beta+2))$. One may apply our result on random lopsided trees with fanout $\beta+2$ to find a new proof of Pittel's theorem on the height.

Theorem 6 (Pittel 1994)**.** *Assume that β is integer-valued. The height $H_n^\star$ of a random*

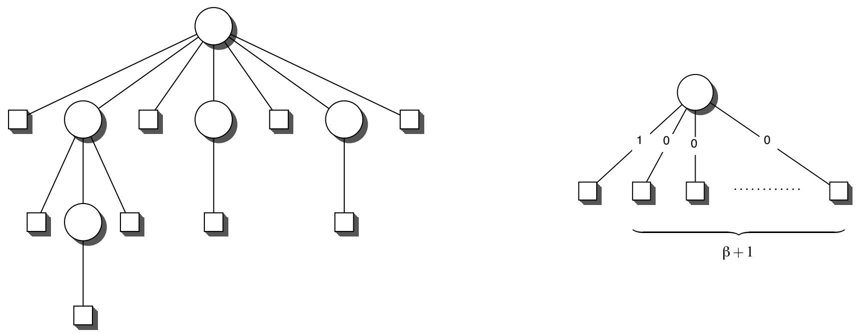

Figure 6: A PORT with the slots represented by squares on the left and the tree pattern on the right, representing the replacement of an external node. The labels on the edges are the costs of crossing them.

linear recursive tree with parameter β and n nodes is such that

$$\frac{H_n^\star}{\log n} \xrightarrow[n\to\infty]{} \frac{c}{\beta+1}$$

in probability, where c is the maximal value of α/ρ along

$$\left\{ \begin{array}{l} \alpha\log((\beta+2)\alpha) + (1-\alpha)\left(\log((\beta+2)(1-\alpha)) - \log(\beta+1)\right) + \rho - 1 - \log\rho = \log(\beta+2) \\ \alpha \geq (\beta+2)^{-1}, \rho \leq 1 \end{array} \right\}.$$

The special case of random recursive trees is obtained for $\beta = 0$ and plane oriented trees for $\beta = 1$ yielding an asymptotic height of $1.7956\ldots\log n$.

4.6 Intersection of random trees

We can also apply Theorem 1 to the intersection of random trees. One can take k independent copies of a certain kind of random b-ary tree on n nodes and ask about the height of the intersection (a node is in the intersection if it is present in all k trees). This model was treated by Baeza-Yates et al. (1992) for random binary search trees in the context of tree matching properties arising in the tree shuffle algorithm (Choppy et al., 1989). The authors were in particular interested in the size of the intersection of two random binary search trees. We will consider the intersection of k binary search trees, and of k plane oriented trees.

Let $S_{k,n}$ be a collection of k independent copies of identically distributed random trees with n nodes, and let $I_{k,n}$ be their intersection. Recall that the shape of the random tree in our framework is related to the random variables E_e in all k copies. The random variables E of Theorem 1 are now k-vectors of independent random variables. From now on, we write E for a coordinate of this vector, and this corresponds to the random variable describing one of the random trees. By independence of the k trees in $S_{k,n}$ the rate function that corresponds to the presence of a node in $I_{n,k}$ is $k\Lambda_E^\star$. We obtain that the curve to be considered is

$$\left\{\Lambda_Z^\star(\alpha) + k\cdot\Lambda_E^\star(\rho) = \log b\right\},$$

where E and Z are the random variables describing the random trees and b is the branching factor. As an example, this yields the following result.

Proposition 4. *The height $H_n^\star$ of the intersection $I_{k,n}$ of k independent copies of*

1- *random binary search trees is asymptotically* $c_{BST}(k)\log n$*, in probability, where* $c_{BST}(k)$ *is the maximal value of* ρ^{-1} *along*

$$\left\{\rho - 1 - \log\rho = \frac{\log 2}{k}, \rho \leq 1\right\};$$

2- *plane oriented trees is asymptotically* $(c_{PORT}(k)/2)\log n$*, where* $c_{PORT}(k)$ *is the maximum value of* α/ρ *along*

$$\{\alpha\log(3\alpha) + (1-\alpha)(\log(3(1-\alpha)) - \log 2) + k(\rho - 1 - \log\rho) = \log 3, \alpha \geq 1/3, \rho \leq 1\}.$$

Remark: Note that $I_{k,n}$ is likely to contain fewer than n nodes, and the height is not given as a function of the size of $I_{k,n}$.

Table 1 gives numerical values of c_1 and c_2 for certain values of k. The limit values as $k \to \infty$ can also be derived.

k	2	5	10	50	100
c_{BST}	2.62729...	1.78088...	1.48726...	1.18680...	1.12760...
c_{PORT}	2.03950...	1.39752...	1.20841...	1.05078...	1.02788...

Table 1: Some numerical values of the asymptotic height of $I_{k,n}$.

Proposition 5. *There exist limits of both constants* $c_{BST}(k)$ *and* $c_{PORT}(k)$ *as* k *goes to infinity and*

$$\lim_{k\to\infty} c_{BST}(k) = \lim_{k\to\infty} c_{PORT}(k) = 1.$$

Proof. For random binary search trees, this is easily seen since $\{\Lambda_E^\star(\rho) = \rho - 1 - \log\rho = \log 2/k, \rho \leq 1\}$ is the intersection of two explicitly defined curves. By continuity of $\Lambda_E^\star$, $\rho \to 1$ as $k \to \infty$. Consider now PORTs. From the properties of $\mathscr{C}_{Z,E}$, $\rho \geq \rho_{\min}$, where $\rho_{\min}$ is the value at $\alpha = \mathbf{E}Z = 1/3$, and $k\Lambda_E^\star(\rho_{\min}) = \log 3$, giving that $\rho_{\min} \to 1$ as $k \to \infty$. As a consequence, we only look at α. Now, the line Δ going through the origin and $(\rho = 1, \alpha = 1)$ crosses $\mathscr{C}_{Z,E}$ because of its concavity and horizontal tangent at $\rho = 1$. Therefore, the slope of the tangent τ at the optimal point (ρ, α) is greater than 1. Writing $(\rho_{\min}, \alpha_\Delta)$ for the intersection of Δ and $\{\rho = \rho_{\min}\}$ (Figure 7), we get that $\alpha \geq \alpha_\Delta = \rho_{\min}$, yielding $c_{\text{PORT}} \to 1$ as $k \to \infty$.

□

4.7 Change of direction in random binary search trees

Given a tree rooted T and a path π from a leaf to the root, we define $D_T(\pi)$ as the number of changes of direction in π. If we let 0 and 1 encode a move down to the left and to the right respectively, then the path encoded by 0100101 will have $D = 5$, that is, a count of each occurrence of the patterns 01 and 10. We are interested in the maximal value over all the paths of the tree $D_T = \max\{D_T(\pi) : \pi \in T\}$. When T is a random binary search tree, this turns into a random variable that may be handled by our framework. It suffices to notice that if we took a left step, the next move will increase D only if we go right. We have of course something similar when the first step was to the right. Thus, we label the edges as follows. For each level $k \geq 2$ of edges, we form the word $(0110)^{k-1}$, and map the binary characters to the edges from left to right. Then D_π corresponds exactly to the sum of these labels along π (Figure 8).

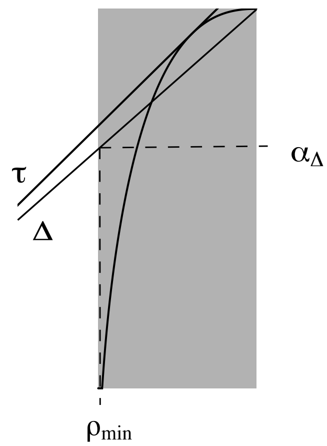

Figure 7: $\mathscr{C}_{Z,E}$ together with the optimal tangent τ and the line Δ through the origin and $(1, 1)$.

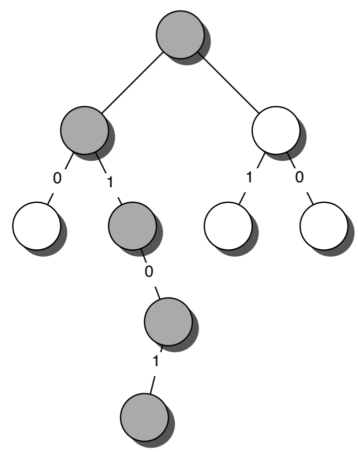

Figure 8: The path consisting of grey nodes is the one with the maximum number of change of direction. Note that the number of changes of direction is the sum of the labels along the path.

This means that for the tree to match our model we need Z to be binomial(1/2), and E exponential(1) because the underlying tree is a binary search tree. Therefore the maximum number of changes of directions along a path in a random binary search tree is asymptotic to the height of random recursive trees.

Proposition 6. *The maximal number of change of direction along a path D_n in a random binary search tree is asymptotic to* $\mathrm{e} \log n$ *in probability.*

4.8 Elements with two lifetimes

Consider a binary tree in which elements have two independent exponential (1) lifetimes, E and Z, and let D_u and G_u keep their meaning from section 2. In the tree T_n, that is, the tree of all nodes u with $G_u \leq n$, it is interesting to ask what the maximal value of D_u when measured with respect to the second lifetimes (Z). Since Z and E have similar Cramér functions, and both have mean one, we have by Theorem 1,

Proposition 7. *The maximal age D_u of any node u in the tree of two lifetimes described* *above, cut off at date of birth $G_u \le n$ is H_n. We have*

$$\frac{H_n}{n} \xrightarrow[n\to\infty]{} c$$

in probability, where $c = 5.82840157\ldots$ is the maximal value of α/ρ along

$$\mathscr{C}_{Z,E} = \{(\rho,\alpha) : \rho - 1 - \log\rho + \alpha - 1 - \log\alpha = \log 2; \rho \le 1, \alpha \ge 1\}.$$

Thus, in spite of the fact that measured by first lifetimes, all have age less than n, there exist elements whose age as measured in the other time scale is almost six times as large!

4.9 Random k-coloring of the edges in a random tree

Assume that we randomly color the edges of a random binary search tree with k colors, and that we ask for the maximal number of similar colors on one path from a root to a leaf. This is equivalent, when k is constant, to studying the maximum number of red colored edges on such paths. But then, this can be studied by attaching to edges independent copies of Z where $Z = 1$ with probability $1/k$ and $Z = 0$ otherwise. That is, Z is Bernoulli $(1/k)$. We have seen already the that the rate function for Bernoulli (Dembo and Zeitouni, 1992) is

$$\Lambda^\star_Z(\alpha) = \alpha\log(k\alpha) + (1-\alpha)\left(\log(1-\alpha) - \log(k-1)\right) + \log k,$$

and the curve of interest is

$$\left\{\begin{array}{l} \Lambda^\star_Z(\alpha) = \alpha\log(k\alpha) + (1-\alpha)\left(\log(1-\alpha) - \log(k-1)\right) + \log k + \rho - 1 - \log\rho = \log 2, \\ k\alpha \ge 1, \rho \le 1 \end{array}\right\}.$$

Note that for $k = 2$, or $p = 1/2$, we have a situation not unlike that of the maximum number of sign changes in random binary search trees, or the random recursive tree, where the asymptotic maximum value is $e\log n$. The maximal path length decreases with the number of colors.

k	1	2	3	4	5
c_k	4.3110...	2.7182...	2.1206...	1.7955...	1.5869...
	6	7	8	9	10
	1.4397...	1.3292...	1.2426...	1.1725...	1.1148...

Table 2: Some numerical values of c_k.

For $k = 1$ and 2 we have the known results for the height of the random binary search trees and random recursive trees, respectively, as one can check in Table 2. Clearly, we may even introduce p values not equal to $1/k$, and even ask on which path we have most red-blue color changes, for example, where red and blue occur with probabilities p and q respectively.

For studying the maximal number of colors of one kind (among k colors) in a random recursive tree, it takes just a moment to see that it suffices to take $Z =$ Bernoulli $(1/k)\times$ Bernoulli $(1/2)$. In other words, Z is Bernoulli $(1/(2k))$.

4.10 The maximum left minus right exceedance

Let the differential depth of a node u be $D_u = \sum_{e\in\pi(u)}(L(e) - R(e))$, where $L(e)$ is the indicator of e being a left edge and $R(e)$ is the indicator of e being a right edge. We want to study the extreme value (differential height) $H_n^\star$ of D_u with an application of Theorem 1, when u ranges over the nodes of a random binary search tree. For this purpose, we may make $Z = 1$ or -1 with probability 1/2. Note that for our Z,

$$\Lambda_Z(\lambda) = \log\left(e^\lambda + e^{-\lambda}\right) - \log 2.$$

And we obtain the Crámer function associated to Z,

$$\Lambda_Z^\star(\alpha) \quad = \quad \begin{cases} \infty & \alpha \geq 1 \\ \frac{\alpha}{2}\log\left(\frac{1+\alpha}{1-\alpha}\right) + \log 2 - \log\left(\sqrt{\frac{1+\alpha}{1-\alpha}} + \sqrt{\frac{1-\alpha}{1+\alpha}}\right) & 0 \leq \alpha < 1. \end{cases}$$

Then, results presented in section 2 allow to conclude that there exists a limit constant c such that $H_n^\star \sim c \log n$ in probability as n tends to infinity. Numerical tools allow to determine $c = 2.07345\ldots$.

5 Concluding remarks

About the type of convergence, one can see that the upper bound of Theorem 1 also holds almost surely. This is a straightforward application of the first Borel-Cantelli Lemma. As for the lower bound, with some extra assumptions on the distribution of E (for example, $E \stackrel{\mathcal{L}}{=}$ exponential(1)), and some minor technical modification, one can show that the proof can be extended to obtain almost sure convergence too.

We considered Z as being a real-valued random variable, but the work can be extented to $Z \in \mathbb{R}^d$. This has been considered by Biggins (1978) in the context of multivariate branching random walks. We may look for the extremes of some multidimensional branching random walk such as the univariate extremes after one has projected the walk on some direction θ. It has been proved by Biggins (1978) that the location of the extremes when θ takes values in the unit ball of $\mathbb{R}^d$ tends to some convex asymptotic shape.

One can also think of a generalized model where the random variables E and Z we have considered are allowed to be dependent for edges that emanate from the same node. This may be handled with a multidimensional version of Cramér's Theorem. In this case, we need the joint distribution of E and Z, and one way to look at it is to consider a unique random vector $X = (Z, E)$. A bivariate rate function $\Lambda_X^\star$ can be defined in a way that is similar to the univariate case. Then, the curve

$$\mathscr{C}_X = \{\Lambda_X^\star(\alpha, \rho) = \log b, \alpha \geq \mathbf{E}Z, \rho \leq \mathbf{E}E\}$$

can be proved to be analogous to $\mathscr{C}_{Z,E}$ in the independent case. Our results could have been stated in terms of a unique random vector, but one would have lost some insight of what is going on: our approach distinguishes the shape of the tree (random variables E) and the weighted depths (random variables Z) for the sake of the presentation.

Acknowledgment

We are grateful to S. Janson for his helpful comments and his improvements of our proofs of Propositions 1 and 2. We would also like to thank both referees for their careful reading.

References

K. B. Athreya and P. E. Ney. *Branching Processes*. Springer, Berlin, 1972.

R. Baeza-Yates, R. Casas, J. Diaz, and C. Martinez. On the average size of the intersection of binary trees. *SIAM Journal on Computing*, 21(1):24–32, 1992.

J. L. Bentley. Multidimensional binary search trees used for associative searching. *Comm. ACM*, 18:509–517, 1975.

J. D. Biggins. Chernoff's theorem in the branching random walk. *Journal of Applied Probability*, 14:630–636, 1977.

J. D. Biggins. The asymptotic shape of the branching random walk. *Advanced Applied Probability*, 10:62–84, 1978.

J. D. Biggins and D. R. Grey. A note on the growth of random trees. *Statistics and Probability letters*, 32:339–342, 1997.

V. Choi and M. J. Golin. Lopsided trees I: a combinatorial analysis. *Algorithmica*, 31: 240–290, 2001.

C. Choppy, S. Kaplan, and M. Soria. Complexity analysis of term-rewriting systems. *Theoretical Computer Science*, 67:261–282, 1989.

Y. S. Chow and H. Teicher. *Probability Theory: Independendence, Interchangeability, Martingales*. Springer, 1978.

K. S. Crump and C. J. Mode. A general age-dependent branching process. *Journal of Mathematical Analysis and its Applications*, 24:494–508, 1968.

A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*. Jones and Bartlett, 1992.

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

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

L. Devroye. Applications of the theory of records in the study of random trees. *Acta Informatica*, 26:123–130, 1988.

L. Devroye. On the expected height of fringe balanced trees. *Acta Informatica*, 30:459–466, 1993.

L. Devroye. Branching processes and their application in the analysis of tree structures and tree algorithms. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Springer Series on Algorithms and Combinatorics*, pages 249–314, Berlin, 1998a. Springer.

L. Devroye. Universal limit laws for depth in random trees. *SIAM Journal on Computing*, 28(2):409–432, 1998b.

A. Duch. *Design and Analysis of Multidimensional Data Structures*. PhD thesis, UPC, Barcelona, 2004.

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.

G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes.* Oxford University Press, second edition, 2001.

T. E. Harris. *The Theory of Branching Processes.* Springer, Berlin, 1963.

S. Kapoor and E. Reingold. Optimum lopsided binary trees. *Journal of the ACM*, 36(3): 573–590, july 1989.

D. E. Knuth. *The Art of Computer Programming: Sorting and Searching*, volume 3. Addison-Wesley, Reading, MA, 1973.

H. Mahmoud. Distances in plane-oriented recursive trees. *J. Comput. Appl. Math.*, 41: 237–245, 1992a.

H. Mahmoud. *Evolution of Random Search Trees.* Wiley, New York, 1992b.

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

P. V. Poblete and J. I. Munro. The analysis of a fringe heuristic for binary search trees. *Journal of Algorithms*, 6:336–350, 1985.

R. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

R. Sedgewick. *Quicksort.* Garland Publishing, New York, 1975.

R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Theor. Probability and Math. Statist.*, 51:1–27, 1995.

B. F. Varn. Optimal variable length codes (arbitrary symbol costs and equal code word probabilities). *Informat. Contr.*, 19:289–301, 1971.

A. Walker and D. Wood. Locally balanced binary trees. *Computer Journal*, 19:322–325, 1976.

Appendix: Review of large deviations

We now review some properties of Cramér's functions that are useful in the proofs of Theorem 1. One can find an introduction to large deviations and Cramér's Theorem in Grimmett and Stirzaker (2001) or more advanced material in the extensive textbook of Dembo and Zeitouni (1992). Let X be a positive random variable and assume that for some $\lambda > 0$,

$$M(\lambda) \stackrel{\text{def}}{=} \mathbf{E}\left\{e^{\lambda X}\right\} < \infty.$$

This implies that $\mathbf{E}\{X^r\} < \infty$ for all $r > 0$, and that $M(\lambda') < \infty$ for all $\lambda' < \lambda$. We define

$$\lambda^\star = \sup\{\lambda \geq 0 : M(\lambda) < \infty\}.$$

Clearly, we may have $\lambda^\star = \infty$ as well as $\lambda^\star = 0$. We introduce the cumulant generating function

$$\Lambda(\lambda) \stackrel{\text{def}}{=} \log(M(\lambda)).$$

Then Cramér's function is defined to be the Fenchel-Legendre dual of Λ: for t such that $t \geq \mathbf{E}X$:

$$\Lambda^\star(t) \stackrel{\text{def}}{=} \sup_{\lambda \geq 0}\{\lambda t - \Lambda(\lambda)\} = \sup_{0 \leq \lambda < \lambda^\star}\{\lambda t - \Lambda(\lambda)\}.$$

Similarly, for $t \leq \mathbf{E}X$, the left-tail Cramér function is

$$\Lambda^\star(t) \stackrel{\text{def}}{=} \sup_{\lambda \leq 0}\{\lambda t - \Lambda(\lambda)\}.$$

Recall that if $X_1, X_2, \ldots, X_n$ are i.i.d. with the same distribution as X, then Cramér's theorem states that (Dembo and Zeitouni, 1992, Theorem 2.2.3, p. 27)

$$\mathbf{P}\left\{X_1 + X_2 + \cdots + X_n \geq nt\right\} = \exp\left(-n\Lambda^\star(t) + o(n)\right),$$

for $t \geq \mathbf{E}X$. Similarly,

$$\mathbf{P}\left\{X_1 + X_2 + \cdots + X_n \leq nt\right\} = \exp\left(-n\Lambda^\star(t) + o(n)\right),$$

for $t \leq \mathbf{E}X$, where $\Lambda^\star$ is now the left-tail Crámer function. This gives sharp estimates for large deviations of a sum of i.i.d. random variables, provided we have some information about the rate $\Lambda^\star$. These functions have been intensively studied, as the Fenchel-Legendre transform is standard in convex analysis (Rockafellar, 1970). An example is shown on Figure 9. We note here that $\Lambda^\star(t)$ may be infinite for all t larger than a finite threshold. Also, we may have $\Lambda^\star(t) = 0$ for all t (this occurs when $\lambda^\star = 0$, a case which is largely uninteresting for us).

Lemma 5 (Properties of Cramér's functions)**.** *Let X be any random variable with $\mathbf{E}X = \mu$ finite. Let M, Λ and $\Lambda^\star$ be the moment, cumulant generating functions and Cramér's function, respectively. Then*

(1) $\Lambda^\star(t)$ is increasing for $t \geq \mu$, $M(\lambda)$ is increasing where $\lambda > 0$ and $M(\lambda) < \infty$, and $M(0) = 1$.

(2) $\Lambda(\lambda) \geq \lambda\mu$.

(3) $\Lambda^\star(\mu) = 0$.

Proof. The first fact is obvious from the definitions. Fact (2) comes from Jensen's inequality

$$\Lambda(\lambda) = \log \mathbf{E}\left\{e^{\lambda X}\right\} \geq \mathbf{E}\left\{\log\left(e^{\lambda X}\right)\right\} = \lambda \mathbf{E}X.$$

Consider now fact (3). By definition we have,

$$\Lambda^\star(\mathbf{E}X) = \sup_\lambda\{\lambda\mathbf{E}X - \Lambda(\lambda)\} \leq \sup_\lambda\{\lambda\mu - \lambda\mu\} = 0.$$

But $\Lambda^\star(t)$ is non-negative:

$$\Lambda^\star(t) = \sup_\lambda\{\lambda t - \Lambda(\lambda)\} \geq \{\lambda t - \Lambda(\lambda)\}_{\lambda=0} = 0,$$

so that $\Lambda^\star(\mu) = 0$. $\square$

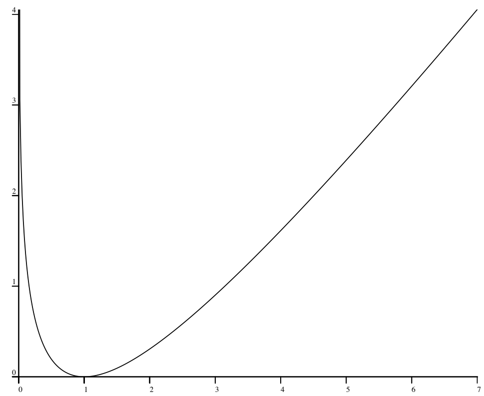

Figure 9: The Crámer function $\Lambda_E^\star$ when E is exponentially distributed with mean one is $\rho - 1 - \log \rho$.

We need one more result to prove the uniqueness of the optimal point on the curve $\mathscr{C}_{Z,E}$. Indeed, $\mathscr{C}_{Z,E}$ is concave and the values of the slope at both ends are of great importance. This is directly dependent on the derivative of the rate functions $\Lambda_Z^\star$ and $\Lambda_E^\star$ at their mean.

Proposition 8. *If X is not a single mass at $\mathbf{E}X = \mu$, and $\lambda^\star > 0$, then*

$$\left.\frac{d\Lambda^\star}{dt}\right|_{[t=\mu]} = 0.$$

The following lemmas will help us to prove Proposition 8 without using fancy arguments.

Lemma 6. *Under the conditions of Lemma 5, $\Lambda(\lambda)/\lambda$ is increasing in λ. Since $\mathbf{P}\{X = \mu\} < 1$, $\Lambda(\lambda)/\lambda$ is in fact strictly increasing.*

Proof. Let $Z = e^X$ and $\lambda < \lambda'$. From Hölder's inequality we have that

$$\mathbf{E}\left\{Z^\lambda\right\}^{1/\lambda} \le \mathbf{E}\left\{Z^{\lambda'}\right\}^{1/\lambda'}, \tag{8}$$

and therefore

$$(M(\lambda))^{1/\lambda} \le (M(\lambda'))^{1/\lambda'}.$$

Taking logarithms shows that $\Lambda(\lambda)/\lambda$ is increasing. Note that equality in (8) occurs if and only if Z puts all its mass at one point, and thus the second statement holds. $\square$

Lemma 7. *Under the conditions of Lemma 5,*

$$\lim_{\lambda \downarrow 0} \frac{\Lambda(\lambda)}{\lambda} = \mathbf{E}X.$$

Proof. Write $\psi(x) = e^{\lambda x} - \lambda x - 1$, then

$$M(\lambda) = \mathbf{E}\left\{e^{\lambda X}\right\} = \mathbf{E}\{1 + \lambda X + \psi(X)\} = 1 + \lambda \mathbf{E}X + \mathbf{E}\{\psi(X)\}.$$

By Taylor's series expansion,

$$\psi(x) \le \frac{\lambda^2 x^2}{2} e^{\lambda x},$$

and for any $\epsilon > 0$, $e^{\epsilon x} \geq \epsilon x$ so that

$$\psi(x) \leq \frac{\lambda^2}{2\epsilon^2} e^{(\lambda+2\epsilon)x}.$$

Choosing ϵ such that $\lambda + 2\epsilon < \lambda^\star$ gives the bounds

$$1 + \lambda \mathbf{E}X \leq M(\lambda) \leq 1 + \lambda \mathbf{E}X + \frac{\lambda^2}{2\epsilon^2} M(\lambda + 2\epsilon).$$

As $M(\lambda + 2\epsilon)$ decreases as $\lambda \downarrow 0$, $\Lambda(\lambda) \sim \lambda \mathbf{E}X$. □

Lemma 8. *Assume $\lambda^\star > 0$. At every $\lambda \in [0, \lambda^\star)$, $\Lambda(\lambda)/\lambda$ is continuous in λ.*

Proof. For $\lambda = 0$, this follows from Lemma 7. If $\lambda > 0$, then we have the result if M is continuous, but M is the moment generating function and is known to be continuous. □

Lemma 9. *If X is not a single mass, and $\lambda^\star > 0$, then there exists $t^\star > \mu = \mathbf{E}X$ (assumed finite) such that $\Lambda^\star(t) < \infty$ for all $t < t^\star$.*

Proof. Assume first that $\Lambda(\lambda)/\lambda$ increases to ∞ (we consider limits in the domain where it is finite). Since $\Lambda(\lambda)/\lambda$ is continuous, for any $t \geq \mu$, there exists a solution $\lambda_t < \lambda^\star$ of $\Lambda(\lambda) = \lambda t$. Thus,

$$\Lambda^\star(t) = \sup \left\{ \lambda \left(t - \frac{\Lambda(\lambda)}{\lambda} \right) \right\} \leq \lambda_t (t - \mu) < \infty. \tag{9}$$

Now if $\Lambda(\lambda)/\lambda$ increases to a finite limit F, we have the similar result $\Lambda^\star(t) < \infty$ for every $\mu \leq t < F$. □

Proof of Proposition 8. Recall that from Lemma 9 that $\Lambda^\star(t) < \infty$, $t \leq t^\star$ for some $t^\star > \mathbf{E}X$. As $t \downarrow 0$, by the strict increasing nature and continuity of $\Lambda(\lambda)/\lambda$, the solution λ_t of

$$\frac{\Lambda(\lambda)}{\lambda} = t$$

tends to 0. But by (9) $\Lambda(t) \leq \lambda_t(t - \mu)$, and thus

$$\lim_{t \downarrow \mu} \frac{\Lambda^\star(t) - \Lambda^\star(\mu)}{t - \mu} \leq \lim_{t \downarrow \mu} \lambda_t = 0.$$ □

Lemma 10. *For any random variable X with finite mean and having $\lambda^\star > 0$, the associated rate function $\Lambda^\star$ is convex and strictly convex inside the interior of the set $\{\Lambda'(\lambda) : \lambda < \lambda^\star\}$. If $\lambda^\star = 0$, then $\Lambda^\star(t) = 0$ for all $t \geq \mathbf{E}X$.*

Proof. We can assume without loss of generality that $\mu \leq t < t^\star$, where $t^\star$ is as in the previous Lemma. Then for $\epsilon > 0$ small enough, it suffices to show that

$$\Lambda^\star(t + 2\epsilon) - \Lambda^\star(t + \epsilon) \geq \Lambda^\star(t + \epsilon) - \Lambda^\star(t).$$

But for functions f, g,

$$\max_x f(x) + \max_x g(x) \geq \max_x \{f(x) + g(x)\}, \tag{10}$$

and so, using the definition of $\Lambda^\star$,

$$\begin{aligned}\Lambda^\star(t+2\epsilon)+\Lambda^\star(t) &= \sup_\lambda\{\lambda(t+2\epsilon)-\Lambda(\lambda)\}+\sup_\lambda\{\lambda t-\Lambda(\lambda)\} \\ &\geq 2\sup_\lambda\{\lambda(t+\epsilon)-\Lambda(\lambda)\} \qquad (11) \\ &= 2\Lambda^\star(t+\epsilon).\end{aligned}$$

We can obtain strict convexity by being more careful. Using the geometric interpretation of $\Lambda^\star$ (Dembo and Zeitouni, 1992), the optimal λ for the supremum is the abscissa of the point where Λ admits a tangent of slope t. Since Λ is convex and has a continuous derivative where it is finite (proof of this is very similar to the one of Lemma 7 using a Taylor series expansion), the values of λ and λ' corresponding to t and t' respectively are distinct whenever $t \neq t'$. As equality in (10) occurs only if both maxima are achieved at the same point x, we actually obtain a strict inequality in (11) and therefore strict convexity where one can find such tangents to Λ with slope t and $t+2\epsilon$, that is in the interior of the set $\{\Lambda'(\lambda) : \lambda < \lambda^\star\}$. $\square$