THE LEFT-RIGHT-IMBALANCE OF BINARY SEARCH TREES

MARKUS KUBA AND ALOIS PANHOLZER

Abstract. We present a detailed study of left-right-imbalance measures for random binary search trees under the random permutation model, i.e., where binary search trees are generated by random permutations of $\{1, 2, \dots, n\}$. For random binary search trees of size n we study (i) the difference between the left and the right depth of a randomly chosen node, (ii) the difference between the left and the right depth of a specified node $j = j(n)$, and (iii) the difference between the left and the right pathlength, and show for all three imbalance measures limiting distribution results.

1. Introduction

A binary tree consists of a distinguished node, the root of the tree, together with a possibly empty left subtree and a possibly empty right subtree, which are both again binary trees. Binary trees are of particular importance in many applications in computer science. The most important probability models for binary trees, i.e., when we assume that the occurrence of trees of a given shape follow a certain distribution, are the so called "random tree model" (also known as Catalan model) and the "random permutation model". Whereas in the random tree model any binary tree of size n appears with equal probability, where size is measured by the number of nodes in the tree, in the random permutation model the trees are generated by random permutations of the numbers $\{1, 2, \dots, n\}$ leading to what is called a "random binary search tree". The binary tree model turns out to be appropriate in formal language theory, computer algebra, etc., whereas the binary search tree model is of importance in sorting and searching algorithms and a lot of combinatorial algorithms. See, e.g., [15, 17, 20, 25] for a detailed description and applications of both tree models.

There are several recent papers devoted to a study of properties of the left and right length of paths in binary trees (see, e.g., [2, 14, 16, 21]), where all these analyzes are using as the underlying probability model the random tree model. In particular "local imbalance measures" as the difference between the right and the left depth of nodes in binary trees, and "global imbalance measures" as the difference between the right and the left pathlength of binary trees are studied. The depth of node v in a tree T (also called altitude or heigth of node v) is measured by the number of edges lying on the unique path from the root of T to v, where for the left (right) depth only left (right) edges are counted, and the pathlength of the tree T is the sum of the depths of all nodes $v \in T$, where again for the left (right) pathlength the left (right) depth of nodes is counted.

An analysis of the left and right length of paths in binary trees was also suggested by Donald Knuth in 2004 during the workshop on *Analysis of Algorithms* at MSRI, Berkeley, USA. Following this suggestion we present here a study of left-right-imbalance measures for the random permutation model and thus for random binary search trees. In particular we analyze in this paper the difference $D_{n,j} := A_{n,j}^{[R]} - A_{n,j}^{[L]}$ between the right depth $A_{n,j}^{[R]}$ and the left depth $A_{n,j}^{[L]}$ of the node labeled by j in a random size-n binary search tree. We also consider the (easier) question of analyzing the difference $D_n := D_{n,U_n}$ of a randomly chosen node in a random binary search tree of size n, where U_n denotes a uniformly and independent of $D_{n,j}$ on $\{1, 2, \dots, n\}$ distributed random variable (r. v.). As a result for these local imbalance parameters we obtain that, suitably normalized, D_n is asymptotically Gaussian, but more important, we get a very detailed description of the imbalance of node $j = j(n)$, possibly growing with the tree size n, namely that, after shifting with $\log j - \log(n + 1 - j)$ and scaling with

Date: October 29, 2006.
2000 *Mathematics Subject Classification.* 05C05,60F05.
Key words and phrases. Binary search trees, Imbalance, Limiting distribution.
This work was supported by the Austrian Science Foundation FWF, grant S9608-N13.

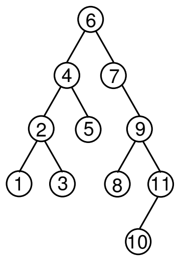

FIGURE 1. A size-11 binary search tree generated, e.g., from the permutation (6 7 9 4 2 8 11 3 10 1 5), where node 1 has a left-right-imbalance $0-3=-3$, node 5 has imbalance $1-1=0$, and node 10 has imbalance $3-1=2$. Furthermore the left-right-imbalance of a randomly chosen node is $\frac{2}{11}$ and the left-right-imbalance of the tree is $13-11=2$.

$\sqrt{\log j + \log(n+1-j)}$, the normalized r. v. $D_{n,j}$ converges in distribution to a normal distributed r. v., for all sequences $\big(j(n)\big)_{n\in\mathbb{N}}$, with $1 \le j = j(n) \le n$ and $n \to \infty$. Moreover, we study the difference $\Delta_n := R_n - L_n$ between the right and the left pathlength of a random binary search tree of size n. For this global imbalance parameter we can also characterize the limiting distribution, here by computing (asymptotically) all integer moments $\mathbb{E}(\Delta_n^r)$, for $r \ge 0$. An example illustrating the parameters studied is given as Figure 1.

Although there is a huge literature devoted to the analysis of parameters in random binary search trees (see again the references given above and in particular the following references, which show corresponding results for the "ordinary" counterparts, i.e., the depth of a randomly chosen node [7, 19], the depth of specified nodes [6, 10] and the pathlength of the trees [13, 23]), it seems that these natural imbalance questions are up to now not considered and this paper might fill this gap.

In our analysis we use the natural decomposition of a binary search tree generated by a random permutation of $\{1,2,\dots,n\}$ according to the first element k in the permutation. Of course, in a random permutation, every element $1 \le k \le n$ can occur as first element with the same probability $\frac{1}{n}$, which gives the root of the binary search tree. Moreover, the left subtree of the root k can be considered as a binary search tree generated by a random permutation of $\{1,2,\dots,k-1\}$, whereas the right subtree can be considered as a binary search tree generated by a random permutation of $\{k+1,k+2,\dots,n\}$.

Throughout this paper we denote by $X_n \xrightarrow{(d)} X$ the weak convergence, i.e., the convergence in distribution, of the sequence of random variables X_n to a random variable X and by $X \stackrel{(d)}{=} Y$ the equality in distribution of the random variables X and Y. The distribution function of the standard normal distribution $\mathcal{N}(0,1)$ is here always denoted by $\Phi(x)$. Furthermore we use the notation $H_n := \sum_{k=1}^{n}\frac{1}{k}$ for the harmonic numbers and $H_n^{(r)} := \sum_{k=1}^{n}\frac{1}{k^r}$ for the higher order harmonic numbers. Moreover $x^{\underline{k}} := x(x-1)\cdots(x-k+1)$ and $x^{\overline{k}} := x(x+1)\cdots(x+k-1)$ denote the falling and rising factorials, respectively. We also use general hypergeometric series (see, e.g., [11]), which are for m upper and n lower parametes defined by

$$ {}_mF_n\left(\begin{matrix} a_1,\dots,a_m \\ b_1,\dots,b_n \end{matrix}\,\middle|\, z\right) := \sum_{k\ge 0} \frac{a_1^{\overline{k}}\cdots a_m^{\overline{k}}}{b_1^{\overline{k}}\cdots b_n^{\overline{k}}}\frac{z^k}{k!}. $$

2. RESULTS

The first result describes the local left-right-imbalance of a randomly chosen node in a random binary search tree.

Theorem 1. *The random variable D_n, which counts the difference between the right and the left depth of a randomly chosen node in a random binary search tree of size n, is asymptotically, for $n \to \infty$,* *Gaussian distributed, where the rate of convergence is of order* $\mathcal{O}\big(\frac{1}{\sqrt{\log n}}\big)$*:*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big\{\frac{D_n}{\sqrt{2\log n}} \le x\Big\} - \Phi(x) \right| = \mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big),$$

and the expectation $\mathbb{E}(D_n)$ *and the variance* $\mathbb{V}(D_n)$ *are given by*

$$\mathbb{E}(D_n) = 0, \qquad \mathbb{V}(D_n) = 2\Big(1+\frac{1}{n}\Big)H_n - 4.$$

As a referee remarks Theorem 1 could be shown alternatively by applying probabilistic techniques, as using the theory of records in sequences of independent identically distributed random variables in combination with Lindeberg-Feller central limit theorems, which has been used to obtain limiting distribution results of the ordinary depth of a randomly chosen node in a random size-n binary search tree in [5].

The second result gives the local left-right-imbalance of specified nodes in a random binary search tree.

Theorem 2. *The centered and scaled random variable* $\tilde{D}_{n,j}$*, where* $D_{n,j}$ *counts the difference between the right and the left depth of node* j *in a random binary search tree of the elements* $\{1,2,\dots,n\}$*, is, for arbitrary sequences* $\big(j(n)\big)_{n\in\mathbb{N}}$*, with* $1 \le j = j(n) \le n$*, asymptotically, for* $n \to \infty$*, Gaussian distributed,*

$$\tilde{D}_{n,j} := \frac{D_{n,j} - \mathbb{E}(D_{n,j})}{\sqrt{\mathbb{V}(D_{n,j})}} \xrightarrow{(d)} \mathcal{N}(0,1),$$

where the expectation $\mathbb{E}(D_{n,j})$ *and the variance* $\mathbb{V}(D_{n,j})$ *are given exact and asymptotically as follows, with* $\mathcal{O}$*-bounds that hold uniformly for all* $1 \le j \le n$*:*

$$\mathbb{E}(D_{n,j}) = H_j - H_{n+1-j} = \log j - \log(n+1-j) + \mathcal{O}(1),$$

$$\begin{aligned}\mathbb{V}(D_{n,j}) &= H_j + H_{n+1-j} - H_j^{(2)} - H_{n+1-j}^{(2)} + \frac{2}{j}\big(H_j + H_{n+1-j} - H_n\big)\\ &\quad + \frac{2}{n+1-j}\big(H_j + H_{n+1-j} - H_n\big) - \frac{2}{j(n+1-j)} - 2 = \log j + \log(n+1-j) + \mathcal{O}(1).\end{aligned}$$

The third result gives an answer to the global left-right-imbalance of a random binary search tree.

Theorem 3. *The suitably scaled random variable* Δ_n*, which counts the difference between the right and the left pathlength in a random binary search tree of size* n*, converges, for* $n \to \infty$*, in distribution to a random variable* Δ*, whose distribution is fully characterized by its* r*-th moments:*

$$\frac{\Delta_n}{n} \xrightarrow{(d)} \Delta, \quad \text{with} \quad \mathbb{E}(\Delta^r) = \tilde{c}_r,$$

where the constants $\tilde{c}_r$*, with* $r \ge 0$*, are defined recursively by using the auxiliary quantities* $\tilde{d}_l$*, with* $l \ge 0$*:*

$$\tilde{c}_r = \frac{1}{r-1}\Big(2\sum_{k=0}^{r-1}\binom{r}{k}\tilde{c}_k + \sum_{l=1}^{r-1}(-1)^l \tilde{d}_l \tilde{d}_{r-l}\Big), \quad \text{for } r \ge 2 \text{ even}, \quad \tilde{c}_0 = 1, \qquad \tilde{c}_r = 0, \quad \text{for } r \text{ odd},$$

$$\tilde{d}_l = \sum_{k=0}^{l}\binom{l}{k}\tilde{c}_k, \quad \text{for } l \ge 0.$$

Furthermore the expectation $\mathbb{E}(\Delta_n)$ *and the variance* $\mathbb{V}(\Delta_n)$ *are given by*

$$\mathbb{E}(\Delta_n) = 0, \qquad \mathbb{V}(\Delta_n) = n^2 - 2(n+1)H_n + 3n.$$

We remark that Theorem 3 can also be obtained by applying the contraction method, see [22, 24] for relevant theorems. This leads to an alternative characterization of the limiting distribution Δ as the unique fixed-point with mean 0 of the fixed-point equation given below in the metric space of probability measures with finite second moment and Mallows d_2-metric:

$$X \stackrel{(d)}{=} UX + (1-U)X^* + 2U - 1,$$

where U, X and X^* are independent, X and X^* are identically distributed and U is uniformly distributed on $[0,1]$.

3. Local imbalance of a random node

We study first the left-right-imbalance $D_n := A_{n,U_n}^{[R]} - A_{n,U_n}^{[L]}$ of a randomly chosen node in a random binary search tree of size n, where the random variables $A_{n,j}^{[R]}$ and $A_{n,j}^{[L]}$ count the right and the left depth of node j in a random binary search tree of size n and U_n denotes a uniformly and independent of $A_{n,j}^{[R]}$ and $A_{n,j}^{[L]}$ on $\{1,2,\dots,n\}$ distributed random variable. Introducing the probability generating functions

$$p_n(v) := \sum_{m=-\infty}^{\infty} \mathbb{P}\{D_n = m\} v^m, \quad \text{for } n \ge 1,$$

we obtain from the decomposition of random binary search trees by conditioning on the root node k as described in Section 1 the following recurrence

$$p_n(v) = \frac{1}{n}\sum_{k=1}^{n}\Big(\frac{k-1}{n}p_{k-1}(v)\frac{1}{v} + \frac{n-k}{n}p_{n-k}(v)v + \frac{1}{n}\Big), \quad \text{for } n \ge 2, \quad p_1(v) = 1. \tag{1}$$

An explanation of the derivation of (1) is given next. We consider binary search trees generated by a random permutation of $\{1,\dots,n\}$ and condition on the root node k, i.e., we compute the probability generating function for all binary search trees whose root node has label k, and thus the root has a left subtree of size $k-1$ generated by a random permutation of $\{1,\dots,k-1\}$ and a right subtree of size $n-k$ generated by a random permutation of $\{k+1,\dots,n\}$, seperately and then we sum up all these contributions. Since the probability that the root node has label k is $\frac{1}{n}$, for $1 \le k \le n$, this explains the factor $\frac{1}{n}$ and the sum appearing in (1). It remains to explain the three summands in (1). We compute the probability generating function of a randomly chosen node in a random binary search tree whose root node is k. We select the root node k with probability $\frac{1}{n}$ and examine its left-right-imbalance. Since the left-right-imbalance of the root node is 0 this leads to the contribution $\frac{v^0}{n} = \frac{1}{n}$. The probability that we choose a node hanging on the left subtree of k and examine its left-right-imbalance is given by $\frac{k-1}{n}$. The probability generating function of the left-right-imbalance of the $k-1$ nodes within the left subtree is given by $p_{k-1}(v)$. Since the left-right-imbalance decreases by one for all nodes on the left subtree of k due to the left edge connecting k with the root node of the left subtree of k we obtain also a factor v^{-1} leading to the contribution $\frac{k-1}{n}p_{k-1}(v)\frac{1}{v}$. Analogous considerations lead to the contribution $\frac{n-k}{n}p_{n-k}(v)v$ for the nodes on the right subtree of k and complete the derivation of (1).

We treat recurrence (1) by introducing the bivariate generating function $N(z,v) := \sum_{n\ge 1} np_n(v)z^n$, which leads to the following first order linear differential equation:

$$\frac{\partial}{\partial z}N(z,v) = \Big(v + \frac{1}{v}\Big)\frac{1}{1-z}N(z,v) + \frac{1}{(1-z)^2}, \quad N(0,v) = 0.$$

The solution of this differential equation as is given next can be obtained easily by standard methods, but can be computed also by using computer algebra systems:

$$N(z,v) = \frac{1}{v+\frac{1}{v}-1}\bigg(\frac{1}{(1-z)^{v+\frac{1}{v}}} - \frac{1}{1-z}\bigg). \tag{2}$$

Extracting coefficients from $N(z,v)$ as given by (2) leads then to the following explicit formula for the probability generating function $p_n(v) = \frac{1}{n}[z^n]N(z,v)$:

$$p_n(v) = \frac{1}{n(v+\frac{1}{v}-1)}\bigg(\binom{n+v+\frac{1}{v}-1}{n} - 1\bigg). \tag{3}$$

Stirling's formula for the factorials leads then from (3) to the following asymptotic expansion of $p_n(v)$, which holds uniformly in a complex neighborhood of $v=1$, where $\epsilon > 0$ denotes an arbitrary constant:

$$p_n(v) = \frac{n^{v+\frac{1}{v}-2}}{(v+\frac{1}{v}-1)\Gamma(v+\frac{1}{v})}\Big(1 + \mathcal{O}\big(n^{-1+\epsilon}\big)\Big).$$

Thus the moment generating function $\mathbb{E}(e^{D_n s}) = p_n(e^s)$ of D_n has the following asymptotic expansion uniformly in a complex neighborhood of $s = 0$:

$$\mathbb{E}(e^{D_n s}) = \exp\Big((e^s + e^{-s} - 2)\log n - \log\big((e^s + e^{-s} - 1)\Gamma(e^s + e^{-s})\big)\Big) \cdot \big(1 + \mathcal{O}(n^{-1+\epsilon})\big). \quad (4)$$

To show a central limit theorem for D_n we can apply the so called quasi-power theorem due to Hwang [12]. It gives a powerful method not only to prove the Gaussian limit law but also to determine the rate of convergence. It is stated below.

Theorem 4. [H. K. Hwang] *Let $\{X_n\}_{n\ge 1}$ be a sequence of integral random variables. Suppose that the moment generating function satisfies the asymptotic expression*

$$M_n(s) := \mathbb{E}\left(e^{X_n s}\right) = \sum_{m\ge 0} \mathbb{P}\{X_n = m\} e^{ms} = e^{H_n(s)}\big(1 + \mathcal{O}(\kappa_n^{-1})\big),$$

the $\mathcal{O}$–term being uniform for $|s| \le \sigma$, $s \in \mathbb{C}$, $\sigma > 0$, where

(i) *$H_n(s) = U(s)\phi(n) + V(s)$, with $U(s)$ and $V(s)$ analytic for $|s| \le \sigma$ and independent of n; $U''(0) \ne 0$,*
(ii) *$\phi(n) \to \infty$,*
(iii) *$\kappa_n \to \infty$.*

Under these assumptions, the distribution of X_n is asymptotically Gaussian with the given convergence rate in the Kolmogorov metric:

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{X_n - \mathbb{E}(X_n)}{\sqrt{\mathbb{V}(X_n)}} \le x \right\} - \Phi(x) \right| = \mathcal{O}\bigg(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\phi(n)}}\bigg).$$

Moreover, the mean and the variance of X_n satisfy

$$\mathbb{E}(X_n) = U'(0)\phi(n) + V'(0) + \mathcal{O}(\kappa_n^{-1}), \qquad \mathbb{V}(X_n) = U''(0)\phi(n) + V''(0) + \mathcal{O}(\kappa_n^{-1}).$$

An application of Theorem 4 to equation (4) immediately shows the limiting distribution result given in Theorem 1, using $U(s) := e^s + e^{-s} - 2$ leading to $U'(0) = 0$ and $U''(0) = 2$. Furthermore from the explicit solution (2) we easily obtain by differentiating with respect to v once and twice and evaluating at $v = 1$ explicit formulæ for the first two moments of D_n as stated in Theorem 1. Of course one obtains $\mathbb{E}(D_n) = 0$, which is known in advance due to symmetry arguments.

4. Local imbalance of a specified node

4.1. **The probability generating function.** Next we study the left-right-imbalance $D_{n,j} := A_{n,j}^{[R]} - A_{n,j}^{[L]}$ of node j, measured by the difference between the right depth and the left depth of node j in a random binary search tree of $\{1, 2, \ldots, n\}$. Introducing for $1 \le j \le n$ the probability generating functions $p_{n,j}(v) := \sum_{m=-\infty}^{\infty} \mathbb{P}\{D_{n,j} = m\} v^m$, we obtain again from the recursive description of random binary search trees by conditioning on the root node k the following recurrence:

$$p_{n,j}(v) = \frac{1}{n}\sum_{k=1}^{j-1} p_{n-k,j-k}(v)v + \frac{1}{n}\sum_{k=j+1}^{n} p_{k-1,j}(v)\frac{1}{v} + \frac{1}{n}, \quad \text{for } n \ge 1,\ \ 1 \le j \le n. \quad (5)$$

To treat recurrence (5) we introduce the trivariate generating function
$N(z, u, v) := \sum_{n\ge 1}\sum_{1\le j\le n} p_{n,j}(v) z^n u^j$, which satisfies the following linear differential equation:

$$\frac{\partial}{\partial z} N(z, u, v) = \frac{1}{v}\frac{1}{1-z} N(z, u, v) + v\frac{u}{1-zu} N(z, u, v) + \frac{u}{(1-z)(1-zu)}, \quad N(0, u, v) = 0.$$

Standard techniques lead to the following solution of this differential equation:

$$N(z, u, v) = \frac{u}{(1-z)^{\frac{1}{v}}(1-zu)^v} \int_0^z (1-t)^{\frac{1}{v}-1}(1-tu)^{v-1} dt. \quad (6)$$

Extracting coefficients from (6) gives then the following explicit formula for the probability generating function $p_{n,j}(v) = [z^n u^j] N(z,u,v)$:

$$\begin{aligned} p_{n,j}(v) &= [z^{n-j+1}(uz)^{j-1}] \frac{1}{(1-z)^{\frac{1}{v}}(1-zu)^v} \int_0^z (1-t)^{\frac{1}{v}-1}(1-tu)^{v-1} dt \\ &= \sum_{k=0}^{n-j+1} \sum_{l=0}^{j-1} \binom{n-j+1-k+\frac{1}{v}-1}{n-j+1-k} \binom{j-1-l+v-1}{j-1-l} [z^k (uz)^l] \int_0^z (1-t)^{\frac{1}{v}-1}(1-tu)^{v-1} dt \\ &= \sum_{k=1}^{n-j+1} \sum_{l=0}^{j-1} \frac{1}{k+l} \binom{n-j+1-k+\frac{1}{v}-1}{n-j+1-k} \binom{j-1-l+v-1}{j-1-l} \binom{k-1-\frac{1}{v}}{k-1} \binom{l-v}{l}. \end{aligned} \tag{7}$$

The double sum appearing in (7) can be reduced to a single sum by using the identity

$$\sum_{k=1}^{n-j+1} \frac{1}{k+l} \binom{n-j+1-k+\frac{1}{v}-1}{n-j+1-k} \binom{k-1-\frac{1}{v}}{k-1} = \frac{v\binom{n-j+1+l+\frac{1}{v}-1}{n-j+1+l}}{\binom{l+\frac{1}{v}}{l}},$$

which can be shown by an application of the Pfaff-Saalschütz identity for hypergeometric functions (see, e.g., [11]):

$${}_3F_2\left(\begin{matrix} a, b, -n \\ c, a+b-c-n+1 \end{matrix} \middle| 1 \right) = \frac{(a-c)^{\underline{n}}(b-c)^{\underline{n}}}{(-c)^{\underline{n}}(a+b-c)^{\underline{n}}}, \quad \text{for an integer } n \geq 0.$$

Thus (7) leads to the following formula of the probability generating function $p_{n,j}(v)$, for $1 \leq j \leq n$ and $n \geq 1$, which is the starting point of our considerations leading to the limiting distribution result:

$$p_{n,j}(v) = v \sum_{l=0}^{j-1} \frac{\binom{j-1-l+v-1}{j-1-l}\binom{n-j+1+l+\frac{1}{v}-1}{n-j+1+l}\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}}. \tag{8}$$

4.2. **Expectation and variance.** From the explicit formula of $N(z,u,v)$ as given by (6) it is an easy, but a bit computational, task to obtain explicit expressions for the first moments. Nevertheless we give here the results, since it turns out that they are essential in our proof of the central limit theorem, because these explicit formulæ immediately lead to uniform estimates of $\mathbb{E}(D_{n,j})$ and $\mathbb{V}(D_{n,j})$ as required. We obtain by differentiating $N(z,u,v)$ with respect to v and evaluating at $v=1$ the expressions

$$\begin{aligned} \frac{\partial}{\partial v} N(z,u,v)\Big|_{v=1} &= \frac{1}{(1-z)(1-uz)} \log \frac{1}{1-uz} - \frac{u}{(1-z)(1-uz)} \log \frac{1}{1-z}, \\ \frac{\partial^2}{\partial v^2} N(z,u,v)\Big|_{v=1} &= \frac{2zu}{(1-z)(1-uz)} + \frac{2zu}{(1-z)(1-uz)} \log \frac{1}{1-z} - \frac{2}{1-z} \log \frac{1}{1-uz} \\ &\quad + \frac{u}{(1-z)(1-uz)} \log^2 \frac{1}{1-z} + \frac{1}{(1-z)(1-uz)} \log^2 \frac{1}{1-uz} \\ &\quad - \frac{2u}{(1-z)(1-uz)} \log \frac{1}{1-z} \log \frac{1}{1-uz} - \frac{2}{1-z} \log \frac{1}{1-z} \log \frac{1}{1-uz} \\ &\quad - \frac{2u}{(1-z)(1-uz)} \int_0^z \log \frac{1}{1-t} \log \frac{1}{1-ut} dt. \end{aligned}$$

Extracting coefficients requires, in particular to obtain the variance, again a bit "computational effort", but eventually leads, for $1 \leq j \leq n$, due to $\mathbb{E}(D_{n,j}) = [z^n u^j] \frac{\partial}{\partial v} N(z,u,v)\big|_{v=1}$ and $\mathbb{V}(D_{n,j}) = [z^n u^j] \frac{\partial^2}{\partial v^2} N(z,u,v)\big|_{v=1} + \mathbb{E}(D_{n,j}) - \big(\mathbb{E}(D_{n,j})\big)^2$ to the explicit formulæ stated in Theorem 2. By using elementary estimates for the harmonic numbers, H_n and $H_n^{(2)}$, one also obtains the following estimates, which hold uniformly for all $1 \leq j \leq n$ and $n \geq 1$:

$$\big|\mathbb{E}(D_{n,j}) - \big(\log j - \log(n+1-j)\big)\big| \leq 1, \qquad \big|\mathbb{V}(D_{n,j}) - \big(\log j + \log(n+1-j)\big)\big| \leq 10. \tag{9}$$

4.3. **Asymptotic expansion of the moment generating function.** The aim of this subsection is to show that the moment generating function $M_{n,j}(s) := \mathbb{E}(e^{\tilde{D}_{n,j}s})$, with $\tilde{D}_{n,j} := \frac{D_{n,j}-\mu_{n,j}}{\sigma_{n,j}}$, where we use the abbreviations $\mu_{n,j} := \mathbb{E}(D_{n,j})$ and $\sigma_{n,j} := \sqrt{\mathbb{V}(D_{n,j})}$, converges for all sequences $\big(j(n)\big)_{n\in\mathbb{N}}$, with $1 \le j \le n$, pointwise for all real s in a neighborhood of $s = 0$ to the moment generating function $e^{\frac{s^2}{2}}$ of the standard normal distribution $\mathcal{N}(0,1)$. Together with Curtiss' theorem [4] this will show the central limit theorem of the r. v. $D_{n,j}$ as given by Theorem 2. To to this we will examine in detail the probability generating function $p_{n,j}(v)$ as given by (8), since

$$M_{n,j}(s) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}s} p_{n,j}(e^{\frac{s}{\sigma_{n,j}}}).$$

It is appropriate to the problem to consider several regions of $j = j(n)$ separately. We mention first that it suffices to consider only the region $1 \le j \le \lceil \frac{n}{2} \rceil$, since the region $\lceil \frac{n}{2} \rceil < j \le n$ can be treated easily by using symmetry arguments as shown at the end of this subsection. We distinguish now between the regions "j small": $1 \le j < \log n$, and "j large": $\log n \le j \le \lceil \frac{n}{2} \rceil$.

• *j large:* First we consider the probability generating function $p_{n,j}(v)$, for j large, $\log n \le j \le \lceil \frac{n}{2} \rceil$, in a real neighborhood of $v = 1$: $|v-1| \le \eta$, with a "sufficiently small" $\eta > 0$, e.g., all computations are valid for $\eta \le \frac{1}{12}$.

To do this we split the sum appearing in (8) into two parts, namely into the regions $l \ge j^{\frac{2}{3}}$ and $1 \le l < j^{\frac{2}{3}}$. For $l \ge j^{\frac{2}{3}}$ we obtain due to standard estimates of the factorials (or the Γ-function), see, e.g., [1], the following estimate, where K denotes some constant, which can be different at every appearance:

$$\begin{aligned}
&\left| v \sum_{l=j^{\frac{2}{3}}}^{j-1} \frac{\binom{j-1-l+v-1}{j-1-l}\binom{n-j+1+l+\frac{1}{v}-1}{n-j+1+l}\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} \right| \le K(n-j+1)^{\frac{1}{v}-1} \sum_{l=j^{\frac{2}{3}}}^{j-1} \left| \frac{\binom{j-1-l+v-1}{j-1-l}\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} \right| \\
&\le K j^{\eta} (n-j+1)^{\frac{1}{v}-1} \sum_{l=j^{\frac{2}{3}}}^{j-1} \frac{1}{l^{2-2\eta}} \le K j^{\eta}(n-j+1)^{\frac{1}{v}-1} \frac{1}{j^{\frac{2}{3}(1-2\eta)}} \le K j^{v-1}(n-j+1)^{\frac{1}{v}-1}\frac{1}{j^{\frac{2}{3}-4\eta}} \\
&= j^{v-1}(n-j+1)^{\frac{1}{v}-1} \cdot \mathcal{O}\big(j^{-\frac{1}{3}}\big). \qquad (10)
\end{aligned}$$

Now we examine the region $0 \le l < j^{\frac{2}{3}}$, which gives the main contribution:

$$\begin{aligned}
&v \sum_{l=0}^{j^{\frac{2}{3}}} \frac{\binom{j-1-l+v-1}{j-1-l}\binom{n-j+1+l+\frac{1}{v}-1}{n-j+1+l}\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} \\
&\quad = \frac{v}{(v-1)!(\frac{1}{v}-1)!} \sum_{l=0}^{j^{\frac{2}{3}}} j^{v-1}\Big(1 - \frac{1+l}{j}\Big)^{v-1}(n-j+1)^{\frac{1}{v}-1}\Big(1 + \frac{l}{n-j+1}\Big)^{\frac{1}{v}-1} \frac{\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} \\
&\quad = \frac{v}{(v-1)!(\frac{1}{v}-1)!} j^{v-1}(n-j+1)^{\frac{1}{v}-1}\big(1 + \mathcal{O}(j^{-\frac{1}{3}})\big) \sum_{l=0}^{j^{\frac{2}{3}}} \frac{\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} \\
&\quad = \frac{v}{(v-1)!(\frac{1}{v}-1)!(1-v+v^2)} j^{v-1}(n-j+1)^{\frac{1}{v}-1}\big(1 + \mathcal{O}(j^{-\frac{1}{3}})\big). \qquad (11)
\end{aligned}$$

We used there the following estimate, where the sum appearing can be evaluated by applying Gauss-hypergeometric identity ${}_2F_1\left(\begin{smallmatrix} a,b \\ c \end{smallmatrix} \middle| 1\right) = \frac{\Gamma(c-a-b)\Gamma(c)}{\Gamma(c-a)\Gamma(c-b)}$, for $\Re c > \Re a + \Re b$:

$$\sum_{l=0}^{j^{\frac{2}{3}}} \frac{\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} = \sum_{l=0}^{\infty} \frac{\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} - \sum_{l=j^{\frac{2}{3}}+1}^{\infty} \frac{\binom{l-v}{l}}{\binom{l+\frac{1}{v}}{l}} = \frac{1}{1-v+v^2} + \mathcal{O}\Big(\frac{1}{j^{\frac{2}{3}-3\eta}}\Big).$$

Combining equations (10) and (11) we obtain, for $\log n \le j \le \lceil \frac{n}{2} \rceil$, uniformly for real v, with $|v-1| \le \eta$ and $\eta \le \frac{1}{12}$, the expansion:

$$p_{n,j}(v) = \frac{v}{(v-1)!(\frac{1}{v}-1)!(1-v+v^2)} j^{v-1}(n-j+1)^{\frac{1}{v}-1}\big(1+\mathcal{O}(j^{-\frac{1}{3}})\big). \tag{12}$$

Expansion (12) together with the estimates (9) for the first moments of $D_{n,j}$ leads for $\log n \le j \le \lceil \frac{n}{2} \rceil$ to the following expansion of the moment generating function $M_{n,j}(s)$, which holds for real s fixed:

$$\begin{aligned} M_{n,j}(s) &= e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}s}\Big(1+\mathcal{O}\Big(\frac{s}{\sigma_{n,j}}\Big)\Big) e^{\big(\frac{s}{\sigma_{n,j}}+\frac{s^2}{2\sigma_{n,j}^2}+\mathcal{O}\big(\frac{s^3}{\sigma_{n,j}^3}\big)\big)\log j} e^{\big(-\frac{s}{\sigma_{n,j}}+\frac{s^2}{2\sigma_{n,j}^2}+\mathcal{O}\big(\frac{s^3}{\sigma_{n,j}^3}\big)\big)\log(n-j+1)} \times \\ &\quad \times \big(1+\mathcal{O}\big((\log n)^{-\frac{1}{3}}\big)\big) \\ &= e^{(\log j - \log(n-j+1)-\mu_{n,j})\frac{s}{\sigma_{n,j}}} e^{\frac{\log j + \log(n+1-j)}{\sigma_{n,j}^2}\frac{s^2}{2}}\big(1+\mathcal{O}\big((\log n)^{-\frac{1}{3}}\big)\big) \\ &= e^{\frac{s^2}{2}}\big(1+\mathcal{O}\big((\log n)^{-\frac{1}{3}}\big)\big). \end{aligned} \tag{13}$$

• *j small:* Next we consider the probability generating function $p_{n,j}(v)$, for j small, $1 \le j < \log n$, where we will obtain suitable asymptotic expansions of $p_{n,j}(e^{\frac{s}{\sigma_{n,j}}})$, for real s fixed, by using Taylor series expansions of the factorials. The following expansions required are not hard to show and thus the computations are omitted:

$$\begin{aligned} (j-1-l+e^{\frac{s}{\sigma_{n,j}}}-1)! &= (j-1-l)!\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big), \\ \binom{j-1-l+e^{\frac{s}{\sigma_{n,j}}}-1}{j-1-l} &= \frac{1}{(e^{\frac{s}{\sigma_{n,j}}}-1)!}\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big), \\ \binom{n-j+1+l+e^{-\frac{s}{\sigma_{n,j}}}-1}{n-j+1+l} &= \frac{1}{(e^{-\frac{s}{\sigma_{n,j}}}-1)!}(n-j+1+l)^{\big(e^{-\frac{s}{\sigma_{n,j}}}-1\big)}\big(1+\mathcal{O}(n^{-1})\big) \\ &= \frac{1}{(e^{-\frac{s}{\sigma_{n,j}}}-1)!}(n-j+1)^{\big(e^{-\frac{s}{\sigma_{n,j}}}-1\big)}\Big(1+\mathcal{O}\Big(\frac{\log n}{n}\Big)\Big), \\ \binom{l+e^{-\frac{s}{\sigma_{n,j}}}}{l} &= \frac{1}{(e^{-\frac{s}{\sigma_{n,j}}})!}(l+1)\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big), \\ \binom{l-e^{\frac{s}{\sigma_{n,j}}}}{l} &= \frac{1}{(-e^{\frac{s}{\sigma_{n,j}}})!}\frac{1}{l}\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big), \quad \text{for } l \ge 1. \end{aligned}$$

Using these expansions the summands of $p_{n,j}(e^{\frac{s}{\sigma_{n,j}}})$, with $1 \le l \le j-1$, where $p_{n,j}(v)$ is given by (8), can be estimated as follows, where we additionally use the estimate $\frac{1}{\big(-e^{\frac{s}{\sigma_{n,j}}}\big)!} = \mathcal{O}(\frac{1}{\sqrt{\log n}})$:

$$\begin{aligned} & e^{\frac{s}{\sigma_{n,j}}}\sum_{l=1}^{j-1}\frac{\binom{j-1-l+e^{\frac{s}{\sigma_{n,j}}}-1}{j-1-l}\binom{n-j+1+l+e^{-\frac{s}{\sigma_{n,j}}}-1}{n-j+1+l}\binom{l-e^{\frac{s}{\sigma_{n,j}}}}{l}}{\binom{l+e^{-\frac{s}{\sigma_{n,j}}}}{l}} \\ &= \frac{(e^{-\frac{s}{\sigma_{n,j}}})!(n-j+1)^{(e^{-\frac{s}{\sigma_{n,j}}}-1)}\big(1+\mathcal{O}\big(\frac{\log\log n}{\sqrt{\log n}}\big)\big)}{(e^{\frac{s}{\sigma_{n,j}}}-1)!(e^{-\frac{s}{\sigma_{n,j}}}-1)!(-e^{\frac{s}{\sigma_{n,j}}})!}\sum_{l=1}^{j-1}\frac{1}{l(l+1)} \\ &= (n-j+1)^{(e^{-\frac{s}{\sigma_{n,j}}}-1)}\cdot\mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big). \end{aligned} \tag{14}$$

Thus we will see that the first summand $l=0$ of (8) gives the main contribution in the expansion of $p_{n,j}(e^{\frac{s}{\sigma_{n,j}}})$:

$$e^{\frac{s}{\sigma_{n,j}}}\binom{j-1+e^{\frac{s}{\sigma_{n,j}}}-1}{j-1}\binom{n-j+1+e^{-\frac{s}{\sigma_{n,j}}}-1}{n-j+1} = (n-j+1)^{\big(e^{-\frac{s}{\sigma_{n,j}}}-1\big)}\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big). \tag{15}$$

Combining equations (14) and (15) we obtain, for $1 \le j < \log n$ and real s fixed, the expansion:

$$p_{n,j}(e^{\frac{s}{\sigma_{n,j}}}) = (n-j+1)^{(e^{-\frac{s}{\sigma_{n,j}}}-1)}\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big). \tag{16}$$

Therefore expansion (16) together with the estimates (9) for the first moments of $D_{n,j}$ leads for $1 \le j < \log n$ to the following expansion of the moment generating function $M_{n,j}(s)$, which holds for real s fixed:

$$M_{n,j}(s) = e^{-(\mu_{n,j}+\log(n-j+1))\frac{s}{\sigma_{n,j}}} e^{\frac{\log(n-j+1)}{\sigma_{n,j}^2}\frac{s^2}{2}}\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big) = e^{\frac{s^2}{2}}\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\sqrt{\log n}}\Big)\Big). \tag{17}$$

• *j in the whole region:* Equations (13) and (17) show for $1 \le j \le \lceil \frac{n}{2} \rceil$ that $M_{n,j}(s)$ converges, for $n \to \infty$, pointwise to $e^{\frac{s^2}{2}}$ for every $s \in \mathbb{R}$:

$$M_{n,j}(s) \to e^{\frac{s^2}{2}}. \tag{18}$$

To treat the region $\lceil \frac{n}{2} \rceil < j \le n$ also we use the relation $p_{n,n+1-j}(v) = p_{n,j}(\frac{1}{v})$, which is obvious due to symmetry arguments. This gives $M_{n,n+1-j}(s) = M_{n,j}(-s)$ and thus for $1 \le j \le \lceil \frac{n}{2} \rceil$ and $n \to \infty$:

$$M_{n,n+1-j}(s) = M_{n,j}(-s) \to e^{\frac{s^2}{2}}. \tag{19}$$

Combining (18) and (19) we have shown that the moment generating function $M_{n,j}(s)$ of $\tilde{D}_{n,j} := \frac{D_{n,j}-\mathbb{E}(D_{n,j})}{\sqrt{\mathbb{V}(D_{n,j})}}$ converges pointwise for all $s \in \mathbb{R}$ to the moment generating function of the standard normal distribution $\mathcal{N}(0,1)$, which shows Theorem 2.

5. Global imbalance of the tree

5.1. **Generating functions for the moments.** Now we study the left-right-imbalance $\Delta_n := R_n - L_n$ of a random binary search tree of size n, which is counted by the difference between the right and the left pathlength. Again we introduce for $n \ge 1$ the probability generating functions $p_n(v) := \sum_{m=-\infty}^{\infty} \mathbb{P}\{\Delta_n = m\} v^m$ and use the decomposition of random binary search trees according to the root node k to obtain the following recurrence

$$p_n(v) = \frac{1}{n}\sum_{k=1}^{n} p_{k-1}(v)p_{n-k}(v)v^{n-2k+1}, \quad \text{for } n \ge 2, \tag{20}$$

with initial value $p_1(v) = 1$. It is advantageous to define additionally $p_0(v) := 1$. To treat recurrence (20) we first introduce the bivariate generating function $F(z,v) := \sum_{n\ge0} p_n(v) z^n$, which leads to the following functional-differential equation for $F(z,v)$:

$$\frac{\partial}{\partial z}F(z,v) = F(\frac{z}{v},v)F(zv,v), \quad F(0,v) = 1.$$

However, it turns out to be advantageous for a moment's study to introduce the generating function $\tilde{F}(z,s) := F(z,e^s)$ and consider the following functional-differential equation:

$$\frac{\partial}{\partial z}\tilde{F}(z,s) = \tilde{F}(ze^s,s)\tilde{F}(ze^{-s},s), \quad \tilde{F}(0,s) = 1. \tag{21}$$

Of course it holds $\tilde{F}(z,0) = \frac{1}{1-z}$.

For a study of the moments $\mathbb{E}(\Delta_n^r)$ we introduce generating functions $M_r(z)$, for $r \ge 0$, and auxiliary functions $N_{l,i}(z)$, $\bar{N}_{l,i}(z)$, for $l,i \ge 0$, as follows, where we use the differential operators with respect to s and z, D_s and D_z, the operator N_s that evaluates at $s=0$, and the operators R_z and $\bar{R}_z$, which substitute z by ze^s and ze^{-s} respectively, $R_z f(z) = f(ze^s)$ and $\bar{R}_z f(z) = f(ze^{-s})$:

$$M_r(z) := N_s D_s^r \tilde{F}(z,s), \quad N_{l,i}(z) := N_s D_s^l R_z D_s^i \tilde{F}(z,s), \quad \bar{N}_{l,i}(z) := N_s D_s^l \bar{R}_z D_s^i \tilde{F}(z,s). \tag{22}$$

Obviously it holds $M_0(z) = \frac{1}{1-z}$ and $N_{0,i}(z) = \bar{N}_{0,i}(z) = M_i(z)$, for $i \ge 0$. We further require the following commutation rules between the operators R_z and $\bar{R}_z$ with D_s and D_z:

$$D_s R_z = R_z D_s + z D_z R_z, \qquad D_s \bar{R}_z = \bar{R}_z D_s - z D_z \bar{R}_z. \tag{23}$$

Using the commutation rules (23) we obtain for the auxiliary functions $N_{l,i}(z)$ and $\bar{N}_{l,i}(z)$, for $l, i \ge 0$, the following recursive description:

$$\begin{aligned} N_{l,i}(z) &= N_{0,i+l}(z) + z\sum_{p=1}^{l} N'_{l-p,i+p-1}(z) = M_{l+i}(z) + z\sum_{p=1}^{l} N'_{l-p,i+p-1}(z), \\ \bar{N}_{l,i}(z) &= \bar{N}_{0,i+l}(z) - z\sum_{p=1}^{l} \bar{N}'_{l-p,i+p-1}(z) = M_{l+i}(z) - z\sum_{p=1}^{l} \bar{N}'_{l-p,i+p-1}(z). \end{aligned} \tag{24}$$

Applying the operator $N_s D_s^r$, with $r \ge 1$, to the left and right side of equation (21) we obtain

$$\begin{aligned} M'_r(z) &= \sum_{l=0}^{r} \binom{r}{l} \big(N_s D_s^l \tilde{F}(ze^s, s)\big)\big(N_s D_s^{r-l} \tilde{F}(ze^{-s}, s)\big) \\ &= \frac{1}{1-z} N_{r,0}(z) + \frac{1}{1-z}\bar{N}_{r,0}(z) + \sum_{l=1}^{r-1}\binom{r}{l} N_{l,0}(z)\bar{N}_{r-l,0}(z) \\ &= \frac{1}{1-z}\Big(M_r(z) + z\sum_{p=1}^{r} N'_{r-p,p-1}(z) + M_r(z) - z\sum_{p=1}^{r}\bar{N}'_{r-p,p-1}(z)\Big) + \sum_{l=1}^{r-1}\binom{r}{l} N_{l,0}(z)\bar{N}_{r-l,0}(z), \end{aligned}$$

and thus for $r \ge 1$ the following linear differential equation with initial value $M_r(0) = 0$:

$$M'_r(z) = \frac{2}{1-z} M_r(z) + \frac{z}{1-z}\sum_{p=1}^{r}\big(N'_{r-p,p-1}(z) - \bar{N}'_{r-p,p-1}(z)\big) + \sum_{l=1}^{r-1}\binom{r}{l} N_{l,0}(z)\bar{N}_{r-l,0}(z). \tag{25}$$

Equation (25) is solved easily and we obtain for $r \ge 1$ the following solution of $M_r(z)$, which requires of course the functions $M_k(z)$, with $0 \le k < r$, and $N_{l,i}(z)$, $\bar{N}_{l,i}(z)$, with $l + i < r$:

$$M_r(z) = \frac{1}{(1-z)^2}\int_0^z \Big(t(1-t)\sum_{p=1}^{r}\big(N'_{r-p,p-1}(t) - \bar{N}'_{r-p,p-1}(t)\big) + (1-t)^2\sum_{l=1}^{r-1}\binom{r}{l} N_{l,0}(t)\bar{N}_{r-l,0}(t)\Big)dt. \tag{26}$$

Thus by combining (24) and (26) and using $M_0(z) = N_{0,0}(z) = \bar{N}_{0,0}(z) = \frac{1}{1-z}$ the generating functions $M_r(z)$ are fully described. Furthermore it is an easy task to compute

$$M_1(z) = 0, \qquad M_2(z) = \frac{2}{(1-z)^3} - \frac{2\log\frac{1}{1-z}}{(1-z)^2} - \frac{2}{(1-z)^2},$$

which leads to the explicit formulæ for the expectation $\mathbb{E}(\Delta_n)$ and the variance $\mathbb{V}(\Delta_n)$ as given in Theorem 3.

5.2. **Asymptotic expansions of the generating functions.** We want to show the following local expansions around the unique dominant singularity $z = 1$ of the functions $M_r(z)$, $N_{l,i}(z)$ and $\bar{N}_{l,i}(z)$, for $r, l, i \ge 0$ (from the recursive description of these functions it follows immediately that $z = 1$ is indeed the only dominant singularity, i.e, the only singularity of smallest modulus):

$$M_r(z) \sim \frac{c_r}{(1-z)^{r+1}}, \quad N_{l,i}(z) \sim \frac{d_{l,i}}{(1-z)^{l+i+1}}, \quad \bar{N}_{l,i}(z) \sim \frac{\bar{d}_{l,i}}{(1-z)^{l+i+1}}, \tag{27}$$

with certain constants c_r, $d_{l,i}$ and $\bar{d}_{l,i}$ that will be specified later. We will show this by induction on r and $l + i$, respectively. Strictly speaking we can show slightly more, namely that in all cases the remainder term can also be quantified: $M_r(z) = \frac{c_r}{(1-z)^{r+1}}\big(1 + \mathcal{O}\big((1-z)^{-1+\epsilon}\big)\big)$, with $\epsilon > 0$; analogous for $N_{l,i}(z)$ and $\bar{N}_{l,i}(z)$. This is remarked here explicitly, since we use so called singular integration and differentiation theorems as described in [8], which are $\mathcal{O}$-transfers, not o-transfers. For the sake of brevity we will use the $\sim$-notation, but we want to point out again that the O-term for the remainder can be quantified, which is necessary to apply the theorems of [8].

For $r = 0$ and $l + i = 0$ the asymptotic expansions (27) trivially holds due to $M_0(z) = N_{0,0}(z) = \bar{N}_{0,0}(z) = \frac{1}{1-z}$ leading to $c_0 = d_{0,0} = \bar{d}_{0,0} = 1$. Furthermore $M_1(z) = 0$ as already mentioned in Subsection 5.1 leading to $c_1 = 0$.

Now we assume that for a given $r \geq 2$ the expansions (27) hold for $M_k(z)$, with $0 \leq k < r$, and $N_{l,i}(z)$, $\bar{N}_{l,i}(z)$, with $0 \leq l+i < r$. To show that (27) also holds for $M_r(z)$ we use theorems for singular differentiation [8] and obtain the following local expansions in a neighborhood of $t = 1$:

$$N'_{r-p,p-1}(t) \sim \frac{r d_{r-p,p-1}}{(1-t)^{r+1}}, \quad \bar{N}'_{r-p,p-1}(t) \sim \frac{r \bar{d}_{r-p,p-1}}{(1-t)^{r+1}}, \quad N_{l,0}(t)\bar{N}_{r-l,0}(t) \sim \frac{d_{l,0}\bar{d}_{r-l,0}}{(1-t)^{r+2}},$$

and thus the following local expansion of the integrand appearing in (26):

$$t(1-t)\sum_{p=1}^{r}\big(N'_{r-p,p-1}(t) - \bar{N}'_{r-p,p-1}(t)\big) + (1-t)^2 \sum_{l=1}^{r-1}\binom{r}{l} N_{l,0}(t)\bar{N}_{r-l,0}(t)$$
$$\sim \frac{1}{(1-t)^r}\Big(r\sum_{p=1}^{r}(d_{r-p,p-1} - \bar{d}_{r-p,p-1}) + \sum_{l=1}^{r-1}\binom{r}{l} d_{l,0}\bar{d}_{r-l,0}\Big).$$

Now singular integration [8] leads to the following local expansion of $M_r(z)$ in a neighborhood of the dominant singularity $z = 1$:

$$M_r(z) \sim \frac{1}{(1-z)^{r+1}}\frac{1}{r-1}\Big(r\sum_{p=1}^{r}(d_{r-p,p-1} - \bar{d}_{r-p,p-1}) + \sum_{l=1}^{r-1}\binom{r}{l} d_{l,0}\bar{d}_{r-l,0}\Big), \quad \text{for } r \geq 2. \tag{28}$$

Thus (28) shows that expansion (27) also holds for $M_r(z)$ and furthermore we obtain the following recurrence for the coefficients c_r, using the auxiliary quantities $d_{l,i}$ and $\bar{d}_{l,i}$, with initial values $c_0 = 1$ and $c_1 = 0$:

$$c_r = \frac{1}{r-1}\Big(r\sum_{p=1}^{r}(d_{r-p,p-1} - \bar{d}_{r-p,p-1}) + \sum_{l=1}^{r-1}\binom{r}{l} d_{l,0}\bar{d}_{r-l,0}\Big), \quad r \geq 2. \tag{29}$$

Now we assume that the asymptotic expansion (27) holds for all $M_k(z)$, with $0 \leq k \leq r$, and $N_{l,i}(z)$, $\bar{N}_{l,i}(z)$, with $0 \leq l+i < r$ and $r \geq 1$. It easily follows that (27) also holds for $N_{l,i}(z)$ and $\bar{N}_{l,i}(z)$ with $l + i = r$:

$$N_{l,i}(z) = M_{l+i}(z) + z\sum_{p=1}^{l} N'_{l-p,i+p-1}(z) \sim \frac{1}{(1-z)^{i+l+1}}\Big(c_{i+l} + (l+i)\sum_{p=1}^{l} d_{l-p,i+p-1}\Big), \tag{30}$$

$$\bar{N}_{l,i}(z) = M_{l+i}(z) - z\sum_{p=1}^{l} \bar{N}'_{l-p,i+p-1}(z) \sim \frac{1}{(1-z)^{i+l+1}}\Big(c_{i+l} - (l+i)\sum_{p=1}^{l} \bar{d}_{l-p,i+p-1}\Big). \tag{31}$$

Furthermore equations (30) and (31) lead to the following recurrence for the coefficients $d_{l,i}$ and $\bar{d}_{l,i}$, valid for all $l, i \geq 0$:

$$d_{l,i} = c_{l+i} + (l+i)\sum_{p=1}^{l} d_{l-p,i+p-1}, \quad \bar{d}_{l,i} = c_{l+i} - (l+i)\sum_{p=1}^{l} \bar{d}_{l-p,i+p-1}. \tag{32}$$

5.3. **Characterization of the limiting distribution.** Using singularity analysis of generating functions [9] we immediately obtain from (27) asymptotic equivalents of the coefficients of the functions $M_r(z)$, and thus in particular asymptotic equivalents of the moments $\mathbb{E}(\Delta_n^r)$, and the auxiliary functions $N_{l,i}(z)$, $\bar{N}_{l,i}(z)$, where it is advantageous to introduce the numbers

$$\tilde{c}_r := \frac{c_r}{r!}, \quad \tilde{d}_{l,i} := \frac{d_{l,i}}{(l+i)!}, \quad \tilde{d}'_{l,i} := \frac{\bar{d}_{l,i}}{(l+i)!}.$$

We get then for $r, l, i \geq 0$:

$$\mathbb{E}(\Delta_n^r) = [z^n]M_r(z) \sim \frac{c_r}{r!}n^r = \tilde{c}_r n^r, \quad [z^n]N_{l,i}(z) \sim \tilde{d}_{l,i} n^{l+i}, \quad [z^n]\bar{N}_{l,i}(z) \sim \tilde{d}'_{l,i} n^{l+i}. \tag{33}$$

Due to recurrences (29) and (32) the numbers $\tilde{c}_r$, $\tilde{d}_{l,i}$ and $\tilde{d}'_{l,i}$ are defined by the following system of recurrences:

$$\tilde{c}_r = \frac{1}{r-1}\Big(\sum_{p=1}^{r}(\tilde{d}_{r-p,p-1} - \tilde{d}'_{r-p,p-1}) + \sum_{l=1}^{r-1}\tilde{d}_{l,0}\tilde{d}'_{r-l,0}\Big), \quad r \ge 2, \quad \tilde{c}_0 = 1, \ \tilde{c}_1 = 0, \tag{34a}$$

$$\tilde{d}_{l,i} = \tilde{c}_{i+l} + \sum_{p=1}^{l}\tilde{d}_{l-p,i+p-1}, \qquad \tilde{d}'_{l,i} = \tilde{c}_{i+l} - \sum_{p=1}^{l}\tilde{d}'_{l-p,i+p-1}. \tag{34b}$$

Thus it holds that for every $r \ge 0$ fixed and $n \to \infty$:

$$\mathbb{E}\Big(\Big(\frac{\Delta_n}{n}\Big)^r\Big) \to \tilde{c}_r, \tag{35}$$

where the constants $\tilde{c}_r$ are described recursively by (34) using the auxiliary quantities $\tilde{d}_{l,i}$ and $\tilde{d}'_{l,i}$. An application of the theorem of Fréchet and Shohat (see, e.g., [18]) shows then convergence in distribution of $\frac{\Delta_n}{n}$ to a random variable Δ, with moments $\mathbb{E}(\Delta^r) = \tilde{c}_r$, provided that the distribution of Δ is fully characterized by the sequence of its moments. To show this and thus to finish the proof of Theorem 3 we require growth estimates of the constants $\tilde{c}_r$ in order to apply Carleman's criterion [3]:

$$\sum_{m=1}^{\infty}\frac{1}{\big(\mathbb{E}(\Delta^{2m})\big)^{\frac{1}{2m}}} = \infty. \tag{36}$$

To obtain the growth estimates of $\tilde{c}_r$ required we will first simplify the recurrence (34). It is an easy task to show the following equations, thus we omit a proof:

$$\tilde{c}_r = 0, \quad \text{for } r \text{ odd}, \qquad \tilde{d}'_{l,i} = (-1)^{l+i}\tilde{d}_{l+i}, \quad \text{for } l, i \ge 0.$$

Furthermore it is not hard to show that the following relation between $\tilde{c}_r$ and $\tilde{d}_{l,i}$ holds:

$$\tilde{d}_{l,i} = \sum_{k=0}^{l}\binom{l}{k}\tilde{c}_{k+i}, \quad \text{for } l, i \ge 0.$$

This gives in particular

$$\tilde{d}_{l,0} = \sum_{k=0}^{l}\binom{l}{k}\tilde{c}_k, \quad \text{and} \quad \sum_{p=1}^{r}\tilde{d}_{r-p,p-1} = \sum_{k=0}^{r-1}\binom{r}{k}\tilde{c}_k,$$

and (34) leads to the following simpler recurrence for $\tilde{c}_r$ using only the auxiliary quantities $\tilde{d}_l := \tilde{d}_{l,0}$:

$$\tilde{c}_r = \frac{1}{r-1}\Bigg(2\sum_{k=0}^{r-1}\binom{r}{k}\tilde{c}_k + \sum_{l=1}^{r-1}(-1)^l\tilde{d}_l\tilde{d}_{r-l}\Bigg), \quad \text{for } r \ge 2 \text{ even}, \quad \tilde{c}_0 = 1, \tag{37a}$$

$$\tilde{c}_r = 0, \quad \text{for } r \text{ odd}, \tag{37b}$$

$$\tilde{d}_l = \sum_{k=0}^{l}\binom{l}{k}\tilde{c}_k, \quad \text{for } l \ge 0. \tag{37c}$$

By means of recurrence (37) it is now easy to obtain crude growth estimates of the constants $\tilde{c}_r$ by constructing a sequence $\hat{c}_r$ and $\hat{d}_l$ of numbers that are majorizing the sequences $\frac{\tilde{c}_r}{r!}$ and $\frac{\tilde{d}_l}{l!}$ via:

$$\hat{d}_l = \sum_{k=0}^{l}\frac{1}{(l-k)!}\hat{c}_k, \quad l \ge 0, \tag{38a}$$

$$\hat{c}_r = 2\sum_{k=0}^{r-1}\frac{1}{(r-k)!}\hat{c}_k + \sum_{l=1}^{r-1}\hat{d}_l\hat{d}_{r-l}, \quad r \ge 2, \quad \hat{c}_0 = 1, \ \hat{c}_1 = 0. \tag{38b}$$

Introducing the generating functions $\hat{D}(z) := \sum_{l\geq 0} \hat{d}_l z^l$ and $\hat{C}(z) := \sum_{r\geq 0} \hat{c}_r z^r$ we obtain from (38) due to $\hat{D}(z) = e^z\hat{C}(z)$ the equation

$$e^{2z}\big(\hat{C}(z)\big)^2 - 3\hat{C}(z) + 2(1-z) = 0,$$

which has the solution

$$\hat{C}(z) = \frac{3 - \sqrt{9 - 8(1-z)e^{2z}}}{2e^{2z}}. \tag{39}$$

Since the dominant singularity of $\hat{C}(z)$ as given by (39) is located at $z \approx 0.126535$, and thus $\frac{1}{z} \approx 7.9029$, we obtain the following estimate for the coefficients (with some $R > 0$):

$$[z^r]\hat{C}(z) = \hat{c}_r \leq 8^r, \quad \text{for } r \geq R.$$

Since $\frac{\tilde{c}_r}{r!} \leq \hat{c}_r$ due to construction this gives the following growth estimate for the coefficients $\tilde{c}_r$, such that Carleman's criterion (36) is applicable:

$$\frac{\tilde{c}_r}{r!} \leq 8^r, \quad \text{for } r \geq R.$$

6. Conclusion

In contrast to recently obtained corresponding results for left-right-imbalance measures of binary trees under the Catalan model (see the references given in the introduction), we have shown for the random permutation model that the difference between the left and right depth of a node resp. the difference between the left and right pathlength of a tree are asymptotically not of a smaller order than the left and right depth resp. pathlength itself.

Acknowledgement. The authors thank the anonymous referees for their valuable comments.

References

[1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions, 10th printing*, National Bureau of Standards Applied Mathematics Series 55, Washington, 1972.
[2] M. Bousquet-Mélou, Limit laws for embedded trees. Applications to the Integrated Superbrownian Excursion, *Random Structures and Algorithms*, to appear, 2005.
[3] T. Carleman, *Les fonctions quasi analytiques*, GauthierVillars, Paris, 1926.
[4] J. Curtiss, A note on the theory of moment generating functions, *Annals of Mathematical Statistics* 13, 430-433, 1942.
[5] L. Devroy, Applications of the theory of records in the study of random trees, *Acta Informatica* 26, 123–130, 1988.
[6] L. Devroye and R. Neininger, Distances and finger search in random binary search trees, *SIAM Journal on Computing* 33, 647–658, 2004.
[7] R. Dobrow and R. Smythe, Poisson approximations for functionals of random trees. *Random Structures & Algorithms* 9, 79-92, 1996.
[8] J. Fill, P. Flajolet and N. Kapur, Singularity analysis, Hadamard products, and tree recurrences, *Journal of Computational and Applied Mathematics* 174, 271–313, 2005.
[9] P. Flajolet and A. Odlyzko, Singularity Analysis of Generating Functions, *SIAM Journal on Discrete Mathematics* 3, 216–240, 1990.
[10] R. Grübel and N. Stefanoski, Mixed Poisson approximation of node depth distributions in random binary search trees, *The Annals of Applied Probability* 15, 279–297, 2005.
[11] R. Graham, D. Knuth and O. Patashnik, *Concrete Mathematics*, Second Edition, Addison-Wesley, Reading, 1994.
[12] H.-K. Hwang, On Convergence Rates in the Central Limit Theorems for Combinatorial Structures, *European Journal of Combinatorics* 19, 329–343, 1998.
[13] H.-K. Hwang and R. Neininger, Phase change of limit laws in the quicksort recurrence under varying toll functions, *SIAM Journal on Computing* 31, 1687–1722, 2002.
[14] S. Janson, Left and right pathlengths in random binary trees, *Algorithmica*, to appear, 2004.
[15] R. Kemp, *Fundamentals of the average case analysis of particular algorithms*, Wiley-Teubner, Stuttgart, 1984.
[16] C. Knessl and W. Szpankowski, Binary trees, left and right paths, WKB expansions, and Painlevé transcendents, in: *Proc. of the Third Workshop on Analytic Algorithmics and Combinatorics (ANALCO06)*, Miami, 2006.
[17] D. Knuth, *The art of computer programming. Volume 3. Sorting and searching*, Addison-Wesley, Reading, 1973.
[18] M. Loève, *Probability Theory I*, 4th Edition, Springer, New York, 1977.
[19] G. Louchard, Exact and asymptotic distributions in digital and binary search trees, *RAIRO Theoretical Informatics and Applications* 21, 479–495, 1987.
[20] H. Mahmoud, *Evolution of random search trees*, Wiley, New York, 1992.

[21] J.-F. Marckert, The rotation correspondence is asymptotically a dilatation, *Random Structures and Algorithms* 24, 118–132, 2004.
[22] R. Neininger and L. Rüschendorf, A general limit theorem for recursive algorithms and combinatorial structures, *Annals of Applied Probability* 14, 378–418, 2004.
[23] U. Rösler, A limit theorem for "Quicksort", *RAIRO Theoretical Informatics and Applications* 25, 85–100, 1991.
[24] U. Rösler, On the analysis of stochastic divide and conquer algorithms, *Algorithmica* 29, 238–261, 2001.
[25] R. Sedgewick and P. Flajolet, *An introduction to the analysis of algorithms*, Addison-Wesley, Reading, 1996.

Markus Kuba, Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria

E-mail address: markus.kuba@tuwien.ac.at

Alois Panholzer, Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria

E-mail address: Alois.Panholzer@tuwien.ac.at