DISTRIBUTION OF DISTANCES IN RANDOM BINARY SEARCH TREES

BY HOSAM M. MAHMOUD AND RALPH NEININGER[1]

George Washington University and McGill University

We investigate random distances in a random binary search tree. Two types of random distance are considered: the depth of a node randomly selected from the tree, and distance between randomly selected pairs of nodes. By a combination of classical methods and modern contraction techniques we arrive at a Gaussian limit law for normed random distances between pairs. The exact forms of the mean and variance of this latter distance are first derived by classical methods to determine the scaling properties, then used for norming, and the normed random variable is then shown by the contraction method to have a normal limit arising as the fixed-point solution of a distributional equation. We identify the rate of convergence in the limit law to be of the order $\Theta(1/\sqrt{\ln n})$ in the Zolotarev metric ζ_3. In the analysis we need the rate of convergence in the central limit law for the depth of a node, as well. This limit law was derived before by various techniques. We establish the rate $\Theta(1/\sqrt{\ln n})$ in ζ_3.

1. Introduction. Distances between nodes in random combinatorial objects is an interesting topic in the study of random structures and algorithms, it is related to the cost of finger searches [see Seidel and Aragon (1996)] and also has applications in many other scientific fields. For example, the collective sum of all such distances in the graph underlying a molecule is known in chemistry as the *Wiener index* [see Gutman and Polansky (1986) and Trinajstić (1992)].

In this paper we look at the distances between distinct pairs of nodes in random binary search trees. The investigation starts with the derivation of mean and variance by classical methods and leads to a normal characterization via the contraction method. The study illustrates in a vivid manner the strength and shortcomings of this method. In the course of the study, one needs to use results on the asymptotic behavior of the depth of a randomly selected node in the tree. The depth coincides up to an additive constant with the number of comparisons required for a random successful search in the tree. While asymptotic normality for the depth has been shown by various different analytic and probabilistic techniques, the contraction method faces some difficulty in characterizing the normal limit distribution underlying this random variable. The difficulty arises in the degenerate nature of the distributional limit equation. Asymptotic normality

with an estimate for the rate of convergence could, however, be proven via a subtle application of the contraction method in Cramer and Rüschendorf (1996) under some (still unverified) assumption. We will close this gap and transfer the rate of convergence for the depth into a central limit law for the distance of random pairs including a rate of convergence.

One might be able to go forward with classical methods in the study of the distance between a randomly chosen pair of nodes in the tree. Indeed, one is able to use such classical methods to find the mean and variance of distances, and in principle higher moments, too, in both exact and asymptotic form. However, soon enough one realizes that there is a combinatorial explosion, and each higher moment is much more difficult to obtain than the one before: classical methods may be prohibitive in this context. The way to surmount such difficulty is to formulate the limit distribution of the (normed) distance as a fixed-point solution of a distributional equation involving the limit law of the (normed) random depth of a randomly selected node serving as a toll distribution. Now, in this limit equation we have an extra leg to go on; the normality of the toll will lead us in a simple way to demonstrate the normality of the normed distance. Moreover, the method will provide a rate for this convergence by making estimates explicit. The entire exercise will then serve as an illustration of how classical techniques can go hand-in-hand with modern contraction methods to arrive at limit laws in combinatorial structures.

2. The framework.   A binary tree is a random structure of nodes each having no children, one left child, one right child or two children (one left and one right). Many sorting and other combinatorial algorithms are modeled by labeled binary trees endowed with a *search property*. According to the search property, the label of any node is larger than the labels in its left subtree and no greater than any label in its right subtree. For definitions and combinatorial properties see Mahmoud (1992), and for applications in sorting see Knuth (1998) or Mahmoud (2000).

Several models of randomness are in common use on binary trees. The uniform model in which all trees are equally likely is useful in formal language studies, compilers, computer algebra, etc. Kemp (1984) is a good source for this subject. The *random permutation model* conforms more closely to sorting applications and several other uses of binary search trees as data structures. In the random permutation probability model, we assume that the tree is built from permutations of $\{1, \ldots, n\}$, where a uniform probability model is imposed on the permutations instead of the trees. When all $n!$ permutations are equally likely or *random*, binary search trees are not equally likely. Several permutations may give the same tree, and generally the model is biased toward shorter and well-balanced trees rather than scrawny and tall ones, which is a desirable property for fast search applications [see Mahmoud (1992)].

We study distances in binary search trees grown from random permutations. The term *random tree* will refer to a tree built from a random permutation. A tree grows from a permutation $(\pi_1, \ldots, \pi_n)$ of $\{1, \ldots, n\}$ as follows. The first element

π_1 is inserted in an empty tree and thus goes into the root of a new nonempty subtree. Each subsequent element π_j is taken to the left or right subtree according as whether it is smaller or larger than the root. In the subtree, the element is inserted recursively. The search continues until an empty subtree is found where the element is inserted into a new node, just like π_1 was initially inserted in the root.

The chief objective is to study the distribution of Δ_n, the distance between two distinct randomly selected nodes in a random tree of size n, where again the pair choice is uniform in the sense that all $\binom{n}{2}$ pairs are equally likely. A recursive formulation for Δ_n goes via D_n, the depth of a randomly selected node in a random tree of size n, with random meaning all nodes are equally likely choices.

We shall use standard notation. From algorithmics, we borrow the notation $H_n^{(j)}$ to denote the nth harmonic number of order j; that is, $\sum_{k=1}^n 1/k^j$. As customary, the superscript will be dropped when it is 1. We also use $\langle x \rangle_j$ to denote the rising factorial $x(x+1)\cdots(x+j-1)$, and ∇ as the backward difference operator, that is, for a sequence (h_n), $\nabla h_n = h_n - h_{n-1}$. From probability theory we need the symbols $\overset{\mathcal{L}}{=}$ and $\overset{\mathcal{L}}{\longrightarrow}$ to denote exact equality and convergence in distribution, respectively. The distribution or law of a random variable X will be denoted by $\mathcal{L}(X)$. The notation $\overset{\mathcal{P}}{\longrightarrow}$ will mean convergence in probability, and $\overset{\text{a.s.}}{\longrightarrow}$ will mean convergence almost surely. The Bernoulli random variable with rate of success p per trial will be denoted by $\text{Ber}(p)$, the normally distributed with mean μ and variance σ^2 will be denoted by $\mathcal{N}(\mu, \sigma^2)$ and the uniform on the interval (a, b) will be denoted by $\text{unif}(a, b)$. We will denote the L_p norm of X by $\|X\|_p$. For measuring distances between probability distributions we use the Zolotarev metric ζ_3 given for distributions $\mathcal{L}(X), \mathcal{L}(Y)$ by

$$\zeta_3\big(\mathcal{L}(X), \mathcal{L}(Y)\big) := \sup_{f \in \mathcal{F}_3} |\mathbf{E}[f(X)] - \mathbf{E}[f(Y)]|,$$

where $\mathcal{F}_3 := \{f \in C^2(\mathbb{R}, \mathbb{R}) : |f''(x) - f''(y)| \le |x - y|\}$ is the space of all twice differentiable functions with second derivative being Lipschitz continuous with Lipschitz constant 1. We will use the brief notation $\zeta_3(X, Y) := \zeta_3(\mathcal{L}(X), \mathcal{L}(Y))$. Finally, C denotes a universal constant, which might be different from place to place.

It is well known that convergence in ζ_3 implies weak convergence. We recall that $\zeta_3(X, Y) < \infty$ if $\mathbf{E}[X] = \mathbf{E}[Y]$, $\mathbf{E}[X^2] = \mathbf{E}[Y^2]$ and $\mathbf{E}[|X|^3], \mathbf{E}[|Y|^3] < \infty$, and we may assume that X, Y satisfy these conditions subsequently. The key property being used is that ζ_3 is a $(3, +)$-ideal metric; that is, for Z independent of (X, Y) with $\mathbf{E}[|Z|^3] < \infty$, and $c \ne 0$ we have

$$\zeta_3(X + Z, Y + Z) \le \zeta_3(X, Y), \qquad \zeta_3(cX, cY) = |c|^3 \zeta_3(X, Y).$$

We will use the following bounds for ζ_3:

(1) $\quad \frac{1}{6}|\mathbf{E}[X^3] - \mathbf{E}[Y^3]| \le \zeta_3(X, Y) \le \frac{1}{2}(\|X\|_3^2 + \|X\|_3 \|Y\|_3 + \|Y\|_3^2)\|X - Y\|_3,$

where the right inequality holds for any choice of X, Y on a joint probability space. For properties of this metric we refer the reader to Zolotarev (1976) and Rachev (1991).

To develop results for Δ_n, one has to go through the necessary catalyst D_n, and therefore we shall take this up first. For the number S_n of comparisons exercised by a successful search in the tree we have $D_n \overset{\mathcal{L}}{=} S_n - 1$. We restate the results known about S_n in terms of D_n and endow the limit law with the correct order of the rate of convergence in ζ_3, which will later be transferred into a rate of convergence for Δ_n. The moments and limit law for D_n were obtained by classic recurrence and generating function techniques [see Mahmoud (1992), Section 2.5], as well as various probabilistic techniques [see Devroye (1988, 1999)]. These results include

$$
(2) \qquad \mathbf{E}[D_n] = \frac{2(n+1)}{n} H_n - 4 = 2 \ln n + O(1),
$$

$$
\mathbf{Var}[D_n] = \left(2 + \frac{10}{n}\right) H_n - 4\left(1 + \frac{1}{n}\right)\left(\frac{H_n^2}{n} + H_n^{(2)}\right) + 4 \sim 2 \ln n,
$$

$$
(3) \qquad \mathbf{E}[z^{D_n}] = \frac{1}{n(2z-1)}\left[\frac{\langle 2z \rangle_n}{n!} - 1\right],
$$

$$
D_n^* := \frac{D_n - 2 \ln n}{\sqrt{2 \ln n}} \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, 1).
$$

An estimate of the rate of convergence in ζ_3 for this limit law was previously discussed in Cramer and Rüschendorf (1996). For later use we extend their results to the following.

THEOREM 1. *The depth D_n of a randomly selected node in a random binary search tree of size n satisfies*

$$
\zeta_3\left(\frac{D_n - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}, \mathcal{N}(0, 1)\right) = \Theta\left(\frac{1}{\sqrt{\ln n}}\right).
$$

A direct onslaught on D_n by the contraction method gives us a tantalizing result. If the limit of D_n^* is Y, say, the contraction method yields the distribution of Y as a fixed-point solution of the equation

$$
Y \overset{\mathcal{L}}{=} IY + (1 - I)\tilde{Y},
$$

where $\tilde{Y} \overset{\mathcal{L}}{=} Y$, I is a Ber($\frac{1}{2}$) random variable and Y, $\tilde{Y}$ and I are independent. It can be easily argued that any distribution satisfies this equation. Such degeneracy of the fixed-point equation typically necessitates a refined use of the contraction method [see Cramer and Rüschendorf (1996)].

Classical methods can aid us to obtain moments of Δ_n in exact and asymptotic form. By contrast, these same classical methods become much too complicated to obtain the limit distribution of (normed) Δ_n. So, the classic methods will take us halfway to determining the scaling properties for appropriate norming of Δ_n. From there, the contraction method will provide a shortcut to the

limit distribution. The method was introduced in the context of Quicksort by Rösler (1991). Several extensions, using various probability metrics, were added by Rachev and Rüschendorf (1995), and recently more universal contraction theorems and multivariate extensions were added by Rösler (2001), Neininger (2001) and Neininger and Rüschendorf (2002b); for a useful survey see Rösler and Rüschendorf (2001). Contexts similar to our current setup where bivariate contraction was used appear in the study of the Wiener index [Neininger (2002)]. The contraction method will give us the distribution of the (normed) distance between random pairs as the fixed-point solution to the equation

$$X \overset{\mathcal{L}}{=} IX + J\tilde{X} + K\frac{1}{\sqrt{2}}(Y + \tilde{Y}),$$

with (X, Y), $(\tilde{X}, \tilde{Y})$ and (I, J, K) being independent, where $(X, Y) \overset{\mathcal{L}}{=} (\tilde{X}, \tilde{Y})$, and (I, J, K) are mutually exclusive indicators, with marginal distributions like that of a $\text{Ber}(\frac{1}{3})$ random variable, and Y, $\tilde{Y}$ are standard normal variates arising as the limits of D_n^*, with convergence rate as in Theorem 1. Although we have several dependencies here, as for example among the indicators, and that between X and Y, the normality of Y, $\tilde{Y}$ and the interplay of these variables through mutually exclusive indicators simplifies the outcome. One is able to get a limiting Gaussian law. Our main result states the following.

THEOREM 2. *The distance Δ_n between two randomly selected nodes in a random binary search tree of size n has mean value*

$$\mathbf{E}[\Delta_n] = \frac{4(n+1)}{n(n-1)}[(n+3)H_{n+1} - 3n - 3] = 4\ln n + O(1),$$

and variance

$$\begin{aligned}
\mathbf{Var}[\Delta_n] = \frac{2}{n^2(n-1)^2}\big[&(2n^4 + 194n^3 + 534n^2 + 294n)H_n \\
&- (44n^3 + 184n^2 + 212n + 72)H_n^2 \\
&- (8n^4 + 20n^3 - 8n^2 - 20n)H_n^{(2)} \\
&+ 15n^4 - 190n^3 - 337n^2\big]
\end{aligned}$$

$$\sim 4\ln n.$$

Furthermore, the normalized random variable is asymptotically Gaussian:

$$\zeta_3\left(\frac{\Delta_n - \mathbf{E}[\Delta_n]}{\sqrt{\mathbf{Var}[\Delta_n]}}, \mathcal{N}(0, 1)\right) = \Theta\left(\frac{1}{\sqrt{\ln n}}\right),$$

in particular,

$$\frac{\Delta_n - 4\ln n}{\sqrt{4\ln n}} \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, 1).$$

The expression for $\mathbf{E}[\Delta_n]$ has already been given in terms of $\mathbf{E}[W_n]$, the mean of the Wiener index of a random binary search tree, in Hwang and Neininger (2002). We have $\mathbf{E}[W_n] = \binom{n}{2}\mathbf{E}[\Delta_n]$.

Several consequences of bounds on the ζ_3 convergence as in our Theorems 1 and 2 regarding other distance measures were discussed in Neininger and Rüschendorf [(2002a), Section 3]. We could also give a convergence proof in the Wasserstein metrics, which are more common in the context of the contraction method. The reason we make use of ζ_3 is to demonstrate how the rate of convergence of Theorem 1 can be transferred into the rate of Theorem 2.

Toward our goals, the paper is organized as follows. Theorem 1 on D_n is proved in Section 3 and used later in the analysis of Δ_n. We return to Δ_n in Section 4, where the mean and variance are handled by classical recurrence methods, but the distribution is obtained as a fixed-point solution of a limiting distributional equation; convergence to that limit is shown via convergence in the ζ_3 metric. The core of the proof of convergence in the ζ_3 metric is relegated to Section 5. Section 6 provides some discussion of the merits of the contraction method.

3. The depth of a random node. In a binary search tree generated from a random permutation of $\{1, \ldots, n\}$, select a node uniformly at random. Let the depth of the random node be D_n. Condition on the label of the root, R. The outcome of the random selection may fall in the left subtree, may be the root or may fall in the right subtree. The left subtree has a recursive structure like a binary search tree built from a permutation of $\{1, \ldots, R - 1\}$, and symmetrically the right subtree has a recursive structure like that of a tree built from a random permutation of $\{1, \ldots, n - R\}$, but with labels upgraded by adding R to each. Thus,

$$D_n = \begin{cases} D_{R-1}, & \text{with probability } (R - 1)/n, \\ 0, & \text{with probability } 1/n, \\ \tilde{D}_{n-R}, & \text{with probability } (n - R)/n, \end{cases}$$

where $(D_1, \ldots, D_n)$ and $(\tilde{D}_1, \ldots, \tilde{D}_n)$ are independent and identically distributed. This formulation will help us put the limit of the normed variable D_n^* in the form of a fixed-point solution of a limiting distributional equation. The lack of a toll term in that equation will put the contraction method in perspective.

To prove Theorem 1, we need an estimate of the third moment of D_n. Differentiating the probability generating function three times [cf. (3)], and evaluating at $z = 1$, gives us the third factorial moment $\mathbf{E}[D_n(D_n - 1)(D_n - 2)]$. However, we also have the first and second moments and can therefore extract the exact third moment. After a lengthy mechanical calculation, one finds

$$\mathbf{E}[D_n^3] = \frac{2(n + 1)}{n}\bigg[4H_{n+1}^3 - 18H_{n+1}^2 + (49 - 12H_{n+1}^{(2)})H_{n+1}$$
$$+ 8H_{n+1}^{(3)} + 18H_{n+1}^{(2)} - \frac{62n + 49}{n + 1}\bigg].$$

The following result ensues.

PROPOSITION 1.

$$\mathbf{E}\left[\left(\frac{D_n - \mathbf{E}[D_n]}{\sqrt{2\ln n}}\right)^3\right] = \frac{1}{\sqrt{2\ln n}} + o(\ln^{-1/2} n).$$

PROOF OF THEOREM 1. *Lower bound.* Using the left inequality in (1) and that $\mathcal{N}(0, 1)$ is a symmetric distribution we have

$$\frac{1}{6}\left|\mathbf{E}\left[\left(\frac{D_n - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}\right)^3\right]\right| \le \zeta_3\left(\frac{D_n - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}, \mathcal{N}(0, 1)\right).$$

From the asymptotic relation of Proposition 1 we obtain $\zeta_3((D_n - \mathbf{E}[D_n])/\sqrt{\mathbf{Var}[D_n]}, \mathcal{N}(0, 1)) = \Omega(1/\sqrt{\ln n})$.

Upper bound. While D_n is the depth of a node chosen uniformly at random we will also consider d_n, the depth of the nth inserted node in the tree. Note that with I_n being uniformly distributed on $\{1, \ldots, n\}$ we have $D_n \overset{\mathcal{L}}{=} d_{I_n}$, where I_n is independent of $d_1, \ldots, d_n$. The upper bound for the rate of convergence of the scaled d_n is known: Cramer and Rüschendorf [(1996), Theorem 4.2] show

(4) $$\qquad \zeta_3\left(\frac{d_n - \mathbf{E}[d_n]}{\sqrt{\mathbf{Var}[d_n]}}, \mathcal{N}(0, 1)\right) \le \frac{C}{\sqrt{\ln n}}, \qquad n \ge 3,$$

with a constant $C > 0$. This rate of convergence can be transposed to the depth D_n of a random node: We define, for $k = 1, \ldots, n$,

$$m_k := \mathbf{E}\left[\frac{d_k - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}\right], \qquad s_k := \left(\mathbf{Var}\left[\frac{d_k - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}\right]\right)^{1/2} = \sqrt{\frac{\mathbf{Var}[d_k]}{\mathbf{Var}[D_n]}}.$$

Let N be a standard normal variate independent of the tree. Note that $s_{I_n}N + m_{I_n}$ has zero mean, variance 1 and a finite absolute third moment. Here, $\mathbf{Var}[s_{I_n}N + m_{I_n}] = 1$ follows from the fact that $(D_n - \mathbf{E}[D_n])/\sqrt{\mathbf{Var}[D_n]} \overset{\mathcal{L}}{=} (d_{I_n} - \mathbf{E}[D_n])/\sqrt{\mathbf{Var}[D_n]}$ and $s_{I_n}N + m_{I_n}$ conditioned on $I_n = k$ both have mean m_k and variance s_k^2 for $k = 1, \ldots, n$. Thus also unconditioned they have equal variances being 1. Therefore, using the triangle inequality we have the estimate

(5)
$$\zeta_3\left(\frac{D_n - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}, \mathcal{N}(0, 1)\right)$$
$$\le \zeta_3\left(\frac{d_{I_n} - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}, s_{I_n}N + m_{I_n}\right) + \zeta_3\left(s_{I_n}N + m_{I_n}, \mathcal{N}(0, 1)\right),$$

where the right-hand side is finite. We show that both summands in (5) are of the order $O(1/\sqrt{\ln n})$ to finish the proof of Theorem 1.

First summand in (5). Conditioning on I_n, using (4), that $s_1 = s_2 = 0$, and that ζ_3 is $(3, +)$ ideal we obtain

$$\zeta_3\left(\frac{d_{I_n} - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}, s_{I_n}N + m_{I_n}\right)$$

$$= \sup_{f \in \mathscr{F}_3}\left|\frac{1}{n}\sum_{k=1}^n \mathbf{E}\left[f\left(\frac{d_k - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}\right) - f(s_kN + m_k)\right]\right|$$

$$\leq \frac{1}{n}\sum_{k=1}^n \zeta_3\left(\frac{d_k - \mathbf{E}[D_n]}{\sqrt{\mathbf{Var}[D_n]}}, s_kN + m_k\right)$$

(6)

$$= \frac{1}{n}\sum_{k=3}^n \zeta_3\left(s_k\frac{d_k - \mathbf{E}[d_k]}{\sqrt{\mathbf{Var}[d_k]}} + m_k, s_kN + m_k\right)$$

$$\leq \frac{1}{n}\sum_{k=3}^n s_k^3 \frac{C}{\sqrt{\ln k}}$$

$$= O\left(\frac{1}{\sqrt{\ln n}}\right),$$

since $\sqrt{\ln n}$ is slowly varying at infinity and the s_k are uniformly bounded.

Second summand in (5). We will show below that $\|s_{I_n}N + m_{I_n}\|_3$ is uniformly bounded in $n \geq 1$. Hence, by the right inequality in (1) we have

$$\zeta_3\big(s_{I_n}N + m_{I_n}, \mathcal{N}(0, 1)\big) \leq C'\|s_{I_n}N + m_{I_n} - N\|_3$$

$$\leq C'(\|s_{I_n} - 1\|_3\, \|N\|_3 + \|m_{I_n}\|_3),$$

with a constant $C' > 0$. We show $\|s_{I_n} - 1\|_3 = O(1/\sqrt{\ln n})$ and $\|m_{I_n}\|_3 = O(1/\sqrt{\ln n})$. Note that this as well implies that $\|s_{I_n}N + m_{I_n}\|_3$ is uniformly bounded. We will use the well-known expansions $\mathbf{E}[d_n] = 2\ln n + O(1)$, $\mathbf{Var}[d_n] = 2\ln n + O(1)$ together with $\mathbf{E}[D_n] = 2\ln n + O(1)$ and $\mathbf{Var}[D_n] = 2\ln n + O(1)$. We have

$$\|m_{I_n}\|_3 = \frac{1}{\sqrt{\mathbf{Var}[D_n]}}\left(\frac{1}{n}\sum_{k=1}^n |\mathbf{E}[d_k] - \mathbf{E}[D_n]|^3\right)^{1/3}$$

$$\sim \frac{1}{\sqrt{2\ln n}}\left\|2\ln\left(\frac{I_n}{n}\right) + O(1)\right\|_3$$

$$= O\left(\frac{1}{\sqrt{\ln n}}\right),$$

since $\|\ln(I_n/n)\|_3 \to \|\ln U\|_3 < \infty$ for a unif$[0, 1]$ distributed random variable U.

Similarly we obtain

$$
\begin{aligned}
\|s_{I_n} - 1\|_3 &= \left\| \frac{1}{n} \sum_{k=1}^{n} \frac{\sqrt{\mathbf{Var}[d_k]} - \sqrt{\mathbf{Var}[D_n]}}{\sqrt{\mathbf{Var}[D_n]}} \right\|_3 \\
&= \frac{1}{\sqrt{\mathbf{Var}[D_n]}} \left\| \frac{1}{n} \sum_{k=1}^{n} (\sqrt{2\ln k} - \sqrt{2\ln n} + O(1)) \right\|_3 \\
&\sim \frac{1}{\sqrt{2\ln n}} \left\| \frac{1}{n} \sum_{k=1}^{n} \left(\frac{2\ln(k/n)}{\sqrt{2\ln k} + \sqrt{2\ln n}} + O(1) \right) \right\|_3 \\
&\leq \frac{1}{\sqrt{2\ln n}} \left\| 2\ln\left(\frac{I_n}{n}\right) + O(1) \right\|_3 \\
&= O\left(\frac{1}{\sqrt{\ln n}}\right).
\end{aligned}
$$

This completes the proof of Theorem 1. $\square$

4. Distribution of distances. In a random binary search tree generated from a random permutation of $\{1, \ldots, n\}$, select two nodes uniformly at random, all $\binom{n}{2}$ choices being equally likely. Let the distance between the two nodes be Δ_n. This distance has a recursive structure that relates to D_n. The two nodes may come from the left subtree, from the right subtree, may involve the root and one node from the left subtree, may involve the root and one node from the right subtree or may be in different subtrees with a path connecting them passing through the root. Conditioned on R being the root label, we have a recurrence:

$$
(7) \quad \Delta_n = \begin{cases}
\Delta_{R-1}, & \text{with probability } \dfrac{\binom{R-1}{2}}{\binom{n}{2}}, \\[2ex]
\tilde{\Delta}_{n-R}, & \text{with probability } \dfrac{\binom{n-R}{2}}{\binom{n}{2}}, \\[2ex]
(D_{R-1} + 1) + (\tilde{D}_{n-R} + 1), & \text{with probability } \dfrac{(R-1)(n-R)}{\binom{n}{2}}, \\[2ex]
D_{R-1} + 1, & \text{with probability } \dfrac{(R-1)}{\binom{n}{2}}, \\[2ex]
\tilde{D}_{n-R} + 1, & \text{with probability } \dfrac{(n-R)}{\binom{n}{2}}.
\end{cases}
$$

Here, $(D_1, \ldots, D_n)$ and $(\tilde{D}_1, \ldots, \tilde{D}_n)$ denote independent copies of random depths; likewise $(\Delta_1, \ldots, \Delta_n)$ and $(\tilde{\Delta}_1, \ldots, \tilde{\Delta}_n)$ are independent. It should be noted, however, that $(D_1, \ldots, D_n)$ and $(\Delta_1, \ldots, \Delta_n)$ are dependent.

By right and left symmetries, we have the recurrence

$$\mathbf{E}[\Delta_n] = \frac{2}{\binom{n}{2}}\mathbf{E}\left[\Delta_{R-1}\binom{R-1}{2}\right] + \frac{2}{\binom{n}{2}}\mathbf{E}[(D_{R-1}+1)(R-1)(n-R)]$$
$$+ \frac{2}{\binom{n}{2}}\mathbf{E}[(D_{R-1}+1)(R-1)].$$

Define $x_n = \binom{n}{2}\mathbf{E}[\Delta_n]$ to obtain the recurrence

$$x_n = \frac{2}{n}\sum_{r=1}^{n}x_{r-1} + \frac{2}{n}\sum_{r=1}^{n}(\mathbf{E}[D_{r-1}]+1)(r-1)(n-r+1).$$

Plug in the exact expression for $\mathbf{E}[D_n]$ from (2). Difference nx_n and $(n-1)x_{n-1}$ and rearrange in the form

$$nx_n = (n+1)x_{n-1} + 2\sum_{r=2}^{n}(2rH_{r-1} - 3r + 3).$$

The transformation $y_n = x_{n-1}/(n+1)$ linearizes the recurrence:

$$y_n = y_{n-1} + \frac{q(n)}{n(n+1)},$$

where $q(n) = \sum_{r=2}^{n}(2rH_{r-1} - 3r + 3)$. We unwind the recurrence under the boundary condition $y_1 = 0$, get the answer and clean up the sums in it and finally obtain $\mathbf{E}[\Delta_n]$ as in Theorem 2.

Going further with second moment recurrence, use the conditional representation (7) to write

$$\binom{n}{2}\mathbf{E}[\Delta_n^2] = \frac{2}{n}\sum_{r=1}^{n}\mathbf{E}[\Delta_{r-1}^2]\binom{r-1}{2} + \frac{2}{n}\sum_{r=1}^{n}\mathbf{E}[(D_{r-1}+1)^2](n-r+1)$$
$$+ \frac{2}{n}\sum_{r=1}^{n}\mathbf{E}[D_{r-1}]\mathbf{E}[D_{n-r}](r-1)(n-r)$$
$$+ \frac{4}{n}\sum_{r=1}^{n}\mathbf{E}[D_{r-1}](r-1)(n-r) + \frac{2}{n}\sum_{r=1}^{n}(r-1)(n-r).$$

Apart from the full-history summation, all the others involve only the first two moments of D_n. These sums vary in complexity. Some of these sums are trivial, like the last one. Other sums involve only $\mathbf{E}[D_n]$ or $\mathbf{E}[D_n^2]$, so they can be obtained in closed form from combinatorial identities for harmonic numbers. The sum involving the cross product $\mathbf{E}[D_{r-1}]\mathbf{E}[D_{n-r}]$ needs special treatment. Upon

plugging in $\mathbf{E}[D_n]$, we want to reduce

$$\sum_{r=1}^{n}[2rH_{r-1}-4(r-1)][2(n-r+1)H_{n-r}-4(n-r)]$$

$$=4\sum_{r=1}^{n}r(n-r+1)H_{r-1}H_{n-r}-8\sum_{r=1}^{n}r(n-r)H_{r-1}$$

$$-8\sum_{r=1}^{n}(r-1)(n-r+1)H_{n-r}+16\sum_{r=1}^{n}(r-1)(n-r).$$

The last of these sums is trivial; the middle two are identical and are both obtained from standard identities for harmonic numbers. The first sum suggests a squared generating function. Let

$$a_n = \sum_{j=0}^{n-1}(j-1)(n-j)H_jH_{n-j-1},$$

and let $A(z)=\sum_{n=0}^{\infty}a_nz^n$ be its generating function. Then

$$A(z)=z\left(\sum_{r=0}^{\infty}(r+1)H_rz^r\right)^2.$$

Hence

$$A(z)=\frac{z}{(1-z)^4}[z-\ln(1-z)]^2.$$

The required sum a_n can be found by extraction of coefficients. [A similar sum appears in Mahmoud and Smythe (1998).] One finds

$$a_n = \frac{1}{6}n(n+1)(n+2)(H_{n-1}^2 - H_{n-1}^{(2)}) + \frac{n-1}{108}(37n^2+112n+60)$$

$$-\frac{n-1}{18}(5n^2+20n+12)H_{n-1}.$$

So, the recurrence for the second moment of Δ_n takes the telescopic form

$$n\binom{n}{2}\mathbf{E}[\Delta_n^2]=2\sum_{r=1}^{n}\binom{r-1}{2}\mathbf{E}[\Delta_{r-1}^2]+\nu_n,$$

where

$$\nu_n = \tfrac{8}{3}(n^2+3n+2)H_n^2 - \tfrac{2}{9}(53n^2+93n+16)H_n + \tfrac{1}{27}(475n^2+492n+5).$$

Following our previous steps in solving such a recurrence, let $w_n=\binom{n}{2}\mathbf{E}[\Delta_n^2]$, and difference nw_n and $(n-1)w_{n-1}$. One obtains the iteratable recurrence

$$\frac{w_n}{n+1}=\frac{w_{n-1}}{n}+\frac{\nabla(n\nu_n)}{n(n+1)},$$

yielding the solution

$$\mathbf{E}[\Delta_n^2] = \frac{n+1}{\binom{n}{2}} \sum_{j=2}^{n} \frac{\nabla(j\nu_j)}{j(j+1)}.$$

Upon going through a lengthy process of combinatorial reduction, all via elementary combinatorics, one gets

$$\mathbf{E}[\Delta_n^2] = \frac{2}{n(n-1)}\big[(8n^2+28n+20)H_n^2 - (46n^2+124n-54)H_n + (87n^2+137n)\big].$$

Finally the variance in Theorem 2 follows from $\mathbf{E}[\Delta_n^2] - \big(\mathbf{E}[\Delta_n]\big)^2$.

Having determined the mean and variance of Δ_n, we are in possession of appropriate norming factors for centering and scaling. We try to go forward with the contraction method. Let $I_{\mathcal{LL}}$, $I_{\mathcal{RR}}$, $I_{\mathcal{LR}}$, $I_{\mathcal{L},\mathcal{ROOT}}$, $I_{\mathcal{R},\mathcal{ROOT}}$ be respectively indicators of the following events; that the two nodes involved in Δ_n are both from the left subtree; both nodes are from the right subtree; the nodes come from different subtrees; one is from the left, the other is the root; and one is from the right, the other is the root. From the basic recurrence (7) we have a distributional equation:

(8)
$$\Delta_n \overset{\mathcal{L}}{=} \Delta_{R-1}I_{\mathcal{LL}} + \tilde{\Delta}_{n-R}I_{\mathcal{RR}} + (D_{R-1} + \tilde{D}_{n-R} + 2)I_{\mathcal{LR}}$$
$$+ (D_{R-1} + 1)I_{\mathcal{L},\mathcal{ROOT}} + (\tilde{D}_{n-R} + 1)I_{\mathcal{R},\mathcal{ROOT}},$$

where, for each n, $(\tilde{\Delta}_n, \tilde{D}_n) \overset{\mathcal{L}}{=} (\Delta_n, D_n)$, and $(I_{\mathcal{LL}}, I_{\mathcal{RR}}, I_{\mathcal{LR}}, I_{\mathcal{L},\mathcal{ROOT}}, I_{\mathcal{R},\mathcal{ROOT}}, R)$ and the sequences $((\Delta_n, D_n))$, $((\tilde{\Delta}_n, \tilde{D}_n))$ are independent. Note that $(\Delta_n, \tilde{\Delta}_n)$ and $(D_n, \tilde{D}_n)$ are *dependent*, because the vector (Δ_n, D_n) [as well as $(\tilde{\Delta}_n, \tilde{D}_n)$] has dependent components being the random distance and the random depth in the *same* random binary search tree of size n. Define the normed random variables $\Delta_0^* := \Delta_1^* := \Delta_2^* := 0$ and, for $n \geq 3$,

$$\Delta_n^* := \frac{\Delta_n - 4\ln n}{2\sqrt{\ln n}},$$

and analogously define $\tilde{\Delta}_n^*$. So, the recurrence can be normed, for $n \geq 3$, in the form

(9)
$$\Delta_n^* \overset{\mathcal{L}}{=} I_{\mathcal{LL}}\sqrt{\frac{\ln(R-1)}{\ln n}}\Delta_{R-1}^* + I_{\mathcal{RR}}\sqrt{\frac{\ln(n-R)}{\ln n}}\tilde{\Delta}_{n-R}^*$$

$$+ (I_{\mathcal{LR}} + I_{\mathcal{L},\mathcal{ROOT}})\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(R-1)}{\ln n}}D_{R-1}^*$$

$$+ (I_{\mathcal{LR}} + I_{\mathcal{R},\mathcal{ROOT}})\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(n-R)}{\ln n}}\tilde{D}_{n-R}^*$$

$$+ T_n,$$

where the independence and dependence relations are handed down from (8). Again $(\Delta_n^*, \tilde{\Delta}_n^*)$ and $(D_n^*, \tilde{D}_n^*)$ are *dependent*; in the latter equation T_n is the expression

$$
\begin{aligned}
T_n = \frac{1}{2\sqrt{\ln n}}\Bigg[& 4I_{\mathcal{L}\mathcal{L}} \ln\left(\frac{R-1}{n}\right) + 4I_{\mathcal{R}\mathcal{R}} \ln\left(\frac{n-R}{n}\right) \\
& + (I_{\mathcal{L}\mathcal{R}} + I_{\mathcal{L},\mathcal{ROOT}})\left(1 + 2\ln\left(\frac{R-1}{n}\right)\right) \\
& + (I_{\mathcal{L}\mathcal{R}} + I_{\mathcal{R},\mathcal{ROOT}})\left(1 + 2\ln\left(\frac{n-R}{n}\right)\right) \\
& - 2(I_{\mathcal{L},\mathcal{ROOT}} + I_{\mathcal{R},\mathcal{ROOT}})\ln n \Bigg].
\end{aligned}
$$

Again we will use the convention $\ln k = 0$, for $k = 0$.

Before proving asymptotic normality for Δ_n^* via the contraction method we will first have a heuristic look at recurrence (9). Several convergence relations suggest that the limit X of Δ_n^* (if it exists) will satisfy a distributional equation. The indicators converge as follows:

$$
I_{\mathcal{L},\mathcal{ROOT}} \xrightarrow{\mathcal{P}} 0; \qquad I_{\mathcal{R},\mathcal{ROOT}} \xrightarrow{\mathcal{P}} 0;
$$

$$
I_{\mathcal{L}\mathcal{L}} \xrightarrow{\mathcal{L}} I \stackrel{\mathcal{L}}{=} \mathrm{Ber}\left(\frac{1}{3}\right); \qquad I_{\mathcal{R}\mathcal{R}} \xrightarrow{\mathcal{L}} J \stackrel{\mathcal{L}}{=} \mathrm{Ber}\left(\frac{1}{3}\right); \qquad I_{\mathcal{L}\mathcal{R}} \xrightarrow{\mathcal{L}} K \stackrel{\mathcal{L}}{=} \mathrm{Ber}\left(\frac{1}{3}\right),
$$

with I, J, K being dependent, but mutually exclusive, with $I + J + K \equiv 1$. The asymptotic behavior of R/n is like that of $U \stackrel{\mathcal{L}}{=} \mathrm{unif}(0, 1)$. Also $\ln((R-1)/\ln n) = [\ln((R-1)/n) + \ln n]/\ln n$, which is asymptotically $\ln U/\ln n + 1 \xrightarrow{\mathcal{P}} 1$. Similarly $\ln(n-R)/\ln n \xrightarrow{\mathcal{P}} 1$. Therefore the expression T_n converges in probability to 0. By Theorem 1, D_n and D_n^* have a Gaussian tendency.

Thus, if the limit X exists, it is plausible that (9) will approach the *limit equation*

$$
(10) \qquad X \stackrel{\mathcal{L}}{=} IX + J\tilde{X} + K\frac{1}{\sqrt{2}}(Y + \tilde{Y}),
$$

where $(\tilde{X}, \tilde{Y})$ is an independent probabilistic copy of (X, Y), the variates $Y, \tilde{Y}$ are standard normal variates and the row vector $(X, \tilde{X}, Y, \tilde{Y})$ is independent of (I, J, K).

We argue next that $\mathcal{N}(0, 1)$ is the unique solution of (10) among all univariate probability measures. Let $\theta(t)$ be the characteristic function of X. Since $W := (Y + \tilde{Y})/\sqrt{2}$ has a standard normal distribution, and $(X, \tilde{X}, W)$ is independent

of (I, J, K), we obtain

$$
\begin{aligned}
\theta(t) =\, & \mathbf{E}\big[\exp\big(it(IX + J\tilde{X} + KW)\big)\big] \\
= & \mathbf{E}\big[\exp\big(it(IX + J\tilde{X} + KW)\big) \mid (I, J, K) = (1, 0, 0)\big] P\{(I, J, K) = (1, 0, 0)\} \\
& + \mathbf{E}\big[\exp\big(it(IX + J\tilde{X} + KW)\big) \mid (I, J, K) = (0, 1, 0)\big] \\
& \quad \times P\{(I, J, K) = (0, 1, 0)\} \\
& + \mathbf{E}\big[\exp\big(it(IX+, J\tilde{X} + KW)\big) \mid (I, J, K) = (0, 0, 1)\big] \\
& \quad \times P\{(I, J, K) = (0, 0, 1)\} \\
= & \tfrac{1}{3}\theta(t) + \tfrac{1}{3}\theta(t) + \tfrac{1}{3}\mathbf{E}[\exp(itW)].
\end{aligned}
$$

Thus $\theta(t) = \mathbf{E}[\exp(itW)]$ and $X \overset{\mathcal{L}}{=} W \overset{\mathcal{L}}{=} \mathcal{N}(0, 1)$. Note that the joint distributions of (X, Y), $(\tilde{X}, \tilde{Y})$ do not matter since I, J, K are mutually exclusive.

In the next section we will establish the asymptotic normality of Δ_n^* based on these heuristic considerations by an application of the contraction method.

5. Asymptotic normality. In this section we prove $\Delta_n^* \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, 1)$ with the rate $\zeta_3((\Delta_n - \mathbf{E}[\Delta_n])/\sqrt{\mathbf{Var}[\Delta_n]}, \mathcal{N}(0, 1)) = \Theta(1/\sqrt{\ln n})$.

We will use the notation $\sigma_n := \sqrt{\mathbf{Var}[\Delta_n^*]}$, $\bar{\sigma}_n := \sqrt{\mathbf{Var}[D_n^*]}$, for $n \geq 0$. Then the expansions of the variances of Δ_n and D_n imply $\sigma_n^2 = 1 + O(1/\ln n)$, and $\bar{\sigma}_n^2 = 1 + O(1/\ln n)$.

Since the finiteness of ζ_3 requires that the second moments of the quantities compared coincide, we show

$$
\tag{11} \zeta_3\big(\Delta_n^*, \sigma_n \mathcal{N}(0, 1)\big) = \Theta\left(\frac{1}{\sqrt{\ln n}}\right).
$$

With $\sigma_n \to 1$ and $\zeta_3((\Delta_n - \mathbf{E}[\Delta_n])/\sqrt{\mathbf{Var}[\Delta_n]}, \mathcal{N}(0, 1)) = \sigma_n^{-3}\zeta_3(\Delta_n^*, \sigma_n \mathcal{N})$ this implies the remaining assertion in Theorem 2.

Upper bound in (11). To use the right inequality in (1) involving the L_3 norm of appropriate representation of the random variables we will model all quantities on a joint probability space. We assume that for a unif$[0, 1]$ variable U we have $R = R_n = \lceil nU \rceil$ and that (I, J, K) is a vector such that a version of its conditional distribution given U is given by

$$
\tag{12} U^2 \delta_{(1,0,0)} + (1 - U)^2 \delta_{(0,1,0)} + 2U(1 - U)\delta_{(0,0,1)},
$$

where δ_p denotes the Dirac measure at point p. In particular this implies $\mathcal{L}(I, J, K) = \frac{1}{3}\delta_{(1,0,0)} + \frac{1}{3}\delta_{(0,1,0)} + \frac{1}{3}\delta_{(0,0,1)}$. We denote by $N, \tilde{N}, N', \tilde{N}'$ standard

normal random variables such that N, $\tilde{N}$, N', $\tilde{N}'$, $(I_{\mathscr{L}\mathscr{L}}, I_{\mathscr{R}\mathscr{R}}, I_{\mathscr{L}\mathscr{R}}, I_{\mathscr{L},\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}, I_{\mathscr{R},\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}, I, J, K, U)$ are independent. We abbreviate $I_{\mathscr{L},\mathscr{R}\vee\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}} := I_{\mathscr{L}\mathscr{R}} + I_{\mathscr{L},\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}$ and $I_{\mathscr{R},\mathscr{L}\vee\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}} := I_{\mathscr{L}\mathscr{R}} + I_{\mathscr{R},\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}$ and define the scaled quantities $N_n := \sigma_n N$, $\tilde{N}_n := \sigma_n \tilde{N}$, $N'_n := \bar{\sigma}_n N'$, $\tilde{N}'_n := \bar{\sigma}_n \tilde{N}'$, which have finite ζ_3 distances to Δ_n^* and D_n^*, respectively. Moreover, we define $Q_2 := Q'_2 := 0$ and (still using the convention $\ln k = 0$ for $k = 0$), for $n \geq 3$,

$$Q_n := I_{\mathscr{L}\mathscr{L}}\sqrt{\frac{\ln(R-1)}{\ln(n)}}N_{R-1} + I_{\mathscr{R}\mathscr{R}}\sqrt{\frac{\ln(n-R)}{\ln n}}\tilde{N}_{n-R}$$

$$+ I_{\mathscr{L},\mathscr{R}\vee\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(R-1)}{\ln n}}D_{R-1}^*$$

$$+ I_{\mathscr{R},\mathscr{L}\vee\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(n-R)}{\ln n}}\tilde{D}_{n-R}^* + T_n,$$

$$Q'_n := I_{\mathscr{L}\mathscr{L}}\sqrt{\frac{\ln(R-1)}{\ln n}}N_{R-1} + I_{\mathscr{R}\mathscr{R}}\sqrt{\frac{\ln(n-R)}{\ln n}}\tilde{N}_{n-R}$$

$$+ I_{\mathscr{L},\mathscr{R}\vee\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(R-1)}{\ln n}}N'_{R-1}$$

$$+ I_{\mathscr{R},\mathscr{L}\vee\mathscr{R}\mathscr{O}\mathscr{O}\mathscr{T}}\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(n-R)}{\ln n}}\tilde{N}'_{n-R} + T_n.$$

Comparing these expressions with the representation of Δ_n^* in (9) and noting that the dependence between Δ_n^* and D_n^* (and $\tilde{\Delta}_n^*$ and $\tilde{D}_n^*$) has no influence on the distribution of the right-hand side in (9), we obtain $\mathbf{E}[Q_n] = \mathbf{E}[Q'_n] = 0$ and $\mathbf{Var}[Q_n] = \mathbf{Var}[Q'_n] = \sigma_n^2$. Therefore, ζ_3 distances between Δ_n^*, Q_n, Q'_n and N_n are finite.

By the triangle inequality we have

$$\zeta_3\big(\Delta_n^*, \sigma_n\mathscr{N}(0,1)\big) \leq \zeta_3(\Delta_n^*, Q_n) + \zeta_3(Q_n, Q'_n) + \zeta_3\big(Q'_n, \sigma_n\mathscr{N}(0,1)\big).$$

We postpone the estimates of $\zeta_3(Q_n, Q'_n)$ and $\zeta_3(Q'_n, \sigma_n\mathscr{N}(0,1))$ to later, where we show

$$(13) \qquad \zeta_3(Q_n, Q'_n) + \zeta_3\big(Q'_n, \sigma_n\mathscr{N}(0,1)\big) = O\bigg(\frac{1}{\sqrt{\ln n}}\bigg).$$

Recall that we have

$$\mathbf{P}(\mathscr{L}\mathscr{L} \mid R = k) = \frac{(k-1)(k-2)}{n(n-1)},$$

which, by conditioning on R and using that ζ_3 is $(3, +)$ ideal, implies

$$\zeta_3(\Delta_n^*, Q_n) \le 2\zeta_3\left(I_{\mathcal{L}\mathcal{L}}\sqrt{\frac{\ln(R-1)}{\ln n}}\Delta_{R-1}^*, I_{\mathcal{L}\mathcal{L}}\sqrt{\frac{\ln(R-1)}{\ln n}}N_{R-1}\right)$$

$$\le \frac{2}{n}\sum_{k=2}^{n}\frac{(k-1)(k-2)}{n(n-1)}\left(\frac{\ln(k-1)}{\ln n}\right)^{3/2}\zeta_3(\Delta_{k-1}^*, N_{k-1}).$$

By (13) there exists a constant $C > 0$ such that $\zeta_3(Q_n, Q_n') + \zeta_3(Q_n', \sigma_n\mathcal{N}(0, 1)) \le C/\sqrt{\ln n}$ for all $n \ge 2$. Then we obtain by induction that $\zeta_3(\Delta_n^*, \sigma_n\mathcal{N}(0, 1)) = \zeta_3(\Delta_n^*, N_n) \le 3C/\sqrt{\ln n}$: The assertion is true for $n = 2$ since $\Delta_2^* = 0$ and $\sigma_2 = 0$ and, for $n \ge 3$, assuming the induction hypothesis, we have

$$\zeta_3\bigl(\Delta_n^*, \sigma_n\mathcal{N}(0, 1)\bigr)$$

$$\le \frac{2}{n^2(n-1)\sqrt{\ln n}}\sum_{k=2}^{n-1}k(k-1)\frac{\ln^{3/2}k}{\ln n}\frac{3C}{\sqrt{\ln k}} + \frac{C}{\sqrt{\ln n}}$$

$$\le \frac{1}{\sqrt{\ln n}}\left(\frac{2}{n^2(n-1)}3C\frac{1}{3}n(n-1)(n-2) + C\right)$$

$$\le \frac{3C}{\sqrt{\ln n}},$$

which gives the lower bound $\zeta_3(\Delta_n^*, \sigma_n\mathcal{N}(0, 1)) = O(1/\sqrt{\ln n})$.

We still have to prove the asymptotic relation (13).

First summand in (13). The first summand $\zeta_3(Q_n, Q_n')$ is estimated using Theorem 1. Since $(\bar{\sigma}_n)$ is bounded we obtain from Theorem 1 the existence of a constant $C > 0$ with $\zeta_3(D_n, N_n') \le C/\sqrt{\ln n}$ for all $n \ge 2$. This implies, conditioning on R and using that ζ_3 is $(3, +)$ ideal,

$$\zeta_3(Q_n, Q_n') \le 2\zeta_3\left(I_{\mathcal{L}, \mathcal{R}\vee\mathcal{ROOT}}\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(R-1)}{\ln n}}D_{R-1}^*,\right.$$

$$\left. I_{\mathcal{L}, \mathcal{R}\vee\mathcal{ROOT}}\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(R-1)}{\ln n}}N_{R-1}'\right)$$

$$\le \frac{2}{n}\sum_{k=2}^{n}\mathbf{P}(\mathcal{L}, \mathcal{R}\vee\mathcal{ROOT})\frac{1}{2^{3/2}}\left(\frac{\ln(k-1)}{\ln n}\right)^{3/2}\zeta_3(D_{k-1}, N_{k-1}')$$

$$\le \frac{1}{n}\sum_{k=2}^{n-1}\zeta_3(D_k, N_k') \le \frac{1}{n}\sum_{k=2}^{n-1}\frac{C}{\sqrt{\ln k}}$$

$$\sim \frac{C}{\sqrt{\ln n}} = O\left(\frac{1}{\sqrt{\ln n}}\right).$$

Second summand in (13). Here we use the distributional identity

$$\sigma_n \mathcal{N}(0, 1) \stackrel{\mathcal{L}}{=} \sigma_n \big(I N + J \tilde{N} + K(N' + \tilde{N}')/\sqrt{2} \big).$$

The third absolute moments of Q'_n and N_n are uniformly bounded. By the right inequality in (1) this implies that their ζ_3 distance can be estimated up to a constant factor by their L_3 distance. Hence, for some $C > 0$,

$$\zeta_3 \big(Q'_n, \sigma_n \mathcal{N}(0, 1) \big) \leq C \big\| Q'_n - \sigma_n \big(I N + J \tilde{N} + K(N' + \tilde{N}')/\sqrt{2} \big) \big\|_3$$

$$\leq C \big\| Q'_n - \big(I N + J \tilde{N} + K(N' + \tilde{N}')/\sqrt{2} \big) \big\|_3$$

$$+ C \|N\|_3 |1 - \sigma_n|.$$

The expansion $\sigma_n = 1 + O(1/\ln n)$ gives $C \|N\|_3 |1 - \sigma_n| = O(1/\ln n)$. For the first summand in the latter expression we have

$$\big\| Q'_n - \big(I N + J \tilde{N} + K(N' + \tilde{N}')/\sqrt{2} \big) \big\|_3$$

$$\leq \left\| I_{\mathcal{L L}} \sqrt{\frac{\ln(R - 1)}{\ln n}} \sigma_{R-1} N - I N \right\|_3 + \left\| I_{\mathcal{R R}} \sqrt{\frac{\ln(n - R)}{\ln n}} \sigma_{n-R} \tilde{N} - I \tilde{N} \right\|_3$$

$$+ \left\| I_{\mathcal{L}, \mathcal{R} \vee \mathcal{R O O T}} \frac{1}{\sqrt{2}} \sqrt{\frac{\ln(R - 1)}{\ln n}} \bar{\sigma}_{R-1} N' - \frac{K}{\sqrt{2}} N' \right\|_3$$

$$+ \left\| I_{\mathcal{R}, \mathcal{L} \vee \mathcal{R O O T}} \frac{1}{\sqrt{2}} \sqrt{\frac{\ln(n - R)}{\ln n}} \bar{\sigma}_{n-R} \tilde{N}' - \frac{K}{\sqrt{2}} \tilde{N}' \right\|_3$$

$$+ \|T_n\|_3$$

$$=: s_1 + s_2 + s_3 + s_4 + s_5.$$

We have to show $s_r = O(1/\sqrt{\ln n})$, for $r = 1, \ldots, 5$. Clearly, $s_1 = s_2$ and $s_3 = s_4$. Since for all terms arising subsequently we only have to reach the order of the worst term, we will occasionally take the liberty to use some simple estimates not being tight.

Estimate of s_1, s_2. We have

$$s_1 = \|N\|_3 \left\| I_{\mathcal{L L}} \sqrt{\frac{\ln(R - 1)}{\ln n}} \sigma_{R-1} - I \right\|_3$$

$$\leq \|N\|_3 \left(\left\| \sqrt{\frac{\ln(R - 1)}{\ln n}} (I_{\mathcal{L L}} \sigma_{R-1} - I) \right\|_3 + \left\| I \sqrt{\frac{\ln(R - 1)}{\ln n}} - I \right\|_3 \right)$$

$$\leq \|N\|_3 \left(\|\sigma_{R-1} - 1\|_3 + \|I_{\mathcal{L L}} - I\|_3 + \left\| \sqrt{\frac{\ln(R - 1)}{\ln n}} - 1 \right\|_3 \right).$$

From $\sigma_n = 1 + O(1/\ln n)$ we obtain

$$\|\sigma_{R-1} - 1\|_3 \leq \left(\frac{1}{n}\sum_{k=0}^{n-1}\left(\frac{C}{1 \vee \ln k}\right)^3\right)^{1/3} = O\left(\frac{1}{\ln n}\right).$$

For the estimate of $\|I_{\mathcal{L}\mathcal{L}} - I\|_3 = \|I_{\mathcal{L}\mathcal{L}} - I\|_1^{1/3}$ we recall that for the conditional expectations we have almost surely

$$\mathbf{E}[I_{\mathcal{L}\mathcal{L}} \mid U] = \frac{(\lceil nU\rceil - 1)(\lceil nU\rceil - 2)}{n(n-1)}, \qquad \mathbf{E}[I \mid U] = U^2,$$

where (12) is used. This implies

$$\|I_{\mathcal{L}\mathcal{L}} - I\|_1^{1/3} = \left(\mathbf{E}\left|\frac{(\lceil nU\rceil - 1)(\lceil nU\rceil - 2)}{n(n-1)} - U^2\right|\right)^{1/3} = O\left(\frac{1}{n^{1/3}}\right).$$

The expansion $\sum_{k=1}^{n-1}\sqrt{\ln k} = n\sqrt{\ln n} + O(n/\ln n)$ implies

$$\left\|\sqrt{\frac{\ln(R-1)}{\ln n}} - 1\right\|_3 = \left(\frac{1}{n}\sum_{k=1}^{n}\left(1 - \sqrt{\frac{\ln(k-1)}{\ln n}}\right)^3\right)^{1/3}$$

$$= \left(\frac{1}{n}\sum_{k=1}^{n}\left(1 - \sqrt{\frac{\ln(k-1)}{\ln n}}\right)\right)^{1/3}$$

$$= O\left(\frac{1}{\ln^{3/2} n}\right)^{1/3}$$

$$= O\left(\frac{1}{\sqrt{\ln n}}\right).$$

Together we obtain $s_1 = s_2 = O(1/\sqrt{\ln n})$.

Estimate of s_3, s_4. Adding and subtracting appropriate terms, estimating as for the case of s_1, s_2 and noting $1/\sqrt{2} \leq 1$ we find

$$s_3 = \left\|I_{\mathcal{L},\mathcal{R}\vee\mathcal{ROOT}}\frac{1}{\sqrt{2}}\sqrt{\frac{\ln(R-1)}{\ln n}}\bar{\sigma}_{R-1}N' - \frac{K}{\sqrt{2}}N'\right\|_3$$

$$\leq \|N'\|_3\left(\|\bar{\sigma}_{R-1} - 1\|_3 + \left\|I_{\mathcal{L},\mathcal{R}\vee\mathcal{ROOT}} - K\right\|_3 + \left\|\sqrt{\frac{\ln(R-1)}{\ln n}} - 1\right\|_3\right).$$

All three summands can be estimated as in the case of s_1, s_2, since we have the analogous expansion $\bar{\sigma}_n = 1 + O(1/\ln n)$. Hence, we obtain $s_3 = s_4 = O(1/\sqrt{\ln n})$.

Estimate of $s_5 = \|T_n\|_3$. A direct calculation yields

$$\left\| \ln\left(\frac{R-1}{n}\right) \right\|_3^3 \le \frac{1}{n} \sum_{k=1}^n \left(\ln\left(\frac{k-1}{n}\right) \right)^3 = \frac{1}{n}(6n + O(1)) = O(1),$$

and in particular $\|1 + \ln((R-1)/n)\|_3 = O(1)$. Furthermore we have $\mathbf{P}(\mathcal{L}, \mathcal{ROOT}) = (n-1)/(2n^2) = O(1/n)$. Altogether this implies

$$\|T_n\|_3 \le \frac{1}{2\sqrt{\ln n}} \left[4 \left\| \ln\left(\frac{R-1}{n}\right) \right\|_3 + 4 \left\| \ln\left(\frac{R-1}{n}\right) \right\|_3 + \left\| 1 + \ln\left(\frac{R-1}{n}\right) \right\|_3 \right.$$

$$\left. + \left\| 1 + \ln\left(\frac{n-R}{n}\right) \right\|_3 + (\mathbf{P}(\mathcal{L}, \mathcal{ROOT}))^{1/3} 4 \ln n \right]$$

$$= O\left(\frac{1}{\sqrt{\ln n}}\right).$$

Putting together the estimates for $s_1, \ldots, s_5$ we obtain $\zeta_3(Q_n', \sigma_n \mathcal{N}(0,1)) = O(1/\sqrt{\ln n})$, which finishes the proof of expansion (13) and thus yields the upper bound $\zeta_3(\Delta_n^*, \sigma_n \mathcal{N}(0,1)) = O(1/\sqrt{\ln n})$ in (11).

Lower bound in (11). Using the left inequality in (1) and that $\sigma_n \mathcal{N}(0,1)$ is a symmetric distribution we obtain

$$\tfrac{1}{6} |\mathbf{E}[(\Delta_n^*)^3]| \le \zeta_3(\Delta_n^*, \sigma_n \mathcal{N}(0,1)).$$

To determine the rate of convergence of Δ_n^*, we need its third moment. Here we shall abandon the general strategy of finding asymptotics via exact solutions, as an exact estimate here is exceptionally demanding. We shall set up a recurrence for this third moment, and it will be sufficient for our purpose to solve it asymptotically.

From the recurrence (7), and various independencies, one finds

$$\binom{n}{2} \mathbf{E}[\Delta_n^3] = \frac{2}{n} \sum_{r=1}^n \binom{r-1}{2} \mathbf{E}[\Delta_{r-1}^3]$$

$$+ \frac{1}{n} \sum_{r=1}^n \left(\mathbf{E}[(D_{r-1}+1)^3] + 3\mathbf{E}[(D_{r-1}+1)^2]\mathbf{E}[D_{n-r}+1] \right.$$

$$+ 3\mathbf{E}[D_{r-1}+1]\mathbf{E}[(D_{n-r}+1)^2] + \mathbf{E}[(D_{n-r}+1)^3] \big)$$

$$\times (r-1)(n-r)$$

$$+ \frac{2}{n} \sum_{r=1}^n (\mathbf{E}[D_{r-1}^3] + 3\mathbf{E}[D_{r-1}^2] + 3\mathbf{E}[D_{r-1}] + 1)(r-1).$$

To prepare for telescoping, set $z_n = \binom{n}{2}\mathbf{E}[\Delta_n^3]$, and difference nz_n and $(n-1)z_{n-1}$, to obtain the iteratable recurrence

(14)
$$\frac{z_n}{n+1} = \frac{z_{n-1}}{n} + \frac{P_n}{n(n+1)},$$

with

$$P_n = 2\sum_{r=1}^{n}\mathbf{E}\big[(D_{r-1}+1)^3\big](r-1) + 6\sum_{r=1}^{n}\mathbf{E}[D_{r-1}+1]\mathbf{E}\big[(D_{n-r}+1)^2\big](r-1).$$

With some effort the first sum can be computed exactly, yielding the asymptotic equivalent:

$$8n^2\ln^3 n + (24\gamma - 36)n^2\ln^2 n + o(n^2\ln^2 n).$$

Here and subsequently $\gamma = 0.57721\ldots$, denotes Euler's constant. Six different sums appear in the second sum in P_n. One can write this latter sum as

$$S_1(n) + 2S_2(n) + S_3(n) + S_4(n) + 2S_5(n) + S_6(n),$$

with

$$S_1(n) := \sum_{r=1}^{n-1}\mathbf{E}[D_{r-1}]\mathbf{E}[D_{n-r}^2](r-1), \qquad S_2(n) := \sum_{r=1}^{n-1}\mathbf{E}[D_{r-1}]\mathbf{E}[D_{n-r}](r-1),$$

$$S_3(n) := \sum_{r=1}^{n-1}\mathbf{E}[D_{r-1}](r-1), \qquad S_4(n) := \sum_{r=1}^{n-1}\mathbf{E}[D_{n-r}^2](r-1),$$

$$S_5(n) := \sum_{r=1}^{n-1}\mathbf{E}[D_{n-r}](r-1), \qquad S_6(n) := \sum_{r=1}^{n-1}(r-1).$$

Of these, $S_3(n)$ and $S_6(n)$ are simple. From their exact form we see that both are $O(n^2\ln n)$.

The other four sums include numerous formidable combinatorial sums involving various powers of the harmonic numbers. In principle we can obtain them exactly, but it strikes us as a daunting task. For our asymptotic purposes, we shall content ourselves with leading terms, which can be found from comparing sums with areas under continuous curves. For example, $S_1(n)$, the most formidable of all six sums, can be assessed asymptotically as follows. Plug in the exact values of the first and second moments of D_n, to obtain

$$S_1(n) = 8\sum_{j=1}^{n-2}\frac{j(n-j)H_jH_{n-j}^2}{n-j+1} + O(n\ln^2 n).$$

Approximating harmonic numbers with their asymptotic expansion (the first two terms suffice here), one gets

$$S_1(n) = 8 \sum_{j=1}^{n-2} j \left[\ln j + \gamma + O\left(\frac{1}{j}\right) \right] \left[\ln(n-j) + \gamma + O\left(\frac{1}{n-j}\right) \right]^2 + O(n \ln^2 n)$$

$$= 8 \sum_{j=1}^{n-2} j \ln j \ln^2(n-j) + 16\gamma \sum_{j=1}^{n-2} j \ln j \ln(n-j) + 8\gamma \sum_{j=1}^{n-2} j \ln^2(n-j)$$

$$+ O(n^2 \ln n).$$

We obtain approximations for these sums by integrals. For instance

$$\sum_{j=1}^{n-2} j \ln j \ln^2(n-j) = \int_1^n x \ln x \ln^2(n-x) \, dx + o(n^2 \ln n)$$

$$= \tfrac{1}{2} n^2 \ln^3 n - \tfrac{1}{4} n^2 \ln^2 n + o(n^2 \ln n).$$

Likewise, the other sums in $S_1(n)$ are obtained:

$$S_1(n) = 4n^2 \ln^3 n + (8\gamma - 2)n^2 \ln^2 n + o(n^2 \ln^2 n).$$

In a like manner we find all the other five sums, yielding

$$S_2(n) = 2n^2 \ln^2 n + o(n^2 \ln^2 n), \qquad S_3(n) = n^2 \ln n + o(n^2 \ln n),$$
$$S_4(n) = 2n^2 \ln^2 n + o(n^2 \ln^2 n), \qquad S_5(n) = n^2 \ln n + o(n^2 \ln n),$$
$$S_6(n) = O(n^2).$$

Putting it together, one finds

$$P_n = 32n^2 \ln^3 n + 72\gamma n^2 \ln^2 n + o(n^2 \ln^2 n).$$

Working our way back, we unwind the recurrence (14):

$$z_n = (n+1) \sum_{j=2}^{n} \frac{P_j}{j(j+1)}.$$

So,

$$z_n = 32n^2 \ln^3 n + 72\gamma n^2 \ln^2 n + o(n^2 \ln^2 n).$$

We now recover $\mathbf{E}[\Delta_n^3]$ in the asymptotic form

$$\mathbf{E}[\Delta_n^3] = \frac{2}{n(n-1)} z_n = 64 \ln^3 n + 144\gamma \ln^2 n + o(\ln^2 n).$$

Now, a calculation of $\mathbf{E}[(\Delta_n - \mathbf{E}[\Delta_n])^3]$, resorting to what we now know about all three moments, finally yields

$$\mathbf{E}\bigg[\bigg(\frac{\Delta_n - \mathbf{E}[\Delta_n]}{2\sqrt{\ln n}}\bigg)^3\bigg] = \frac{48(12 - \gamma)}{\sqrt{\ln n}} + o(\ln^{-1/2} n).$$

This implies $\zeta_3(\Delta_n^*, \sigma_n \mathcal{N}(0, 1)) = \Omega(1/\sqrt{\ln n})$ and finishes the proof of Theorem 2.

6. Concluding remarks. We have investigated random distances in a random binary search tree. Two types of distance have been considered, the depth of a random node and the distance between random pairs of nodes. Classical methods lead to a Gaussian limit law for normed random distances from the root. By a combination of classical methods and modern contraction techniques we arrived at a Gaussian limit law for normed random distances between random pairs. All proofs of convergence in distribution are accompanied by rates of convergence.

The heuristic approach to the limiting random variable of D_n^* by the contraction method would give the following. The normalization of D_n induces a distributional equation for D_n^*:

(15) $$D_n^* \stackrel{\mathcal{L}}{=} D_{R-1}^* I_{\mathcal{L}} \sqrt{\frac{\ln(R - 1)}{\ln n}} + \tilde{D}_{n-R}^* I_{\mathcal{R}} \sqrt{\frac{\ln(n - R)}{\ln n}} + S_n,$$

where S_n is the expression

$$S_n = \frac{1}{\sqrt{\ln n}}\big(I_{\mathcal{L}} + I_{\mathcal{R}} + 2I_{\mathcal{L}} \ln(R - 1) + 2I_{\mathcal{R}} \ln(n - R) - 2 \ln n\big).$$

We have the following convergences:

$$I_{\mathcal{L}} \stackrel{\mathcal{L}}{\longrightarrow} I; \qquad I_{\mathcal{R}} \stackrel{\mathcal{L}}{\longrightarrow} 1 - I;$$

$$\sqrt{\frac{\ln(R - 1)}{\ln n}} \stackrel{\text{a.s.}}{\longrightarrow} 1; \qquad \sqrt{\frac{\ln(n - R)}{\ln n}} \stackrel{\text{a.s.}}{\longrightarrow} 1; \qquad S_n \stackrel{\text{a.s.}}{\longrightarrow} 0,$$

where $I \stackrel{\mathcal{L}}{=} \text{Ber}(\frac{1}{2})$. Passing formally to the limits in (15) we arrive at

$$Y \stackrel{\mathcal{L}}{=} IY + (1 - I)\tilde{Y},$$

where Y, $\tilde{Y}$ and I are independent, with $Y \stackrel{\mathcal{L}}{=} \tilde{Y}$. The additive term S_n is asymptotically annihilated.

The characteristic property of this fixed-point equation is that it is degenerate as it is satisfied by *any* distribution. By contrast to this degenerate behavior of the depth D_n the limit distribution of Δ_n came about as a fixed-point of the equation

$$X \stackrel{\mathcal{L}}{=} IX + J\tilde{X} + K\frac{1}{\sqrt{2}}(Y + \tilde{Y}),$$

where $(\tilde{X}, \tilde{Y})$ is an independent probabilistic copy of (X, Y), the variates $Y, \tilde{Y}$ are standard Gaussian, the row vector $(X, \tilde{X}, Y, \tilde{Y})$ is independent of (I, J, K) and (I, J, K) has a uniform distribution on $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.

As we have seen the unique distribution satisfying the equation is the standard normal distribution. The additive term $K(Y + \tilde{Y})/\sqrt{2}$, even though it depends on the recurring terms, guides the solution to be the standard normal distribution. These dependent random variables are disengaged via mutually exclusive indicators.

The contrast between degenerate distributional equations satisfied by any distribution, and those which distinguish a distribution as their unique solution under certain constraints demonstrates a significant difference between the corresponding problems from the point of view of the application of the contraction method.

REFERENCES

CRAMER, M. and RÜSCHENDORF, L. (1996). Analysis of recursive algorithms by the contraction method. *Athens Conference on Applied Probability and Time Series Analysis. Lecture Notes in Statist.* **114** 18–33. Springer, New York.

DEVROYE, L. (1988). Applications of the theory of records in the study of random trees. *Acta Inform.* **26** 123–130.

DEVROYE, L. (1999). Universal limit laws for depths in random trees. *SIAM J. Comput.* **28** 409–432.

GUTMAN, I. and POLANSKY, O. (1986). *Mathematical Concepts in Organic Chemistry.* Springer, Berlin.

HWANG, H.-K. and NEININGER, R. (2002). Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.* **31** 1687–1722.

KEMP, R. (1984). *Fundamentals of the Average Case Analysis of Particular Algorithms.* Wiley, New York.

KNUTH, D. (1998). *The Art of Computer Programming* **3**. *Sorting and Searching*, 2nd ed. Addison-Wesley, Reading, MA.

MAHMOUD, H. (1992). *Evolution of Random Search Trees.* Wiley, New York.

MAHMOUD, H. (2000). *Sorting*: *A Distribution Theory*. Wiley, New York.

MAHMOUD, H. and SMYTHE, R. (1998). Probabilistic analysis of MULTIPLE QUICK SELECT. *Algorithmica* **22** 569–584.

NEININGER, R. (2001). On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms* **19** 498–524.

NEININGER, R. (2002). The Wiener index of random trees. *Combin. Probab. Comput.* To appear.

NEININGER, R. and RÜSCHENDORF, L. (2002a). Rates of convergence for Quicksort. *J. Algorithms.* To appear.

NEININGER, R. and RÜSCHENDORF, L. (2002b). A general contraction theorem and asymptotic normality in combinatorial structures. Technical Report 01-25, Univ. Freiburg.

RACHEV, S. (1991). *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York.

RACHEV, S. and RÜSCHENDORF, L. (1995). Probability metrics and recursive algorithms. *Adv. in Appl. Probab.* **27** 770–799.

RÖSLER, U. (1991). A limit theorem for "Quicksort." *RAIRO Inform. Théor. Appl.* **25** 85–100.

RÖSLER, U. (1992). A fixed point theorem for distributions. *Stochastic Process. Appl.* **42** 195–214.

RÖSLER, U. (2001). On the analysis of stochastic divide and conquer algorithms. *Algorithmica* **29** 238–261.

RÖSLER, U. and RÜSCHENDORF, L. (2001). The contraction method for recursive algorithms. *Algorithmica* **29** 3–33.

SEIDEL, R. and ARAGON, C. R. (1996). Randomized search trees. *Algorithmica* **16** 464–497.

TRINAJSTIĆ, N. (1992). *Chemical Graph Theory*. CRC Press, Boca Raton, FL.

ZOLOTAREV, V. M. (1976). Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory Probab. Appl.* **21** 721–737.

DEPARTMENT OF STATISTICS
GEORGE WASHINGTON UNIVERSITY
2201 G. STREET NW
WASHINGTON, DC 20052
E-MAIL: hosam@gwu.edu

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
3480 UNIVERSITY STREET
MONTREAL H3A 2K6
CANADA
E-MAIL: neiningr@cs.mcgill.ca