On Weighted Depths in Random Binary Search Trees

Rafik Aguech[1] · **Anis Amri**[2] · **Henning Sulzbach**[3]

Abstract Following the model introduced by Aguech et al. (Probab Eng Inf Sci 21:133–141, 2007), the weighted depth of a node in a labelled rooted tree is the sum of all labels on the path connecting the node to the root. We analyse weighted depths of nodes with given labels, the last inserted node, nodes ordered as visited by the depth first search process, the weighted path length and the weighted Wiener index in a random binary search tree. We establish three regimes of nodes depending on whether the second-order behaviour of their weighted depths follows from fluctuations of the keys on the path, the depth of the nodes or both. Finally, we investigate a random distribution function on the unit interval arising as scaling limit for weighted depths of nodes with at most one child.

✉ Henning Sulzbach
henning.sulzbach@gmail.com

Rafik Aguech
rafik.aguech@ksu.edu.sa

Anis Amri
anis.amri@hotmail.com

[1] Department of Statistics and Operation Research, College of Sciences, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

[2] University of Monastir, Avenue Taher Hadded, B.P 56, 5000 Monastir, Tunisia

[3] University of Birmingham, School of Mathematics, Edgbaston B15 2TT, Birmingham, United Kingdom
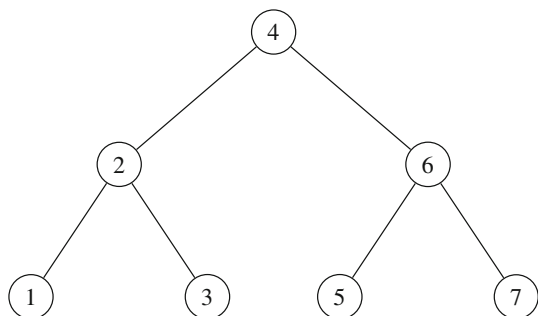
1 Introduction

The binary search tree is an important data structure in computer science allowing for efficient execution of database operations such as insertion, deletion and retrieving of data. Given a list of elements $x_1, x_2, \ldots, x_n$ from a totally ordered set, it is the unique labelled rooted binary tree with n nodes constructed by successive insertion of all elements satisfying the following property: for each node in the tree with label (or key), say y, all keys stored in its left (right) subtree are at most equal to (strictly larger than) y. For an illustration, see Fig. 1.

Properties of binary search trees are typically analysed under the random permutation model where the data $x_1, \ldots, x_n$ are generated by a uniformly chosen permutation of the first n integers. Among the quantities studied in binary search trees, one finds depths of and distances between nodes related to the performance of search queries and finger searches in the database, the (total) path length measuring the cost of constructing the tree as well as the Wiener index. Further, more complex parameters such as the height corresponding to worst case search times, the saturation level and the profile have been studied thoroughly. We review the literature relevant in the context of our work below.

In this note, we complement the wide literature on random binary search trees by the analysis of depths of nodes, path length and Wiener index in their weighted versions as introduced by Aguech et al. [1]. Here, the weighted depth of a node is the sum of all keys stored on the path to the root. In [1], results about weighted depths of extremal paths have been obtained. Kuba and Panholzer [19,20] studied the problem in random increasing trees covering the random recursive tree and the random plane-oriented recursive tree. Weighted depths of nodes and the weighted height were also studied by Broutin and Devroye [3] in a more general tree model, which relies on assigning weights to the edges of the tree. Further, the weighted path length in this model was investigated by Rüschendorf and Schopp [29]. Note that we deviate from the notation introduced in [1,19] using the term weighted depth for what is called weighted path length there since we also study a weighted version of the (total) path length of binary search trees.

Fig. 1 Binary search tree constructed from the list 4, 2, 6, 5, 7, 3, 1

2 Preliminaries

We introduce some notation. By the *size* of a finite binary tree, we refer to its number of nodes. Upon embedding a finite rooted binary tree in the complete infinite binary tree, a node is called *external* if its graph distance to the binary tree is one. Any node on level $k \geq 1$ in a rooted binary tree is associated a vector $v_1 v_2 \ldots v_k \in \{0, 1\}^k$ where $v_i = 0$ if and only if the path from the root to the node continues in the left subtree upon reaching level $i - 1$.

Let $n \geq 1$ and $1 \leq k \leq n$. Under the random permutation model (short: *permutation model*), let $D_k(n)$ be the depth of the node labelled k. By $W_k(n)$ we denote the sum of all keys on the path from the root to the node labelled k including the labels of both endpoints. For $x = x_1 x_2 \ldots \in \{0, 1\}^\infty$, let $B_n(x)$ be the maximal depth among nodes of the form $x_1 \ldots x_k$, $k \geq 0$. We use X_n ($\mathbb{X}_n$) to denote the (weighted) depth of the nth inserted node. Finally, we define the height of the tree by $H_n = \sup\{k \in \mathbb{N} : D_k(n) > 0\}$.

Throughout the paper, we denote by $\mathscr{L}(X)$ the distribution of a random variable X. For real-valued X with finite second moment, we write σ_X for its standard deviation. By $\mathscr{N}$ we denote a random variable with the standard normal distribution, and by μ the Dickman distribution on $[0, \infty)$ characterized by its Fourier transform,

$$\int e^{i\lambda x} d\mu(x) = \exp\left(\int_0^1 \frac{e^{i\lambda x} - 1}{x} dx\right), \quad \lambda \in \mathbb{R}. \tag{1}$$

The origins of the Dickman distribution go back to Dickman's [10] classical result on large prime divisors. Compare Hildebrandt and Tenenbaum [15] for a survey on the problem. In the probabilistic analysis of algorithm, μ first arose in Hwang and Tsai's [17] study of the complexity of Hoare's selection algorithm. We refer to this work for a discussion of more details on the distribution, historical background and further references.

Finally, we use the Landau notations little–o, big–O, little–ω, big–Ω and big–Θ as $n \to \infty$.

2.1 Depths and Height

We recall the following fundamental property of random binary search trees going back to Devroye [6]: in probability and with respect to all moments, we have

$$\frac{H_n}{\log n} \to c^*, \tag{2}$$

where $c^* = 4.31 \ldots$ is the larger of the two solutions to the transcendent equation $e = (\frac{2e}{x})^x$. Next, by classical results due to Brown and Shubert [4] and Devroye [7], for any $x \in \{0, 1\}^\infty$, in distribution,

$$\frac{B_n(x) - \log n}{\sqrt{\log n}} \to \mathscr{N}, \quad \frac{X_n - 2\log n}{\sqrt{2\log n}} \to \mathscr{N}. \tag{3}$$

(In [7, Theorem O1], the first convergence in the last display is formulated for $x = \mathbf{0} := 00\ldots$ The general case follows, since, by symmetry, $\mathscr{L}(B_n(x)) = \mathscr{L}(B_n(\mathbf{0}))$ for all x. The second convergence was also claimed in a footnote by Mahmoud and Pittel [23].) Grübel [13] studied the process $\{B_n(x) : x \in \{0, 1\}^\infty\}$, the so-called *silhouette*, thereby obtaining a functional limit theorem for its integrated version. The asymptotic behaviour of depths of nodes with given labels has been analysed by Devroye and Neininger [9]: uniformly in $1 \leq k \leq n$ and as $n \to \infty$,

$$\mathbf{E}[D_k(n)] = \log(k(n-k)) + O(1), \quad \text{Var}(D_k(n)) = \log(k(n-k)) + O(1). \quad (4)$$

Moreover, for any $1 \leq k \leq n$, which may depend on n, in distribution

$$\frac{D_k(n) - \mathbf{E}[D_k(n)]}{\sigma_{D_k(n)}} \to \mathscr{N}. \quad (5)$$

Here, one should also compare Grübel and Stefanoski [14] for stronger results in the context of the corresponding Poisson approximation. For a survey on depths and distances in binary search trees, we refer to Mahmoud's book [21]. Finally, the asymptotic behaviour of the weighted depths of the nodes associated with the vectors $\mathbf{0}$ and $\mathbf{1} := 11\ldots$ denoted by $\mathscr{L}_n$ and $\mathscr{R}_n$ ($\mathscr{L}$ and $\mathscr{R}$ stand for left and right) were studied in [1]. In distribution,

$$\frac{\mathscr{L}_n}{n} \to \mathscr{Y}, \quad \frac{\mathscr{R}_n - nB_n(\mathbf{1})}{n\sqrt{\log n}} \to 0, \quad (6)$$

where $\mathscr{Y}$ has the Dickman distribution. The first convergence is closely related to the limit law in Theorem 3.1 in [17].

2.2 Path Length and Wiener Index

In a rooted tree, the path length is defined as the sum over all depths of nodes. Moreover, the Wiener index is obtained by summing all distances of unordered pairs of vertices. For a random binary search tree of size n, we denote its path length by P_n and its Wiener index by W_n. Denoting by γ the Euler–Mascheroni constant, we have

$$\mathbf{E}[P_n] = 2n \log n + (2\gamma - 4)n + o(n), \quad \text{Var}(P_n) = \frac{21 - 2\pi^2}{3}n^2 + o(n^2), \quad (7)$$

going back to Hoare [16] and Knuth [18]. Further, by [25],

$$\mathbf{E}[W_n] = 2n^2 \log n + (2\gamma - 6)n^2 + o(n^2), \quad \text{Var}(W_n) = \frac{20 - 2\pi^2}{3}n^4 + o(n^4). \quad (8)$$

Central limit theorems for the path length go back to Régnier [27] and Rösler [28], for the Wiener index to Neininger [25]. More precisely, by [25, Theorem 1.1], there

exists a non-trivial random variable Z^* on $\mathbb{R}^2$ characterized by a stochastic fixed-point equation, such that, in distribution,

$$\left(\frac{W_n - \mathbb{E}[W_n]}{n^2}, \frac{P_n - \mathbb{E}[P_n]}{n} \right) \to Z^*. \tag{9}$$

2.3 The i.i.d Model

We also consider binary search trees of size n where the data are chosen as the first n values of a sequence of independent random variables $U_1, U_2, \ldots$ each having the uniform distribution on $[0, 1]$. Since the vector $(\text{rank}(U_1), \ldots, \text{rank}(U_n))$ constitutes a uniformly chosen permutation, in distribution, both the permutation model and the i.i.d. model lead to the same unlabelled tree. We use the same notation as in the permutation model for quantities not involving the labels of nodes, that is, X_n, h_n, H_n, P_n, W_n and $B_n(x)$. Further, we define the weighted path length $\mathscr{P}_n$ as the sum of all weighted depths, and the weighted Wiener index $\mathscr{W}_n$ as the sum over all pairs of weighted distances. Here, the weighted distance between two nodes equals the sum of all labels on the path connecting them, labels of endpoints included. (Notice that the weighted distance between a node and itself is equal to its label.) Finally, analogously to $B_n(x)$, we define $\mathscr{B}_n(x)$ as the weighted depth of the node of largest depth on the path x. We call $\{\mathscr{B}_n(x) : x \in \{0, 1\}^\infty\}$, the weighted silhouette of the tree (at time n).

3 Main Results

Our main results are divided into two groups: Theorems 1 and 2 hold in the permutation model, while Theorems 3 and 4 are formulated in the i.i.d. model.

3.1 Results in the Permutation Model

We start with the expansions of the first two moments of the weighted depth $W_k(n)$. Uniformly in $1 \le k \le n$, as $n \to \infty$,

$$\mathbb{E}[W_k(n)] = k \log(k(n - k + 1)) + n + O(k + \log n), \tag{10}$$

$$\text{Var}(W_k(n)) = k^2 \log(k(n - k + 1)) + \frac{n^2}{2} + O(kn). \tag{11}$$

It turns out that the asymptotic distributional behaviour of $W_k(n)$ with respect to terms of second order is entirely described by that of $k D_k(n)$ if and only if $k = \omega(n/\sqrt{\log n})$. Accordingly, in the remainder of this paper, we call nodes with labels of order $\omega(n/\sqrt{\log n})$ *large* and of order $O(n/\sqrt{\log n})$ *small*.

Theorem 1 (Weighted depths of large nodes) *For* $k = \omega(n/\sqrt{\log n})$,

$$\mathbb{E}[|W_k(n) - k D_k(n)|] = o(\sigma_{k D_k(n)}). \tag{12}$$

In particular, for $0 < \alpha < 1$ and $|k/n - \alpha| = o((\log n)^{-1/2})$, in distribution,

$$\left(\frac{D_k(n) - 2\log n}{\sqrt{2\log n}}, \frac{W_k(n) - 2\alpha n \log n}{\alpha n \sqrt{2\log n}} \right) \to (\mathcal{N}, \mathcal{N}). \tag{13}$$

For the last inserted node, in distribution,

$$\left(\frac{X_n - 2\log n}{\sqrt{2\log n}}, \frac{\mathbb{X}_n}{2n\log n} \right) \to (\mathcal{N}, \xi), \tag{14}$$

where $\mathcal{N}$ and ξ are independent and ξ is uniformly distributed on $[0, 1]$.

The asymptotic behaviour of weighted depths of small nodes is to be compared with the corresponding results in [19]. Here, another phase transition occurs when $k = o(n/\sqrt{\log n})$.

Theorem 2 (Weighted depths of small nodes) *Let $k = O(n/\sqrt{\log n})$. Then, in distribution,*

$$\left(\frac{D_k(n) - \mathbf{E}[D_k(n)]}{\sigma_{D_k(n)}}, \frac{W_k(n) - kD_k(n)}{n} \right) \to (\mathcal{N}, \mathcal{Y}), \tag{15}$$

where $\mathcal{N}$ and $\mathcal{Y}$ are independent and $\mathcal{Y}$ has the Dickman distribution. Thus, if $k\sqrt{\log n}/n \to \beta \geq 0$, in distribution,

$$\left(\frac{D_k(n) - 2\log n}{\sqrt{2\log n}}, \frac{W_k(n) - \mathbf{E}[W_k(n)]}{n} \right) \to (\mathcal{N}, \mathcal{Y} + \sqrt{2}\beta\mathcal{N} - 1).$$

In particular, if $|k\sqrt{\log n}/n - \beta| = o((\log n)^{-1/2})$ with $\beta > 0$, then, in distribution,

$$\left(\frac{D_k(n) - 2\log n}{\sqrt{2\log n}}, \frac{W_k(n) - 2\beta n\sqrt{\log n}}{n} \right) \to (\mathcal{N}, \mathcal{Y} + \sqrt{2}\beta\mathcal{N}).$$

3.2 Results in the i.i.d Model

Any $x \in \{0, 1\}^\infty$ corresponds to a unique value $x \in [0, 1]$ by $x = \sum_{i=0}^\infty x_i 2^{-i}$. This identification becomes one-to-one upon allowing only those $x \in \{0, 1\}^\infty$ which contain infinitely many zeros and $x = \mathbf{1}$. In the i.i.d. model, for any $x \in \{0, 1\}^\infty$, $k \geq 1$, the node $x_1 \ldots x_k$ eventually appears in the sequence of binary search trees and we write $\Xi_k(x)$ for its ultimate label. The following theorem about the behaviour of $\mathscr{B}_n(x)$ involves a random continuous distribution function arising as the almost sure limit of $\Xi_k(x)$, $x \in [0, 1]$, as $k \to \infty$. We believe that this process is of independent interest and state some of its properties in Proposition 1 in Sect. 3.3. The simulations of Ξ_{15} presented in Fig. 2 illustrate the scaling limit.
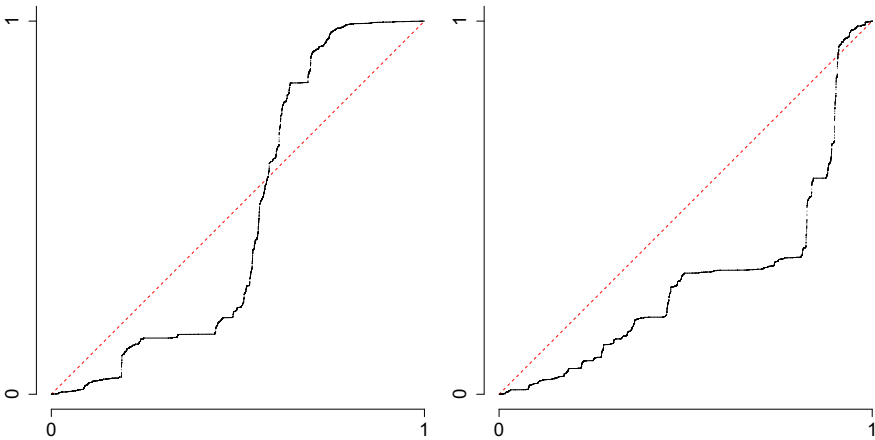
Fig. 2 Two simulations of Ξ_{15}, the *dotted line* being the graph of the identity function

Theorem 3 (Weighted silhouette) *There exists a random continuous and strictly increasing bijection $\Xi(x)$, $x \in [0, 1]$, such that, almost surely, uniformly on the unit interval, $\Xi_k(x) \to \Xi(x)$. For any $x \in [0, 1]$, in probability,*

$$\frac{\mathscr{B}_n(x)}{\log n} \to \Xi(x). \tag{16}$$

Also, for any $m \geq 1$, in probability

$$\int_0^1 \left| \frac{\mathscr{B}_n(x)}{\log n} - \Xi(x) \right|^m dx \to 0. \tag{17}$$

Further, in probability,

$$\sup_{x \in [0,1]} \frac{\mathscr{B}_n(x)}{\log n} \to c^* = 4.31 \ldots \tag{18}$$

with c^* as in (2). Finally, for any $x \in [0, 1]$, in distribution,

$$\left(\frac{B_n(x) - \log n}{\sqrt{\log n}}, \frac{\mathscr{B}_n(x)}{\log n} \right) \to (\mathscr{N}, \Xi(x)), \tag{19}$$

where $\mathscr{N}$ and $\Xi(x)$ are independent.

The next theorem extends the distributional convergence result in Theorem 1.1 in [25], that is (9), by central limit theorems for the weighted path length and the weighted Wiener index.

Theorem 4 (Weighted path length and Wiener index) *In the i.i.d. model, we have*

$$\mathbf{E}[\mathscr{P}_n] = n \log n + (\gamma - 3/2)n + o(n), \quad \mathbf{E}[\mathscr{W}_n] = n^2 \log n + (\gamma - 11/4)n^2 + o(n^2),$$

and

$$\mathrm{Var}(\mathscr{P}_n) = \frac{65 - 6\pi^2}{36}n^2 + o(n^2), \quad \mathrm{Var}(\mathscr{W}_n) = \frac{2413 - 240\pi^2}{1440}n^4 + o(n^4).$$

The leading constants in the expansions of the covariances between P_n, W_n, $\mathscr{P}_n$ and $\mathscr{W}_n$ are given in (36)–(38). (The leading constant for $\mathrm{Cov}(P_n, W_n)$ was already given in [25].) As $n \to \infty$, with convergence in distribution and with respect to the first two moments in $\mathbb{R}^4$, we have

$$\left(\frac{\mathscr{W}_n - \mathbb{E}[\mathscr{W}_n]}{n^2}, \frac{W_n - \mathbb{E}[W_n]}{n^2}, \frac{\mathscr{P}_n - \mathbb{E}[\mathscr{P}_n]}{n}, \frac{P_n - \mathbb{E}[P_n]}{n} \right) \to Z,$$

where the limiting distribution $\mathscr{L}(Z)$ is the unique fixed point of the map T in (35).

Conclusions We have seen that there exist three types of nodes showing significantly different behaviour with respect to their weighted depths. By Theorem 1, for $k = \omega(n/\sqrt{\log n})$, second-order fluctuations of weighted depths are due to variations of the depth of nodes. In the second regime, when $k = \Theta(n/\sqrt{\log n})$, variations of weighted depths are determined by two independent contributions, one for the depths and one for the keys on the paths. Finally, when $k = o(n/\sqrt{\log n})$ only fluctuations of labels on paths influence second-order terms of weighted depths. The third regime can be further subdivided with respect to the first-order terms of $W_k(n)$ and $kD_k(n)$: for $k = \omega(n/\log n)$, they coincide, for $k = \Theta(n/\log n)$, they are of the same magnitude, whereas, for $k = o(n/\log n)$, they are of different scale. By Theorem 3, the weighted silhouette behaves considerably different. Here, the lack of concentration around the mean leads to an interesting random distribution function on the unit interval as scaling limit.

3.3 Further Results and Remarks

Model comparison We decided to present Theorems 3 and 4 in the i.i.d. model rather than in the permutation model since this allows for a stronger mode of convergence in (16), (17) and a clearer presentation of the proof of Theorem 4. In the i.i.d. model, denoting by $\mathscr{W}_{(k)}(n)$ the weighted depth of the node of rank k among the first n inserted keys, Theorems 1 and 2 remain valid upon replacing $W_k(n)$ by $n\mathscr{W}_{(k)}(n)$. Similarly, Theorems 3 and 4 hold in the permutation model where weighted depths and the weighted path length are to be scaled down by a factor n and the weighted Wiener index by a factor n^2. The convergences in (16) and (17) then only hold in distribution. This can be deduced most easily from the following coupling of the two models: starting with the binary search tree in the i.i.d. model, also consider the random binary search tree in the permutation model relying on the permutation $(\mathrm{rank}(U_1), \ldots, \mathrm{rank}(U_n))$. Then, for all $1 \le k \le n$,

$$\left| \mathscr{W}_{(k)}(n) - \frac{W_k(n)}{n} \right| \le H_n \max_{1 \le i \le n} \left| U_i - \frac{\mathrm{rank}(U_i)}{n} \right|. \tag{20}$$

It is well known that the second factor on right-hand side grows like $n^{-1/2}$, compare, e.g. Donsker's theorem for empirical distribution functions or the Dvoretzky–Kiefer–Wolfowitz inequality [11]. Combining this, (20) and (2) is sufficient to transfer all results in Sect. 3 between the two models.

The depth first search process In the permutation model, let $v_1, \ldots, v_{n+1}$ be the external nodes as discovered by the depth first search process from left to right. By $D_k^*(n)$ and $W_k^*(n)$, $1 \le k \le n+1$, we denote depth and weighted depth of the external node v_k. Then, at the end of Sect. 4.1, we show that, uniformly in $1 \le k \le n$,

$$\mathbf{E}\left[|D_k(n) - D_k^*(n)|^2\right] = o(\log n), \quad \mathbf{E}\left[|W_k(n) - W_k^*(n)|^2\right] = o(\mathrm{Var}(W_k(n))). \tag{21}$$

Thus, the results in Theorems 1 and 2 also cover the second-order analysis of the sequences $D_k^*(n)$ and $W_k^*(n)$.

Weighted distances In the permutation model, let $D_{k,\ell}(n)$ be the graph distance between the nodes labelled $1 \le k \le \ell \le n$ and $W_{k,\ell}(n)$ be the sum of all labels on the path from k to ℓ, labels at the endpoints included. Asymptotic normality for the sequence $(D_{k,\ell}(n))$ (after rescaling) under the optimal condition $\ell - k \to \infty$ has been obtained in [9]. For uniformly chosen nodes, distributional convergence results date back to Mahmoud and Neininger [22] and Panholzer and Prodinger [26]. Analogously to Theorem 1, it is straightforward to prove central limit theorems jointly for weighted and non-weighted distances. We only state the results. If $\ell - k = \Omega(n)$ and $k = \omega(n/\sqrt{\log n})$, then

$$\mathbf{E}\left[|W_{k,\ell}(n) - k D_k(n) - \ell D_\ell(n)|\right] = \sigma_{\ell D_\ell(n)}.$$

In particular, for $0 < s < t < 1$ and $|k/n - s| = o((\log n)^{-1/2})$, $|\ell/n - t| = o((\log n)^{-1/2})$, we have, in distribution,

$$\left(\frac{D_k(n) - 2\log n}{\sqrt{2\log n}}, \frac{D_\ell(n) - 2\log n}{\sqrt{2\log n}}, \frac{D_{k,\ell}(n) - 4\log n}{\sqrt{4\log n}}, \frac{W_{k,\ell}(n) - 2(s+t)n\log n}{n\sqrt{2\log n}}\right)$$
$$\to \left(\mathscr{N}_1, \mathscr{N}_2, \frac{\mathscr{N}_1 + \mathscr{N}_2}{\sqrt{2}}, s\mathscr{N}_1 + t\mathscr{N}_2\right).$$

Here, $\mathscr{N}_1, \mathscr{N}_2$ are independent random variables both with the standard normal distribution.

The limit process Ξ The process Ξ in Theorem 3 is a random distribution function. In particular, it can be regarded as an element in the set of càdlàg functions $\mathscr{D}[0,1]$ consisting of all $f : [0,1] \to \mathbb{R}$, such that, for all $t \in [0,1]$, $f(t) = \lim_{s \downarrow t} f(s)$ and $\lim_{s \uparrow t} f(s)$ exists. The absolute value of f is defined by $\sup\{|f(t)| : t \in [0,1]\}$. Endowed with Skorokhod's topology J_1, $\mathscr{D}[0,1]$ becomes a Polish space. We refer to Chapter 3 in Billingsley's book [2] for detailed information on this matter.

Proposition 1 (Properties of Ξ) *The process Ξ is unique (in distribution) among all càdlàg processes with finite absolute second moment satisfying*

$$\mathscr{L}((\Xi(t))_{t\in[0,1]}) = \mathscr{L}\Big(\big(\mathbf{1}_{[0,1/2)}(t)U\,\Xi(2t)$$
$$+ \mathbf{1}_{[1/2,1)}(t)\big((1-U)\Xi'(2t-1)+U\big)\big)_{t\in[0,1]}\Big). \quad (22)$$

Here, Ξ, Ξ', U are independent, U has the uniform distribution on $[0,1]$, and Ξ' is distributed like Ξ. We have

(i) $\mathbf{E}[\Xi(t)] = t$ *for all $t \in (0,1)$;*

(ii) $\mathscr{L}((\Xi(t))_{t\in[0,1]}) = \mathscr{L}((1 - \Xi(1-t))_{t\in[0,1]})$;

(iii) $\Xi(\xi)$ *has the arcsine distribution with density*

$$\frac{1}{\pi\sqrt{x(1-x)}}, \quad x \in (0,1),$$

where Ξ, ξ are independent and ξ has the uniform distribution on $[0,1]$;

(iv) *for $t \in (0,1)$, $\mathscr{L}(\Xi(t))$ has a smooth density $f_t : (0,1) \to (0,\infty)$;*

(v) *for $t \in (0,1/2)$, $xf_t'(x) = -f_{2t}(x)$, $x \in (0,1)$, f_t is strictly monotonically decreasing and $\lim_{x\uparrow 1} f_t(x) = 0$;*

(vi) *with $\alpha_t^{(i)} := \lim_{x\downarrow 0} f_t^{(i)}(x)$, $i = 0, 1$, $t \in (0,1/2)$ and $\gamma_0 = 1/4$, $\gamma_1 = 5/16$, we have $\alpha_t^{(i)} = (-1)^i\infty$ for $0 < t \le \gamma_i$, $|\alpha_t^{(i)}| < \infty$ for $\gamma_i < t < 1/2$ and $|\alpha_t^{(i)}| \uparrow \infty$ as $t \downarrow \gamma_i$.*

Random recursive trees A random recursive tree is constructed as follows: starting with the root labelled one, in the kth step, $k \ge 2$, a node labelled k is inserted in the tree and connected to an already existing node chosen uniformly at random. Weighted depths in random binary search trees differ substantially from those in random recursive trees analysed in [19] where all nodes show an asymptotic behaviour comparable to that of nodes labelled $k = o(n/\sqrt{\log n})$ in the binary search tree. The difference is highlighted by the weighted path length. Being of the same order as the path length in binary search trees, it follows from results in [19] that the weighted path length $\mathscr{Q}_n$ in a random recursive tree of size n is of order n^2. The same is valid for its standard deviation. We conjecture that the sequence $(n^{-2}\mathscr{Q}_n)$ converges in distribution to a non-trivial limit; however, the recursive approach worked out in the proof of Theorem 4, which also applies to the analysis of the path length in random recursive trees, seems not to be fruitful in this context.

Outline All results are proved in Sect. 4 starting with the proofs of Theorems 1 and 2 as well as (21) in Sect. 4.1. Here, most arguments are based on representations of (weighted) depths as sums of bounded independent random variables which go back to Devroye and Neininger [9]. Theorem 3 and Proposition 1 are proved in Sect. 4.2. In this part, the construction of the limiting process relies on suitable uniform L_1-bounds on the increments of the process $\Xi_k(x)_{x\in[0,1]}, k \ge 1$, while the properties of the limit laws formulated in Proposition 1 follow from the distributional fixed-point equation

(22). Finally, the proof of Theorem 4 relying on the contraction method is worked out in Sect. 4.3.

4 Proofs

4.1 Weighted Depths of Labelled Nodes

In the permutation model, let $A_{j,k}$ be the event that the node labelled k is in the subtree of the node labelled j. Then, $D_k(n) = \sum_{j=1}^n \mathbf{1}_{A_{j,k}} - 1$ and $W_k(n) = \sum_{j=1}^n j\mathbf{1}_{A_{j,k}}$. It is easy to see that $A_{1,k}, \ldots, A_{k-1,k}$ and $A_{k+1,k}, \ldots, A_{n,k}$ are two families of independent events; however, there exist subtle dependencies between the sets. Following the approach in [9], let $B_{j,k} = A_{j,k-1}$ for $j < k$ and $B_{j,k} = A_{j,k+1}$ for $j > k$. For convenience, let $B_{k,k}$ be an almost sure event. The following lemma summarizes results in [9], and we refer to this paper for a proof. In this context, note that Devroye [8] gives distributional representations as sums of independent (or m-dependent) indicator variables for quantities growing linearly in n, such as the number of leaves.

Lemma 1 *Let* $1 \le k \le n$. *Then, the events* $B_{j,k}$, $j = 1, \ldots, n$, *are independent. For* $j \ne k$, *we have*

$$\mathbb{P}\left(A_{j,k}\right) = \frac{1}{|k-j|+1}, \quad \mathbb{P}\left(B_{j,k}\right) = \frac{1}{|k-j|}.$$

From the lemma, it follows that

$$\mathbf{E}\left[\sum_{j=1}^n \mathbf{1}_{B_{j,k}\setminus A_{j,k}}\right] \le 2, \quad \text{and} \quad \mathbf{E}\left[\sum_{j=1}^n j\mathbf{1}_{B_{j,k}\setminus A_{j,k}}\right] \le 2k + \log n.$$

The ideas in [9] can also be used to analyse second (mixed) moments. Straightforward calculations show the following bounds:

$$\mathbf{E}\left[\sum_{i,j=1}^n \mathbf{1}_{B_{j,k}}\mathbf{1}_{B_{i,k}\setminus A_{i,k}}\right] = O(1), \quad \text{and}$$

$$\mathbf{E}\left[\sum_{i,j=1}^n ij\mathbf{1}_{B_{j,k}}\mathbf{1}_{B_{i,k}\setminus A_{i,k}}\right] = O(k^2 + k(\log n)^2).$$

Here, both O-terms are uniform in $1 \le k \le n$. Define $\bar{D}_k(n) = \sum_{j=1}^n \mathbf{1}_{B_{j,k}} - 1$ and $\bar{W}_k(n) = \sum_{j=1}^n j\mathbf{1}_{B_{j,k}}$. We make the following observation:

O: The asymptotic statements in (10), (11), Theorem 1 and Theorem 2 are correct if and only if they are correct upon replacing $D_k(n)$ by $\bar{D}_k(n)$ and/or $W_k(n)$ by $\bar{W}_k(n)$.

For $i = 1, 2, n \geq 0$ and $1 \leq k \leq n$, set $H_n^{(i)} := \sum_{j=1}^{n} j^{-i}$ and $H_{k,n}^{(i)} := H_{k-1}^{(i)} + H_{n-k}^{(i)}$. Using Lemma 1, one easily computes

$$\mathbf{E}\left[\bar{W}_k(n)\right] = k(H_{k,n}^{(1)} - 1) + n + 1,$$

$$\mathrm{Var}(\bar{W}_k(n)) = k^2 \left(H_{k,n}^{(1)} - H_{k,n}^{(2)} - 3\right) + \frac{n^2}{2} + kn + 2k\left(H_{k-1}^{(1)} - H_{n-k}^{(1)}\right) - \frac{n}{2} + k + 1.$$

As $H_n^{(1)} = \log(n+1) + O(1)$ and $H_n^{(2)} = O(1)$, both expansions (10) and (11) follow from observation **O**.

4.1.1 Weighted Depths of Large Nodes

We prove Theorem 1. First, (12) follows from (4) and

$$\mathbf{E}\left[|kD_k(n) - W_k(n)|\right] \leq k + \sum_{j=1}^{n} |k - j| \mathbb{P}\left(A_{j,k}\right) \leq k + n. \tag{23}$$

For $k = \omega(n/\sqrt{\log n})$, combining (4), (5) and (10), in distribution,

$$\left(\frac{D_k(n) - \mathbf{E}\left[D_k(n)\right]}{\sigma_{D_k(n)}}, \frac{W_k(n) - \mathbf{E}\left[W_k(n)\right]}{\sigma_{W_k(n)}}\right) \to (\mathcal{N}, \mathcal{N}).$$

From here, statement (13) follows from (4) and (10).

Considering the last inserted node with value Y_n, note that, conditionally on $Y_n = k$, the correlations between the events $A_{j,k}, j < k$ and $A_{j,k}, j > k$ vanish. More precisely, given $Y_n = k$, the family $\{\mathbf{1}_{A_{j,k}}, j = 1, \ldots, n\}$ is distributed like a family of independent Bernoulli random variables $\{V_{j,k} : j = 1, \ldots, n\}$ with $\mathbb{P}\left(V_{j,k} = 1\right) = |k - j|^{-1}$ for $j \neq k$ and $\mathbb{P}\left(V_{k,k} = 1\right) = 1$. Thus,

$$\mathbf{E}\left[|Y_n(X_n + 1) - \mathbb{X}_n|\right] \leq \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}\left[\sum_{j=1}^{n} |k - j| \mathbf{1}_{A_{j,k}} \middle| Y_n = k\right]$$

$$= \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}\left[\sum_{j=1}^{n} |k - j| V_{j,k}\right] \leq n.$$

By (3), we have $X_n/\log n \to 2$ in probability. Hence, in order to prove (14), it suffices to show that, in distribution,

$$\left(\frac{X_n - 2\log n}{\sqrt{2\log n}}, \frac{Y_n}{n}\right) \to (\mathcal{N}, \xi). \tag{24}$$

For a sequence (k_n) satisfying $sn \leq k_n \leq tn$ for $0 < s < t < 1$, let us condition on the event $Y_n = k_n$. Then, by the central limit theorem for triangular arrays of row-wise

independent uniformly bounded random variables with diverging variance applied to V_{j,k_n}, $j = 1, \ldots, n$, in distribution,

$$\frac{X_n - 2\log n}{\sqrt{2\log n}} \to \mathcal{N}.$$

Hence, (24) follows from an application of the theorem of dominated convergence noting that Y_n is uniformly distributed on $\{1, \ldots, n\}$.

4.1.2 Weighted Depths of Small Nodes

We prove Theorem 2. Let $\bar{D}_k^>(n) = \sum_{j=k+1}^n \mathbf{1}_{B_{j,k}}$ and $\bar{W}_k^>(n) = \sum_{j=k+1}^n j\mathbf{1}_{B_{j,k}}$. Since $k = O(n/\sqrt{\log n})$, the same calculation as in (23) shows that,

$$\frac{\mathbf{E}\left[|\bar{W}_k(n) - \bar{W}_k^>(n) - k(\bar{D}_k(n) - \bar{D}_k^>(n))|\right]}{n} \leq \frac{k}{n} \to 0, \quad n \to \infty. \quad (25)$$

For $\lambda, \mu \in \mathbb{R}$, we have

$$\log\mathbf{E}\left[\exp\left(i\lambda\left(\bar{D}_k^>(n) - \log n\right)/\sqrt{\log n} + i\mu\left(\bar{W}_k^>(n) - k\bar{D}_k^>(n)\right)/n\right)\right]$$

$$= -i\lambda\sqrt{\log n} + \log\mathbf{E}\left[\exp\left(i\sum_{j=k+1}^n \left(\frac{\lambda}{\sqrt{\log n}} + \mu\frac{j-k}{n}\right)B_{j,k}\right)\right]$$

$$= -i\lambda\sqrt{\log n} + \sum_{j=k+1}^n \log\left(1 + \frac{\exp\left(i\left(\frac{\lambda}{\sqrt{\log n}} + \mu\frac{j-k}{n}\right)\right) - 1}{j-k}\right).$$

By a standard Taylor expansion, the last display equals

$$-i\lambda\sqrt{\log n} + \sum_{j=k+1}^n \frac{\exp\left(i\left(\frac{\lambda}{\sqrt{\log n}} + \mu\frac{j-k}{n}\right)\right) - 1}{j-k} + o(1)$$

$$= -i\lambda\sqrt{\log n} + \sum_{j=k+1}^n \frac{\exp\left(i\mu\frac{j-k}{n}\right)\left(1 + \frac{i\lambda}{\sqrt{\log n}} - \frac{\lambda^2}{2\log n}\right) - 1}{j-k} + o(1)$$

$$= -\lambda^2/2 + \left(1 + \frac{i\lambda}{\sqrt{\log n}} - \frac{\lambda^2}{2\log n}\right)\sum_{j=0}^{n-1} \frac{\exp\left(i\mu\frac{j+1}{n}\right) - 1}{j+1} + o(1)$$

$$= -\lambda^2/2 + \int_0^1 \frac{e^{i\mu x} - 1}{x}dx + o(1).$$

Here, in the last step, we have used that the sum on the right-hand side is a Riemann sum over the unit interval whose mesh size n^{-1} tends to zero. Thus, using the notation

of the theorem, (1) and Lévy's continuity theorem, in distribution,

$$\left(\frac{\bar{D}_k^>(n) - \log n}{\sqrt{\log n}}, \frac{\bar{W}_k^>(n) - k\bar{D}_k^>(n)}{n}\right) \to (\mathcal{N}, \mathcal{Y}). \qquad (26)$$

In order to deduce (15) note that, by Lemma 1, $\bar{D}_k(n) - \bar{D}_k^>(n)$ and $(\bar{D}_k^>(n), \bar{W}_k^>(n))$ are independent while

$$\frac{\bar{D}_k(n) - \bar{D}_k^>(n) - \mathbf{E}\left[\bar{D}_k(n) - \bar{D}_k^>(n)\right]}{\sigma_{\bar{D}_k(n) - \bar{D}_k^>(n)}} \to \mathcal{N},$$

in distribution if and only if $k \to \infty$ using the central limit theorem for sums of independent and uniformly bounded random variables. Since

$$\frac{\bar{D}_k(n) - \mathbf{E}\left[\bar{D}_k(n)\right]}{\sigma_{\bar{D}_k(n)}} = \frac{\bar{D}_k^>(n) - \mathbf{E}\left[D_k^>(n)\right]}{\sqrt{\log n}} \frac{\sqrt{\log n}}{\sigma_{\bar{D}_k(n)}}$$

$$+ \frac{\bar{D}_k(n) - \bar{D}_k^>(n) - \mathbf{E}\left[\bar{D}_k(n) - \bar{D}_k^>(n)\right]}{\sigma_{\bar{D}_k(n) - \bar{D}_k^>(n)}} \frac{\sigma_{\bar{D}_k(n) - \bar{D}_k^>(n)}}{\sigma_{\bar{D}_k(n)}},$$

we deduce

$$\left(\frac{\bar{D}_k(n) - \mathbf{E}\left[\bar{D}_k(n)\right]}{\sigma_{\bar{D}_k(n)}}, \frac{\bar{W}_k^>(n) - k\bar{D}_k^>(n)}{n}\right) \to (\mathcal{N}, \mathcal{Y}),$$

from (26) upon treating the cases $k = O(1)$ and $k = \omega(1)$ separately. From here, the assertion (15) follows with the help of (25) and observation **O**.

4.1.3 Proof of (21)

The main observation is that the kth external node visited by the depth first search process is always contained in the subtree rooted at the node labelled k. This can be proved by induction exploiting the decomposition of the tree at the root. Thus, denoting by $H_k(n)$ the height of the subtree rooted at the node labelled k, we have

$$D_k(n) \leq D_k^*(n) \leq D_k(n) + H_k(n),$$
$$W_k(n) \leq W_k^*(n) \leq W_k(n) + M_k(n)H_k(n).$$

Here, $M_k(n)$ stands for the largest label in the subtree rooted at the node labelled k. Let $T_k(n)$ be the size of the subtree rooted at k. Then $T_k(n) = 1 + T_k^<(n) + T_k^>(n)$ where $T_k^<(n)$ denotes the number of elements in the subtree rooted at k with values smaller than k. By Lemma 1, for $\ell \leq n - k$, we have $\mathbb{P}\left(T_k^>(n) \geq \ell\right) = \mathbb{P}\left(A_{k,k+\ell}\right) = 1/(\ell+1)$. Using the same arguments for the quantity $T_k^<(n)$, we deduce that, uniformly in $1 \leq k \leq n$,

$$\mathbf{E}\left[T_k(n)\right] = \Theta(\log n), \quad \mathbf{E}\left[(T_k(n))^2\right] = \Theta(n^{1/2}), \quad \mathbf{E}\left[(\log T_k(n))^2\right] = O(1).$$

Thus, by an application of (2), for some $C_1 > 0$,

$$\mathbf{E}\left[|D_k(n) - D_k^*(n)|^2\right] \leq \mathbf{E}\left[(H_k(n))^2\right] \leq C_1 \mathbf{E}\left[(\log T_k(n))^2\right] = O(1).$$

By the same arguments, for some $C_2 > 0$, we have

$$\begin{aligned}
\mathbf{E}\left[|W_k(n) - W_k^*(n)|^2\right] &\leq \mathbf{E}\left[(M_k(n) H_k(n))^2\right] \leq \mathbf{E}\left[(k + T_k(n))^2 (H_k(n))^2\right] \\
&\leq C_2 k^2 + C_1 \left(2k \mathbf{E}\left[T_k(n)(\log T_k(n))^2\right]\right. \\
&\quad \left. + \mathbf{E}\left[(T_k(n))^2 (\log T_k(n))^2\right]\right) \\
&= O(k^2 + (\log n)^2 n^{1/2}).
\end{aligned}$$

From here, (21) follows from (10).

4.2 The Weighted Silhouette

We prove Theorem 3 and Proposition 1.

Proof of Theorem 3 We start with the uniform convergence of $(\varXi_k)$. For all $x \in [0, 1]$, $|\varXi_k(x) - \varXi_{k-1}(x)|$ is distributed like the product of $k+1$ independent random variables, each of which having the uniform distribution on $[0, 1]$. In particular, by the union bound and Markov's inequality, for any $m \geq 1$,

$$\mathbb{P}\left(\sup_{x \in [0,1]} |\varXi_k(x) - \varXi_{k-1}(x)| \geq t\right) \leq 2^k \mathbb{P}\left(\prod_{i=1}^{k+1} U_i \geq t\right) \leq \left(\frac{2}{m+1}\right)^k t^{-m}.$$

For $k \geq 1$, let $\mathscr{D}_k = \{\ell 2^{-k} : \ell = 1, \ldots, 2^k - 1\}$. By construction, for $k \geq 1$, the map $x \to \varXi_k(x)$ is a right continuous step function. Further, it is continuous at x if and only if $x \notin \mathscr{D}_k$. Next, for $0 < q < 1$,

$$\begin{aligned}
\mathbf{E}\left[\sup_{x \in [0,1]} |\varXi_k(x) - \varXi_{k-1}(x)|\right] &= \int_0^\infty \mathbb{P}\left(\sup_{x \in [0,1]} |\varXi_k(x) - \varXi_{k-1}(x)| \geq t\right) dt \\
&\leq q^k + \int_{q^k}^\infty \left(\frac{2}{m+1}\right)^k t^{-m} dt \\
&= q^k + \frac{1}{m-1} \left(\frac{2}{m+1}\right)^k q^{-k(m-1)}.
\end{aligned}$$

With $m = 2$ and $q = \sqrt{2/3}$, the latter expression is bounded by $2q^k$. By Markov's inequality, it follows that $\sup_{m \geq n} \sup_{x \in [0,1]} |\varXi_m(x) - \varXi_n(x)| \to 0$ in

probability as $n \to \infty$. An application of the triangle inequality shows that $\sup_{m,p \geq n} \sup_{x \in [0,1]} |\varXi_m(x) - \varXi_p(x)| \to 0$ in probability as $n \to \infty$. By monotonicity, this convergence is almost sure. Thus, almost surely, $(\varXi_k)$ is uniformly Cauchy in the space of càdlàg functions endowed with the uniform topology. By completeness, $(\varXi_k)$ converges to a limit denoted by $\varXi$ with càdlàg paths. Moreover, $\varXi$ is continuous at $x \notin \mathscr{D}$ where $\mathscr{D} = \cup_{m \geq 1} \mathscr{D}_m$ since this is true for all $\varXi_k$, $k \geq 1$. For $x \in \mathscr{D}$, let $\varPhi(x)$ be the key of the node associated with $x_1 \ldots x_{k-1}$ where $k \geq 1$ is chosen minimal with $x \in \mathscr{D}_k$. Then, $\lim_{y \uparrow x} \varXi(x) = \varPhi(x) = \varXi(x)$. Thus, $x \mapsto \varXi(x)$ is continuous. By the construction of the tree, it is clear that $\varXi(x) < \varXi(y)$ for any $x, y \in \mathscr{D}$ with $x < y$. As $\mathscr{D}$ is dense in $[0, 1]$, the process $\varXi$ is strictly monotonically increasing. Obviously, $\varXi(0) = 0$ and $\varXi(1) = 1$; hence, $\varXi$ is the distribution function of a probability measure on $[0, 1]$. □

We turn to the convergence of $\mathscr{B}_n(x)$. For any fixed $x \in [0, 1]$, display (3) implies that, as $n \to \infty$, in probability, $B_n(x)/\log n \to 1$. Thus, (16) follows from the convergence $\varXi_k(x) \to \varXi(x)$. The convergence (16) is with respect to all moments since $B_n(x) \leq H_n$ and we have convergence of all moments in (2). By the theorem of dominated convergence, for any $m \geq 1$, again using (2), we have

$$\int_0^1 \mathbf{E}\left[\left|\frac{\mathscr{B}_n(x)}{\log n} - \varXi(x)\right|^m\right] dx \to 0.$$

This shows (17). To prove (18), note that, for any $k \geq 1$, $\sup_{x \in [0,1]} \mathscr{B}_n(x)$ is larger than the product of the height of the subtree rooted at the node $w_k := 1 \ldots 1$ on level k and $\varXi_{k-1}(\mathbf{1})$. Let $\varepsilon > 0$. Fix k large enough such that $\mathbb{P}(\varXi_{k-1}(\mathbf{1}) < 1 - \varepsilon) < \varepsilon$. Conditional on its size, the subtree rooted at w_k is a random binary search tree. Since its size grows linearly in n as $n \to \infty$, it follows from (2) that, for all n sufficiently large, its height exceeds $(c^* - \varepsilon) \log n$ with probability at least $1 - \varepsilon$. For these values of n, we have $\sup_{x \in [0,1]} \mathscr{B}_n(x) \geq (c^* - 6\varepsilon) \log n$ with probability at least $1 - 2\varepsilon$. As ε was chosen arbitrarily, this shows (18).

For the joint convergence of $B_n(x)$ and $\mathscr{B}_n(x)$ for fixed $x \in [0, 1]$, we abbreviate $B_n := B_n(x)$, $\mathscr{B}_n := \mathscr{B}_n(x)$, $\varXi_k := \varXi_k(x)$, $\varXi := \varXi(x)$ and $\bar{B}_n = (B_n - \log n)/\sqrt{\log n}$. Note that $\varXi$ and B_n are not independent which causes the proof to be more technical. Denote by N_k the time when the node associated with $x_1 \ldots x_k$ is inserted in the binary search tree. For any $\varepsilon > 0$, we can choose $k, L \geq 1$ such that, for all n sufficiently large,

$$\mathbb{P}\left(|\varXi_k - \varXi| \geq \varepsilon\right) + \mathbb{P}\left(N_k \geq L\right) + \mathbb{P}\left(\left|\frac{\mathscr{B}_n}{\log n} - \varXi\right| \geq \varepsilon\right) \leq \varepsilon.$$

Further, there exists $\delta > 0$ such that $\mathbb{P}(|\varXi_k - \varXi_{k-1}| \leq \delta) \leq \varepsilon$. Then, for $r, y \in \mathbb{R}$ with $\mathbb{P}(\varXi = y) = 0$, and n large enough,

$$\mathbb{P}\left(\bar{B}_n \leq r, \frac{\mathscr{B}_n}{\log n} \leq y\right) \leq 2\varepsilon + \mathbb{P}\left(\bar{B}_n \leq r, \varXi_k \leq y + 2\varepsilon, |\varXi_k - \varXi_{k-1}| \geq \delta, N_k < L\right).$$

Let $\bar{x} = x_{k+1}x_{k+2}\ldots$, $(V_1, V_2, \ldots)$ be an independent copy of $(U_1, U_2, \ldots)$ and

$$\text{Bin}(n, p) := \sum_{i=1}^{n} \mathbf{1}_{\{V_i \leq p\}}, \quad n \geq 0, \, p \in [0, 1].$$

Given $\Xi_k, |\Xi_k - \Xi_{k-1}|, N_k$, on $N_k < n$, $\bar{B}_n$ is distributed like $\bar{B}^*_{\text{Bin}(n-N_k,|\Xi_k-\Xi_{k-1}|)}(\bar{x}) + k/\sqrt{\log n}$ where $(B^*_n(\bar{x}))$ is distributed like $(B_n(\bar{x}))$ and independent from the remaining quantities. We deduce

$$\mathbb{P}\left(\bar{B}_n \leq r, \frac{\mathscr{B}_n}{\log n} \leq y\right)$$

$$\leq 2\varepsilon + \mathbb{P}\left(\frac{k}{\sqrt{\log n}} + \bar{B}^*_{\text{Bin}(n-L,\delta)}(\bar{x}) \leq r, \, \Xi_k \leq y + 2\varepsilon, \, |\Xi_k - \Xi_{k-1}| \geq \delta, \, N_k < L\right)$$

$$\leq 3\varepsilon + \mathbb{P}\left(\frac{k}{\sqrt{\log n}} + \bar{B}^*_{\text{Bin}(n-L,\delta)}(\bar{x}) \leq r\right) \mathbb{P}\left(\Xi \leq y + 2\varepsilon\right).$$

Using the asymptotic normality of $(\bar{B}^*_n(\bar{x}))$ (after rescaling) in (3), taking the limit superior as $n \to \infty$ and then letting ε tend to zero, we obtain

$$\limsup_{n\to\infty} \mathbb{P}\left(\bar{B}_n \leq r, \frac{\mathscr{B}_n}{\log n} \leq y\right) \leq \mathbb{P}\left(\mathscr{N} \leq r\right) \mathbb{P}\left(\Xi \leq y\right).$$

The proof of the converse direction establishing (19) is easier. It runs along the same lines upon using the trivial bounds $|\Xi_k - \Xi_{k-1}| \leq 1$ and $N_k \geq 0$.

Proof of Proposition 1 We start with the characterization of the distribution of the process. For a deterministic sequence of pairwise different numbers $u_1, u_2, \ldots$ on the unit interval, we define $\xi_k(x)$ analogously to $\Xi_k(x)$ in the infinite binary search tree constructed from this sequence. Here, we abbreviate $\xi_k(x) = 0$ if the node $x_1 \ldots x_k$ is not in the tree. Let $n_m^-, m \geq 1$, be the subsequence defined by the elements $u_{n_m^-} < u_1$ and $u_m^+, m \geq 1$, be the subsequence defined by the elements $u_{n_m^+} > u_1$. At least one of these sequences is infinite. For $m \geq 1$, let $y_m^- = u_{n_m^-}/u_1$ and $y_m^+ = (u_{n_m^+} - u_1)/(1 - u_1)$. Next, define ξ_k^- (ξ_k^+, respectively) analogously to ξ_k based on the sequence (y_m^-) $((y_m^+),$ respectively). By construction, for $k \geq 1$,

$$\xi_k(x) = \mathbf{1}_{[0,1/2)}(x)u_1\xi_{k-1}^-(2x) + \mathbf{1}_{[1/2,1]}(x)((1 - u_1)\xi_{k-1}^+(2x - 1) + u_1).$$

Applying the construction to the sequence $U_1, U_2, \ldots$ yields

$$\Xi_k(x) = \mathbf{1}_{[0,1/2)}(x)U_1\Xi_{k-1}^-(2x) + \mathbf{1}_{[1/2,1]}(x)((1 - U_1)\Xi_{k-1}^+(2x - 1) + U_1).$$

Almost surely, the random sequences y_m^- and y_m^+ are both infinite and $(\Xi_k^-), (\Xi_k^+)$ are independent copies of (Ξ_k). Further, both sequences are independent of U_1. Hence,

letting $k \to \infty$ in the last display, we obtain (22) on an almost sure level. The characterization of $\mathscr{L}(\varXi)$ by (22) follows from a standard contraction argument, and the argument on page 267 in [12] applies to our setting without any modifications. $\qquad\square$

We move on to the statements $(i) - (vi)$ on the marginal distributions of the process. Here, we use notation that was introduced in the proof of Theorem 3. By continuity, it suffices to show (i) for $x \in \mathscr{D}$. Let $k \geq 1$. By symmetry, for $1 \leq i \leq 2^k - 1$, we have $\mathbf{E}\left[\varPhi(i2^{-k})\right] = i2^{-k}$. Thus, the assertion follows for $x \in \mathscr{D}$ since $\varPhi(x) = \varXi(x)$. The symmetry statement (ii) is reminiscent of the fact that the uniform distribution on $[0, 1]$ is symmetric around $1/2$. More precisely, we apply the reflection argument from [1] which is at the core of the proof of the second assertion in (6). Let $U_1^* = 1 - U_1, U_2^* = 1 - U_2, \ldots$ and define $\varXi^*$ analogously to $\varXi$ in the binary search tree process relying on the sequence $U_1^*, U_2^*, \ldots$. Then, $\varXi^*(t) + \varXi(1 - t) = 1$ for all $t \in [0, 1]$ which proves (ii). With $Y = \varXi(\xi)$, (22) yields

$$\mathscr{L}(Y) = \mathscr{L}(UY + \mathbf{1}_A(1 - U)),$$

where $\mathbf{1}_A, U, Y$ are independent and $\mathbb{P}(A) = 1/2$. From [5], it follows that Y has the arcsine distribution, proving (iii). We move on to the statements about the distribution of $\varXi(t)$. Let $t \in (0, 1/2)$. Since $\varXi$ is strictly increasing, we have $\varXi(2t) \in (0, 1)$ almost surely. By (22), $\mathscr{L}(\varXi(t)) = \mathscr{L}(U\varXi(2t))$ with conditions as in (22). Therefore, $\mathscr{L}(\varXi(t))$ admits a density. By symmetry, the same is true for $t \in (1/2, 1)$. For $t \in (0, 1/2)$, by conditioning on the value of U, one finds the density

$$f_t(x) = \mathbf{E}\left[\frac{\mathbf{1}_{[x,1]}(\varXi(2t))}{\varXi(2t)}\right], \quad x \in (0, 1]. \tag{27}$$

$f_t(x)$ is monotonically decreasing and continuous on $(0, 1]$ with $f(1) = 0$. For $t \in (1/2, 1)$, $f_t(x) = f_{1-t}(1 - x), x \in (0, 1)$ is a density of $\mathscr{L}(\varXi(t))$ by (ii). By (27), for $t \in (0, 1/2), x \in (0, 1)$,

$$f_t(x) = \int_x^1 \frac{f_{2t}(y)}{y}dy, \quad \text{or} \quad xf_t'(x) = -f_{2t}(x). \tag{28}$$

Upon setting $f_0 = f_1 = 0$, the last identity also holds for $t = 0$ and $t = 1/2$ since $f_{1/2} = \mathbf{1}_{[0,1]}$ is a density of $\mathscr{L}(\varXi(1/2))$. Thus, for any $t \in (0, 1)$, f_t is smooth on $(0, 1)$. Since the uniform distribution takes values arbitrarily close to one, it follows that, for all $\delta > 0, t \in (0, 1)$, we have $\mathbb{P}(\varXi(t) > 1 - \delta) > 0$. Hence, for all $t \in (0, 1)$, the density f_t is strictly positive on $(0, 1)$. Thus, for $t \in (0, 1/2)$, f_t is strictly monotonically decreasing. Summarizing, we have shown (iv) and (v). For $t \in (0, 1/4]$, the assertion $\alpha_t^{(0)} = \infty$ in (vi) follows immediately from (28) since $\alpha_{2t}^{(0)} > 0$. Let $1/4 < t < 1/2$. Assume $\alpha_{1-2(1-2t)}^{(0)} < \infty$. Then, $f_{2(1-2t)}(1) < \infty$. By (28), it follows that $f_{1-2t}'(1)$ is finite and hence $f_{2t}'(0)$ is finite. Thus, $f_{2t}(y)/y$ is bounded in a neighbourhood of zero and $\alpha_t^{(0)} < \infty$. For $t > 3/8$, we have $1 - 2(1 - 2t) > 1/2$; thus, $\alpha_t^{(0)} < \infty$. Iterating this argument leads to $\alpha_t^{(0)} < \infty$ for all $1/3 < t < 1/2$.

In order to proceed further, note that, for $t > 1/4$, there exists $k \in \mathbb{N}$, such that, in probability, $\Xi(t) \geq Z := U_1(U_2 + (1 - U_2) \prod_{\ell=1}^{k} U_{2+\ell})$. Z admits a density f_Z given by

$$f_Z(x) = 1 + \int_x^1 r(y)dy - xr(x), \quad r(x) = \frac{1}{x^2} \int_0^x \mathbb{P}\left(\prod_{\ell=1}^{k} U_{2+\ell} \leq \frac{x-v}{1-v} \right) dv.$$

Thus,

$$\lim_{x \downarrow 0} f_Z(x) = 1 + \int_0^1 r(y)dy < \infty.$$

It follows that $\alpha_t^{(0)} \leq 1 + \int_0^1 r(x)dx < \infty$. Since Ξ is increasing, the function $t \mapsto \alpha_t^{(0)}$ is decreasing. Thus, by monotonicity and continuity, it follows $\alpha_t^{(0)} \uparrow \infty$ as $t \downarrow 1/4$. For $t \leq 1/4$, $\alpha_t^{(0)} = \infty$ follows immediately from (28) since $\alpha_{2t}^{(0)} < \infty$. For $1/4 < t < 1/2$, the remaining statements about $\alpha_t^{(1)}$ are direct corollaries of the results for $\alpha_t^{(0)}$ since $\alpha_t^{(1)} = \alpha_{1-2(1-2t)}^{(0)}$. This finishes the proof of (vi).

The curvature We make a concluding remark about the curvature of f_t, $t \in (0, 1/2)$. First, since $x f_t''(x) = -f_{2t}'(x) - f_t'(x)$, for $0 < t \leq 1/4$, the function f_t is convex. From (28) it is easy to deduce $f_{1/3}(x) = 2(1 - x)$. Since $f_{1/3}'' = f_{1/2}'' = 0$, it is plausible to conjecture that f_t is convex for $t \leq 1/3$ and concave for $1/3 \leq t < 1/2$. Concavity at rational points with small denominator such as $t = 3/8$ or $t = 5/12$ can be verified by hand using (28).

4.3 Weighted Path Length and Wiener Index

In order to obtain mean and variance for the weighted path length and the weighted Wiener index, we use the reflection argument from the proof of Proposition 1 (ii). To this end, let $\mathscr{P}_n^*$ and $\mathscr{W}_n^*$ denote weighted path length and weighted Wiener index in the binary search tree built from the sequence $U_1^* = 1 - U_1$, $U_2^* = 1 - U_2, \ldots$ Then, $\mathscr{P}_n + \mathscr{P}_n^* = P_n + n$ and $\mathscr{W}_n + \mathscr{W}_n^* = W_n + \binom{n}{2}$ providing the claimed expansions for $\mathbf{E}[\mathscr{P}_n]$ and $\mathbf{E}[\mathscr{W}_n]$ upon recalling (7) and (8).

For a finite rooted labelled binary tree T, denote by $p(T)$ its path length, by $\mathbf{p}(T)$ its weighted path length, by $w(T)$ its Wiener index and by $\mathbf{w}(T)$ its weighted Wiener index. Let T_1, T_2 be its left and right subtree and x the label of the root. Then, denoting by $|T|$ the size of T, for $|T| \geq 1$,

$$p(T) = p(T_1) + p(T_2) + |T| - 1, \tag{29}$$
$$w(T) = w(T_1) + w(T_2) + (|T_2| + 1)p(T_1) + (|T_1| + 1)p(T_2) + |T| + 2|T_1||T_2| - 1. \tag{30}$$

The first statement is obvious, the argument for the second can be found in [25]. For the weighted quantities, one obtains

$$\mathbf{p}(T) = \mathbf{p}(T_1) + \mathbf{p}(T_2) + |T|x, \tag{31}$$

$$\mathbf{w}(T) = \mathbf{w}(T_1) + \mathbf{w}(T_2) + (|T_2| + 1)\mathbf{p}(T_1) + (|T_1| + 1)\mathbf{p}(T_2) + (|T| + |T_1||T_2|)x. \tag{32}$$

Again, the first assertion is easy to see and we only justify the second. The terms $\mathbf{w}(T_1)$ and $\mathbf{w}(T_2)$ account for weighted distances within the subtrees. The sum of all weighted distances between nodes in the left subtree and the root equals $\mathbf{p}(T_1) + |T_1|x$. Replacing T_1 by T_2, we obtain the analogous sum in the right subtree. The sum of all distances between nodes in different subtrees equals $|T_1|\mathbf{p}(T_2) + |T_2|\mathbf{p}(T_1) + |T_1||T_2|x$. Finally, we need to add x for the weighted distance of the root to itself. Adding up the terms and simplifying leads to (32). For $\alpha, \beta > 0$ let $\alpha T + \beta$ be the tree obtained from T where each label y is replaced by $\alpha y + \beta$. Obviously, $p(T) = p(\alpha T + \beta)$ with the analogous identity for the Wiener index. For the weighted quantities, we have

$$\mathbf{p}(\alpha T + \beta) = \alpha \mathbf{p}(T) + (p(T) + |T|)\beta, \tag{33}$$

$$\mathbf{w}(\alpha T + \beta) = \alpha \mathbf{w}(T) + (w(T) + |T|(|T| + 1)/2)\beta. \tag{34}$$

Let T be the binary search tree of size n in the i.i.d. model. Then, given $I_n := \mathrm{rank}(U_1)$, $U := U_1$, in distribution, the trees $\frac{1}{U}T_1$ and $\frac{1}{1-U}T_2 - \frac{U}{1-U}$ are independent binary search trees of size $I_n - 1$ and $n - I_n$, constructed from independent sequences of uniformly distributed random variables on $[0, 1]$. Thus, combining (29)–(34), for the vector $Y_n = (\mathscr{W}_n, W_n, \mathscr{P}_n, P_n)^T$, we have

$$Y_n \overset{d}{=} \begin{bmatrix} U & 0 & (n+1-I_n)U & 0 \\ 0 & 1 & 0 & n+1-I_n \\ 0 & 0 & U & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} Y_{I_n-1}$$

$$+ \begin{bmatrix} 1-U & U & I_n(1-U) & I_nU \\ 0 & 1 & 0 & I_n \\ 0 & 0 & 1-U & U \\ 0 & 0 & 0 & 1 \end{bmatrix} Y'_{n-I_n}$$

$$+ \begin{pmatrix} (2n + (n-I_n)(3I_n + n - 2))U/2 \\ n - 1 + 2(I_n - 1)(n - I_n) \\ (2n - I_n)U \\ n - 1 \end{pmatrix},$$

where (Y'_n), (Y_n), (I_n, U) are independent and (Y'_n) is distributed like (Y_n). Here, $\overset{d}{=}$ indicates that left- and right-hand side are identically distributed.

We consider the sequence $(Z_n)_{n\geq 0}$ defined by

$$Z_n := \left(\frac{\mathscr{W}_n - \mathbb{E}[\mathscr{W}_n]}{n^2}, \frac{W_n - \mathbb{E}[W_n]}{n^2}, \frac{\mathscr{P}_n - \mathbb{E}[\mathscr{P}_n]}{n}, \frac{P_n - \mathbb{E}[P_n]}{n} \right)^T, \quad n \geq 1,$$

and $Z_0 = 0$. Let $\alpha_n = \mathbf{E}[\mathcal{W}_n]$, $\beta_n = \mathbf{E}[W_n]$, $\gamma_n = \mathbf{E}[\mathcal{P}_n]$ and $\delta_n = \mathbf{E}[P_n]$. Further, let

$$
A_1^{(n)} =
\begin{bmatrix}
\left(\frac{I_n-1}{n}\right)^2 U & 0 & \left(1 - \frac{I_n-1}{n}\right)\frac{I_n-1}{n}U & 0 \\
0 & \left(\frac{I_n-1}{n}\right)^2 & 0 & \left(1 - \frac{I_n-1}{n}\right)\frac{I_n-1}{n} \\
0 & 0 & \frac{I_n-1}{n}U & 0 \\
0 & 0 & 0 & \frac{I_n-1}{n}
\end{bmatrix},
$$

$$
A_2^{(n)} =
\begin{bmatrix}
\left(1 - \frac{I_n}{n}\right)^2(1-U) & \left(1 - \frac{I_n}{n}\right)^2 U & \frac{I_n}{n}\left(1 - \frac{I_n}{n}\right)(1-U) & \frac{I_n}{n}\left(1 - \frac{I_n}{n}\right)U \\
0 & \left(1 - \frac{I_n}{n}\right)^2 & 0 & \frac{I_n}{n}\left(1 - \frac{I_n}{n}\right) \\
0 & 0 & \left(1 - \frac{I_n}{n}\right)(1-U) & \left(1 - \frac{I_n}{n}\right)U \\
0 & 0 & 0 & 1 - \frac{I_n}{n}
\end{bmatrix},
$$

and $C^{(n)} = (C_1^{(n)}, C_2^{(n)}, C_3^{(n)}, C_4^{(n)})^T$ with

$$
\begin{aligned}
C_1^{(n)} &= \frac{U}{n^2}\alpha_{I_n-1} + \frac{1-U}{n^2}\alpha_{n-I_n} + \frac{U}{n^2}\beta_{n-I_n} + U\frac{(n+1-I_n)}{n^2}\gamma_{I_n-1} + (1-U)\frac{I_n}{n^2}\gamma_{n-I_n} \\
&\quad + U\frac{I_n}{n^2}\delta_{n-I_n} + U\frac{2n+(n-I_n)(3I_n+n-2)}{2n^2} - \frac{1}{n^2}\alpha_n, \\
C_2^{(n)} &= \frac{1}{n^2}\beta_{I_n-1} + \frac{1}{n^2}\beta_{n-I_n} + \left(1 - \frac{I_n-1}{n}\right)\frac{1}{n}\delta_{I_n-1} + \frac{I_n}{n^2}\delta_{n-I_n} \\
&\quad + \frac{n-1+2(n-1)(n-I_n)}{n^2} - \frac{1}{n^2}\beta_n, \\
C_3^{(n)} &= \frac{U}{n}\gamma_{I_n-1} + \frac{1-U}{n}\gamma_{n-I_n} + \frac{U}{n}\delta_{n-I_n} + \left(2 - \frac{I_n}{n}\right)U - \frac{1}{n}\gamma_n, \\
C_4^{(n)} &= \frac{1}{n}\delta_{I_n-1} + \frac{1}{n}\delta_{n-I_n} + 1 - \frac{1}{n} - \frac{1}{n}\delta_n.
\end{aligned}
$$

Then, from the recurrence for (Y_n), it follows

$$
Z_n \overset{d}{=} A_1^{(n)} Z_{I_n-1} + A_2^{(n)} Z'_{n-I_n} + C^{(n)}, \quad n \geq 1,
$$

where (Z_n), (Z'_n), (I_n, U) are independent and (Z'_n) is distributed like (Z_n). We prove convergence of Z_n in distribution by an application of the contraction method. To this end, note that $I_n/n \to U$ almost surely by the strong law of large numbers. Thus, with convergence in L_2 and almost surely,

$$
A_1^{(n)} \to A_1 :=
\begin{bmatrix}
U^3 & 0 & U^2(1-U) & 0 \\
0 & U^2 & 0 & U(1-U) \\
0 & 0 & U^2 & 0 \\
0 & 0 & 0 & U
\end{bmatrix},
$$

$$A_2^{(n)} \to A_2 := \begin{bmatrix} (1-U)^3 & U(1-U)^2 & U(1-U)^2 & U^2(1-U) \\ 0 & (1-U)^2 & 0 & U(1-U) \\ 0 & 0 & (1-U)^2 & U(1-U) \\ 0 & 0 & 0 & 1-U \end{bmatrix},$$

and

$$C^{(n)} \to C := \begin{pmatrix} U^2 \log U + (1-U^2) \log(1-U) + U(-14U^2 + 9U + 5)/4 \\ 2U \log U + 2(1-U) \log(1-U) + 6U(1-U) \\ U^2 \ln U + (1-U^2) \ln(1-U) + U \\ 2U \ln U + 2(1-U) \ln(1-U) + 1 \end{pmatrix}.$$

For a quadratic matrix A, denote by $\|A\|_{\mathrm{op}}$ its spectral radius. By calculating the eigenvalues of $A_1 A_1^T$ and $A_2 A_2^T$, one checks that $\|A_1\|_{\mathrm{op}} = U$ and $\|A_2\|_{\mathrm{op}} = 1 - U$. Thus,

$$\mathbf{E}\left[\|A_1 A_1^T\|_{\mathrm{op}}\right] + \mathbf{E}\left[\|A_2 A_2^T\|_{\mathrm{op}}\right] \le \mathbf{E}\left[\|A_1\|_{\mathrm{op}}^2\right] + \mathbf{E}\left[\|A_2\|_{\mathrm{op}}^2\right] < 1.$$

Moreover, we have $\mathbb{P}(I_n \in \{1, \ldots, \ell\} \cup \{n\}) \to 0$ for all fixed ℓ. Thus, by Theorem 4.1 in [24], in distribution and with convergence of the first two moments, we have $Z_n \to (\mathscr{W}, W, \mathscr{P}, P)$ where $\mathscr{L}(\mathscr{W}, W, \mathscr{P}, P)$ is the unique fixed-point of the map:

$$T : \mathscr{M}_2^4(0) \longrightarrow \mathscr{M}_2^4(0), \quad T(\mu) = \mathscr{L}\left(A_1 Z + A_2 Z' + C\right), \qquad (35)$$

with A_1, A_2, C defined above, where Z, Z', U are independent and $\mathscr{L}(Z) = \mathscr{L}(Z') = \mu$. Here, $\mathscr{M}_2^4(0)$ denotes the set of probability measures on $\mathbb{R}^4$ with finite absolute second moment and zero mean. Variances and covariances can be computed successively using the fixed-point equation, e.g. in the following order: $\mathbf{E}\left[P^2\right]$, $\mathbf{E}[PW]$, $\mathbf{E}\left[W^2\right]$, $\mathbf{E}[P\mathscr{P}]$, $\mathbf{E}\left[\mathscr{P}^2\right]$, $\mathbf{E}[\mathscr{P}W]$, $\mathbf{E}[P\mathscr{W}]$, $\mathbf{E}[W\mathscr{W}]$, $\mathbf{E}[\mathscr{P}\mathscr{W}]$, $\mathbf{E}\left[\mathscr{W}^2\right]$. Additionally to the variances given in the theorem, one obtains

$$\mathrm{Cov}(P_n, \mathscr{P}_n) \sim \frac{21 - 2\pi^2}{6} n^2, \quad \mathrm{Cov}(P_n, W_n) \sim \frac{20 - 2\pi^2}{3} n^3, \qquad (36)$$

$$\mathrm{Cov}(\mathscr{P}_n, W_n) \sim \frac{10 - \pi^2}{3} n^3, \quad \mathrm{Cov}(P_n, \mathscr{W}_n) \sim \frac{10 - \pi^2}{3} n^3, \qquad (37)$$

$$\mathrm{Cov}(W_n, \mathscr{W}_n) \sim \frac{10 - \pi^2}{3} n^4, \quad \mathrm{Cov}(\mathscr{P}_n, \mathscr{W}_n) \sim \frac{481 - 48\pi^2}{288} n^3. \qquad (38)$$

References

1. Aguech, R., Lasmar, N., Mahmoud, H.: Extremal weighted path lengths in random binary search trees. Probab. Eng. Inform. Sci. **21**(1), 133–141 (2007)
2. Billingsley, P.: Convergence of Probability Measures. Wiley Series in Probability and Statistics: Probability and Statistics. Wiley, New York (1999)
3. Broutin, N., Devroye, L.: Large deviations for the weighted height of an extended class of trees. Algorithmica **46**(3–4), 271–297 (2006)
4. Brown, G.G., Shubert, B.O.: On random binary trees. Math. Oper. Res. **9**(1), 43–65 (1984)
5. Chen, R., Lin, E., Zame, A.: Another arc sine law. Sankhyā Ser. A **43**(3), 371–373 (1981)
6. Devroye, L.: A note on the height of binary search trees. J. Assoc. Comput. Mach. **33**(3), 489–498 (1986)
7. Devroye, L.: Applications of the theory of records in the study of random trees. Acta Inform. **26**(1–2), 123–1301 (1988)
8. Devroye, L.: Limit laws for local counters in random binary search trees. Random Struct. Algorithms **2**(3), 303–315 (1991)
9. Devroye, L., Neininger, R.: Distances and finger search in random binary search trees. SIAM J. Comput. **33**(3), 647–658 (2004)
10. Dickman, K.: On the frequency of numbers containing prime factors of a certain relative magnitude. Arkiv för Matematik, Astronomi och Fysik **22A**(10), 1–14 (1930)
11. Dvoretzky, A., Kiefer, J., Wolfowitz, J.: Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Stat. **27**, 642–669 (1956)
12. Grübel, R., Rösler, U.: Asymptotic distribution theory for Hoare's selection algorithm. Adv. Appl. Probab. **28**(1), 252–269 (1996)
13. Grübel, R.: On the silhouette of binary search trees. Ann. Appl. Probab. **19**(5), 1781–1802 (2009)
14. Grübel, R., Stefanoski, N.: Mixed Poisson approximation of node depth distributions in random binary search trees. Ann. Appl. Probab. **15**(1A), 279–297 (2005)
15. Hildebrand, A., Tenenbaum, G.: Integers without large prime factors. J. Théor. Nombres Bordx. **5**(2), 411–484 (1993)
16. Hoare, C.A.R.: Quicksort. Comput. J. **5**, 10–15 (1962)
17. Hwang, H.-K., Tsai, T.-H.: Quickselect and the Dickman function. Combin. Probab. Comput. **11**(4), 353–371 (2002)
18. Knuth, D.E.: The Art of Computer Programming: Sorting and Searching, vol. 3. Addison-Wesley, Reading (1973)
19. Kuba, M., Panholzer, A.: On weighted path lengths and distances in increasing trees. Probab. Eng. Inf. Sci. **21**(3), 419–433 (2007)
20. Kuba, M., Panholzer, A.: On edge-weighted recursive trees and inversions in random permutations. Discrete Math. **308**(4), 529–540 (2008)
21. Mahmoud, H.M.: Evolution of Random Search Trees. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, New York (1992)
22. Mahmoud, H.M., Neininger, R.: Distribution of distances in random binary search trees. Ann. Appl. Probab. **13**(1), 253–276 (2003)
23. Mahmoud, H.M., Pittel, B.: On the most probable shape of a search tree grown from a random permutation. SIAM J. Algebr. Discrete Methods **5**(1), 69–81 (1984)
24. Neininger, R.: On a multivariate contraction method for random recursive structures with applications to quicksort. Random Struct. Algorithms **19**(3–4), 498–524 (2001). Analysis of algorithms (Krynica Morska, 2000)
25. Neininger, R.: The Wiener index of random trees. Combin. Probab. Comput. **11**(6), 587–597 (2002)
26. Panholzer, A., Prodinger, H.: Spanning tree size in random binary search trees. Ann. Appl. Probab. **14**(2), 718–733 (2004)
27. Régnier, M.: A limiting distribution for quicksort. RAIRO Inform. Théor. Appl. **23**(3), 335–343 (1989)
28. Rösler, U.: A limit theorem for "Quicksort". RAIRO Inform. Théor. Appl. **25**, 85–100 (1991)
29. Rüschendorf, L., Schopp, E.-M.: Note on the weighted internal path length of b-ary trees. Discrete Math. Theor. Comput. Sci. **9**(1), 1–6 (2007)