STOCHASTIC ANALYSIS OF TREE-LIKE DATA STRUCTURES

MICHAEL DRMOTA

ABSTRACT. The purpose of this article is to present two types of data structures, binary search trees and usual (combinatorial) binary trees. Although they constitute the same set of (rooted) trees they are constructed via completely different rules and thus the underlying probabilitic models are different, too. Both kinds of data structures can be analyzed by probabilistic and stochastic tools, binary search trees (more or less) with martingales and binary trees (which can be considered as a special case of Galton-Watson trees) with stochastic processes.

It is also an aim of this article to demonstrate the strength of analytic methods in specific parts of probabilty theory related to combinatorial problems, especially we make use of the concept of generating functions. One reason is that that recursive combinatorial descriptions can be translated to relations for generating functions, and second analytic properties of these generating functions can be used to derive asymptotic (probabilistic) relations.

1. Introduction

Probably the most widely used sorting algorithm is the algorithm *Quicksort* which has been invented by C. A. R. Hoare [28, 29]. It is the standard sorting procedure in Unix systems, and the basis idea can be described as follows:

> A list of n (different) real numbers $A = (x_1, x_2, \dots, x_n)$ is given. Select an (pivot) element x_j from this list. Divide the remaining numbers into sets $A_\le, A_>$ of numbers smaller (or equal) and larger than x_j. Next apply the same procedure to each of these two sets if they contain more than one element. Finally, we end up with a sorted list of the original numbers.

This sorting procedure can be encoded with a binary tree with n (internal) nodes.[1] The first selected element x_j is put to the root, whereas recursively $A_\le$ produces a left subtree of x_j and $A_>$ the right subtree of x_j. (An empty string produces an empty tree which is usually encoded as an external node.)

This kind of binary trees are also called *binary search trees* and are also quite common as a data structure to store data represented by keys which can be totally ordered (compare with [35, 40]). It is then easy to search for an item by comparing it with the root and then proceeding to the left subtree if it is smaller resp. to the right subtree if it is larger.

In what follows we will discuss several parameters which are important in the analysis of Quicksort resp. in searching in binary search trees. In doing this we assume that the *data* $x_1, \dots, x_n$ constitute iid random continuous variables. We start with the number of comparisions needed for Quicksort which is the same as the total path length of the corresponding binary search tree. We then present a martingale approach to describe the profile of binary search trees, and finally a generating function approach to the height of binary search trees which equals the maximal number of comparisions needed for searching in binary search trees or the maximal number of recursive calls of Quickstep. [2]

Date: June 10, 2002.

Key words and phrases. binary search tree, binary tree, height, path length, profile, generating functions.

Institut für Geometrie, TU Wien, Wiedner Hauptstrasse 8-10/118, A-1040 Wien, Austria, email: drmota@tuwien.ac.at.

[1]The nodes of a (rooted) binary tree can be divided into *internal* nodes with two descendents and *external* nodes with no descendents.

[2]There are also quite similar search trees, e.g. digital search trees or tries, which appear as data structures (see [40]) but will not be discussed in this article.

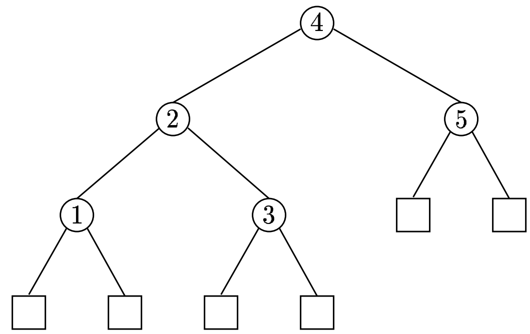

FIGURE 1. Binary search tree generated by the list $(4, 2, 3, 5, 1)$, where the pivot element is always the first element.

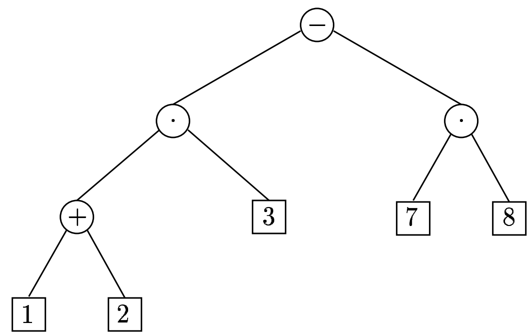

FIGURE 2. Binary tree for the arithmetic expression $((1 + 2) \cdot 3) - (7 \cdot 8)$.

Binary trees appear also in other application. For example, consider an arithmetic expression $((1 + 2) \cdot 3) - (7 \cdot 8)$. which can be stored as a binary tree in an obvious way.

Although this kind of data structure looks like binary search trees, its *average behaviour* is completely different. In this context it is natural to assume that every binary tree (with n internal nodes) appears with equal probability. This induces a completely different probabilistic model for binary trees with n internal nodes (which is more or less a translation of a combinatorial problem for trees). For example, the number of binary trees with n internal nodes is exactly the n-th Catalan number

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

and hence every binary tree (of size n) has probability $1/C_n$.

Interestingly, there is natural (probablistic) generalization of binary trees (considered in that way), namely *Galton-Watson branching processes* (resp. Galton-Watson trees) conditioned on the total progeny.

Thus, the second part of this paper is devoted to (critical) Galton-Watson trees (with finite offspring variance). As for random binary search trees we will discuss the path length, the profile and the height. The results will be of a completely different flavour. We will use again the concept of generating functions, however, from the probabilistic side we are confronted with stochastic processes, e.g. with Brownian excursion and its (total) local time.

Notation. Consider a (finite) binary tree t. Then for every internal node x the distance to the root is denoted by $h(x)$. Then the (internal) *path length* is given by

$$l = l_t = \sum_{x \text{ internal node of } t} h(x)$$

and the *height* by

$$h = h_t = \max_{x \text{ internal node of } t} h(x).$$

We will also consider the *profile* $(v_k)_{k\ge 0}$, where v_k denotes the number of (internal) nodes at level k:

$$v_k = \#\{x : h(x) = k\}.$$

Observe that the path length is also given by

$$l_t = \sum_{k\ge 0} k v_k$$

and the height by

$$h_t = \max\{k \ge 0 : v_k > 0\}.$$

Finally, the *width* of t is defined by

$$w = w_t = \max_{k\ge 0} v_k.$$

2. Binary Search Trees

2.1. **Probabilistic Model.** When analyzing Quicksort or binary search trees it is standard to assume that the data $x_1 = X_1, x_2 = X_2, \dots, x_n = X_n$ are iid real random variables with a (common) continuous probability distribution. For example, this implies that their ranks form a random permutation of $\{1, 2, \dots, n\}$. Thus, the kind of the distribution of X_j $(1 \le j \le n)$ has no influence on the distribution of the parameters of Quicksort or the corresponding binary search tree. It is therefore no loss of generality to assume that X_j are uniformly distributed on $[0, 1]$. Even the choice of the pivot element (in Quicksort) does not change the probabilistic structure. It is therefore common to assume that the pivot elment is always the first in corresponding list.[3]

In this context it is also natural to consider an infinte sequence $(X_n)_{n\ge 1}$ of iid random variables (uniformly distributed on $[0, 1]$) which induce a random sequence $(T_n)_{n\ge 0}$ of binary search trees. This means that T_n contains n internal nodes, where the data $X_1, X_1, \dots, X_n$ are *stored* in a way that all (internal) nodes of the left subtree of X_j have smaller values than X_j and all nodes of the right subtree are larger than X_j. Furthermore, T_{n+1} is generated from T_n by inserting X_{n+1} at one of the $n + 1$ external nodes of T_n in the following way. One starts at the root and goes to the left subtree if X_{n+1} is smaller that the value of the root and to the right subtree if X_{n+1} is larger. This procedure is recursively applied until one reaches an external node where X_{n+1} is inserted. By assumption each of these $n + 1$ external nodes (or *free places*) is replaced by X_{n+1} with equal probability $1/(n + 1)$. Thus, we also have a kind of Markov property: T_{n+1} just depends on T_n (and this in a very simple way).

In what follows we will discuss just parameters in (random) binary search trees. As explained above there is a direct correspondance to Quicksort.

2.2. **Internal Path Length.** Let L_n denote the internal path length of random binary search trees T_n (or equivalently the number of comparisions which are needed to sort a random permutation of $\{1, 2, \dots, n\}$ with Quicksort). The recursive description of binary search trees (or Quicksort) immediately translates to[4]

$$\mathcal{L}(L_n) = \mathcal{L}\left(L_{Z_n - 1} + \overline{L}_{n - Z_n} + n - 1\right), \qquad n \ge 2, \tag{1}$$

where $L_0 = L_1 = 0$, $L_2 = 1$, Z_n is uniformly distributed on $\{1, 2, \dots, n\}$, $\mathcal{L}(L_j) = \mathcal{L}(\overline{L}_j)$, and Z_n, L_j, $\overline{L}_j$ $(1 \le j \le n)$ are independent.

Equivalently we can consider the generating functions

$$l_n(u) = \mathbf{E}\, u^{L_n}$$

[3]In Unix systems the pivot element is always an element in the middle position.

[4]We denote by $\mathcal{L}(X)$ the distribution function of X.

and get the recurrence

$$l_n(u) = \frac{u^{n-1}}{n} \sum_{j=1}^{n} l_{j-1}(u) l_{n-j}(u).$$

Consequently the differential equation for the double generating function $L(x,u) = \sum_{n\ge 0} l_n(u) x^n$:

$$\frac{\partial L(x,u)}{\partial x} = L(xu,u)^2$$

with the side conditions

$$\frac{\partial L(0,u)}{\partial x} = 1 \quad \text{and} \quad L(x,1) = \frac{1}{1-x}.$$

We will not use these relations here but we will see that corresponding generating functions for the profile and the height have a similar structure.

It is now an easy exercise to obtain explicit representations for the expeced value $\mathbf{E}L_n$, e.g. the recurrence

$$\mathbf{E}L_n = n - 1 + \frac{1}{n} \sum_{j=1}^{n} (\mathbf{E}L_{j-1} + \mathbf{E}L_{n-j})$$

can be explicitly solved to

$$\begin{aligned} \mathbf{E}L_n &= 2(n+1) \sum_{h=1}^{n+1} \frac{1}{h} - 4(n+1) + 2 \\ &= 2n \log n + n(2\gamma - 4) + 2 \log n + 2\gamma + 1 + \mathcal{O}\left((\log n)/n\right) \end{aligned}$$

with $\gamma = 0.57721...$ beeing Euler's constant.

In fact, much more is known about this random variable.[5]

Theorem 1. *The normalized path length*

$$Y_n = \frac{L_n - \mathbf{E}\, L_n}{n}$$

converges weakly to a random variable Y:

$$Y_n \Longrightarrow Y,$$

which distribution is defined by the fixed point equation

$$\mathcal{L}(Y) = \mathcal{L}\big(UY + (1-U)\overline{Y} + c(U)\big), \tag{2}$$

where U is uniformly distributed on on $[0,1]$, $\mathcal{L}(\overline{Y}) = \mathcal{L}(Y)$, $U, \overline{Y}, Y$ are independent, and

$$c(x) = 2x \log x + 2(1-x)\log(1-x) + 1.$$

The existence of a limiting distribution (the *Quicksort distribution*) was first observed by Régnier [50] via a martingale approach, whereas the characterization of Y with a fixed point equation is due to Rösler [54]. It is now also known that there exists a density ([58]), which is a bounded C^∞ function, tail estimates are available, and orders of convergence are estimated (compare with [21, 22, 23, 34]). However, no explicit representations for the limiting distribution are known. In the meantime various properties of the limiting distribution,

From the fixed point equation (2) it also possible to calculate moments step by step, e.g. the variance of Y is given by

$$\mathbf{Var}\, Y = 7 - \frac{2}{3}\pi^2.$$

We briefly describe Rösler's approach which has developed to a powerful method in the probablistic analysis of algorithms, the *contraction method* (compare with [55]). His main obervation was that (2) has a unique fixed point because (because it constitutes a contraction with respect to the Wasserstein metric d_2).

[5]We will always denote weak convergence of with "$\Longrightarrow$".

Let D denote the space of distribution functions with finite second moment and zero first moment. Then the Wasserstein metric d_2 is defined as

$$d_2(F,G) = \inf \|X - Y\|_2,$$

where $\|\cdot\|_2$ denotes the L_2-norm and the infimum is taken over all random variables X with distributions function F and all Y with distribution function G. It is well known that (D, d_2) constitutes a Polish space.[6]

Let us consider the random variables $Y_n = (L_n - \mathbf{E}L_n)/n$. From (1) we directly get

$$\mathcal{L}(Y_n) = \mathcal{L}\left(Y_{Z_n-1}\frac{Z_n-1}{n} + \overline{Y}_{n-Z_n}\frac{n-Z_n}{n} + c_n(Z_n)\right), \quad n \geq 2,$$

where $Y_0 = Y_1 = 0$, Z_n is uniformly distributed on $\{1, 2, \ldots, n\}$, and $\mathcal{L}(Y_j) = \mathcal{L}(\overline{Y}_j)$, and Z_n, Y_j, $\overline{Y}_j$ $(1 \leq j \leq n)$ are independent. Furthermore,

$$c_n(j) = \frac{n-1}{n} + \frac{1}{n}\left(\mathbf{E}L_{j-1} + \mathbf{E}L_{n-j} - \mathbf{E}L_n\right).$$

Thus, if Y_n has a limiting distribution Y then it has to satisfy (2).

The first step is to show that (2) has actually a unique solution with $\mathbf{E}Y = 0$.

Lemma 1. *Let $S : D \to D$ be a map defined by*

$$S(F) := \mathcal{L}(UX + (1-U)\overline{X} + c(U)),$$

where $X, \overline{X}, U$ are independent, $\mathcal{L}(\overline{X}) = \mathcal{L}(X) = F$, and U is uniformly distributed on $[0,1]$. *Then S is a contraction with respect to the Wasserstein metric d_2 and, thus, there is a unique fixed point $F \in D$ with $S(F) = F$.*

Proof. Let $F, G \in D$ and suppose that $\mathcal{L}(\overline{X}) = \mathcal{L}(X) = F$, $\mathcal{L}(\overline{Y}) = \mathcal{L}(Y) = G$, and U is ud on $[0,1]$ such that $U, \overline{X}, X$ and $U, \overline{Y}, Y$ are independent. Then $S(F) = \mathcal{L}(UX + (1-U)\overline{X} + c(U))$ and $S(G) = \mathcal{L}(UY + (1-U)\overline{Y} + c(U))$ and consequently

$$\begin{aligned} d_2^2(S(F), S(G)) &\leq \|UX + (1-U)\overline{X} - UY - (1-U)\overline{Y}\|_2^2 \\ &= \|U(X-Y) + (1-U)(\overline{X} - \overline{Y})\|_2^2 \\ &= \mathbf{E}(X-Y)^2 \cdot \mathbf{E}U^2 + \mathbf{E}(\overline{X} - \overline{Y})^2 \cdot \mathbf{E}(1-U)^2 \\ &= \frac{2}{3}\mathbf{E}(X-Y)^2. \end{aligned}$$

Taking the infimum over all possible X, Y we obtain

$$d_2(S(F), S(G)) \leq \sqrt{\frac{2}{3}}\, d_2(F,G),$$

which completes the proof of the lemma. □

The next step is to show that Y_n actually converges to Y. (Recall that $d_2(\mathcal{L}(Y_n), \mathcal{L}(Y)) \to 0$ implies $Y_n \Longrightarrow Y$.)

Lemma 2. *We have*

$$d_2^2(\mathcal{L}(Y_n), \mathcal{L}(Y)) \leq \frac{2}{n}\sum_{j=1}^{n}\left(\frac{j-1}{n}\right)^2 d_2^2(\mathcal{L}(Y_{j-1}), \mathcal{L}(Y)) + \mathcal{O}\left(\frac{\log^2 n}{n}\right).$$

and consequently $\lim_{n\to\infty} d_2(\mathcal{L}(Y_n), \mathcal{L}(Y)) = 0$.

[6] A sequence F_n converges to F in D if and only if F_n converges weakly to F and if the second moments of F_n converge to the second moment of F.

Proof. Let $G_n = \mathcal{L}(Y_n)$, where Y_n is the normalized path length, and let Y and $\overline{Y}$ be independent with $\mathcal{L}(Y) = \mathcal{L}(\overline{Y}) = F$, where F is the unique fixed point of $S(F) = F$. Next choose versions Y_j, $\overline{Y}_j$ (which are independent for $1 \le j \le n-1$) with

$$\mathbf{Var}(Y_j - Y) = d_2^2(G_j - F) \quad \text{and} \quad \mathbf{Var}(\overline{Y}_j - \overline{Y}) = d_2^2(G_j - F)$$

and set $V_x = Y_j$ and $\overline{V}_x = \overline{Y}_j$ for $x \in (\frac{j}{n}, \frac{j+1}{n}]$. Then, for U ud on $[0,1]$ and independent of Y_j and $\overline{Y}_j$ we have

$$G_n = \mathcal{L}\left(\frac{\lceil nU\rceil - 1}{n} V_U + \frac{n - \lceil nU\rceil}{n}\overline{V}_U + c_n(\lceil nU\rceil)\right).$$

Since

$$\sup_{0\le x\le 1} |c_n(\lceil nx\rceil) - c(x)| = \mathcal{O}\left(\frac{\log n}{n}\right)$$

we directly obtain

$$\begin{aligned}
d_2^2(G_n, F) &\le \mathbf{E}\left(\frac{\lceil nU\rceil - 1}{n} V_U - UY + \frac{n - \lceil nU\rceil}{n}\overline{V}_U - (1-U)Y + c_n(\lceil nU\rceil) - c(U)\right)^2 \\
&\le \mathbf{E}\left(\sum_{j=1}^{n} \mathbf{1}_{(\frac{j-1}{n}, \frac{j}{n}]}(U)\left(\frac{j-1}{n} Y_{j-1} - UY\right)\right)^2 \\
&\quad + \mathbf{E}\left(\sum_{j=1}^{n} \mathbf{1}_{(\frac{j-1}{n}, \frac{j}{n}]}(U)\left(\frac{n-j}{n} Y_{j-1} - (1-U)Y\right)\right)^2 \\
&\quad + \mathbf{E}\left(c_n(\lceil nU\rceil) - c(U)\right)^2 \\
&\le \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}\left(\frac{j-1}{n}(Y_{j-1} - Y)\right)^2 \\
&\quad + \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}\left(\frac{n-j}{n}(Y_{j-1} - Y)\right)^2 + \mathcal{O}\left(\frac{\log^2 n}{n}\right) \\
&\le \frac{2}{n}\sum_{j=1}^{n}\left(\frac{j-1}{n}\right)^2 d_2^2(G_{j-1}, F) + \mathcal{O}\left(\frac{\log^2 n}{n}\right).
\end{aligned}$$

Thus, with $a_j = d_2^2(G_j, F)$ one has

$$a_n \le \frac{2}{3}\max_{0\le j\le n-1} a_j + \mathcal{O}\left(\frac{\log^2 n}{n}\right),$$

which implies that $a_n \to 0$. □

2.3. **Profile.** We now consider the sequence $(T_n)_{n\ge 0}$ of binary search trees generated by an iid sequence $(X_n)_{n\ge 1}$ of random variables which are ud on $[0,1]$. The profile will be denoted $V_{k,n}$, i.e. $V_{k,n}$ equals the number of internal nodes of T_n at level k. We will also consider the *external profile* $U_{k,n}$, where $U_{k,n}$ counts the number of external nodes of T_n at level k. In fact it turns out that $U_{k,n}$ is much easier to handle. Furthermore we have

$$V_{k,n} = \sum_{j>k} 2^{k-j} U_{j,n}. \tag{3}$$

Thus, it is sufficient to work with $U_{k,n}$.

Next, let us introduce the generating functions

$$Y_k(x,u) := \sum_{n\ge 0} \mathbf{E}\, u^{U_{k,n}} \cdot x^n. \tag{4}$$

Then we have $Y_0(x,u) = u + x/(1-x)$ and recursively

$$\frac{\partial Y_{k+1}(x,u)}{\partial x} = Y_k(x,u)^2$$

with $Y_k(0,u) = 1$.

There is no method known to solve this kind or recurrence (explicitly or asymptotically). Nevertheless it can be used to derive the expected profile By definition we have

$$Z_k(x) := \frac{\partial Y_k(x,1)}{\partial u} = \sum_{n\ge 0} \mathbf{E}\, U_{k,n} \cdot x^n.$$

Furthermore, $Z_0(x) = 1$ and by (4)

$$Z'_{k+1}(x) = 2Y_k(x,1)Z_k(x) = \frac{2}{1-x} Z_k(x)$$

with $Z_{k+1}(0) = 0$ (for $k \ge 0$). Hence,

$$Z_k(x) = \frac{2^k}{k!}\left(\frac{1}{1-x}\right)^k$$

and one obtains

$$\mathbf{E}\, U_{k,n} = \frac{2^k}{n!} s_{n,k}, \tag{5}$$

where $s_{n,k}$ are the (absolute) Stirling number of the first kind, in other words the number of permutations σ of n elements such that the canonical cyclic representation of σ has exactly k cycles. (It seems that this explicit formula was first observed by Lynch [39], compare also with [40]). By well known asymptotics for Stirling numbers (see [43]) we derive (for $k = \mathcal{O}(\log n)$)

$$\mathbf{E} U_{k,n} = \frac{2^k (\log n)^k}{k!\, n\, \Gamma(\alpha_{n,k})} \sim \frac{n^{\alpha_{n,k}(1-\log(\alpha_{n,k}/2))-1}}{\sqrt{2\pi k}}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right), \tag{6}$$

where $\alpha_{n,k} = \frac{k}{\log n}$. Especially, if we just consider a local expansion which is tight for k close to $2\log n$ we obtain

$$\mathbf{E} U_{k,n} = \frac{n}{\sqrt{4\pi \log n}}\left(e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)\right). \tag{7}$$

By (3) we get the same for $V_{k,n}$:

$$\mathbf{E} V_{k,n} = \frac{n}{\sqrt{4\pi \log n}}\left(e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)\right).$$

This indicates that the *mass* of a binary search tree T_n is concentrated around level $2\log n$. There exist no distributional results for $U_{k,n}$ or $V_{k,n}$. However it is possible to obtain an almost sure result of the following kind.

Theorem 2. *We have a.s.*

$$U_{k,n} = \frac{n}{\sqrt{4\pi \log n}}\left(e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)\right) \quad \textit{and}$$
$$V_{k,n} = \frac{n}{\sqrt{4\pi \log n}}\left(e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)\right)$$

as $n \to \infty$, where the error term $\mathcal{O}(1/\sqrt{\log n})$ is uniform for all $k \ge 0$.

This shows that the profile of T_n is more or less given by its expected value. The *probability* is hidden in the $\mathcal{O}$-term. However, one can be a little bit more precise.

Theorem 3. *There exists a random analytic function $M(z)$ for $|z-1| < (\sqrt{2})^{-1}$ with $M(1) = 1$ such that for any given $\varepsilon > 0$ we have a.s.*

$$\frac{U_{k,n}}{\mathbf{E}U_{k,n}} - M\left(\frac{k}{2\log n}\right) \to 0 \quad \text{and}$$
$$\frac{V_{k,n}}{\mathbf{E}V_{k,n}} - M\left(\frac{k}{2\log n}\right) \to 0$$

as $n \to \infty$, uniformly for all k with $1.2\log n \le k \le 2.8\log n$.

In fact, $M(z)$ is the limit of a certain martingale of analytic functions as we will describe below.

The uniformity property of Theorem 2 can be used to obtain also a result for the width and for the path length.

Theorem 4. *Let $\overline{U}_n = \max_{k\ge 0} U_{k,n}$ and $\overline{V}_n = \max_{k\ge 0} V_{k,n}$. Then we have a.s.*

$$\frac{\overline{U}_n}{n/\sqrt{4\pi\log n}} = 1 + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right),$$
$$\frac{\overline{V}_n}{n/\sqrt{4\pi\log n}} = 1 + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \quad \text{and}$$
$$\frac{L_n}{2n\log n} = 1 + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)$$

as $n \to \infty$.

We will now present an outline of the proof of Theorems 2 and 3 which are due to Chauvin, Drmota, and Jabbour-Hattab [8].

The basic tool for the proofs are the so-called *profile polynomials*

$$W_n(z) = \sum_{k\ge 0} U_{k,n} z^k.$$

By (5) it is easy to derive that

$$\mathbf{E}W_n(z) = (-1)^n \binom{-2z}{n}. \tag{8}$$

The basic property of $W_n(z)$ is that the normalized version is a martingale (see [31]).

Lemma 3. *The polynomials*

$$M_n(z) = \frac{W_n(z)}{\mathbf{E}W_n(z)}$$

constitute a martingale with respect to the natural filtration $(\mathcal{F}_n)_{n\ge 0}$ associated to the sequence of trees $(T_n)_{n\ge 0}$.

Proof. With help of the above description how T_{n+1} evolves from T_n one has

$$\begin{aligned}\mathbf{E}(U_{k,n+1}|\mathcal{F}_n) &= (U_{k,n}+2)\frac{U_{k-1,n}}{n+1} + (U_{k,n}-1)\frac{U_{k,n}}{n+1} + U_{k,n}\left(1 - \frac{U_{k-1,n}+U_{k,n}}{n+1}\right)\\ &= \frac{2U_{k-1,n}}{n+1} + \frac{nU_{k,n}}{n+1}.\end{aligned}$$

Hence

$$\mathbf{E}(W_{n+1}(z)|\mathcal{F}_n) = \frac{2z+n}{n+1} W_n(z)$$

and consequently

$$\mathbf{E}(M_{n+1}(z)|\mathcal{F}_n) = M_n(z),$$

which completes the proof of the martingale property. □

Hence, for positive values of z, the martingale converges to an almost sure limit $M(z)$. Interestingly this property is also true for certain complex values of z.

Lemma 4. *For any compact set* $C \subseteq \{z \in \mathbf{C} : |z-1| < 1/\sqrt{2}\}$ *the martingale* $M_n(z)$ *converges a.s. uniformly to its limit* $M(z)$ *(which is again an analytic function).*

We note that $M(z)$ is exactly the random analytic function appearing in Theorem 3. We also note that $M_n(1) = 1$. So there is no *probability* at $z = 1$.

The proof of Lemma 4 is based on an L^2-study. With help of an explicit expression for $\mathbf{E}(W_n(z_1)W_n(z_2))$ it is shown that $M_n(z)$ is bounded in L^2 for z with $|z-1| < 1/\sqrt{2}$, from which it follows that $M_n(z) \to M(z)$ a.s. and in L^2 for some random variable $M(z)$. Next an estimate of the kind

$$\mathbf{E}|M(z_1) - M(z_2)|^2 \le c\,|z_1 - z_2|^2$$

(for some constand $c > 0$) and Kolmogoroff's criterion show that for every continuously differentialbe curve $\gamma : [0,1] \to \{z \in \mathbf{C} : |z-1| < 1/\sqrt{2}\}$ there is s continuous version $\tilde{M}_\gamma$ of $M(\gamma(t))$ ($t \in [0,1]$) such that

$$\mathbf{E}\left(\sup_{0\le t\le 1} |\tilde{M}_\gamma(t)|^2\right) < \infty.$$

Finally, which help of a theorem for vector martingales (see [44, Proposition V-2-6, p. 104]) it follows that

$$\sup_{0\le t\le 1} |M_n(\gamma(t)) - \tilde{M}_\gamma(t)| \to 0 \qquad \text{a.s.}$$

and

$$\mathbf{E}\left(|M_n(\gamma(t)) - \tilde{M}_\gamma(t)|^2\right) \to 0 \qquad (n \to \infty).$$

Thus, by Cauchy's formula $M_n(z)$ converges uniformly to its limit which is again an analytic function. (For details see [8]).

Overall, we know that

$$W_n(z) \sim M(z) \cdot \mathbf{E}W_n(z) \tag{9}$$

if $|z-1| < 1/\sqrt{2}$. The idea of the proof of Theorems 2 (and completely similar for Theorems 3) is to use Cauchy's formula to evaluate $U_{k,n}$:

$$U_{k,n} = \frac{1}{2\pi i}\int_{|z|=1} \frac{W_n(z)}{z^{k+1}}\,dz.$$

For z with $|z-1| < 1/\sqrt{2}$ we can use Lemma 4 (resp. (9)).

For z with $|z-1| \ge 1/\sqrt{2}$ we need some information, too.

Lemma 5. *For any* $K > 0$ *there exists* $\delta > 0$ *such that a.s.*

$$\sup_{|z|=1, |z-1|\ge 1/\sqrt{2}-\delta} |W_n(z)| = \mathcal{O}\left(\frac{n}{(\log n)^K}\right)$$

as $n \to \infty$.

The proof of Lemma 5 is not difficult. It relies on explicit expressions for the second moment of $W_n(z)$, Markov's inequality, and the Borel-Cantelli lemma (compare with [8]).

It is now easy to complete the proof of Theorem 2. We have

$$\begin{aligned} U_{k,n} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} W_n(e^{it})e^{-kit}\,dt \\ &= \frac{1}{2\pi}\left(\int\limits_{|t|\le \arccos\frac{3}{4}-\delta} + \int\limits_{\arccos\frac{3}{4}-\delta\le t\le\pi}\right) W_n(e^{it})e^{-kit}\,dt \\ &= I_1 + I_2. \end{aligned}$$

With help of Lemma 5 we can easily estimate I_2 from above. A.s. we have

$$\begin{aligned} |I_2| &\le \frac{1}{2\pi} \int\limits_{\arccos\frac{3}{4}-\delta\le t\le\pi} |W_n(e^{it})|\,dt \\ &\ll \frac{n}{(\log n)^K}. \end{aligned}$$

For $|t| \le \arccos\frac{3}{4} - \delta$, $M_n(e^{it})$ is uniformly bounded a.s. Hence, we have by (8)

$$|W_n(e^{it})| \ll ne^{2(\cos t-1)\log n} \ll ne^{-c't^2\log n}$$

for some constant $c' > 0$. Now fix some (sufficiently small) $\eta > 0$. Then we have

$$\begin{aligned} \frac{1}{2\pi} \int\limits_{(\log n)^{-(1-\eta)/2}\le|t|\le\arccos\frac{3}{4}-\delta} |W_n(e^{it})|\,dt &\ll n \int\limits_{(\log n)^{(1-\eta)/2}}^{\infty} e^{-c't^2\log n}\,dt \\ &\ll ne^{-c'(\log n)^\eta}. \end{aligned}$$

So it remains so consider the integral

$$I_1' := \frac{1}{2\pi} \int\limits_{|t|\le(\log n)^{-(1-\eta)/2}} W_n(e^{it})e^{-kit}\,dt.$$

For $|t| \le (\log n)^{-(1-\eta)/2}$ we have (8) a.s. and uniformly in k

$$W_n(e^{it})e^{-kit} = (n+1)e^{it(2\log n-k)-t^2\log n}\left(1+\mathcal{O}\left(t+t^3\log n\right)\right).$$

Since

$$\int_{-\infty}^{\infty} e^{-t^2\log n}(|t|+|t|^3\log n)\,dt \ll \frac{1}{\log n}$$

and

$$\int\limits_{|t|\ge(\log n)^{-(1-\eta)/2}} e^{-t^2\log n}(1+|t|+|t|^3\log n) \ll e^{-(\log n)^\eta}$$

it follows that

$$\begin{aligned} \frac{I_1'}{n+1} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it(2\log n-k)-t^2\log n}\,dt + \mathcal{O}\left(\frac{1}{\log n}\right) \\ &= \frac{1}{\sqrt{2\pi 2\log n}} e^{-\frac{(k-2\log n)^2}{2\log n}} + \mathcal{O}\left(\frac{1}{\log n}\right), \end{aligned}$$

which completes the proof of Theorem 2.

2.4. **Height.** The distribution of the height H_n of binary search trees has turned out to be an interesting (and difficult) problem. We start with some history.

In 1986 Devroye [10] proved that the expected value $\mathbf{E}H_n$ satisfies the asymptotic relation $\mathbf{E}H_n \sim c\log n$ (as $n \to \infty$), where $c = 4.31107\ldots$ is the (largest real) solution of the equation $c\log\left(\frac{2e}{c}\right) = 1$. (Earlier Pittel [47] had shown that $H_n/\log n \to \gamma$ almost surely as $n \to \infty$, where $\gamma \le c$, compare also with Robson [51]. Later Devroye [11] provided a first bound for the error term, he proved $H_n - c\log n = \mathcal{O}(\sqrt{\log n\,\log\log n})$ in probability.) Based on numerical data Robson conjectured that the variance $\mathbf{Var}H_n$ is bounded. In fact, he could prove (see [52]) that there is an infinite subsequence for which

$$\mathbf{E}|H_n - \mathbf{E}H_n| = \mathcal{O}(1),$$

and that his conjecture is equivalent to the assertion that the expected value of the number of nodes at level $k = H_n$ is bounded (see [53]). The best bounds (before 1999) were given by two completely different methods by Devroye and Reed [12] and later by Drmota [15]. They (both) proved

$$\mathbf{E}H_n = c\log n + \mathcal{O}(\log\log n) \tag{10}$$

and

$$\mathbf{Var}H_n = \mathbf{E}(H_n - \mathbf{E}H_n)^2 = \mathcal{O}((\log\log n)^2).$$

Eventually, Reed [49] settled Robson's conjecture saying that

$$\mathbf{Var}H_n = \mathcal{O}(1) \qquad (n \to \infty).$$

His approach is related to that of [12], moreover he could also show that

$$\mathbf{E}H_n = c\log n - \frac{3c}{2(c-1)}\log\log n + \mathcal{O}(1). \tag{11}$$

A second proof of Robson's conjecture was given (independently) by the author [16] (just a few months later than Reed).

Before stating the result on the distribution of the height of binary search trees we want to present a first flavour of this problem. It is clear that $U_{k,n} = 0$ implies $H_n > k$ (the converse statement is *almost true*). By (7) it follows that we have $\mathbf{E}\,U_{n,k} < 1$ if $k > c\log n - \frac{c}{2(c-1)}\log\log n + \mathcal{O}(1)$, where $c = 4.31107\ldots$ is the (largest real) solution of the equation $c\log\left(\frac{2e}{c}\right) = 1$. Hence, one might expect that H_n is *concentrated* around $c_n := c\log n - \frac{c}{2(c-1)}\log\log n$. We can be even more precise. Since

$$\mathbf{Pr}[H_n > k] \le \mathbf{Pr}[U_{k,n} = 0] \le \mathbf{E}U_{k,n}$$

we get (with help of (6))

$$\begin{aligned}\mathbf{E}H_n &= \sum_{k\ge 0}\mathbf{Pr}[H_n > k]\\ &\le c_n + \sum_{k\ge c_n}\mathbf{E}U_{k,n}\\ &\le c_n + \mathcal{O}(1).\end{aligned}$$

This estimate would be optimal if $\mathbf{E}\,U_{k,n}^2 = \mathcal{O}(1)$ for $k > c_n$. However, this is not true. And this is really the crux of the matter. As mentioned above, see (11), the expected height is definitely smaller.

In what follows we want to present an approach to the height of binary search trees which is totally based on generating functions (and is completely different to Reed's method [49]). We mainly follow [17].

Let

$$y_k(x) = \sum_{n\ge 0}\mathbf{Pr}[H_h \le k]\cdot x^n.$$

Then $y_0(x) \equiv 1$ and

$$y'_{k+1}(x) = y_k(x)^2$$

with initial condition $y_{k+1}(0) = 1$. Obviously, $y_k(x)$ are polynomials of degree $2^k - 1$ and have a limit $y(x) = 1/(1-x)$ (for $0 \le x < 1$). Thus, there is a singularity for $x = 1$, and in fact everything can be formulated in terms of the (singular) sequence $y_k(1)$.

Theorem 5. *There exists a monotonically decreasing function $\Psi(y)$, $y \ge 0$, with $\Psi(0) = 1$ and $\lim_{y\to\infty}\Psi(y) = 0$ satisfying the integral equation*

$$y\Psi(y/e^{1/c}) = \int_0^y \Psi(z)\Psi(y-z)\,dz$$

such that

$$\mathbf{Pr}[H_n \le k] = \Psi(n/y_k(1)) + o(1) \quad (n \to \infty), \tag{12}$$

where the $o(1)$-error term is uniform for all $k \ge 0$. Furthermore, there exist constants $C, \eta > 0$ such that

$$\mathbf{Pr}[|H_n - h_n| \ge y] \le Ce^{-\eta y}, \qquad (y > 0), \tag{13}$$

where $h_n = \max\{k : y_k(1) \le n\}$.

Especially, it follows from Theorem 5 that the expected value of the height H_n of binary search trees of size n is given by

$$\mathbf{E}H_n = \max\{k : y_k(1) \le n\} + \mathcal{O}(1) \qquad (n \to \infty). \tag{14}$$

and that all centralized moments are bounded:[7]

$$\mathbf{E}|H_n - \mathbf{E}H_n|^r = \mathcal{O}(1) \qquad (n \to \infty). \tag{15}$$

If one combines (11) and (14) one gets[8]

$$y_k(1) = e^{k/c + \frac{3}{2(c-1)}\log k + \mathcal{O}(1)}. \tag{16}$$

It would be (very) interesting to find a direct proof of (16) or even tighter estimates.

The first step of the proof is to solve a proper fixed point equation.

Lemma 6. *Let $\alpha = e^{1/c}$. Then there uniquely exists a function $\Psi(y)$, $y \ge 0$, with the following properties:*

1. $\Psi(y) - 1 \sim c_1 y^{c-1}\log y$ *as $y \to 0+$ for some constant c_1.*
2. $\Psi(y) = \mathcal{O}\left(e^{-Cy^\gamma}\right)$ *as $y \to \infty$ for some $C > 0$ and some $\gamma > 1$.*
3. $\Psi(y)$, $0 \le y < \infty$, *is decreasing.*
4. $\displaystyle\int_0^\infty \Psi(y)\,dy = 1.$
5. $\displaystyle y\Psi(y/\alpha) = \int_0^y \Psi(z)\Psi(y-z)\,dz$

Proof. We first suppose that $1 < \rho < \alpha = e^{1/c}$ and consider $\beta < c-1$ satisfying $2\rho^\beta = \beta + 1$. (Note that then $2\rho^{c-1} < c$.) Let D denote the space of all decreasing functions Ψ on $[0,\infty)$ with $\Psi(y) = 1 - y^\beta + \mathcal{O}\left(y^{c-1}\right)$ (for all $y \ge 0$) and $\Psi(y) \to 0$ as $y \to \infty$. Furthermore, set

$$d(\Psi_1, \Psi_2) = \sup_{y>0}\left|y^{1-c}(\Psi_1(y) - \Psi_2(y)\right|.$$

Then (D, d) is a complete metric space. Next consider the mapping $S : D \to D$ defined by

$$S(\Psi)(y) = \frac{1}{\rho y}\int_0^{\rho y} \Psi(z)\Psi(\rho y - z)\,dz.$$

Suppose that $d(\Psi_1, \Psi_2) = \delta$ for some $\Psi_1, \Psi_2 \in D$. From

$$\begin{aligned}|\Psi_1(z)\Psi_1(\rho y - z) - \Psi_2(z)\Psi_2(\rho y - z)| &\le |\Psi_1(z) - \Psi_2(z)| + |\Psi_1(\rho y - z) - \Psi_2(\rho y - z)| \\ &\le \delta\left(z^{c-1} + (\rho y - z)^{c-1}\right)\end{aligned}$$

we obtain

$$|S(\Psi_1)(y) - S(\Psi_2)(y)| \le \frac{2\delta}{\rho y}\int_0^{\rho y} z^{c-1}\,dz = \frac{2\rho^{c-1}}{c}\delta y^{c-1}.$$

and consequently $d(S(\Psi_1)(y), S(\Psi_2)(y)) \le L \cdot d(\Psi_1, \Psi_2)$ with $L = 2\rho^{c-1}/c < 1$. Thus, $S : D \to D$ is a contraction and one gets a unique fixed point in D.

---

[7] Note that (15) solves Robson's conjecture, however, in a quite implicit way. Interestingly, Theorem 5 does not provide any information on the magnitude of $y_k(1)$ and thus (via (14)) no quantitative bound for the expected height $\mathbf{E}H_n$. It just uses the property that the sequence $y_{k+1}(1)/y_k(1)$ converges (see Lemma 9). The exact order of $\mathbf{E}H_n$ was given by Reed [49] by improving the previous bound (10) by Devroye and Reed [12] and by the author [15].

[8] Interestingly the asymptotic behaviour of $y_k(1)$ was posed as an unsolved problem by C. Ponder [48] without stating any connection to binary search trees.

By keeping track of the iteration (leading to the fixed point Ψ) it is easy to show that there exists constants $C > 0, \gamma > 1$ such that $\Psi(y) = \mathcal{O}\left(e^{-Cy^\gamma}\right)$ as $y \to \infty$ (see [17]). Thus, the integral $\int_0^\infty \Psi(y)\, dy$ exists and by proper scaling of the argument we can assume that this integral equals 1. (We will denote this function by $\Psi(y, \rho)$.

Unfortunately, we cannot use the same argument for $\rho = \alpha$ since $2\alpha^{c-1}/c = 1$ and we have no contraction. However, it is possible to consider the limit $\lim\limits_{\rho\to\alpha-} \Psi(y, \rho)$ though not in a direct way.

Let $\Phi(u) = \Phi(u, \rho)$ denote the Laplace transform

$$\Phi(u, \rho) = \int_0^\infty \Psi(y, \rho) e^{-uy}\, dy$$

of $\Psi(z, \rho)$. Then $\Phi(u) = \Phi(u, \rho)$ satisfies the differential equation

$$\Phi'(u) = -\frac{1}{\rho^2} \Phi\left(\frac{u}{\rho}\right)^2$$

with the initial condition $\Phi(0) = 1$. It is now easy to show that for every $\rho > 0$ this differential equation has a unique (entire) solution of the form

$$\Phi(u, \rho) = \sum_{k\ge 0} (-1)^k c_k(\rho) u^k,$$

where $c_0(\rho) = 1$ and

$$c_{k+1}(\rho) = \frac{\rho^{-k}}{k+1} \sum_{l=0}^{k} c_l(\rho) c_{k-l}(\rho),$$

i.e. $c_k(\rho)$ are polynomials in $1/\rho$ with non-negative coefficients. For $1 < \rho < \alpha$ these solutions coincide with the Laplace transforms of $\Psi(y, \rho)$. It is no problem to consider the (obvious) limit $\lim_{\rho\to\alpha-} \Phi(u, \rho) = \Phi(u, \alpha)$ and then to use the inverse Laplace transform to ensure that the limit $\lim_{\rho\to\alpha-} \Psi(y, \rho)$ actually exists and has all proposed properties (for details see [17]). □

With help of the function $\Psi(y)$ of Lemma 6 we define auxiliary functions

$$\tilde{y}_k(x) := \int_0^\infty \Psi(y/\alpha^k) e^{-y(1-x)}\, dy = \alpha^k \Phi(\alpha^k(1-x), \alpha), \tag{17}$$

where k is an arbitrary real (not necessarily an integral) number. In some sense these functions *simulate* the above polynomials $y_k(x)$.

Lemma 7. *The functions $\tilde{y}_k(x)$, $k \ge 0$, $x \ge 0$, defined by (17) satisfy*

1. $\tilde{y}'_{k+1}(x) = \tilde{y}_k(x)^2$.
2. $0 < \tilde{y}_k(0) < 1$.
3. $1 - \tilde{y}_k(0) \sim \dfrac{C_1}{c} k \left(\dfrac{2}{c}\right)^k \quad (k \to \infty)$.
4. $\tilde{y}_k(1) = \alpha^k$.
5. $\tilde{y}_k(x)$ *has a power series expansion* $\tilde{y}_k(x) = \sum_{n\ge 0} a_{nk} x^n$ *with positive coefficients* $a_{nk} > 0$ *which are asymptotically given by*
$$a_{nk} = \Psi(n/\alpha^k) + o(1),$$
where the $o(1)$ error term is uniform for all integers $n \ge 0$ and all real numbers $k \ge 0$.
6. *For every integer $k \ge 0$ and for every real number l the difference $y_k(x) - \tilde{y}_l(x)$ has exactly one zero for $x_{k,l}$ on the positive real line. Furthermore, these zeros satisfy $x_{k+1,l+1} > x_{k,l}$.*

Proof. The first 4 properties are direct translations of corresponding properties of $\Psi(y)$.

Next, by definition we have

$$\tilde{y}_k(x) = \sum_{n\ge 0} \left(\frac{1}{n!} \int_0^\infty y^n e^{-y} \Psi(y\alpha^{-k})\, dy\right) x^n.$$

If $C_1 \le n\alpha^{-k} \le C_2$ for some constants $C_1, C_2 > 0$. then by the Laplace method and by using the relation

$$\frac{1}{n!}\int_0^\infty y^n e^{-y}\,dy = 1,$$

we obtain (by expanding locally around $y = n$) $a_{nk} \sim \Psi(n/\alpha^k)$ uniformly for this kind of range. However, this directly implies $a_{nk} = \Psi(n/\alpha^k) + o(1)$ for the whole range.

In order to prove the last property we proceed by induction on k. For convenience we set $\delta_{k,l}(x) = y_k(x) - \tilde{y}_l(x)$. Since $\tilde{y}_l(x)$ is strictly increasing and satisfies $0 < \tilde{y}_l(0) < 1$ and $\lim_{x\to\infty} \tilde{y}_l(x) = \infty$ the assertion is surely true for $k = 0$. Now suppose that $\delta_{k,l}(x)$ has exatly one zero $x_{k,l} > 0$. Since

$$\begin{aligned}\delta'_{k+1,l+1}(x) &= y'_{k+1}(x) - \tilde{y}'_{l+1}(x)\\ &= y_k(x)^2 - \tilde{y}_l(x)^2\\ &= \delta_{k,l}(x)(y_k(x) + \tilde{y}_l(x)).\end{aligned}$$

it follows that $\delta_{k+1,l+1}(x)$ is increasing for $0 \le x < x_{k,l}$ and decreasing for $x > x_{k,l}$. Since $\delta_{k+1,l+1}(0) > 0$ and $\lim_{x\to\infty}\delta_{k+1,l+1}(x) = -\infty$ there exists a unique zero $x_{k+1,l+1} > x_{k,l}$ of $\delta_{k+1,l+1}(x)$. □

With help of these auxiliary function $\tilde{y}_k(x)$ we obtain proper tail estimates of the distribution of H_n.

Lemma 8. *Set $e_k = c\log y_k(1)$. Then $e_{k+1} \ge e_k + 1$ and there exist a constant $C < 0$ such that*

$$\mathbf{Pr}[H_n \le k] \le C\alpha^{-(c\log n - e_k)}$$

and

$$\mathbf{Pr}[H_n > k] \le C\alpha^{-(e_k - c\log n)}.$$

Proof. By definition $\tilde{y}_{e_k}(1) = y_k(1)$. Thus, by 6. of Lemma 7 we have $\tilde{y}_{e_k}(x) \le y_k(x)$ for $0 \le x \le 1$ and $\tilde{y}_{e_k}(x) \ge y_k(x)$ for $x \ge 1$. Especially it follows that $\tilde{y}_{e_k+1}(x) \le y_{k+1}(x)$ for $0 \le x \le 1$ and consequently $\alpha^{e_k+1} \le \alpha^{e_{k+1}}$ which gives $e_{k+1} \ge e_k + 1$.

Suppose that $x \ge 1$. Then we get (by using the trivial inequality $\mathbf{Pr}[H_n \le k] \le \mathbf{Pr}[H_{n+1} \le k]$)

$$\tilde{y}_{e_k}(x) \ge y_k(x) \ge \sum_{l=0}^{n} \mathbf{Pr}[H_l \le k]\, x^l \ge \mathbf{Pr}[H_n \le k]\,\frac{x^{n+1}-1}{x-1}.$$

Choosing $x = 1 + \alpha^{-e_k}$ and using the definition of $\tilde{y}_{e_k}(x)$ we obtain the upper bound

$$\mathbf{Pr}[H_n \le k] \le \frac{1}{(1+\alpha^{-e_k})^{n+1} - 1}\Phi(-1) \ll \frac{1}{n\alpha^{-e_k}} = \alpha^{-(c\log n - e_k)}. \tag{18}$$

In the same fashion we have for $0 < x < 1$

$$\begin{aligned}\frac{1}{1-x} - \tilde{y}_{e_k}(x) &\ge \frac{1}{1-x} - y_k(x)\\ &\ge \sum_{l=n}^{\infty} (1 - \mathbf{Pr}[H_l \le k])\, x^l\\ &\ge (1 - \mathbf{Pr}[H_n \le k])\,\frac{x^n}{1-x}.\end{aligned}$$

Finally, setting $x = 1 - 1/n$ we directly get

$$1 - \mathbf{Pr}[H_n \le k] \ll 1 - \alpha^{e_k - c\log n}\Phi(\alpha^{e_k - c\log n}) \ll \alpha^{-(e_k - c\log n)}. \tag{19}$$

□

Obviously, the tail estimate (13) follows from Lemma 8. In order to complete the proof of Theorem 5 we have to refine the methods a little bit. The idea is to approximate $y_k(x)$ by $\tilde{y}_{e_k}(x)$. Recall that e_k was defined such that $y_k(1) = \tilde{y}_{e_k}(1)$. It turns out that the next property is crucial in order to get more.

Lemma 9. *We have*

$$\frac{y_{k+2}(1)}{y_{k+1}(1)} \le \frac{y_{k+1}(1)}{y_k(1)}. \tag{20}$$

Consequently the sequence $y_{k+1}(1)/y_k(1)$ *converges and its limit is given by*

$$\lim_{k\to\infty} \frac{y_{k+1}(1)}{y_k(1)} = \alpha = e^{1/c}. \tag{21}$$

Proof. Let $\gamma \in (0,1)$ be a fixed constant and let $z_k(x,\gamma)$ be defined by $z_0(x,\gamma) = 1/(1-x)$ for $x \le 1-\gamma$, by $z_0(x,\gamma) = 1/\gamma$ for $x > 1-\gamma$, and recursively by

$$z_{k+1}(x,\gamma) = 1 + \int_0^x z_k(t,\gamma)^2\,dt \quad \text{for } h \ge 0.$$

Of course, by induction it follows that $z_k(x,\gamma) = 1/(1-x)$ for $x \le 1-\gamma$ and that

$$z_k(x,\gamma) = \frac{1}{\gamma} y_k\left((x-1+\gamma)/\gamma\right) \quad \text{for } x > 1-\gamma.$$

Now we proceed as in the proof of 6. of Lemma 7 and obtain that the difference $z_k(x,\gamma) - y_{k+1}(x)$ has exactly one zero $\xi_k(\gamma)$ in the range $x > 0$, i.e. $z_k(x,\gamma) > y_{k+1}(x)$ for $0 < x < \xi_k(\gamma)$ and $z_k(x) < y_{k+1}(x)$ for $x > \xi_k(\gamma)$. Furthermore $\xi_{k+1}(\gamma) > \xi_k(\gamma)$. We now apply this property for $\gamma = y_k(1)/y_{k+1}(1)$. Since $z_k(1,\gamma) = y_k(1)/\gamma$ it follows that

$$z_k\left(1, \frac{y_k(1)}{y_{k+1}(1)}\right) = y_{k+1}(1)$$

or $\xi_k\big(y_k(1)/y_{k+1}(1)\big) = 1$. Consequently $\xi_{k+1}(y_k(1)/y_{k+1}(1)) > 1$ and thus

$$z_{k+1}\left(1, \frac{y_k(1)}{y_{k+1}(1)}\right) = \frac{y_{k+1}(1)^2}{y_k(1)} \ge y_{k+2}(1)$$

as proposed.

Since the sequence $y_{k+1}(1)/y_k(1)$ is decreasing and non-negative is it thus convergent. We already know that $\log y_k(1) \sim k/c$ (compare with (16)). Hence the limit is given by (21). □

For example, it follows from Lemma 9 that

$$y_k'(1) = y_{k-1}(1)^2 \sim y_k(1)^2\alpha^{-2} = \tilde{y}_{e_k-1}(1)^2 = \tilde{y}_{e_k}'(1).$$

Inductively, one gets for every fixed $l \ge 0$ that $y_k^{(l)}(1) \sim \tilde{y}_{e_k}^{(l)}(1)$ as $k \to \infty$. Thus, $y_k(x)$ can be properly approximated by $\tilde{y}_{e_k}(x)$ in a complex neighbourhood of $x = 1$. Together with some further (technical but easy) estimates (compare with [17]) it follows via Cauchy's formula that

$$\begin{aligned}
\mathbf{Pr}[H_n \le k] &= \frac{1}{2\pi i}\int_{|x|=1} \frac{y_k(x)}{x^{n+1}}\,dx \\
&= \frac{1}{2\pi i}\int_{|x|=1} \frac{\tilde{y}_{e_k}(x)}{x^{n+1}}\,dx + o(1) \\
&= \Psi(n/\alpha^{e_k}) + o(1) \\
&= \Psi(n/y_k(1)) + o(1),
\end{aligned}$$

which completes the proof of Theorem 5.

2.5. **Branching Random Walks.** Devroye and others (see [10, 12, 47, 49]) have used a slightly modified apporach to tackle the height H_n of binary search trees. Let T_∞ denote the infinte binary tree rooted at v_{root}. Each node v of T_∞ has a rights son $r(v)$ and a left son $l(v)$. For every node v we consider iid random variable $U(v)$, which are ud on $[0,1]$, and label the edge $(v, r(v))$ by $U(v)$ and the edge $(v, l(v))$ by $1 - U(v)$. Furthermore, let $f(v)$ denote the product of the labels of the edges on the unique path from the root v_{root} to v. If the labels on the path from the root to v are $U_1, U_2, \ldots, U_k$ then we also define

$$h_n(v) = \lfloor \cdots \lfloor \lfloor nU_1 \rfloor U_2 \rfloor \cdots U_k \rfloor.$$

One observation from Devroye [10] was that the sequence of random trees

$$T_n = \{v \in T_\infty : h_n(v) \ge 1\}$$

constitute exactly random binary search tree. In this context it is also quite natural to consider another sequence of random trees:

$$\overline{T}_n = \{v \in T_\infty : f(v) \ge 1/n\}.$$

Since

$$nf(v) - k \le h_n(v) \le nf(v)$$

for every node v of distance k to the root the random trees T_n and $\overline{T}_n$ are quite close, especially their heights H_n and $\overline{H}_n$.[9] Note that Reed [49] has used this connection, in fact, we first provded that

$$\mathbf{E}\,\overline{H}_n = c\log n - \frac{3c}{2(c-1)}\log\log n + \mathcal{O}(1) \tag{22}$$

and $\mathbf{Var}\,\overline{H}_n = \mathcal{O}(1)$ as $n \to \infty$, where $c = 4.31107...$ is the same constant as in (11). The corresponding result for binary search trees was (more or less) a *corollary* from these facts.

The height distribution of $\overline{H}_n$ is also very closely connected to the *Random Bisection Problem* or *Random Fragmentation Problem* (see [37]). There, an interval of length x is cut into two halves of length $x_1 = Ux$ and $x_2 = (1-U)x$, where U is ud on $[0,1]$. Next, each of these two fragments is cut again into two parts (independently of the other and previous cuts). After the k-th step, there are 2^k fragments whose lengths are correlated random variables. Given the initial length x, the problem is to determine the probability $P_k(x,l)$ that each of the 2^k fragments is shorter (or equal) than l. (Obviously this problem is homegeneous in x and l, thus we can write $P_k(x,l) = \overline{P}_k(x/l)$ or set $l = 1$.) By definition we have $\overline{P}_0(x) = 1$ for $0 \le x < 1$ and $\overline{P}_0(x) = 0$ for $x \ge 1$ and recursively

$$\overline{P}_{k+1}(x) = \frac{1}{x}\int_0^x \overline{P}_k(y)\overline{P}_k(x-y)\,dy. \tag{23}$$

For later use we also define

$$x_k := \int_0^\infty \overline{P}_k(x)\,dx, \tag{24}$$

which is surely finite because $P_k(x) = 0$ for $x \ge 2^k$.

The connection between the height $\overline{H}_n$ and the random bisection problem (which is valid for non-integral n, too) is given by

$$\mathbf{Pr}[\overline{H}_n \le k] = P_k\left(1, \frac{1}{n}\right) = \overline{P}_k(n) \tag{25}$$

(which is obvious by definition).

There is even another interpretation of $\overline{H}_n$. Consider the discrete *branching random walk* defined by the point process

$$Z = \delta_{X_1} + \delta_{X_2},$$

where $X_1 = \log(1/U)$ and $X_2 = \log(1/(1-U))$ with U ud on $[0,1]$.[10] Let L_k denote the left most particle at stage k (and R_k the corresponding right most particle). Then the distribution of L_k is given by

$$\mathbf{Pr}[L_k > x] = \overline{P}_k(e^x).$$

[9] In what follows it will be convenient to consider n in the defintion of $\overline{T}_n$ and $\overline{H}_n$ also as a continuous parameter $n > 0$. Of course, this can be also done for T_n and H_n but then there is no interpretation as usual random binary search trees for non integral n.

[10] The branching random walk is a sequence of point processes Z_k, where $Z_0 = \delta_0$ and Z_{k+1} evolves from Z_k by splitting each particle of Z_k (independently of the others) into (a random number) N of points with displacements given by a fixed point process $Z = \delta_{X_1} + \cdots + \delta_{X_N}$.

Thus, we have three different but equivalent problems. We will show that the height $\overline{H}_n$ can be treated in completely the same way as above H_n. The result reads as follows.

Theorem 6. *Let $\Psi(y)$, $y \ge 0$, be the same function as in Theorem 5. Then*

$$\mathbf{Pr}[\overline{H}_n \le k] = \Psi(n/x_k) + o(1) \quad (n \to \infty), \tag{26}$$

where the $o(1)$-error term is uniform for all $k \ge 0$. Furthermore, there exist constants $C, \eta > 0$ such that

$$\mathbf{Pr}[|\overline{H}_n - \overline{h}_n| \ge y] \le Ce^{-\eta y}, \qquad (y > 0), \tag{27}$$

where $\overline{h}_n = \max\{k : x_k \le n\}$.

As in the case of binary search trees it follows from Theorem 6 that the expected value of the height $\overline{H}_n$ is given by

$$\mathbf{E}\overline{H}_n = \max\{k : x_k \le n\} + \mathcal{O}(1) \qquad (n \to \infty). \tag{28}$$

and that all centralized moments are bounded:

$$\mathbf{E}|\overline{H}_n - \mathbf{E}\overline{H}_n|^r = \mathcal{O}(1) \qquad (n \to \infty). \tag{29}$$

If one combines (22) and (28) one gets as above

$$x_k = e^{k/c + \frac{3}{2(c-1)}\log k + \mathcal{O}(1)}. \tag{30}$$

Equivalently, we have solved the random bisection problem:

$$P_k(x,l) = \overline{P}_k(x/l) = \Psi\left(\frac{x}{lx_k}\right) + o(1).$$

Finally, the distribution of the left most particle L_k at stage k in the above discrete branching random walk is given by

$$\mathbf{Pr}[L_k > x] = w(x - m(k)) + o(1), \tag{31}$$

where $w(x) = \Psi(e^x)$ and

$$m(k) = \log x_k = \frac{k}{c} + \frac{3}{2(c-1)}\log k + \mathcal{O}(1)$$

is closely related the $\Psi(1)$-quantil of the distibution of L_k. Note that the function $w(x)$ is known as a *travelling wave* in the context of branching random walks. The existence of travelling waves is an important problem in branching random walks and only known for special cases (see [7] for the classical *Branching Brownian motion* and [4] for discrete branching random walks with iid displacements X_j (of the defining point process Z) with a log-concave density.) The relation (31) seems to be the first travelling wave solution for a discrete branching random walk with non-independent and unbounded increments of the defining point process Z.

The proof of Theorem 6 is (as already mentioned) very close to that of Theorem 5. If we define the *Laplace transforms*

$$\begin{aligned}\overline{y}_k(x) &= \int_0^\infty \mathbf{Pr}[\overline{H}_n \le k]\, e^{(x-1)n}\, dn \\ &= \int_0^\infty \overline{P}_k(y) e^{(x-1)y}\, dy\end{aligned}$$

then (23) directly translates to

$$\overline{y}'_{k+1}(x) = \overline{y}_k(x)^2$$

with initial conditions

$$\overline{y}_0(x) = \frac{1}{x-1}\left(e^{x-1} - 1\right)$$

and

$$\overline{y}_k(0) = \int_0^\infty \overline{P}_k(y) e^{-y}\, dy \sim 1.$$

Hence, we have a completely similar situation as for the case of binary search trees. Note also that

$$\overline{y}_k(1) = x_k.$$

As above, the *priciple idea* is to approximate $\overline{y}_k(x)$ by $\tilde{y}_{\overline{e}_k}(x)$, where $\overline{e}_k = c\log \overline{y}_k(1)$ and

$$\tilde{y}_k(x) = \int_0^\infty \Psi(y/\alpha^k)e^{(x-1)y}\,dy$$

is the same function as above.

3. Galton Watson Trees

3.1. **Probabilistic Model.** Let ξ be a non-negative integer valued random variable with $\mathbf{E}\,\xi = 1$, $0 < \mathbf{Var}\,\xi = \sigma^2 < \infty$.[11] The Galton-Watson branching process $(Z_k)_{k\ge 0}$ is now given by $Z_0 = 1$, and for $k \ge 1$ by

$$Z_k = \sum_{j=1}^{Z_{k-1}} \xi_j^{(k)},$$

where the $(\xi_j^{(k)})_{k,j}$ are iid random variables distributed as ξ.

It is well known that Galton-Watson branching processes can be represented by ordered (finite or infinite) rooted trees T such that the sequence $(Z_k)_{k\ge 0}$ is just the profile sequence and $\sum_{k\ge 0} Z_k$, which is called the *total progeny*, is just the number of nodes $|T|$ of T.[12] We will denote $\nu(T)$ the probability that T occurs. The generating function $y(x) = \sum_{n\ge 1} y_n x^n$ of the numbers

$$y_n = \mathbf{Pr}[|T| = n] = \sum_{|T|=n} \nu(T)$$

satisfies the functional equation

$$y(x) = x\varphi(y(x)),$$

where $\varphi(t) = \mathbf{E}\,t^\xi = \sum \varphi_i t^i$ with $\varphi_i = \mathbf{Pr}[\xi = i]$. If $\mathcal{T}_n$ denotes the set of rooted trees T of size $|T| = n$ then

$$\nu_n(T) := \frac{\nu(T)}{y_n}$$

is a probability distribution on $\mathcal{T}_n$ which we will use in the sequel. Note that

$$y_n \sim \frac{d}{\sqrt{2\pi}\sigma} n^{-3/2} \qquad (n \equiv 1 \bmod d), \tag{32}$$

where $d = \gcd\{i > 0 : \varphi_i > 0\}$.[13]

For example, for $\varphi(t) = \mathbf{E}\,t^\xi = (1+t)^2/4 = \frac14 + \frac t2 + \frac{t^2}{4}$ we just recover the class of binary trees with n (internal) nodes, where each binary tree (of size n) has equal probability. Another well know expample are planted plane trees (again with uniform distribution on trees of size n). They are induced by $\varphi(t) = \mathbf{E}\,t^\xi = 1/(2-t) = \frac12 + \frac t4 + \frac{t^2}{8} + \cdots$.

In order to analyze a rooted tree we consider the so-called *depth-first search*. It can be described as a walk $(v(i), 1 \le i \le 2n-1)$ around the vertices. Let $v(1)$ be the root. Given $v(i)$ choose (if possible) the first (in the ordering) edge at $v(i)$ leading away from the root which has not already been travered, and let $(v(i), v(i+1))$ be that edge. If this is not possible, let $(v(i), v(i+1))$ be the edge from $v(i)$ leading towards the root. This walk terminates with $v(2n-1)$, which equals (again) the root.

The search depth $x(i)$ is now defined as the distance from the root to $v(i)$ plus 1 and $x(0) = x(2n) = 0$. (For non-ingegral i we use linear interpolation and thus $x(t)$, $0 \le t \le 2n$, can be considered as a continuous *excursion*).

[11] Usually it is not assumed that $\mathbf{E}\,\xi = 1$ (which characterize so-called *critical branching processes*). However, for our purposes it is no loss of generality to make this assumption (see [32]).

[12] For critical branching processes the probability that the total progeny is finite equals 1.

[13] In what follows we will always assume that $d = 1$. The case $d > 1$ is completely analogue.

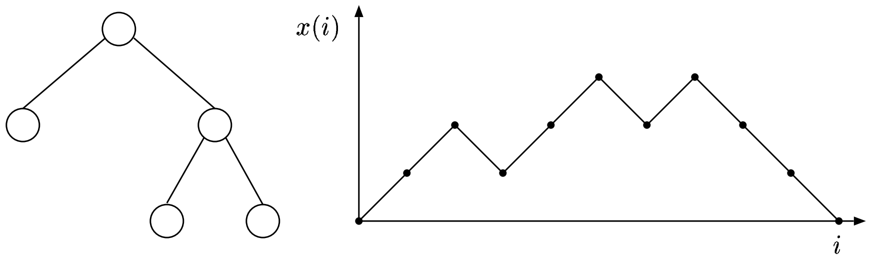

FIGURE 3. Depth-first search of a rooted tree

Note that the height h and the path lenght l can be expressed with help of $x(i)$:

$$h = \max_{0\le i\le 2n} x(i) - 1 \quad \text{and} \quad l = \frac{1}{2}\sum_{i=0}^{2n} x(i) - \frac{n}{2}.$$

If we consider planted plane trees, i.e. $\mathbf{E}\, t^{\xi} = 1/(2-t)$, then this excursion is nothing else than the *Dyck path*, i.e. the *standard random walk* starting at the origin with probabilty $\frac{1}{2}$ going one step up resp. one step down conditioned on $x(2n) = 0$ and $x(j) \ne 0$ for $1 \le j \le 2n-1$.

We will denote by $X_n(t)$, $0 \le t \le 2n$, the random search depth of trees in $\mathcal{T}_n$ resp. of Galton-Watson trees conditioned on the total progeny. As for binary search trees, the random profile (for trees in $\mathcal{T}_n$) will be denoted by $(V_{k,n})_{k\ge 0}$ (which can be considered as the branching process $(Z_k)_{k\ge 0}$ conditions on the total progeny); L_n denotes the total path length, and H_n the height.

3.2. **Search Depth.** The following general result on the search depth process has been proved by Aldous [1, 2, 3].

Theorem 7. *Suppose that the offspring distribution ξ of a Galton-Watson branching process is critical, i.e. $\mathbf{E}\,\xi = 1$, and that the variance $\sigma^2 = \mathbf{Var}\,\xi$ is non-zero and finite. Then the rescaled search depth process converges weakly to $\frac{2}{\sigma}e(t)$, where $e(t)$ denotes the standard Brownian excursion of duration 1:*

$$\left(\frac{1}{\sqrt{n}}X_n(2nt),\ 0\le t\le 1\right) \Longrightarrow \left(\frac{2}{\sigma}e(t), 0\le t\le 1\right).$$

In view of the usual discrete excursion (correponding to the case $\mathbf{E}t^{\xi} = 1/(2-t)$), where the weak convergence to Brownian excursion is a well known result, Theorem 7 is a natural generalization.

Since the height is given by

$$H_n = \max_{0\le t\le 2n} X_n(t) - 1$$

and the internal path length by

$$L_n = \frac{1}{2}\sum_{i=0}^{2n} X_n(i) - \frac{n}{2} \sim \frac{1}{2}\int_0^{2n} X_n(t)\,dt$$

we directly obtain the following weak convergence results.

Theorem 8. *We have*

$$\frac{1}{\sqrt{n}}H_n \Longrightarrow \frac{2}{\sigma}\max_{0\le t\le 1} e(t)$$

and

$$\frac{1}{n^{3/2}}L_n \Longrightarrow \frac{2}{\sigma}\int_0^1 e(t)\,dt.$$

The distribution function of $M = \max_{0\le t\le 1} e(t)$ is given by

$$\mathcal{L}(M) = \mathbf{Pr}[M \le x] = 1 - 2\sum_{k=1}^{\infty}(4x^2k^2-1)e^{-2x^2k^2}$$

and the moments by

$$\mathbf{E}(M^r) = 2^{-r/2}r(r-1)\Gamma(r/2)\zeta(r),$$

where $\zeta(s)$ denotes the Riemann Zeta-function and $(r-1)\zeta(r) = 1$ for $r = 1$. The moments of $I = \int_0^1 e(t)\,dt$ determined by

$$\mathbf{E}(I^r) = K_r\frac{4\sqrt{\pi}r!}{\Gamma\left(\frac{3r-1}{2}\right)2^{r/2}},$$

where K_r is recursively given by

$$K_r = \frac{3r-4}{4}K_{r-1} + \sum_{j=1}^{r-1}K_jK_{r-j}, \qquad (r\ge 2),$$

with initial values $K_0 = -\frac{1}{2}$ and $K_1 = \frac{1}{8}$.

Thus, one might expect that $\mathbf{E}(H_n^r) \sim \left(\frac{2}{\sigma}\right)^r \mathbf{E}(M^r)\cdot n^{r/2}$ and $\mathbf{E}(L_n^r) \sim \left(\frac{2}{\sigma}\right)^r \mathbf{E}(I^r)\cdot n^{3r/2}$ (which is fact true and has been proved by Flajolet and Odlyzko [24] resp. by Takács [56]) but this does follow from Theorem 7.

Aldous' proof of Theorem 7 (see [1, 2, 3]) is based on a very elegant concept of the *continuum random tree* (which will not be described here). But it *only* provides weak convergence, i.e. we get

$$\lim_{n\to\infty}\mathbf{E}\,F\left(X_n(2nt)/\sqrt{n}\right) = \mathbf{E}F\left(\frac{2}{\sigma}e(t)\right)$$

for bounded continuous functionals $F: C[0,1]\to\mathbb{R}$. However, those functionals we are interested in (projections, maximum, integral) are not bounded.

In what follows we will describe an alternate proof of Theorem 7 which is based on analytic methods (singularity analysis) applied to generating functions and which is suitable to prove convergence of moments of functionals of polynomial growth. In order to this we have to assume a little bit more that the existence of the second moment of ξ. We suppose that there exits $\gamma > 0$ such that

$$\mathbf{E}\,e^{\gamma\xi} < \infty \tag{33}$$

In order to simplify notation we set

$$e_n(t) := \frac{\sigma}{2\sqrt{n}}X_n(2nt).$$

Since $e_n(t)$ and $e(t)$ are stochastic processes on the space $C[0,1]$ weak convergence of $e_n(t) \Longrightarrow e(t)$ is equivalent to finite dimensions convergence and to tightness (see [5, Theorem 12.3]). These two conditions can be directly checked. (The proof can be worked out in the same way as in Gittenberger [27], where the *contour*, i.e. the sequence of heights of the leaves, has been considered.)

Lemma 10. *Suppose that (33) is satisfied for some $\gamma > 0$ and let $0 \le s_1 < s_2 < \cdots < s_k \le$ be fixed (with $k \ge 1$). Then the random vector $(e_n(s_1), e_n(s_2), \ldots, e_n(s_k))$ converges weakly to $(e(s_1), e(s_2), \ldots, e(s_k))$.*

Lemma 11. *Suppose that (33) is satisfied for some $\gamma > 0$ Then there exist constants $C, D > 0$ such that for all $s,t\in[0,1]$ and $\epsilon > 0$,*

$$\mathbf{Pr}\{|e_n(s)-e_n(t)|\ge\varepsilon\} \le \frac{C}{\varepsilon^4|s-t|}\exp\left(-D\frac{\varepsilon}{\sqrt{|s-t|}}\right).$$

Hence it follows that

$$\mathbf{Pr}\{|e_n(s) - e_n(t)| \ge \varepsilon\} \le C\frac{|s-t|^\alpha}{\varepsilon^{2\alpha}}$$

for every $\alpha > 1$. Thus, by [5, Theroem 12.3] the sequence of processes $e_n(t)$ is tight. It would have also sufficient to show

$$\mathbf{E}\left(|e_n(t) - e_n(s)|^\beta\right) \le C|t-s|^\alpha \quad \text{for all } s,t \in [0,1]. \tag{34}$$

for some $\alpha > 1$ and some $\beta > 0$. Of course, this follows from Lemma 11, too. For example, we can choose $\beta = 2\alpha$ for every $\alpha > 1$.

The idea of the proof is to use generating functions and analytic methods. The *basic* generating function is $y(x) = \sum_{n\ge1} \mathbf{Pr}[|T| = n]\, x^n$, the solution of the function equation $y = x\varphi(y)$. It is well known that (conditioned on (33)) the radius of convergence of $y(x)$ equals 1, there is a local expansion of the form

$$y(x) = 1 - \frac{\sqrt{2}}{\sigma}\sqrt{1-x} + c_2(1-x) + \mathcal{O}\left(|1-x|^{3/2}\right) \tag{35}$$

which is valid for a complex neighbourhood of $x = 1$ (without the half real axis $x \ge 1$), and $y(x)$ has an analytic continuation to the region $\{x \in \mathbb{C} : |x| < 1+\kappa, \arg x - 1 \ne 0\}$ for some $\kappa > 0$ (if one assumes that $d = 1$ — for $d > 1$ the situation is similar). For example, (35) directly implies (32) even with an error term of order $\mathcal{O}\left(n^{-5/2}\right))$ by an application of the *Transfer Lemma* of Flajolet and Odlyzko [25]:

Lemma 12. *Suppose that* $F(z) = \sum_{n\ge0} f_n z^n$ *is analytic in the domain*

$$\Delta = \{z : |z| < 1+\kappa,\ |\arg(z-1)| > \gamma\}$$

for some $\kappa > 0$ *and* $0 < \gamma < \frac{\pi}{2}$. *If*

$$|F(z)| \le C|1-z|^\alpha$$

for $z \in \Delta$ *(and some* $C > 0$ *and some real* α*) then, as* $n \to \infty$

$$f_n = \mathcal{O}\left(n^{-\alpha-1}\right).$$

Next consider the probabilities $a_{km,n} = \mathbf{Pr}[|T| = n,\ x(m) = k]$ (where $x(\cdot)$ denotes the search depth of T). Then the generating function $A_k(x,u) = \sum_{m,n} a_{km,n} x^n u^k$ of $a_{km,n}$ is given by (compare with [27])

$$A_k(x,u) = A(x,u)\phi_1(x,u,1)^{k-1}, \tag{36}$$

where

$$A(x,u) = xu\frac{y(xu^2)\varphi(y(xu^2)) - y(x)\varphi(y(x))}{y(xu^2) - y(x)}$$

and

$$\phi_1(x,u,v) = xuv\frac{\varphi(y(xu^2)) - \varphi(y(xv^2))}{y(xu^2) - y(xv^2)}.$$

Similarly one also gets explicit expressions for the generating function of the numbers

$$b_{k_1m_1,k_2m_2,n} = \mathbf{Pr}[|T| = n,\ x(m_1) = k_1,\ x(m_2) = k_2].$$

For $k_1 < k_2$ we obtain

$$\begin{aligned} B_{k_1,k_2}(x,u_1,u_2) &= \sum_{m_1,m_2,n\ge0} b_{k_1m_1,k_2m_2,n} x^n u_1^{m_1} u_2^{m_2} \\ &= A(x(u_1u_2)^2,1)A(x,u)\phi_1(x,u_1u_2,u_2)^{k_1-1}\phi_1(x,u_2,1)^{k_2-1} \\ &\times \phi_2(x,u_1u_2,u_2)\frac{1-q^{\min(k_1,k_2)}}{1-q}, \end{aligned}$$

where

$$\phi_2(x,u,v) = x\sum_{i\ge 2}\varphi_i \sum_{j_1+j_2+j_3=i-2} y(xu^2)^{j_1}y(xv^2)^{j_2}y(x)^{j_3}$$

and

$$q = \frac{\phi_1(x,u_1u_2,1)}{\phi_1(x,u_1u_2,u_2)\phi_1(x,u_2,1)}.$$

(The general case is even more involved way, compare again with [27]).

With help of Cauchy's formula it is possible to extract

$$a_{km,n} = \frac{1}{(2\pi i)^2}\int_{|x|=r_1}\int_{|u|=r_2}\frac{A_k(x,u)}{x^{n+1}u^{m+1}}\,du\,dx$$

The advantage of (36) is that one exactly knows the behaviour of the singularities of $A_k(x,u)$, and it is no doubt that the asymptotic leading term of $a_{km,n}$ comes from that part of the Cauchy integral, where x and u are close to the singularities. More precisely, one can use a *double Hankel contour* (compare with [13, 27]) locally around the singularities and gets

$$a_{km,n} = \frac{\sigma^2k^2}{\pi(m(2n-n))^{3/2}}\exp\left(-\frac{\sigma^2k^2n}{2m(2n-m)}\right) + \mathcal{O}\left(\frac{k}{n^3}\right)$$

uniformly for $\varepsilon \le \frac{m}{n} \le 2-\varepsilon$ and $k \le C\sqrt{n}$, as $n\to\infty$, where $\varepsilon>0$ and $C>0$ are arbitrary constants. Thus, if $m \sim tn$ then the distribution of $a_{km,n}/y_n$ behaves like the density of a Maxwell distribution, which is exactly the marginal distributions of the Brownian excursion. (The multi dimensional case is technically more involved (see [13, 27] but finally one gets an alternate proof of Lemma 10).

The proof of Lemma 11 is in some sense similar to that of Lemma 10. One uses explicit representation for

$$\sum_{|k_1-k_2|\ge l} B_{k_1k_2}(x,u_1,u_2)$$

in terms of $y(xu^2)$ and $y(x)$ (which is easy) and then has to estimate the coeffients of $x^nu_1^{m_1}u_2^{m_2}$. For this purpose one uses (again) Cauchy's formula, estimates the integrand absolutely and applies the following technical estimate:

Lemma 13. *Suppose that $F(z)=\sum_{n\ge 0}f_nz^n$ is analytic in the domain*

$$\Delta = \{z : |z| < 1+\kappa,\ |\arg(z-1)| > \gamma\}$$

for some $\kappa>0$ and $0<\gamma<\frac{\pi}{2}$. Suppose further that

$$|F(z)| \le \left|e^{-C\sqrt{1-z}}\right|$$

for $z\in\Delta$ and for some $C>0$, and let $f_n^{[k]}$ denote the coefficients of $F(z)^k=\sum_{n\ge0}f_n^{[k]}z^n$. Then there exists a constant $C'>0$ such that, as $n\to\infty$,

$$f_n^{[k]} = \mathcal{O}\left(\frac{1}{n}\exp\left(-C'\frac{k}{\sqrt{n}}\right)\right),$$

uniformly for all $k\ge 0$.

The proof of Lemma 13 is quite similar to that of the Tranfer Lemma 12 (for details see [27]).

3.3. **Convergence of Moments.** The methods of Lemma 10 and 11 constitute an alternate proof of Theorem 7. Of course, they need a little bit more restrictive assumption, namely (33), however, they can be used to prove convergence of moment. Especially Lemma 11 (resp. an estimate of the kind (34)) is important in this context.

Theorem 9. *Suppose that a sequence of stochastic processes $x_n(t)$ defined on $C[0,1]$ converges weakly to $x(t)$. Furthermore suppose that there exists $s_0 \in [0,1]$ such that for all $r \ge 0$*

$$\sup_{n\ge 0} \mathbf{E}\left(|x_n(s_0)|^r\right) < \infty, \tag{37}$$

and that for every $\alpha > 1$ there is $\beta > 0$ and $C > 0$ with

$$\mathbf{E}\left(|x_n(t) - x_n(s)|^\beta\right) \le C|t-s|^\alpha \quad \textit{for all } s,t \in [0,1]. \tag{38}$$

Let $F : C[0,1] \to \mathbb{R}$ be a continuous functional of polynomial growth, i.e. there exits $r \ge 0$ with

$$|F(y)| \le \|y\|_\infty^r$$

for all $y \in C[0,1]$. Then

$$\lim_{n\to\infty} \mathbf{E}\, F(x_n) = \mathbf{E}\, F(x). \tag{39}$$

For example, Theorem 9 implies that all finite dimensional moments of $x_n(t)$ converge to that of $x(t)$, i.e. for all fixed $0 \le s_1 < s_2 < \cdots < s_k \le$ and $r_1, r_2, \ldots, r_k \ge 0$

$$\lim_{n\to\infty} \mathbf{E}\left(x_n(s_1)^{r_1} x_n(s_2)^{r_2} \cdots x_n(s_k)^{r_k}\right) = \mathbf{E}\left(x(s_1)^{r_1} x(s_2)^{r_2} \cdots x(s_k)^{r_k}\right).$$

Furthermore, we have

$$\lim_{n\to\infty} \mathbf{E}\left(\max_{0\le t\le 1} x_n(t)\right)^r = \mathbf{E}\left(\max_{0\le t\le 1} x(t)\right)^r$$

and similarly for the integral $\int_0^1 x_n(t)\,dt$.

Especially we get these properties for the rescaled search depth process $e_n(t)$; we can use $s_0 = 0$ and apply (34).

The key of the proof of Theorem 9 is the following observation.

Lemma 14. *Suppose that $x_n(t)$ and $x(t)$ are stochastic processes satisfying the assumptions of Theorem 9. Then for every $\alpha > 1$ there exists a constant $K > 0$ such that for $\varepsilon > 0$ and $0 < \delta < 1$*

$$\mathbf{P}\left\{\sup_{|s-t|\le\delta} |x_n(s) - x_n(t)| \ge \varepsilon\right\} \le K\frac{\delta^{\alpha-1}}{\varepsilon^\beta} \tag{40}$$

and consequently

$$\mathbf{E}\left(\sup_{|s-t|\le\delta} |x_n(s) - x_n(t)|^r\right) = \mathcal{O}\left(\delta^{r\frac{\alpha-1}{\beta}}\right) \tag{41}$$

for every fixed $r \le \beta - 1$.

Note that if (38) holds for all $\alpha > 1$ then $\beta = \beta(\alpha)$ is unbounded as a function in α. Thus, the *restriction* $r \le \beta - 1$ is not that serious.

Proof. First (40) follows from (38) by using the methods of Billingsley [5, pp. 95]. Finally, (41) is an easy consequence of (40). □

It is now quite easy to complete the proof of Theorem 9:

Proof. By combining (37) and (38) it follows that for every $s_0 \in [0,1]$

$$\sup_{n\ge 0} \mathbf{E}\left(|x_n(s_0)|^r\right) < \infty.$$

Next, by a direct combination of (37) with $s_0 = \frac{1}{2}$ and (41) with $\delta = \frac{1}{2}$ we obtain (for all $r \ge 0$)

$$\sup_{n\ge 0} \mathbf{E}\left(\max_{0\le t\le 1} |x_n(t)|^r\right) < \infty$$

and consequently (for any $\varepsilon > 0$)

$$\sup_{n\ge 0} \mathbf{E}\, F(x_n)^{1+\varepsilon} < \infty.$$

By weak convergence we also have $F(x_n) \Longrightarrow F(x)$. Thus, by [6, p. 338] all moments of $F(x)$ exist and

$$\lim_{n\to\infty} \mathbf{E}\, F(x_n) = \mathbf{E}\, F(x).$$

□

3.4. **Height.** The distribution of the height H_n was already established in Theorem 8. With help of Lemma 11 and Theorem 9 we directly obtain convergence of moments, too:

Theorem 10. *Suppose that the offspring distribution ξ of a critial Galton-Watson branching process satisfies (33) for some $\gamma > 0$. Then, for every $r \ge 0$ we have, as $n \to \infty$,*

$$\mathbf{E}(H_n^r) \sim 2^{r/2}\sigma^{-r} r(r-1)\Gamma(r/2)\zeta(r) \cdot n^{r/2}$$

where $\zeta(s)$ denotes the Riemann Zeta-function and $(r-1)\zeta(r) = 1$ for $r = 1$.

This kind of approach is quite general, but we do not get any error term. The only method known, which provides an error term, is due to Flajolet and Odlyzko [24]. In fact, they provided the first proof of Theorem 10 (by the use of generating functions and analytic methods.

Let $y_k(x)$ denote the generating functions

$$y_k(x) = \sum_{n\ge 1} \mathbf{Pr}[|T| = n,\, h_T \le k]\, x^n,$$

where h_T denotes the height of T, then $y_0(x) = \varphi_0 x$, and recursively

$$y_{k+1}(x) = x\varphi(y_k(x)), \qquad (k \ge 0). \tag{42}$$

With help of these function one gets the generating function of the expected height:

$$H(x) := \sum_{n\ge 1} \mathbf{E}\, H_n \cdot y_n \cdot x^n = \sum_{k\ge 0} (y(x) - y_k(x)).$$

After a subtle analysis of the above recurrence (42) it is possible to derive a local representation of the form

$$H(x) = \frac{1}{\sigma^2} \log \frac{1}{1-x} + K + \mathcal{O}\left(|1-x|^{\frac14 - \kappa}\right)$$

for some constant K and every (fixed) $\kappa > 0$ (see [24]). Thus, with help of the Transfer Lemma 12 and (32) we directly get, as $n \to \infty$,

$$\mathbf{E}\, H_n = \frac{\sqrt{2\pi}}{\sigma} \cdot \sqrt{n} + \mathcal{O}\left(n^{\frac14 + \kappa}\right).$$

In a similar way one gets corresponding asymptotic equivalents for higher moments which characterize again the distribution of $n^{-1/2} H_n$ of Theorem 8 (see [24]).

3.5. **Internal Path Length.** As for the height we can apply Lemma 11 and Theorem 9 to obtain (without any further work) convergence of moments for the internal path length, too:

Theorem 11. *Suppose that the offspring distribution ξ of a critial Galton-Watson branching process satisfies (33) for some $\gamma > 0$. Then, for every $r \ge 0$ we have, as $n \to \infty$,*

$$\mathbf{E}(L_n^r) \sim K_r \frac{4\sqrt{\pi} r! 2^{r/2}}{\Gamma\left(\frac{3r-1}{2}\right) \sigma^r} \cdot n^{3r/2},$$

where K_r is recursively given by

$$K_r = \frac{3r-4}{4} K_{r-1} + \sum_{j=1}^{r-1} K_j K_{r-j}, \qquad (r \ge 2),$$

with initial values $K_0 = -\frac12$ and $K_1 = \frac18$.

Again there is no error term. However, (in contrary to the height) there is an easy method to reprove Theorem 11 even with an error term (but then the relation to Brownian excursion is completely hidden.)

The corresponding generating function $p(x,u) = \sum_{n,k} \mathbf{Pr}[|T| = n,\, l_T = k] x^n u^k$ (where l_T denote the path length of T) satisfies the functional equation

$$p(x,u) = x\varphi(p(xu,u)). \tag{43}$$

If we are interested, for example, in the expected value $\mathbf{E}\, L_n$ then we consider the partial derivative with respect to u and set $u = 1$:

$$\frac{\partial p(x,1)}{\partial u} = \sum_{n\geq 0} \mathbf{E}\, L_n \cdot y_n \cdot x^n.$$

From (43) we directly get

$$\frac{\partial p(x,1)}{\partial u} = \frac{x^2\varphi'(y(x))y'(x)}{1 - x\varphi'(y(x))}.$$

If we use also use the local expansions $y'(x) = 1/(\sqrt{2}\sigma\sqrt{1-x}) + \mathcal{O}\left(|1-x|^{-1/2}\right)$ and $x\varphi'(y(x)) = 1 - \sqrt{2}\sigma\sqrt{1-x} + \mathcal{O}\left(|1-x|\right)$ one gets

$$\frac{\partial p(x,1)}{\partial u} = \frac{1}{2\sigma^2(1-x)} + \mathcal{O}\left(|1-x|^{-1/2}\right)$$

and with help of the Transfer Lemma 12 and (32)

$$\mathbf{E}L_n = \sqrt{\frac{\pi}{2\sigma^2}} \cdot n^{3/2} + \mathcal{O}\left(n\right).$$

In the same way one can evaluate all higher moments (compare with Takács [56]) and obtain also an error term of the form $\left(1 + \mathcal{O}\left(n^{-1/2}\right)\right)$.

3.6. **Profile.** Let $(V_{k,n})_{k\geq 0}$ denote the profile of Galton-Watson trees, i.e. the branching process $(Z_k)_{k\geq 0}$ conditioned on the total progeny $\sum_k Z_k = n$. Then the following result holds.

Theorem 12. *Suppose that the offspring distribution ξ of a Galton-Watson branching process is critical, i.e. $\mathbf{E}\,\xi = 1$, and that the variance $\sigma^2 = \mathbf{Var}\,\xi$ is non-zero and finite. Then the rescaled profile process converges weakly to $\left(\frac{\sigma}{2} l(\frac{\sigma}{2}s), s \geq 0\right)$, where $(l(s), s \geq 0)$ denotes the total local time of the Brownian excursion of duration 1:*

$$\left(\frac{1}{\sqrt{n}} V_{[s\sqrt{n}],n},\ s \geq 0\right) \Longrightarrow \left(\frac{\sigma}{2} l(\frac{\sigma}{2}s), s \geq 0\right).$$

The (total) local time $l(s)$ can be defined by

$$l(s) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^1 \mathbf{1}_{[s,s+\varepsilon]}(e(t))\, dt.$$

It might be interpreted as the time a Brownian excursion stays at level s. (We can assume that $l(s)$ is continuous a.s.)

In view of Theorem 7 this result is not unexpected. However, it does not follow from it.[14] The first proof of Theorem 12 is due to Drmota and Gittenberger [18]. It uses generating functions and analytic methods — we will comment on their proof in the sequel —, however, one has to assume (again) that $\mathbf{E}\, e^{\gamma\xi}$ exists for some $\gamma > 0$. The full version of Theorem 12 was given by Pitman [46] and Kersting [33]. Partial results go back to [9, 26, 30, 38, 57].

The advantage of Drmota and Gittenberger's approach is that one gets convergence of moments, too, as already discussed for the depth search process. Thus, we get, for example, convergence of moments of the width, see the next section.

[14] In fact, Theorem 12 has been formulated as a conjecture by Aldous [2].

The method of [18] is to use the generating functions

$$y_k(x,u) = \sum_{n,m} \mathbf{Pr}[|T| = n,\ v_k = m]\, x^n u^m = \sum_{n\ge 1} \mathbf{E}\, u^{V_{k,n}} \cdot y_n \cdot x^n.$$

We have $y_0(x,u) = uy(x)$ and recursively

$$y_{k+1}(x,u) = x\varphi(y_k(x,u)), \qquad (k \ge 0). \tag{44}$$

Thus, with help of Cauchy's formula one can extract $\mathbf{E}\, u^{V_{k,n}}$ which the characteristic function of $V_{k,n}$ if $u = e^{it}$. If $u = 1$ then $y_k(x,u) = y(x)$. Thus, if u is close to 1 then $y_k(x,u)$ is close to $y(x)$. More precisely, if $|u-1| = \mathcal{O}\left(n^{-1/2}\right)$ and $|1-\sqrt{1-x}| \le 1 + \mathcal{O}\left(n^{-1/2}\right)$, then $y_k(x,u)$ can be represented as

$$y_k(x,u) = y(x) + \frac{(u-1)y(x)\alpha^k}{\frac{\sqrt{1-x}+\sigma(1-uy(x))/\sqrt{2}}{2\sqrt{1-x}} + \frac{\sqrt{1-x}-\sigma(1-uy(x))/\sqrt{2}}{2\sqrt{1-x}}\alpha^k + \mathcal{O}\left(\sqrt{|1-x|}\right)},$$

where α is an abbreviation for $\alpha = x\varphi'(y(x))$. It turns out that this representation for $y_k(x,u)$ is suitable to prove one dimensional weak convergence. The multi dimensional case is more technical but of the same flavour (for details see [18]).

If one is only interested in the expected profile $\mathbf{E}\, V_{k,n}$ then (44) gives

$$\frac{\partial y_k(x,1)}{\partial u} = \sum_{n\ge 0} \mathbf{E}\, V_{k,n} \cdot y_n \cdot x^n = y(x)(x\varphi'(y(x))^k.$$

Since $x\varphi'(y(x)) = 1 - \sqrt{2}\sigma\sqrt{1-x} + \mathcal{O}\left(|1-x|\right)$ one can use [14] and obtains

$$\mathbf{E}\, V_{k,n} = 2\sigma k e^{-(\sigma^2k^2)/(2n)} + \mathcal{O}\left(1\right)$$

uniformly for $k = \mathcal{O}\left(\sqrt{n}\right)$. (This has already been observed by Meir and Moon [42]).

In order to complete the proof of Theorem 12 one needs a tightness result for the profile.

Lemma 15. *For every integer $r \ge 1$ there exist constants $C_1, C_2 > 0$ such that*

$$\mathbf{E}\left(|V_{k,n} - V_{k+h,n}|^{2r}\right) \le C_1 e^{-C_2 k/\sqrt{n}} h^r n^{r/2}$$

for all $k, h \ge 0$.

Proof. Set $y_{k,0}(x,u,v) = vy_k(x,u)$ and recursively

$$y_{k,h+1}(x,u,v) = x\varphi(y_{k,h}(x,u,v)), \qquad (h \ge 0).$$

Then

$$H_{k,h}(x) = \left(u\frac{\partial}{\partial u}\right)^{2r} y_{k,h}(x,u,u^{-1})\Big|_{u=1}.$$

is the generating function

$$H_{k,h}(x) = \sum_{n\ge 0} \mathbf{E}\left(|V_{k,n} - V_{k+h,n}|^{2r}\right) \cdot y_n \cdot x^n.$$

It is not too difficult to show that $H_{k,h}(x)$ can be represented as

$$H_{k,h}(x) = \alpha^k \sum_{j=0}^{r} G_{j,k,h}(x) \frac{(1-\alpha^h)^j}{(1-\alpha)^{r-1+j}}, \tag{45}$$

where $\alpha = x\varphi'(y(x))$ (as above) and the functions $G_{j,k,h}(x)$ satisfy

$$\max_{x\in\Delta} |G_{j,k,h}(x)| = \mathcal{O}\left(1\right)$$

with $\Delta = \{x : |x| < 1 + \kappa,\ |\arg(z-1)| > \gamma\}$ for some $\kappa > 0$ and $0 < \gamma < \frac{\pi}{2}$. Thus, a mixture of the Transfer Lemma 12 and Lemma 13 completes the proof of Lemma 15. □

3.7. **Width.** A direct consequence of Theorem 12 is a weak convergence result for the width $W_n = \max_{k\ge 0} V_{k,n}$. Furthermore, if we apply the tightness inequality from Lemma 15 to a variant of Theorem 9 we also obtain convergence of moments (see [19]).

Theorem 13. *Suppose that the offspring distribution ξ of a critical Galton-Watson branching has finite and non-zero variance $\mathbf{Var}\,\xi = \sigma^2$. Then*

$$\frac{1}{\sqrt{n}} W_n \Longrightarrow \frac{\sigma}{2} \sup_{s\ge 0} l(s).$$

If there is an $\gamma > 0$ such that $\mathbf{E}\, e^{\gamma\xi} < \infty$ then we also have (for every $r \ge 0$)

$$\mathbf{E}(W_n^r) \sim 2^{-r/2}\sigma^r r(r-1)\Gamma(r/2)\zeta(r)\cdot n^{r/2}$$

where $\zeta(s)$ denotes the Riemann Zeta-function and $(r-1)\zeta(r) = 1$ for $r = 1$.

Note that the distribution of $\sup_{s\ge 0} l(s)$ and $\max_{0\le t\le 1} e(t)$ are almost the same:

$$\mathcal{L}\left(\sup_{s\ge 0} l(s)\right) = \mathcal{L}\left(2 \max_{0\le t\le 1} e(t)\right).$$

The width of Galton-Watson trees has attracted the interest of many authors. First, Odlyzko and Wilf [45] became interested in this tree parameter when studying the bandwidth

$$\beta(T) = \min_f \left(\max_{(u,v)\in E(T)} |f(u) - f(v)| \right)$$

of a tree T, where f is an assignment of distinct integers to the vertices of the tree. They showed for a tree with n vertices and height $h(T)$ and width $w(T)$ that

$$\frac{n-1}{2h(T)} \le \beta(T) \le 2w(T) - 1$$

and proved that there exist positive constants c_1 and c_2 such that

$$c_1\sqrt{n} < \mathbf{E} w_n < c_2\sqrt{n\log n}. \tag{46}$$

The exact order of magnitude was left as an open problem.

Marckert and Chassaing [41] used the relation of parking functions and rooted trees as well as the strong convergence theorem of Komlos, Major and Tusnady [36] to derive tight bounds for the moments of the width (even with error terms) for the special case of Cayley trees ($\mathbf{E}\, t^\xi = e^{t-1}$). Theorem 13 (which is due to Drmota and Gittenberger [19]) does not provide an error term but it applies for the general case.

4. Conclusions

There are several natural probablistic models for binary trees which appear as data structures. In this paper we have concentrated on (so-called) random binary search trees which are closely related to Quicksort and on combinatorial binary trees which can be considered as a special case of Galton-Watson trees conditioned on the total progeny. Both probabilistic models are very well discussed in the literature (most papers are from last two decades of the 20^{th} century, and some problems are still open).

The purpose of this paper was to present some very recent result on these kind of data structures. Especially we have described the internal path length, the profile and the height. We have also tried to present the major ideas of the proofs.

The internal path length of binary search trees is treated with help of a contraction mapping for distribution functions, the profile with help of a martingale of analytic functions, and the height with help of another contraction mapping and analytic methods for generating functions. We present a complete description of the distribution of the height and give a proof of Robson's conjecture (saying that the variance of the height remains bounded).

By a variation of the method for the height we also show that the distribution of the left most particle in a specific discrete branching random walk can be described with a travelling wave. Equivalently we have solved the Random Bisection Problem.

The analysis of Galton-Watson trees has a completely different flavour. Here the basis is a weak convergence result of the depth-first search to Brownian excursion (due to Aldous [1, 2, 3]). This general theorem directly implies weak convergence properties for the height and the path length (but not for the profile and width for which we need another approach). In this paper we report on an alternate proof of Aldous' result which is based on analytic methods for generating functions. This approach also allows to transfer convergence of moments of unbounded functionals (which is impossible with weak convergence). A similar approach is used for the profile and consequently for the width.

Thus, many interesting parameters of these kind of tree structures can be described in a satisfactory way although some problems are still open, for example, no distributional results for the profile and width of binary search trees are known, for the width we even do not know the exact behaviour of the expected value.

It is also a purpose of this paper to demonstrate the strength of (complex) analytic methods in the context of probabilistic limit theorems for recursive combinatorial structures (like rooted trees), especially if they are combined with probabilistic methods (e.g. convergence of martingales, tightness of sequences of stochastic processes).

References

[1] D. J. Aldous, The continuum random tree I, Ann. Prob. **19** (1991), 1–28.

[2] D. J. Aldous, The continuum random tree II: an overview, *Stochastic Analysis,* M. T. Barlow and N. H. Bingham, Eds., Cambridge University Press 1991, 23–70.

[3] D. J. Aldous, The continuum random tree III, Ann. Prob. **21** (1993), 248–289.

[4] M. Bachmann, Limit theorems for the minimal position in a branching random walk with independent logconcave displacements, Adv. Appl. Probab. **32** (2000), 159–176.

[5] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[6] P. Billingsley, *Probability and Measure, 3rd ed.* Chichester, John Wiley & Sons, 1995.

[7] M. Bramson, The convergence of solutions of the Kolmogorov nonlinear diffusion equation to travelling waves, Memoirs Am. Math. Soc. **44** (1983), 1–190.

[8] B. Chauvin, M. Drmota, and J. Jabbour-Hattab, The profile of binary search trees, Ann. Applied Probab. **11** (2001), 1042–1062.

[9] J. W. Cohen and G. Hooghiemstra, Brownian excursion, the $M/M/1$ queue and their occupation times, Mathematics of Operations Research **6** (1981), 608–629.

[10] L. Devroye, A note on the height of binary search trees, J. Assoc. Comput. Mach. **33** (1986), 489–498.

[11] L. Devroye, Branching processes in the analysis of the height of trees, Acta Inform. **24** (1987), 277–298.

[12] L. Devroye and B. Reed, On the variance of the height of random binary search trees, SIAM J. Comput. **24** (1995), 1157–1162.

[13] M. Drmota, The height distribution of leaves in rooted trees, Discrete Mathematics and Applications **4** (1994), 45–58.

[14] M. Drmota, A bivariate asymptotic expansion of coefficients of powers of generating functions, Europ. J. Combinatorics **15** (1994), 139–152.

[15] M. Drmota, An analytic approach to the height of binary search trees, Algorithmica **29** (2001), 89-119.

[16] M. Drmota, The variance of the height of binary search trees, Theoret. Comput. Sci. **270** (2002), 913–919.

[17] M. Drmota, An analytic approach to the height of binary search trees II, J. Assoc. Comput. Mach., to appear.

[18] M. Drmota and B. Gittenberger, On the profile of random trees, Random Struct. Algorithms **10** (1997), 421–451.

[19] M. Drmota and B. Gittenberger, The width of Galton-Watson trees, J. Combinat. Theory, submitted.

[20] M. Drmota and J.-F. Marckert, Reinforced weak convergence of stochastic processes, Statistics Probab. Letters, submitted.

[21] J. A. Fill and S. Janson, Smoothness and decay properties of the limiting Quicksort density function. in *Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities* (eds.: D. Gardy and A. Mokkadem), Trends in Mathematics, Birkhäuser, 2000. 53–64.

[22] J. A. Fill and S. Janson, Approximating the limiting Quicksort distribution. Random Structures Algorithms **19** (2001), 376–406.

[23] J. A. Fill and S. Janson, Quicksort asymptotics, J. Algorithms, to appear.

[24] P. Flajolet and A. M. Odlyzko, The average height of binary trees and other simple trees, J. Comput. Syst. Sci. **25** (1982), 171–213.

[25] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating functions, SIAM J. Discrete Math. **3** (1990), 216–240.
[26] R. K. Getoor and M. J. Sharpe, Excursions of Brownian motion and Bessel processes, Z. Wahrsch. verw. Gebiete **47** (1979), 83–106.
[27] B. Gittenberger , On the contour of random trees, SIAM J. Discrete Math. **12** (1999), 434–458.
[28] C. A. R. Hoare, Algorithm 64: Quicksort, Communic. Assoc. Comput. Masch. **4** (1961), 321.
[29] C. A. R. Hoare, Quicksort, Computer Journal **5** (1962), 10–15.
[30] G. Hooghiemstra, On the explicit form of the density of Brownian excursion local time, Proc. Amer. Math. Soc. **84** (1982), 127–130.
[31] J. Jabbour-Hattab, Martingales and large deviations for binary search trees, Random Struct. Algorithms **19** (2001), 112–127.
[32] D. P. Kennedy, The Galton-Watson process conditioned on the total progeny, *J. Appl. Prob. 12* (1975), 800–806.
[33] G. Kersting, On the height profile of a conditioned Galton-Watson tree, manuscript, 1998.
[34] C. Knessl and W. Szpankowski, Quicksort algorithms again revisited, Discete Math. Theor. Comput. Sci. **3** (1999), 43–64.
[35] D. E. Knuth, The Art of Computer Programming, Vol. 3, 2nd ed., Addison-Wesley, 1998.
[36] J. Komlos, P. Major, and G. Tusnady, An approximation of partial sums of independent RV's and the sample DF. II, Z. Wahrsch. Verw. Geb. **34** (1976), 33–58.
[37] P. L. Krapivsky and S. N. Majumdar, Traveling waves, front selection, and exact nontrivial exponents in a random fragmentation problem, Phys. Rev. Lett. **85** (2002), 5492–5495..
[38] G. Louchard, Kac's formula, Levy's local time and Brownian excursion, J. Appl. Prob. **21** (1984), 479–499.
[39] W. Lynch, More combinatorial problems on certain trees. The Computer J. **7** (1965), 299–302.
[40] H. M. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York, 1992.
[41] J.-F. Marckert and P. Chassaing, Parking functions, empirical processes, and the width of rooted labeled trees, Electronic J. Combinatorics **8** (2001), R14.
[42] A. Meir and J. W. Moon, On the altitude of nodes in random trees, Can. J. Math. **30** (1978), 997–1015.
[43] L. Moser and M. Wyman, Asymptotic development of the Stirling numbers of the first kind. J. London Math. Soc. **33** (1958), 133–146.
[44] J. Neveu, *Martingales à temps discret.* Masson & Cie, 1972.
[45] A. M. Odlyzko and H. Wilf, Bandwiths and profiles of trees, J. Combin. Theory, Ser. B **42** (1987), 348–370.
[46] J. Pitman, The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest, Ann. Prob. **27** (1999), 261–283.
[47] B. Pittel, On growing random binary trees, J. Math. Anal. Appl. **103** (1984), 461–480.
[48] C. Ponder, Unsolved Problems (R. Guy ed.), Amer. Math. Monthly **93** (1986), 280–281.
[49] B. Reed, The height of a random binary search tree, J. Assoc. Comput. Mach., to appear.
[50] M. Régnier, A limit distribution for Quicksort, Informatique théorique et Applications/Theoretical Informatics and Applications **23** (1989), 335–343.
[51] J. M. Robson, The height of binary search trees, Austral. Comput. J. **11** (1979), 151–153.
[52] J. M. Robson, On the concentration of the height of binary search trees. ICALP 97 Proceedings, LNCS **1256** (1997), 441–448.
[53] J. M. Robson, Constant bounds on the moments of the height of binary search trees, Theoret. Comput. Sci, to appear.
[54] U. Rösler, A limit theorem for "Quicksort", Informatique théorique et Applications/Theoretical Informatics and Applications **25** (1991), 85–100.
[55] U. Rösler and L. Rüschendorf, The contraction method for recursive algorithms, Algorithmica **29** (2001), 3–33.
[56] L. Takács, The asymptotic distribution of the total heights of random rooted trees, Acta Sci. Math. **57** (1993), 613–625.
[57] L. Takács, Brownian local times, J. Appl. Math. Stoch. Analysis **8** (1995), 209–232.
[58] K. H. Tan and P. Hadjicostas, Some properties of a limiting distribution of Quicksort, Statistics Probab. Letters **25** (1995), 87–94.