AISM (2008) 60:319–343
DOI 10.1007/s10463-006-0107-z

Limit laws for the Randić index of random binary tree models

Qunqiang Feng · Hosam M. Mahmoud · Alois Panholzer

Received: 20 February 2006 / Revised: 18 July 2006 / Published online: 10 February 2007

Abstract We investigate the Randić index of random binary trees under two standard probability models: the one induced by random permutations and the Catalan (uniform). In both cases the mean and variance are computed by recurrence methods and shown to be asymptotically linear in the size of the tree. The recursive nature of binary search trees lends itself in a natural way to application of the contraction method, by which a limit distribution (for a suitably normalized version of the index) is shown to be Gaussian. The Randić index (suitably normalized) is also shown to be normally distributed in binary Catalan trees, but the methodology we use for this derivation is singularity analysis of formal generating functions.

Keywords Random trees · Binary search trees · Catalan trees · Recurrence · Moments · Contraction method · Functional equation · Computational chemistry · Chemical index · Topological index

Q. Feng
Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, China
e-mail: fengqq@ustc.edu.cn

H. M. Mahmoud (✉)
Department of Statistics, The George Washington University, Washington, DC 20052, USA
e-mail: hosam@gwu.edu

A. Panholzer
Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr., 8-10/104, 1040 Wien, Austria
e-mail: Alois.Panholzer@tuwien.ac.at

1 Introduction

For quite some time there has been rising interest in the field of computational chemistry in topological indexes that capture the structural essence of compounds; see for example Trinajstić (1992), who traces the history back to more than 150 years, and mentions the existence of over 120 topological indexes. Indexes focused on distances within the graph of the molecule include the popular Wiener index, which is considered a measure for the spread of the shape of the molecule. This index was introduced toward an understanding of the boiling points of paraffin in Wiener (1947). Measures such as the Randić (Randić, 1975), the Balaban (Balaban 1982, 1983), and the Zagreb (Gutman and Trinajstić 1972) indexes are focused on combinations of degree sequences of the molecular graph, with interpretations relating to the overall connectivity of the molecule. Additional types of topological indexes bear on information theory (see Bonchev and Trinajstić, 1983). Numerous papers in the chemistry literature can be found on these topics. For examples of use in specific areas of applied research, such as pharmacology and toxicology see Basak (1987) and Kier and Hall (1976). For a textbook style presentation and discussion of the broader context, we refer the reader to the books by Devillers and Balaban (1999), Gutman and Polansky (1986), and Trinajstić (1992). For a view from the other end, Harary (1969) is more specialized in graph theory, but makes connections to chemistry.

We are interested in the distributional properties of chemical indexes in classes of random trees. Some handful of works have been written on the Wiener index of a few classes of random trees including Christophi and Mahmoud (2005), Devroye and Neininger (2004), Janson (2003), Janson and Chassaing (2004), Mahmoud and Neininger (2002), Neininger (2002), and Panholzer and Prodinger (2004). However, we are aware of only one investigation of the Randić index in random trees by Clark and Moon (2000). Our aim in this note is to add a contribution on the probabilistic behavior of the Randić index for binary trees.

2 The Randić index of a binary tree

The degree of vertex v in a graph $G = (V, \mathcal{E})$ with vertex set V and edge set $\mathcal{E}$ is the number of edges that are incident with V. We shall denote the degree of v by $\deg(v)$. The *Randić index*, also called the *connectivity index*, $R(G)$ (with parameter α) of a molecule with (undirected) graph G is the sum

$$R(G) = \sum_{\{u,v\}\in\mathcal{E}} \big(\deg(u)\ \deg(v)\big)^{\alpha}.$$

In other words, the Randić index is

$$\sum_{i,j} (ij)^{\alpha} M_{ij},$$

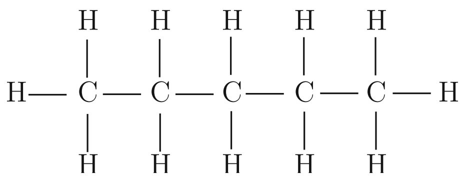

Fig. 1 The tree structure of a Pentane molecule

where M_{ij} is a count of the number of edges (bonds) extending between pairs of vertices (atoms) with degrees (valance) i and j. The Randić index with parameter $\alpha = -1/2$ has been the most popular in the chemistry literature (see for example Estrada, 2002), but other variations have been discussed, too (see for example Miličević and Nikolić, 2004).

Remark The case $\alpha = 0$ is a trivially *degenerate* case in which $R(G) = |\mathcal{E}|$. In a tree of size n, the Randić index will always be $n - 1$. We exclude this choice in the sequel.

Many chemical compounds have a tree-like structure. For example, The entire family of alkanes (saturated hydrocarbons) have tree graphs. This family has many compounds common in our daily life such as Methane (CH_4), Ethane (C_2H_6), Propane (C_3H_8), Butane(C_4H_{10}), which we use in fuels for heating, cooking, transportation, and many industrial processes. Figure 1 illustrates the tree structure of a Pentane (C_5H_{12}) molecule, a common commercially available liquid solvent, and an additive to automotive and aviation fuels.

Connections of chemistry to random binary trees have been noted, for example in Quintas and Szymański (1992), where they studied molecules with a binary tree structure, where nodes of valence 3 are saturated and the unsaturated nodes have affinity that is inversely proportional to their valance. The technical name that Quintas and Szymański (1992) used for their tree model is “recursive trees with bounded degrees,” but they do coincide with “binary search trees” defined below (even though this is not mentioned explicitly in that source).

A *binary tree* is a hierarchy of nodes. If the tree is not empty, it has one distinguished node as its root, and each node has up to two substructures of nodes (called subtrees) positionally distinguished by their orientation as left and right. Thus, a binary tree is either empty, or has left and right subtrees that are themselves recursively binary trees.

Several models of randomness are in common use on binary trees. Binary search trees grown from the insertion of a random permutation are of paramount importance in sorting, searching and a myriad of other combinatorial algorithms (see Knuth, 1998 or Mahmoud, 2000). As noted already, this model is of relevance to chemistry (Quintas and Szymański, 1992). The Catalan probability model is a natural alternative in which all trees of a given size are equally likely. The model is considered appropriate for formal languages and computer algebra (see Kemp, 1984).

In this note we investigate the distribution of the Randić index of a random binary tree under both the random permutation model and the Catalan model.

The sequel is organized as follows. In Sect. 3 we discuss two sets of stochastic recurrence equations for the Randić index of a binary tree: one collective and one refined by conditioning on root degrees. These are recurrences on the trees (as points in a sample space) and thus are valid regardless of the probability model. They apply to the binary search trees as well as binary Catalan trees. The point of departure is the splitting probabilities at the root. Section 4 is dedicated to binary search trees. In these trees, each type of recurrence is more suitable for a particular purpose. The recurrence refined by root degree is a transparent tool for the computation of exact moments: The mean is taken up in Sect. 4.1, and the variance is taken up in Sect. 4.2. However, it is the collective recurrence that is directly amenable to the univariate contraction method, as is demonstrated in Sect. 4.3. Section 5 is dedicated to Catalan trees, where the methodology is the singularity analysis of formal moment generating functions.

3 Stochastic setup

We shall set up recurrence equations for the Randić index, valid for any binary tree, regardless of its probability model. Let the random variable R_n be the Randić index and D_n be the degree of the root of a random binary tree T of size n. We employ a global decomposition at the root of T. When the left and right subtrees are connected to the root of T, adjustments in their root degrees need to be made. These adjustments depend on the structure (particularly the distribution of degrees of nodes) of the first three levels in T. Take for example the standard Randić index with parameter $\alpha = -1/2$. Suppose we computed the Randić indexes of the left and right subtrees, let us say they are respectively R' and R'', and we would like now to compute it for T. If $R' = 0$, the left subtree is empty, and suppose the right subtree has root x, with two children, each of degree 2 (i.e. each has one child). When we adjoin the left (empty) and right subtrees to the root, we render the degree of x equal to 3, and the two children of x are now connected to a degree–3 parent. The Randić index of T needs adjustment by the new contribution of the root of T, and the two computed indexes for the left and right subtrees also need to be adjusted. The two children of x now contribute $2/\sqrt{6}$, and no longer 1 as is their contribution to the index for the right subtree. The connection to the root adds $1/\sqrt{3}$. That is, the global index is $R'' + 1/\sqrt{3} + 2/\sqrt{6} - 1$. However, if the root of the left subtree has degree 1, and the first two levels in the right subtree have the same degrees as in the previous example, the adjoining of the left and right subtrees renders the Randić index of T equal to $R'' + 3/\sqrt{6} + 1/\sqrt{2} - 1$.

It is easier to work with a *modified* Randić index that does not get affected by adjoining the subtrees, but rather only requires changes based on the degree of the root. That is, a modification in which the contribution of the modified index in each subtree enters the formula in a purely additive manner. Such a modified index Y_n is defined in exactly the same way as R_n, except that the degree $\deg(r)$ of the root r of the whole tree is always enhanced to $\deg(r) + 1$ in the computation with any pair of nodes involving the root. It is as if we

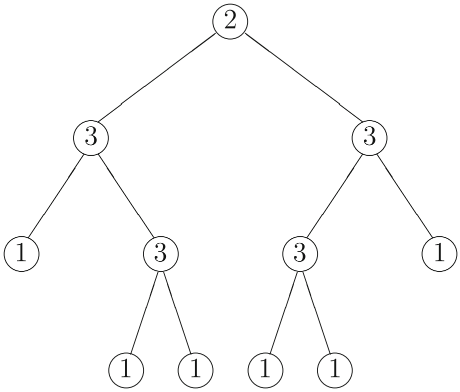

Fig. 2 A binary tree labeled with node degrees

are computing the usual Randić index for a tree where the root is connected to an extra (incoming) fictitious edge that does not have a vertex at its other end, thus accounting for an enhanced root degree, without introducing any new pairs. Figure 2 illustrates the notion of the Randić index. In the figure, the nodes are labeled by their degree. In the tree of Fig. 2, the standard Randić index (with $\alpha = -\frac{1}{2}$) is

$$R_{11} = \frac{6}{\sqrt{3}} + \frac{2}{\sqrt{6}} + \frac{2}{3}.$$

Let us set up some notation to formulate our problem recursively. Let $\mathbf{1}_A$ be the indicator of the event A, i.e., the function that assumes the value 1, if A occurs, and is 0 otherwise. Let the size (number of nodes) of the left subtree be L_n. The variable $R_n(T)$ is the Randić index of a given binary tree T of size n. As usual, we hide the sample point T in this (and in all similar notation).

For notions of probability, we use the notation $\stackrel{\mathcal{L}}{=}$ to indicate *exact* equality in distribution, and $\stackrel{\mathcal{D}}{\longrightarrow}$ is reserved for convergence in distribution. Likewise, $\stackrel{P}{\longrightarrow}$ is reserved for convergence in probability. To use probabilistic copies from the substructures, consistently any random variable pertaining to the left subtree will carry a prime symbol, and any one pertaining to the right subtree will carry double primes. These copies will receive notation such as the hat or tilde to indicate that they are *conditionally independent*. For example, we may refer to the Randić index of the left subtree by R'_{L_n}, and that of the right subtree by $\hat{R}''_{n-1-L_n}$. Even though R'_{L_n} and $\hat{R}''_{n-1-L_n}$ are dependent, through their dependency on L_n, they are conditionally independent, that is $R'_j \stackrel{\mathcal{L}}{=} R_j$ and $\hat{R}''_k \stackrel{\mathcal{L}}{=} R_k$, and R'_j and $\hat{R}''_k$, are independent for all fixed values of j and k.

Set

$$I_n(k) = \mathbf{1}_{\{1 \le k \le n-2\}},$$

and

$$J_n(k) = \mathbf{1}_{\{D_n=k\}}.$$

For $n \geq 1$, the indicators $J_n(0)$, $J_n(1)$, and $J_n(2)$ are mutually exclusive in the sense that

$$J_n(0) + J_n(1) + J_n(2) = 1.$$

The boundary values are $Y_0 = Y_1 = 0$, and so are R_0 and R_1 equal to 0; therefore, $J_0(0)$, $J_0(1)$, and $J_0(2)$ must be identically equal to 0.

The left subtree has modified Randić index Y'_{L_n}, and that of the right subtree is Y''_{n-1-L_n}. When adjoined to a root (with a fictitious incoming edge) the degrees of the two roots of the subtrees are upgraded by 1. So, if the roots of the subtrees are r' and r'', respectively, the collective Randić index of the whole tree is the sum of the two Randić indexes, upgraded by the amount $((\deg(r)+1)(\deg(r')+1))^\alpha + ((\deg(r)+1)(\deg(r'')+1))^\alpha$. So, we see that by conditioning on the degree of the roots of the two subtrees we obtain the following distributional recurrence, for Y_n, for $n \geq 2$,

$$\begin{aligned} Y_n \overset{\mathcal{L}}{=} Y'_{L_n} + \tilde{Y}''_{n-1-L_n} + \left(3^\alpha I_n(L_n) + 2^\alpha (1 - I_n(L_n)\right) \\ \times \sum_{i=0}^{2} (i+1)^\alpha \left(J'_{L_n}(i) + J''_{n-1-L_n}(i)\right). \end{aligned} \tag{1}$$

This recurrence is in a form suitable for an application of the univariate contraction method, as we shall discuss later on in Sect. 4.3. Figure 3 shows the decomposition of the tree of Fig. 2, into a root and two subtrees with modified Randić index $\frac{3}{\sqrt{3}} + \frac{1}{3}$ each. In the figure the nodes are labeled by their degrees, except the roots of the tree and the subtrees; instead these are labeled by the enhanced degrees (including a fictitious incoming edge for each).

When we hook up the two subtrees to the root of the whole tree, we can simply assemble R_n from the modified Randić indexes of the left and right subtrees with some extra adjustments (argued in a manner similar to the rationale for the recurrence of Y_n). It is given by the recurrence

$$\begin{aligned} R_n \overset{\mathcal{L}}{=} Y'_{L_n} + \tilde{Y}''_{n-1-L_n} + \left(2^\alpha I_n(L_n) + (1 - I_n(L_n)\right) \\ \times \sum_{i=0}^{2} (i+1)^\alpha \left(J'_{L_n}(i) + J''_{n-1-L_n}(i)\right), \end{aligned} \tag{2}$$

which is valid for $n \geq 2$.

The recurrence for R_n includes dependent random variables on the right-hand side. For example, Y'_{L_n} and $J'_{L_n}(i)$ are dependent. While this does not influence the mean computation, it introduces some inconvenience in the calculation

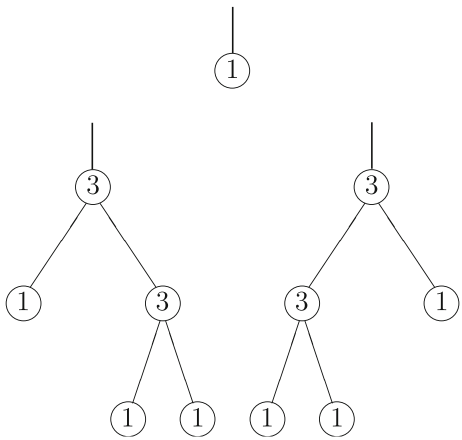

Fig. 3 The decomposition of the tree of Fig. 2, with enhanced root degree for the computation of the modified Randić index

of higher moments. When the right-hand side of (1) is squared or raised to higher powers, cross-products of dependent random variables like $Y'_{L_n} J'_{L_n}(2)$, and $Y''_{n-1-L_n} J''_{n-1-L_n}(1)$ appear. Therefore, toward the moments, we work with more convenient recurrences that are conditional refinements of (1)–(2), where the conditioning is according to the root degree.

Let $R_n^{(i)}$ denote $R_n \mid \{\deg(r) = i\}$, that is, R_n conditioned on the event that the size-n binary tree has a root with i children. The conditioned random variables $Y_n^{(0)}$, $Y_n^{(1)}$, $Y_n^{(2)}$ are defined analogously.

Then, for all $n \geq 1$, R_n and Y_n are given by:

$$R_n = \sum_{i=0}^{2} J_n(i) R_n^{(i)},$$

and

$$Y_n = \sum_{i=0}^{2} J_n(i) Y_n^{(i)}.$$

The conditioned random variable $R_n^{(i)}$ and $Y_n^{(i)}$, with $0 \leq i \leq 2$, satisfy the distributional recurrences

$$R_n^{(1)} \stackrel{\mathcal{L}}{=} \sum_{i=0}^{2} J'_{S_n}(i) \big(\big(Y_{n-1}^{(i)}\big)' + (i+1)^{\alpha} \big), \quad \text{for } n \geq 2,$$

$$R_n^{(2)} \stackrel{\mathcal{L}}{=} \sum_{i=0}^{2} J'_{S_n}(i)\big((\hat{Y}_{S_n}^{(i)})' + 2^\alpha (i+1)^\alpha\big)$$
$$+ \sum_{i=0}^{2} J''_{n-1-S_n}(i)\big((\tilde{Y}_{n-1-S_n}^{(i)})'' + 2^\alpha (i+1)^\alpha\big), \qquad \text{for } n \geq 3,$$
$$Y_n^{(1)} \stackrel{\mathcal{L}}{=} \sum_{i=0}^{2} J'_{S_n}(i)\big((Y_{n-1}^{(i)})' + 2^\alpha (i+1)^\alpha\big), \quad \text{for } n \geq 2,$$
$$Y_n^{(2)} \stackrel{\mathcal{L}}{=} \sum_{i=0}^{2} J'_{S_n}(i)\big((\hat{Y}_{S_n}^{(i)})' + 3^\alpha (i+1)^\alpha\big)$$
$$+ \sum_{i=0}^{2} J''_{n-1-S_n}\big((\tilde{Y}_{n-1-S_n}^{(i)})'' + 3^\alpha (i+1)^\alpha\big), \quad \text{for } \ n \geq 3,$$

with boundary values

$$R_n^{(0)} = Y_n^{(0)} = 0, \quad \text{for } \ n \geq 1,$$
$$R_1^{(1)} = Y_1^{(1)} = R_1^{(2)} = Y_1^{(2)} = R_2^{(2)} = Y_2^{(2)} = 0;$$

here S_n denotes $L_n \mid L_n \in \{1, \ldots, n-2\}$, i.e. the size of the left subtree conditioned on the event that we do not have an extremal case of a root of degree 1.

4 The Randić index of random binary search trees

A binary search tree is constructed from the permutation $(\pi_1, \ldots, \pi_n)$ of the set $\{1, 2, \ldots, n\}$ by a standard search algorithm. The first element of the permutation is inserted in an empty tree, a root node is allocated for it. A subsequent element π_j (with $j \geq 2$) is guided to the left subtree if $\pi_j < \pi_1$, otherwise it is taken into the right subtree. In the recipient subtree π_j is treated recursively by the same insertion algorithm, until it is inserted in an empty subtree, at which point a node is allocated for it and adjoined appropriately as a left (right) child if its label is less than (at least as much as) the label of the last node on the search path.

In the probability model induced by random permutations we assume that the tree is built from permutations of $\{1, \ldots, n\}$, where a uniform probability is imposed on the *permutations* rather than on the trees. When all $n!$ permutations are uniformly *random*, occurring equally likely, binary search trees are not equally likely. Several permutations give rise to the same tree, favoring shorter balanced trees to tall linear shapes (see, Mahmoud, 1992). Henceforth the term *random binary search tree* will refer to a binary search tree built from a random permutation.

For $n \geq 2$, the random permutations induce the distribution:

$$P(D_n = i) = \begin{cases} 0, & i = 0; \\ \frac{2}{n}, & i = 1; \\ 1 - \frac{2}{n}, & i = 2, \end{cases}$$

whereas for $n = 1$ one has $P(D_1 = 0) = 1$. Under the random permutation model L_n is uniformly distributed on the set $\{0, 1, \ldots, n-1\}$. For only aesthetic reasons, we refer to L_n as U_n until the end of this section, to reflect this probabilistic structure in binary search trees. The random variable $S_n = L_n \,|\, L_n \in \{1, \ldots, n-2\}$ is uniformly distributed on $\{1, \ldots, n-2\}$.

4.1 The expectation

From the above stochastic recurrences we get the following recurrence for the expected values of Y_n:

$$\mathbf{E}(Y_n) = \frac{2}{n}\mathbf{E}\left(Y_n^{(1)}\right) + \left(1 - \frac{2}{n}\right)\mathbf{E}\left(Y_n^{(2)}\right), \quad \text{for } n \geq 2, \tag{3}$$

with

$$\begin{aligned} \mathbf{E}(Y_n^{(1)}) &= \left(\mathbf{E}\left(Y_{n-1}^{(0)}\right) + 2^\alpha\right)\mathbf{1}_{\{n=2\}} + \left[\frac{2}{n-1}\left(\mathbf{E}\left(Y_{n-1}^{(1)}\right) + 4^\alpha\right)\right. \\ &\quad \left. + \left(1 - \frac{2}{n-1}\right)\left(\mathbf{E}\left(Y_{n-1}^{(2)}\right) + 6^\alpha\right)\right]\mathbf{1}_{\{n\geq 3\}}, \quad \text{for } n \geq 2, \end{aligned} \tag{4}$$

$$\begin{aligned} \mathbf{E}(Y_n^{(2)}) &= \frac{2}{n-2}\left(\mathbf{E}\left(Y_1^{(0)}\right) + 3^\alpha\right) + \frac{2}{n-2}\sum_{k=2}^{n-2}\frac{2}{k}\left(\mathbf{E}\left(Y_k^{(1)}\right) + 6^\alpha\right) \\ &\quad + \frac{2}{n-2}\sum_{k=2}^{n-2}\left(1 - \frac{2}{k}\right)\left(\mathbf{E}\left(Y_k^{(2)}\right) + 9^\alpha\right), \quad \text{for } n \geq 3. \end{aligned} \tag{5}$$

In what follows the sth harmonic number is denoted by H_s; that is,

$$H_s = \sum_{j=1}^{s}\frac{1}{j}.$$

Plugging in the recurrences for $\mathbf{E}(Y_n^{(1)})$ and $\mathbf{E}(Y_n^{(2)})$ in (3) leads to the recurrence

$$\begin{aligned}\mathbf{E}(Y_n) &= \frac{2}{n}\sum_{k=1}^{n-2}\mathbf{E}(Y_k) + \frac{4}{n(n-1)}4^\alpha + \frac{2}{n}\Big(1 - \frac{2}{n-1}\Big)6^\alpha + \frac{2}{n}3^\alpha \\ &\quad + \frac{4}{n}\big(H_{n-2} - 1\big)6^\alpha + \frac{2}{n}\big(n - 1 - 2H_{n-2}\big)9^\alpha, \quad \text{for } n \geq 3, \end{aligned} \tag{6}$$

with boundary values $Y_1 = 0$, $Y_2 = 2^\alpha$.

This is a "Quicksort"-like recurrence, which has been treated in various works (see Hwang and Neininger, 2002; Kirschenhofer and Prodinger, 1998; Kirschenhofer et al., 1997; Panholzer, 2003; Panholzer and Prodinger, 1998; Prodinger, 1995). We only outline here the strategy: One begins by differencing $(n-1)\,\mathbf{E}(Y_{n-1})$ from $n\,\mathbf{E}(Y_n)$. We have a recurrence of the form

$$\mathbf{E}(Y_n) = \mathbf{E}(Y_{n-1}) + \xi(n),$$

where the so-called *toll function* $\xi(n)$ collects all the nonrecursive terms. So, a solution in the form of a sum of $\xi(j)$ for j running from a suitable initial value up to n is imminently available. What takes considerable effort is the cleaning up to reduce these sums into simple closed form expressions in terms of standard functions. This latter step is facilitated by some heavy use of a computer algebra system, but still requires substantial nontrivial human guidance. In the sequel we shall suppress some steps, where the computation is too lengthy but remains straightforward.

Using the distributional equations (1)–(2) one can show the following relation between $\mathbf{E}(R_n)$ and $\mathbf{E}(Y_n)$:

$$\begin{aligned}\mathbf{E}(R_n) &= \mathbf{E}(Y_n) + \frac{4}{n(n-1)}(2^\alpha - 4^\alpha) + \frac{2}{n}\Big(1 - \frac{2}{n-1}\Big)(3^\alpha - 6^\alpha) \\ &\quad + \frac{2}{n}\big(2^\alpha - 3^\alpha\big) + \frac{4}{n}\big(H_{n-2} - 1\big)\big(4^\alpha - 6^\alpha\big) \\ &\quad + \frac{2}{n}\big(n - 1 - 2H_{n-2}\big)\big(6^\alpha - 9^\alpha\big), \quad \text{for } n \geq 3, \end{aligned}$$

with boundary values $R_1 = 0$, $R_2 = 1$. Solving this Quicksort-like recurrence by the strategy outlined above leads to the following explicit formulas:

$$\begin{aligned}\mathbf{E}(R_n) &= \frac{(3\cdot 2^\alpha + 3\cdot 3^\alpha + 2\cdot 4^\alpha + 5\cdot 6^\alpha + 5\cdot 9^\alpha)}{18}\,n \\ &\quad + \frac{3\cdot 2^\alpha + 3\cdot 3^\alpha + 2\cdot 4^\alpha + 41\cdot 6^\alpha - 67\cdot 9^\alpha}{18} \\ &\quad + \frac{4(4^\alpha - 2\cdot 6^\alpha + 9^\alpha)}{n}\,H_{n-2} + \frac{2(2^\alpha - 2\cdot 4^\alpha - 6^\alpha + 2\cdot 9^\alpha)}{n} \\ &\quad + \frac{4(3\cdot 2^\alpha - 3\cdot 3^\alpha - 2\cdot 4^\alpha + 6^\alpha + 9^\alpha)}{3n(n-1)}, \quad \text{for } n \geq 3, \end{aligned}$$

and

$$\mathbf{E}(Y_n) = \frac{(3\cdot 2^\alpha + 3\cdot 3^\alpha + 2\cdot 4^\alpha + 5\cdot 6^\alpha + 5\cdot 9^\alpha)}{18} n + \frac{3\cdot 2^\alpha + 3\cdot 3^\alpha + 2\cdot 4^\alpha + 5\cdot 6^\alpha - 31\cdot 9^\alpha}{18} + \frac{2(9^\alpha - 6^\alpha)}{n} + \frac{4(4^\alpha - 2\cdot 6^\alpha + 9^\alpha)}{3n(n-1)}, \quad \text{for } n \geq 3.$$

If our interest is only in average values, we could have solved the recurrence (6), without necessarily going via the recurrence system (4)–(5) for $\mathbf{E}(Y_n^{(1)})$ and $\mathbf{E}(Y_n^{(2)})$. However, the refined formulas for the latter two variables are helpful in variance computations.

In the following we use the abbreviation $f_n := \mathbf{E}(Y_n^{(1)})$, and $g_n := \mathbf{E}(Y_n^{(2)})$. Subsequently, we have to study also the following system of recurrences:

$$f_n = \frac{2}{n-1} f_{n-1} + \left(1 - \frac{2}{n-1}\right) g_{n-1} + a_n, \quad n \geq 3,$$
$$g_n = \frac{2}{n-2} \sum_{k=2}^{n-2} \frac{2}{k} f_k + \frac{2}{n-2} \sum_{k=2}^{n-2} \left(1 - \frac{2}{k}\right) g_k + b_n, \quad n \geq 3,$$

where a_n and b_n are given by:

$$a_n = \frac{2}{n-1} 4^\alpha + \left(1 - \frac{2}{n-1}\right) 6^\alpha,$$
$$b_n = \frac{2}{n-2} 3^\alpha + \frac{2}{n-2} 6^\alpha \sum_{k=2}^{n-2} \frac{2}{k} + \frac{2}{n-2} 9^\alpha \sum_{k=2}^{n-2} \left(1 - \frac{2}{k}\right).$$

Despite the fact that this system of recurrences is no longer the well-studied Quicksort-recurrence, we can reduce it to that by eliminating g_n. One obtains

$$f_n = \frac{2}{n-1} \sum_{k=3}^{n-1} f_k + a_n + \left(1 - \frac{2}{n-1}\right) b_{n-1} - \frac{2}{n-1} \sum_{k=3}^{n-2} a_k, \quad n \geq 4.$$

This gives the solution

$$f_n = n\left(\frac{f_4}{4} + \sum_{k=5}^{n} s_k\right), \quad n \geq 5,$$

with

$$s_n := \frac{(n-1)a_n - (n-2)a_{n-1}}{n(n-1)} + \frac{(n-3)b_{n-1} - (n-4)b_{n-2}}{n(n-1)} - \frac{2a_{n-2}}{n(n-1)},$$

and boundary values $f_4 = \frac{2}{3}2^\alpha + \frac{2}{3}3^\alpha + \frac{4}{3}4^\alpha + \frac{1}{3}6^\alpha, f_3 = 2^\alpha + 4^\alpha, f_2 = 2^\alpha, f_1 = 0$.
We further get

$$g_n = \frac{n}{n-2}\left(f_{n+1} - \frac{2}{n}f_n - a_{n+1}\right), \quad n \geq 4,$$

which yields the solution

$$g_n = \frac{n}{n-2}\left((n+1)s_{n+1} - a_{n+1} + (n-1)\left(\frac{f_4}{4} + \sum_{k=5}^{n} s_k\right)\right), \quad n \geq 4,$$

with boundary values $g_4 = 2^\alpha + 3^\alpha + 6^\alpha, g_3 = 2 \cdot 3^\alpha$, $g_2 = 0, g_1 = 0$.
Plugging in the values of a_n and b_n leads after simplification to the following explicit formulas:

$$\begin{aligned}
\mathbf{E}(Y_n^{(1)}) &= \frac{(3 \cdot 2^\alpha + 3 \cdot 3^\alpha + 2 \cdot 4^\alpha + 5 \cdot 6^\alpha + 5 \cdot 9^\alpha)n}{18} + 6^\alpha - 2 \cdot 9^\alpha \\
&\quad + \frac{2(4^\alpha - 2 \cdot 6^\alpha + 9^\alpha)}{n-1} + \frac{4(4^\alpha - 2 \cdot 6^\alpha + 9^\alpha)}{3(n-1)(n-2)}, \quad n \geq 4, \\
\mathbf{E}(Y_n^{(2)}) &= \frac{(3 \cdot 2^\alpha + 3 \cdot 3^\alpha + 2 \cdot 4^\alpha + 5 \cdot 6^\alpha + 5 \cdot 9^\alpha)n}{18} \\
&\quad + \frac{3 \cdot 2^\alpha + 3 \cdot 3^\alpha + 2\,4^\alpha + 5 \cdot 6^\alpha - 31 \cdot 9^\alpha}{18} \\
&\quad + \frac{3 \cdot 2^\alpha + 3 \cdot 3^\alpha + 2 \cdot 4^\alpha - 31 \cdot 6^\alpha + 23 \cdot 9^\alpha}{9(n-2)} \\
&\quad - \frac{8(4^\alpha - 2 \cdot 6^\alpha + 9^\alpha)}{3(n-2)^\alpha}, \quad n \geq 4.
\end{aligned}$$

4.2 The variance

Upon squaring the stochastic recurrences (1) then taking expectations, we obtain the following recurrence for the second moment of Y_n:

$$\mathbf{E}(Y_n^2) = \frac{2}{n}\mathbf{E}\left(\left(Y_n^{(1)}\right)^2\right) + \left(1 - \frac{2}{n}\right)\mathbf{E}\left(\left(Y_n^{(2)}\right)^2\right), \quad \text{for } n \geq 2, \tag{7}$$

with

$$\begin{aligned}
\mathbf{E}\left(\left(Y_n^{(1)}\right)^2\right) &= \frac{2}{n-1}\mathbf{E}\left(\left(Y_{n-1}^{(1)}\right)^2\right) + \left(1 - \frac{2}{n-1}\right)\mathbf{E}\left(\left(Y_{n-1}^{(2)}\right)^2\right) \\
&\quad + \frac{2}{n-1}\left(2 \cdot 4^\alpha \mathbf{E}\left(Y_{n-1}^{(1)}\right) + 4^{2\alpha}\right) \\
&\quad + \left(1 - \frac{2}{n-1}\right)\left(2 \cdot 6^\alpha \mathbf{E}\left(Y_{n-1}^{(2)}\right) + 6^{2\alpha}\right), \quad \text{for } n \geq 3,
\end{aligned} \tag{8}$$

$$\begin{aligned}
\mathbf{E}\left(\left(Y_n^{(2)}\right)^2\right) &= \frac{2}{n-2}\sum_{k=2}^{n-2}\frac{2}{k}\mathbf{E}\left(\left(Y_k^{(1)}\right)^2\right) + \frac{2}{n-2}\sum_{k=2}^{n-2}\left(1-\frac{2}{k}\right)\mathbf{E}\left(\left(Y_k^{(2)}\right)^2\right) \\
&\quad + \frac{2}{n-2}3^{2\alpha} + \frac{2}{n-2}\sum_{k=2}^{n-2}\frac{2}{k}\left(2\cdot 6^{\alpha}\mathbf{E}\left(Y_k^{(1)}\right) + 6^{2\alpha}\right) \\
&\quad + \frac{2}{n-2}\sum_{k=2}^{n-2}\left(1-\frac{2}{k}\right)\left(2\cdot 9^{\alpha}\mathbf{E}\left(Y_k^{(2)}\right) + 9^{2\alpha}\right) \\
&\quad + \frac{4}{n-2}3^{\alpha}\left(\frac{2}{n-2}\left(\mathbf{E}\left(Y_{n-2}^{(1)}\right) + 6^{\alpha}\right)\right. \\
&\quad \left. + \left(1-\frac{2}{n-2}\right)\left(\mathbf{E}\left(Y_{n-2}^{(2)}\right) + 9^{\alpha}\right)\right) \\
&\quad + \frac{2}{n-2}\sum_{k=2}^{n-3}\left(\frac{2}{k}\left(\mathbf{E}\left(Y_k^{(1)}\right) + 6^{\alpha}\right) + \left(1-\frac{2}{k}\right)\left(\mathbf{E}\left(Y_k^{(2)}\right) + 9^{\alpha}\right)\right) \\
&\quad \times\left(\frac{2}{n-1-k}\left(\mathbf{E}\left(Y_{n-1-k}^{(1)}\right) + 6^{\alpha}\right)\right. \\
&\quad \left. + \left(1-\frac{2}{n-1-k}\right)\left(\mathbf{E}\left(Y_{n-1-k}^{(2)}\right) + 9^{\alpha}\right)\right), \qquad (9)
\end{aligned}$$

which is valid for $n \geq 4$. Combining the recurrences the system (8)–(9) leads again to a Quicksort-like recurrence in (7) for the required second moment:

$$\mathbf{E}\left(Y_n^2\right) = \frac{2}{n}\sum_{k=1}^{n-1}\mathbf{E}\left(Y_k^2\right) + t_n, \quad \text{for } n \geq 4,$$

with toll function

$$\begin{aligned}
t_n &= \frac{2}{n}\left(\frac{2}{n-1}\left(2\cdot 4^{\alpha}\mathbf{E}\left(Y_{n-1}^{(1)}\right) + 4^{2\alpha}\right) + \left(1-\frac{2}{n-1}\right)\left(2\cdot 6^{\alpha}\mathbf{E}\left(Y_{n-1}^{(2)}\right) + 6^{2\alpha}\right)\right) \\
&\quad + \frac{2}{n}3^{2\alpha} + \frac{2}{n}\sum_{k=2}^{n-2}\frac{2}{k}\left(2\cdot 6^{\alpha}\mathbf{E}\left(Y_k^{(1)}\right) + 6^{2\alpha}\right) \\
&\quad + \frac{2}{n}\sum_{k=2}^{n-2}\left(1-\frac{2}{k}\right)\left(2\cdot 9^{\alpha}\mathbf{E}\left(Y_k^{(2)}\right) + 9^{2\alpha}\right) \\
&\quad + \frac{2}{n}2\cdot 3^{\alpha}\left(\mathbf{E}\left(Y_{n-2}\right) + \frac{2}{n-2}6^{\alpha} + \left(1-\frac{2}{n-2}\right)9^{\alpha}\right)
\end{aligned}$$

$$+\frac{2}{n}\sum_{k=2}^{n-3}\left(\mathbf{E}(Y_k)+\frac{2}{k}6^\alpha+\left(1-\frac{2}{k}\right)9^\alpha\right)$$
$$+\left(\mathbf{E}\left(Y_{n-1-k}\right)+\frac{2}{n-1-k}6^\alpha+\left(1-\frac{2}{n-1-k}\right)9^\alpha\right).$$

Subsequently, we obtain the solution

$$\mathbf{E}\left(Y_n^2\right)=(n+1)\sum_{k=7}^{n}q_k+\frac{n+1}{7}\mathbf{E}\left(Y_6^2\right),\quad \text{for } n\geq 6,$$

with

$$q_n:=\frac{nt_n-(n-1)t_{n-1}}{n(n+1)}.$$

The stochastic recursions (1)–(2) also lead to a relation between the second moments of R_n and Y_n:

$$\mathbf{E}\left(R_n^2\right)=\mathbf{E}\left(Y_n^2\right)+s_n,\quad \text{for } n\geq 4,$$

with a rather lengthy function s_n.

Plugging in the boundary values and the already computed formulas for the expectation of Y_n, $Y_n^{(1)}$ and $Y_n^{(2)}$, we obtain the following result (the lengthy explicit expression for the variance is relegated to appendix A).

Proposition 1 *The expectation and variance of the Randić index R_n of a randomly grown binary search tree of size n are asymptotically (as $n \to \infty$) given by*

$$\mathbf{E}(R_n)=\mu_\alpha\, n+O(1),$$
$$\mathbf{Var}(R_n)=\sigma_\alpha^2\, n+O(1),$$

where

$$\mu_\alpha:=\frac{3\cdot 2^\alpha+3\cdot 3^\alpha+2\cdot 4^\alpha+5\cdot 6^\alpha+5\cdot 9^\alpha}{18},$$

and

$$\sigma_\alpha^2:=\frac{1}{283,500}\Big(15525\cdot 4^\alpha+27173\cdot 81^\alpha-35160\cdot 18^\alpha+35948\cdot 16^\alpha$$
$$-54902\cdot 54^\alpha+40725\cdot 9^\alpha-43650\cdot 6^\alpha+3420\cdot 8^\alpha$$
$$+31890\cdot 27^\alpha+13833\cdot 36^\alpha-25350\cdot 12^\alpha-9452\cdot 24^\alpha\Big).$$

In view of the small $O(n)$ variance, a concentration law is a natural corollary.

Corollary 1 *For a nondegenerate Randić index,*

$$\frac{R_n}{n} \xrightarrow{P} \mu_\alpha, \quad \textit{as } n \to \infty.$$

Proof By Chebyshev's inequality, for any fixed $\varepsilon > 0$, we have

$$\mathbf{P}\Big(\Big|R_n - \mathbf{E}[R_n]\Big| > \varepsilon\, \mathbf{E}[R_n]\Big) \le \frac{\mathbf{Var}[R_n]}{\varepsilon^2 \mathbf{E}^2[R_n]}.$$

The orders of magnitude of the mean and variance in Proposition 1 yield

$$\mathbf{P}\left(\left|\frac{R_n}{\mathbf{E}[R_n]} - 1\right| > \varepsilon\right) = O\left(\frac{1}{n}\right).$$

We thus have

$$\frac{R_n}{\mathbf{E}[R_n]} \xrightarrow{P} 1,$$

which can subsequently be adjusted with the aid of the convergence $\mathbf{E}[R_n]/n \to \mu_\alpha$. □

4.3 Asymptotic normality

In principle, one can continue pumping higher moments by the methods utilized for the mean and variance, and appeal to the method of recursive moments (see Chern et al., 2002). However, the explosive complexity is forbidding; we saw that the variance $\mathbf{Var}[R_n]$ is given by a formula that is a page long (in the small font), and the formulas get bigger for every higher moment (compare the exact formulas for the mean and variance)!

The contraction method provides a shortcut. The method was crafted by Rösler (1991) to deal with the distribution of Quicksort. Rachev and Rüschendorf (1995) added several extensions. Recently general contraction theorems and multivariate extensions were added by Neininger (2001), Neininger and Rüschendorf (2004), and Rösler (2001). Rösler and Rüschendorf (2001) provide a lucid survey. What makes the contraction method appealing for an application like the one we have at hand is that now we have in our possession several general theorems that greatly reduce the effort in applying it.

Our starting recurrence (1) can be written in the form

$$Y_n = Y'_{U_n} + Y''_{n-1-U_n} + h(n),$$

where $h(n)$ is a toll function given by

$$h(n) = \big(3^\alpha I_n(U_n) + 2^\alpha(1 - I_n(U_n)\big) \sum_{i=0}^{2} (i+1)^\alpha \big(J'_{U_n}(i) + J''_{n-1-U_n}(i)\big).$$

The function $h(n)$ is clearly $O(1)$; an obvious easy uniform bound is obtained when all the indicators and their complements are replaced by 1, namely

$$0 \le h(n) \le 2(2^\alpha + 3^\alpha)(1 + 2^\alpha + 3^\alpha).$$

So, $h(n)/\sqrt{n} = O(1/\sqrt{n}\,) \to 0$.

The rates of growth of the mean and variance in Proposition 1 comport with the rates required for an application of Corollary 5.2 in Neininger and Rüschendorf (2004), who mention that their method has the underlying theme that "Two moments and a recurrence may be the clues" for normality, a method championed by Pittel (1999). The latter method is based on approximative assimilation by normal distributions. So, the latter method may be used as an alternative.

Taking $f(n) = g(n) = n$ in Corollary 5.2 of Neininger and Rüschendorf (2004), we see that $U_n/g(n)$ converges in L_3 to U, a standard Uniform $(0, 1)$ random variable, and $(n - U_n - 1)/g(n)$ converges in L_3 to $1 - U$; all the other terms resulting from normalizing the toll function vanish in L_3, as is required for the convergence

$$\frac{R_n - \mu_\alpha\, n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\alpha^2\right).$$

Further, R_n and Y_n have he same asymptotic behavior, yielding the main result of this note.

Theorem 1 *In a randomly grown binary search tree of size n a nondegenerate Randić index R_n has a Gaussian limit:*

$$\frac{R_n - \mu_\alpha\, n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\alpha^2\right),$$

where μ_α and whenever $\sigma_\alpha^2 \ne 0$ are specified in Proposition 1.

5 The Randić index of random binary Catalan trees

In the Catalan probability model all binary trees of size n are considered equally likely. It is so called because when research started on binary trees, it was natural to assume a uniform model, and binary trees are counted by Catalan numbers:

The number B_n of binary trees of size n is given by the Catalan number

$$B_n = \frac{1}{n+1}\binom{2n}{n},$$

see Knuth (1998).

The uniform distribution in the Catalan model induces the distribution

$$P(D_n = i) = \begin{cases} 0, & i = 0, \\ \frac{2B_{n-1}}{B_n}, & i = 1, \\ 1 - \frac{2B_{n-1}}{B_n}, & i = 2, \end{cases}$$

for the root degree, for $n \geq 2$, whereas for $n = 1$ one has $P(D_1 = 0) = 1$. We thus have for $n \geq 3$:

$$P(L_n = k \mid D_n = 2) = \frac{B_k B_{n-1-k}}{B_n - 2B_{n-1}}, \quad \text{for } 1 \leq k \leq n-2.$$

We are going to study the moment generating functions $\mathbf{E}\big(e^{Y_n s}\big)$, and the refinements $\mathbf{E}\big(e^{Y_n^{(1)} s}\big)$ and $\mathbf{E}\big(e^{Y_n^{(2)} s}\big)$. Defining the functions

$$\begin{aligned} \phi_n(s) &:= B_n\, \mathbf{E}\big(e^{Y_n s}\big), & \psi_n(s) &:= B_n\, \mathbf{E}\big(e^{R_n s}\big), \\ \phi_n^{(1)}(s) &:= 2B_{n-1}\, \mathbf{E}\big(e^{Y_n^{(1)} s}\big), & \psi_n^{(1)}(s) &:= 2B_{n-1}\, \mathbf{E}\big(e^{R_n^{(1)} s}\big), \\ \phi_n^{(2)}(s) &:= (B_n - 2B_{n-1})\, \mathbf{E}\big(e^{Y_n^{(2)} s}\big), & \psi_n^{(2)}(s) &:= (B_n - 2B_{n-1})\, \mathbf{E}\big(e^{R_n^{(2)} s}\big). \end{aligned}$$

the basic stochastic recurrences (1) can be translated into to the following recurrences for the moment generating functions:

$$\begin{aligned} \phi_n(s) &= \phi_n^{(1)}(s) + \phi_n^{(2)}(s), \\ \phi_n^{(1)}(s) &= 2e^{4^\alpha s}\phi_{n-1}^{(1)}(s) + 2e^{6^\alpha s}\phi_{n-1}^{(2)}(s), \quad n \geq 3, \\ \phi_n^{(2)}(s) &= 2e^{3^\alpha s}\left(e^{6^\alpha s}\phi_{n-2}^{(1)}(s) + e^{9^\alpha s}\phi_{n-2}^{(2)}(s)\right) \\ &\quad + \sum_{k=2}^{n-3}\left(e^{6^\alpha s}\phi_k^{(1)}(s) + e^{9^\alpha s}\phi_k^{(2)}(s)\right)\left(e^{6^\alpha s}\phi_{n-1-k}^{(1)}(s)\right. \\ &\quad \left. + e^{9^\alpha s}\phi_{n-1-k}^{(2)}(s)\right), \quad n \geq 4, \end{aligned}$$

$$
\begin{aligned}
\psi_n(s) &= \psi_n^{(1)}(s) + \psi_n^{(2)}(s),\\
\psi_n^{(1)}(s) &= 2e^{2^\alpha s}\phi_{n-1}^{(1)}(s) + 2e^{3^\alpha s}\phi_{n-1}^{(2)}(s), \quad n \geq 3,\\
\psi_n^{(2)}(s) &= 2e^{2^\alpha s}\left(e^{4^\alpha s}\phi_{n-2}^{(1)}(s) + e^{6^\alpha s}\phi_{n-2}^{(2)}(s)\right)\\
&\qquad + \sum_{k=2}^{n-3}\left(e^{4^\alpha s}\phi_k^{(1)}(s) + e^{6^\alpha s}\phi_k^{(2)}(s)\right)\left(e^{4^\alpha s}\phi_{n-1-k}^{(1)}(s)\right.\\
&\qquad\quad \left.+ e^{6^\alpha s}\phi_{n-1-k}^{(2)}(s)\right), \qquad n \geq 4.
\end{aligned}
$$

To solve this recurrences we introduce the following generating functions:

$$
\begin{aligned}
F(z,s) &:= \sum_{n\geq 2}\phi_n z^n, \quad G(z,s) := \sum_{n\geq 2}\psi_n z^n,\\
F_1(z,s) &:= \sum_{n\geq 2}\phi_n^{(1)} z^n, \quad G_1(z,s) := \sum_{n\geq 2}\psi_n^{(1)} z^n,\\
F_2(z,s) &:= \sum_{n\geq 2}\phi_n^{(2)} z^n, \quad G_2(z,s) := \sum_{n\geq 2}\psi_n^{(2)} z^n.
\end{aligned}
$$

Together with the boundary values this leads to the following system of functional equations:

$$
\begin{aligned}
F(z,s) &= F_1(z,s) + F_2(z,s),\\
F_1(z,s) &= 2z\big(e^{4^\alpha s}F_1(z,s) + e^{6^\alpha s}F_2(z,s)\big) + 2e^{2^\alpha s}z^2,\\
F_2(z,s) &= z\big(e^{6^\alpha s}F_1(z,s) + e^{9^\alpha s}F_2(z,s)\big)^2 + 2z^2e^{3^\alpha s}\big(e^{6^\alpha s}F_1(z,s)\\
&\qquad + e^{9^\alpha s}F_2(z,s)\big) + e^{2\cdot 3^\alpha s}z^3,\\
G(z,s) &= G_1(z,s) + G_2(z,s),\\
G_1(z,s) &= 2z\big(e^{2^\alpha s}F_1(z,s) + e^{3^\alpha s}F_2(z,s)\big) + 2e^s z^2,\\
G_2(z,s) &= z\big(e^{4^\alpha s}F_1(z,s) + e^{6^\alpha s}F_2(z,s)\big)^2 + 2z^2e^{2^\alpha s}\big(e^{4^\alpha s}F_1(z,s)\\
&\qquad + e^{6^\alpha s}F_2(z,s)\big) + e^{2\cdot 2^\alpha s}z^3.
\end{aligned}
$$

This system of functional equations can be reduced to a single quadratic equation for the required generating function $G(z,s)$ by successively eliminating the other functions. Solving this quadratic equation (with the aid of a computer algebra system) leads to a lengthy expression, which has the form:

$$
G(z,s) = H(z,s) - K(z,s)\sqrt{B(z,s)},
$$

where the functions $H(z,s)$ and $K(z,s)$ are given as follows:

$$H(z,s) := \frac{C(z,s)}{N(z,s)}, \quad K(z,s) := \frac{A(z,s)}{N(z,s)},$$
$$N(z,s) := 2z\Bigg(e^{2\cdot 9^\alpha s} + 4z\left(e^{(2\cdot 6^\alpha+9^\alpha)s} - e^{(4^\alpha+2\cdot 9^\alpha)s}\right)$$
$$+4z^2\left(e^{2(4^\alpha+9^\alpha)s} - 2e^{(4^\alpha+2\cdot 6^\alpha+9^\alpha)s} + e^{4\cdot 6^\alpha s}\right)\Bigg)^2,$$
$$C(z,s) := \sum_{k=0}^{7} c_k(s)z^k, \quad A(z,s) := \sum_{k=0}^{4} a_k(s)z^k.$$

The very lengthy functions $a_k(s)$ and $c_k(s)$ are given in Appendix B.

It follows that the functions $H(z,s)$ and $K(z,s)$ are analytic in a circle in the complex z plane, with radius $|z| \leq \frac{1}{4} + \epsilon$ for s in a neighborhood of 0, i. e. for $|s| \leq \delta$, with some $\delta, \epsilon > 0$.

For the asymptotic behavior of the coefficients of $G(z,s)$, and thus of the distribution of R_n, only $B(z,s)$ is of relevance, which is given here:

$$B(z,s) = 1 - 4ze^{4^\alpha s} + 4z^2\left(e^{2\cdot 4^\alpha s} - e^{(3^\alpha+9^\alpha)s}\right)$$
$$+8z^3\left(2e^{(3^\alpha+4^\alpha+9^\alpha)s} - e^{(3^\alpha+2\cdot 6^\alpha)s} - e^{(2^\alpha+6^\alpha+9^\alpha)s}\right)$$
$$+16z^4\left(e^{(2^\alpha+4^\alpha+6^\alpha+9^\alpha)s} + e^{(3^\alpha+4^\alpha+2\cdot 6^\alpha)s} - e^{(3^\alpha+2\cdot 4^\alpha+9^\alpha)s} - e^{(2^\alpha+3\cdot 6^\alpha)s}\right).$$

The dominant singularity $z = \rho(s)$ of $G(z,s)$ can be found in a neighborhood of $s = 0$ by solving the equation

$$B(z,s) = 0$$

for z. When $B(\rho(s),s) = 0$, an algebraic singularity manifests itself. For $s = 0$ we obtain $\rho(0) = \frac{1}{4}$.

Expanding $G(z,s)$ around the dominant singularity $z = \rho(s)$ gives uniformly around $s = 0$:

$$G(z,s) = C(s)\,\left[1 + O\big(z-\rho(s)\big)\right] - A(s)\sqrt{1-\frac{z}{\rho(s)}}\quad\left[1 + O\big(z-\rho(s)\big)\right],$$

with the functions $A(s)$ and $C(s)$ defined by:

$$C(s) := H\big(\rho(s),s\big),$$
$$A(s) := K\big(\rho(s),s\big)\sqrt{-\rho(s)\,\frac{\partial}{\partial z}B(z,s)\Big|_{z=\rho(s)}}.$$

Extracting coefficients gives:

$$[z^n]G(z,s) = A(s)\frac{n^{-\frac{3}{2}}}{2\sqrt{\pi}}\left(\frac{1}{\rho(s)}\right)^n\left(1+O\left(\frac{1}{n}\right)\right).$$

We now have $[z^n]G(z,s) = B_n\,\mathbf{E}\big(e^{R_n s}\big)$; by using the well known asymptotic expansion of the Catalan numbers,

$$B_n = \frac{n^{-\frac{3}{2}}}{\sqrt{\pi}}4^n\left(1+O\left(\frac{1}{n}\right)\right),$$

we obtain the following expansion for the moment generating function of R_n, which holds uniformly around $s = 0$:

$$\begin{aligned}\mathbf{E}\left(e^{R_n s}\right) &= \frac{A(s)}{2}\left(\frac{1}{4\rho(s)}\right)^n\left(1+O\left(\frac{1}{n}\right)\right)\\ &= e^{n(-\log\rho(s)-\log 4)+\log A(s)-\log 2}\left(1+O\left(\frac{1}{n}\right)\right).\end{aligned}$$

Now an immediate application of the quasi-power theorem (see Hwang, 1998) leads to a central limit theorem for R_n (provided that $U''(0)\neq 0$), where the expectation and the variance are asymptotically given by

$$\begin{aligned}\mathbf{E}(R_n) &= U'(0)\,n + O(1),\\ \mathbf{Var}(R_n) &= U''(0)\,n + O(1).\end{aligned}$$

Here the function $U(s)$ is the function

$$U(s) = -\log\rho(s) - \log 4,$$

where $\rho(s)$ is determined by the right branch solution of $B(\rho(s),s) = 0$.

To compute the constants we use

$$\mu_\alpha := U'(0) = -\frac{\rho'(0)}{\rho(0)},\quad \sigma_\alpha^2 := U''(0) = \frac{(\rho'(0))^2-\rho(0)\rho''(0)}{(\rho(0))^2},$$

and differentiate the equation $B(\rho(s),s) = 0$ once then twice with respect to s, then evaluate at $s = 0$. This finally leads to

$$\mu_\alpha = \frac{1}{8}2^\alpha + \frac{1}{8}3^\alpha + \frac{1}{4}4^\alpha + \frac{3}{8}6^\alpha + \frac{1}{8}9^\alpha, \tag{10}$$

$$\begin{aligned}\sigma_\alpha^2 = {}& \frac{3}{64}4^\alpha - \frac{3}{32}6^\alpha - \frac{1}{16}8^\alpha + \frac{7}{64}9^\alpha - \frac{5}{32}12^\alpha + \frac{5}{16}16^\alpha - \frac{3}{16}24^\alpha\\ & -\frac{1}{64}36^\alpha + \frac{3}{32}27^\alpha - \frac{5}{32}54^\alpha + \frac{7}{64}81^\alpha.\end{aligned} \tag{11}$$

Theorem 2 *In a random binary Catalan tree of size n a nondegenerate Randić index R_n has a Gaussian limit:*

$$\frac{R_n - \mu_\alpha n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\alpha^2\right),$$

where μ_α and whenever $\sigma_\alpha^2 \neq 0$ are specified in (10)–(11) .

The structure of the mean and variance of the Randić index in random binary Catalan trees is similar to that in the binary search tree case (only the coefficients of linearity are different). So, one also gets a concentration law here, too (cf. Corollary 1), only differing in the specification of μ_α.

From the explicit formula for $G(z,s)$ one can easily obtain an explicit result for $\mathbf{E}(R_n)$, which is for the sake of completeness given here:

$$\begin{aligned}\mathbf{E}(R_n) = {} & \frac{2^\alpha + 3^\alpha + 2\cdot 4^\alpha + 3\cdot 6^\alpha + 9^\alpha}{8} n \\ & + \frac{5\cdot 2^\alpha + 2\cdot 3^\alpha + 4\cdot 4^\alpha - 9\cdot 6^\alpha - 10\cdot 9^\alpha}{8} \\ & + \frac{15(3\cdot 2^\alpha - 3^\alpha - 2\cdot 4^\alpha - 7\cdot 6^\alpha + 7\cdot 9^\alpha)}{16(2n-1)} \\ & + \frac{105(2^\alpha - 3^\alpha - 2\cdot 4^\alpha + 3\cdot 6^\alpha - 9^\alpha)}{16(2n-1)(2n-3)}, \quad n \geq 3,\end{aligned}$$

with boundary conditions $\mathbf{E}(R_2) = 1$, and $\mathbf{E}(R_1) = 0$.

The asymptotic result for the expectation appears as a special instance of the simply generated tree model already in Clark and Moon (2000). However, their approach does not seem easily extendible to a distributional analysis of R_n. We also remark that the enunciation that $\mathbf{Var}(R_n) \sim c_\alpha n^{\frac{3}{2}}$ in Clark and Moon (2000) turns out to be incorrect.

Appendix A: The explicit expression for Var(R_n)

The exact variance of the Randić index of a random binary search tree is given by the exact formula

$$\begin{aligned}\mathbf{Var}(R_n) = {} & \frac{n}{283500}\Big(15525\cdot 4^\alpha + 27173\cdot 81^\alpha - 35160\cdot 18^\alpha + 35948\cdot 16^\alpha - 54902\cdot 54^\alpha + 40725\cdot 9^\alpha \\ & -43650\cdot 6^\alpha + 3420\cdot 8^\alpha + 31890\cdot 27^\alpha + 13833\cdot 36^\alpha - 25350\cdot 12^\alpha - 9452\cdot 24^\alpha\Big) - \frac{169}{1890}12^\alpha \\ & -\frac{27451}{141750}54^\alpha + \frac{8987}{70875}16^\alpha + \frac{19}{1575}8^\alpha + \frac{23}{420}4^\alpha - \frac{97}{630}6^\alpha + \frac{27173}{283500}81^\alpha + \frac{1537}{31500}36^\alpha - \frac{586}{4725}18^\alpha\end{aligned}$$

$$+\frac{181}{1260}9^\alpha+\frac{1063}{9450}27^\alpha-\frac{2363}{70875}24^\alpha-\frac{4}{135n(n-1)}\Big(6270\cdot 12^\alpha-2924\cdot 54^\alpha+5261\cdot 16^\alpha-3420\cdot 8^\alpha$$
$$+495\cdot 4^\alpha-540\cdot 6^\alpha+86\cdot 81^\alpha+10401\cdot 36^\alpha-3180\cdot 18^\alpha+45\cdot 9^\alpha+330\cdot 27^\alpha-12824\cdot 24^\alpha\Big)$$
$$+\frac{4}{15n^2}\Big(585\cdot 12^\alpha-331\cdot 54^\alpha+379\cdot 16^\alpha-275\cdot 8^\alpha+45\cdot 4^\alpha-60\cdot 6^\alpha+14\cdot 81^\alpha+1044\cdot 36^\alpha$$
$$-285\cdot 18^\alpha+5\cdot 27^\alpha-1121\cdot 24^\alpha\Big)-\frac{16}{45n^2(n-1)^2}\Big(360\cdot 12^\alpha-272\cdot 54^\alpha+158\cdot 16^\alpha-150\cdot 8^\alpha$$
$$+45\cdot 4^\alpha-90\cdot 6^\alpha+53\cdot 81^\alpha+543\cdot 36^\alpha-270\cdot 18^\alpha+45\cdot 9^\alpha+60\cdot 27^\alpha-482\cdot 24^\alpha\Big)$$
$$-\frac{8}{225n(n-1)(n-2)(n-3)}\Big(2775\cdot 12^\alpha-8173\cdot 54^\alpha+2902\cdot 16^\alpha-945\cdot 8^\alpha+1777\cdot 81^\alpha$$
$$+13917\cdot 36^\alpha-2715\cdot 18^\alpha+885\cdot 27^\alpha-10423\cdot 24^\alpha\Big)-\frac{16}{n^4}(-4\cdot 54^\alpha+16^\alpha+81^\alpha-4\cdot 24^\alpha+6\cdot 36^\alpha)$$
$$+\frac{128}{5n^2(n-1)^2(n-2)^2}\Big(-4\cdot 54^\alpha+16^\alpha+81^\alpha-4\cdot 24^\alpha+6\cdot 36^\alpha\Big)$$
$$+\frac{2}{9n}\Big(-62\cdot 12^\alpha-391\cdot 54^\alpha+18\cdot 16^\alpha+20\cdot 8^\alpha+3\cdot 4^\alpha-3\cdot 6^\alpha+125\cdot 81^\alpha+407\cdot 36^\alpha+34\cdot 18^\alpha$$
$$+3\cdot 9^\alpha+2\cdot 27^\alpha-156\cdot 24^\alpha\Big)+\frac{16}{n^2}\Big(4\cdot 54^\alpha-16^\alpha-81^\alpha-6\cdot 36^\alpha+4\cdot 24^\alpha\Big)H_n^2$$
$$-256\cdot\frac{-4\cdot 54^\alpha+16^\alpha+81^\alpha-4\cdot 24^\alpha+6\cdot 36^\alpha}{5n^2(n-1)^2(n-2)^2(n-3)^2}-\frac{8}{15n(n-1)(n-2)}\Big(180\cdot 12^\alpha-427\cdot 54^\alpha$$
$$+193\cdot 16^\alpha-65\cdot 8^\alpha+83\cdot 81^\alpha+798\cdot 36^\alpha-165\cdot 18^\alpha+50\cdot 27^\alpha-647\cdot 24^\alpha\Big)$$
$$+\frac{16}{3n^3}\Big(12\cdot 12^\alpha-37\cdot 54^\alpha+4\cdot 16^\alpha-3\cdot 8^\alpha+10\cdot 81^\alpha+48\cdot 36^\alpha-15\cdot 18^\alpha+6\cdot 27^\alpha-25\cdot 24^\alpha\Big)$$
$$+\Big[\frac{4}{9n}\Big(3\cdot 12^\alpha+41\cdot 54^\alpha+7\cdot 16^\alpha-3\cdot 8^\alpha-14\cdot 81^\alpha-33\cdot 36^\alpha+3\cdot 18^\alpha-3\cdot 27^\alpha-24^\alpha\Big)$$
$$-\frac{16}{3n(n-1)}\Big(-18\cdot 12^\alpha+34\cdot 54^\alpha-13\cdot 16^\alpha+6\cdot 8^\alpha-7\cdot 81^\alpha-60\cdot 36^\alpha+18\cdot 18^\alpha-6\cdot 27^\alpha+46\cdot 24^\alpha\Big)$$
$$-64\cdot\frac{4\cdot 54^\alpha-16^\alpha-81^\alpha-6\cdot 36^\alpha+4\cdot 24^\alpha}{3n(n-1)(n-2)}+128\cdot\frac{4\cdot 54^\alpha-16^\alpha-81^\alpha-6\cdot 36^\alpha+4\cdot 24^\alpha}{15n(n-1)(n-2)(n-3)}$$
$$+\frac{16}{3n^2}\Big(-12\cdot 12^\alpha+25\cdot 24^\alpha+3\cdot 8^\alpha+15\cdot 18^\alpha-48\cdot 36^\alpha-6\cdot 27^\alpha-10\cdot 81^\alpha-4\cdot 16^\alpha+37\cdot 54^\alpha\Big)$$
$$-32\cdot\frac{4\cdot 54^\alpha-16^\alpha-81^\alpha-6\cdot 36^\alpha+4\cdot 24^\alpha}{n^3}\Big]H_n-\frac{32}{3n(n-1)(n-2)(n-3)(n-4)}\Big(2\cdot 16^\alpha-5\cdot 54^\alpha$$
$$-8^\alpha+27^\alpha-7\cdot 24^\alpha+3\cdot 12^\alpha+9\cdot 36^\alpha+81^\alpha-3\cdot 18^\alpha\Big).$$

Appendix B: The functions $a_k(s)$ and $c_k(s)$

The functions $a_k(s)$, $k=0,\ldots,4$, and $c_k(s)$, $k=0,\ldots,7$, that appear in the intermediate steps of the calculations for the Catalan trees are listed here:

$$a_0(s) = e^{2\cdot 6^\alpha s},$$
$$a_1(s) = 2\,e^{s(3^\alpha+2\cdot 9^\alpha)}-2\,e^{s(2\cdot 6^\alpha+4^\alpha)},$$
$$a_2(s) = -12\,e^{s(3^\alpha+2\cdot 9^\alpha+4^\alpha)}+6\,e^{s(3^\alpha+2\cdot 6^\alpha+9^\alpha)}+6\,e^{s(2^\alpha+6^\alpha+2\cdot 9^\alpha)},$$
$$a_3(s) = 20\,e^{s(3\cdot 6^\alpha+2^\alpha+9^\alpha)}-20\,e^{s(2^\alpha+6^\alpha+4^\alpha+2\cdot 9^\alpha)}-28\,e^{s(9^\alpha+3^\alpha+4^\alpha+2\cdot 6^\alpha)}$$
$$+4\,e^{s(3^\alpha+4\cdot 6^\alpha)}+24\cdot e^{s(3^\alpha+2\cdot 9^\alpha+2\cdot 4^\alpha)},$$

$$
\begin{aligned}
a_4(s) &= 32\, e^{s(2\cdot 4^\alpha+3^\alpha+2\cdot 6^\alpha+9^\alpha)} + 16\, e^{s(2^\alpha+5\cdot 6^\alpha)} - 32\, e^{s(4^\alpha+3\cdot 6^\alpha+2^\alpha+9^\alpha)}\\
&\quad + 16\, e^{s(2^\alpha+6^\alpha+2\cdot 4^\alpha+2\cdot 9^\alpha)} - 16\, e^{s(3^\alpha+3\cdot 4^\alpha+2\cdot 9^\alpha)} - 16\, e^{s(3^\alpha+4^\alpha+4\cdot 6^\alpha)},\\
c_0(s) &= e^{2\cdot 6^\alpha s},\\
c_1(s) &= 2\, e^{s(3^\alpha+2\cdot 9^\alpha)} - 4\, e^{s(2\cdot 6^\alpha+4^\alpha)},\\
c_2(s) &= 4\, e^{2s(6^\alpha+4^\alpha)} + 6\, e^{s(2^\alpha+6^\alpha+2\cdot 9^\alpha)} - 16\, e^{s(3^\alpha+2\cdot 9^\alpha+4^\alpha)} + 4\, e^{s(3^\alpha+2\cdot 6^\alpha+9^\alpha)},\\
c_3(s) &= -32\, e^{s(9^\alpha+3^\alpha+4^\alpha+2\cdot 6^\alpha)} - 32\, e^{s(2^\alpha+6^\alpha+4^\alpha+2\cdot 9^\alpha)} + 4\, e^{s(1+4\cdot 9^\alpha)}\\
&\quad - 4\, e^{s(2\cdot 3^\alpha+3\cdot 9^\alpha)} + 16\, e^{s(3\cdot 6^\alpha+2^\alpha+9^\alpha)} + 48\, e^{s(3^\alpha+2\cdot 9^\alpha+2\cdot 4^\alpha)},\\
c_4(s) &= 10\, e^{2s(2^\alpha+2\cdot 9^\alpha)} + 80\, e^{s(2\cdot 4^\alpha+3^\alpha+2\cdot 6^\alpha+9^\alpha)} - 20\, e^{s(6^\alpha+3^\alpha+2^\alpha+3\cdot 9^\alpha)}\\
&\quad +56\, e^{s(2^\alpha+6^\alpha+2\cdot 4^\alpha+2\cdot 9^\alpha)} - 64\, e^{s(4^\alpha+3\cdot 6^\alpha+2^\alpha+9^\alpha)} - 64\, e^{s(3^\alpha+3\cdot 4^\alpha+2\cdot 9^\alpha)}\\
&\quad - 32\, e^{s(1+4\cdot 9^\alpha+4^\alpha)} + 32\, e^{s(2\cdot 3^\alpha+3\cdot 9^\alpha+4^\alpha)} + 32\, e^{s(1+3\cdot 9^\alpha+2\cdot 6^\alpha)}\\
&\quad - 22\, e^{2s(3^\alpha+6^\alpha+9^\alpha)} - 16\, e^{s(3^\alpha+4^\alpha+4\cdot 6^\alpha)} + 8\, e^{s(2^\alpha+5\cdot 6^\alpha)},\\
c_5(s) &= -32\, e^{s(2^\alpha+6^\alpha+3\cdot 4^\alpha+2\cdot 9^\alpha)} + 136\, e^{s(2\cdot 3^\alpha+2\cdot 6^\alpha+2\cdot 9^\alpha+4^\alpha)} + 64\, e^{s(2\cdot 4^\alpha+3\cdot 6^\alpha+2^\alpha+9^\alpha)}\\
&\quad -64\, e^{s(3\cdot 4^\alpha+9^\alpha+3^\alpha+2\cdot 6^\alpha)} - 112\, e^{s(2^\alpha+3^\alpha+3\cdot 6^\alpha+2\cdot 9^\alpha)} + 112\, e^{s(2^\alpha+6^\alpha+3\cdot 9^\alpha+3^\alpha+4^\alpha)}\\
&\quad -192\, e^{s(1+3\cdot 9^\alpha+4^\alpha+2\cdot 6^\alpha)} - 96\, e^{s(2\cdot 3^\alpha+3\cdot 9^\alpha+2\cdot 4^\alpha)} + 96\, e^{s(1+4\cdot 9^\alpha+2\cdot 4^\alpha)}\\
&\quad -56\, e^{s\left(2^{1+\alpha}+4\cdot 9^\alpha+4^\alpha\right)} + 96\, e^{s(1+2\cdot 9^\alpha+4\cdot 6^\alpha)} - 32\, e^{s(2^\alpha+5\cdot 6^\alpha+4^\alpha)} + 32\, e^{s\left(3^\alpha+4^{1+\alpha}+2\cdot 9^\alpha\right)}\\
&\quad +32\, e^{s(3^\alpha+2\cdot 4^\alpha+4\cdot 6^\alpha)} + 56\, e^{s\left(2^{1+\alpha}+3\cdot 9^\alpha+2\cdot 6^\alpha\right)} - 40\, e^{s(2\cdot 3^\alpha+4\cdot 6^\alpha+9^\alpha)},\\
c_6(s) &= -208\, e^{s(2^\alpha+3^\alpha+5\cdot 6^\alpha+9^\alpha)} - 208\, e^{s\left(2^{1+\alpha}+3\cdot 9^\alpha+4^\alpha+2\cdot 6^\alpha\right)} + 176\, e^{s(2\cdot 3^\alpha+4\cdot 6^\alpha+9^\alpha+4^\alpha)}\\
&\quad -208\, e^{s(2^\alpha+6^\alpha+3\cdot 9^\alpha+3^\alpha+2\cdot 4^\alpha)} + 416\, e^{s(2^\alpha+3^\alpha+3\cdot 6^\alpha+2\cdot 9^\alpha+4^\alpha)} - 384\, e^{s(1+2\cdot 9^\alpha+4^\alpha+4\cdot 6^\alpha)}\\
&\quad + 128\, e^{s\left(1+6^{1+\alpha}+9^\alpha\right)} + 104\, e^{2s(9^\alpha+2^\alpha+2\cdot 6^\alpha)} + 128\, e^{s(3\cdot 9^\alpha+2\cdot 3^\alpha+3\cdot 4^\alpha)}\\
&\quad -128\, e^{s(1+4\cdot 9^\alpha+3\cdot 4^\alpha)} + 104\, e^{2s(2^\alpha+2\cdot 9^\alpha+4^\alpha)} - 24\, e^{2s(3^\alpha+3\cdot 6^\alpha)}\\
&\quad -280\, e^{2s(3^\alpha+6^\alpha+9^\alpha+4^\alpha)} + 384\, e^{s(1+3\cdot 9^\alpha+2\cdot 4^\alpha+2\cdot 6^\alpha)},\\
c_7(s) &= -64\, e^{s\left(2^{1+\alpha}+4\cdot 9^\alpha+3\cdot 4^\alpha\right)} - 64\, e^{s\left(3\cdot 9^\alpha+2\cdot 3^\alpha+4^{1+\alpha}\right)} - 128\, e^{s(2^\alpha+3^\alpha+7\cdot 6^\alpha)} + 64\, e^{s\left(6^{1+\alpha}+2\cdot 3^\alpha+4^\alpha\right)}\\
&\quad +64\, e^{s\left(2^{1+\alpha}+6^{1+\alpha}+9^\alpha\right)} + 64\, e^{s\left(1+4\cdot 9^\alpha+2^{2+2\alpha}\right)} + 64\, e^{s(1+8\cdot 6^\alpha)} + 192\, e^{s\left(2^{1+\alpha}+3\cdot 9^\alpha+2\cdot 6^\alpha+2\cdot 4^\alpha\right)}\\
&\quad -256\, e^{s(1+3\cdot 9^\alpha+3\cdot 4^\alpha+2\cdot 6^\alpha)} - 192\, e^{s(9^\alpha+2\cdot 3^\alpha+2\cdot 4^\alpha+4\cdot 6^\alpha)} - 192\, e^{s\left(2\cdot 9^\alpha+2^{1+\alpha}+4\cdot 6^\alpha+4^\alpha\right)}\\
&\quad +128\, e^{s(2^\alpha+6^\alpha+3\cdot 9^\alpha+3^\alpha+3\cdot 4^\alpha)} + 384\, e^{s(1+2\cdot 9^\alpha+2\cdot 4^\alpha+4\cdot 6^\alpha)} + 192\, e^{s(2\cdot 9^\alpha+2\cdot 3^\alpha+3\cdot 4^\alpha+2\cdot 6^\alpha)}\\
&\quad -256\, e^{s\left(1+9^\alpha+4^\alpha+6^{1+\alpha}\right)} - 384\, e^{s(2^\alpha+3^\alpha+3\cdot 6^\alpha+2\cdot 9^\alpha+2\cdot 4^\alpha)} + 384\, e^{s(2^\alpha+3^\alpha+5\cdot 6^\alpha+9^\alpha+4^\alpha)}.
\end{aligned}
$$

Acknowledgment The work of the first author is supported by the National Natural Science Foundation of China (Grant No. 10671188) and the Special Foundation of USTC. The work of the third author is supported by the Austrian Science Foundation FWF (Grant No. S9608-N13).

References

Balaban, A. (1982). Highly discriminating distance-based topological index. *Chemical Physics Letters*, *89*, 399–404.

Balaban, A. (1983). Topological indices based on topological distances in molecular graphs. *Pure and Applied Chemistry*, *55*, 199–206.

Basak, S. (1987). Use of molecular complexity indices in predictive pharmacology and toxicology: A QSAR approach. *Medical Science Research*, *15*, 605–609.

Bonchev, D., Trinajstić, N. (1983). Information theoretic indices for characterization of chemical structure. Research Studies Press, Chichester: Wiley.

Chern, H., Hwang, H., Tsai, T. (2002). An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms. *Journal of Algorithms*, *44*, 177–225.

Christophi, C., Mahmoud, H. (2005). The oscillatory distribution of distances in random tries. *The Annals of Applied Probability*, *15*, 1536–1564.

Clark, L., Moon, J. (2000). On the general Randić index for certain families of trees. *Ars Combinatoria*, *54*, 223–235.

Devillers, J., Balaban, A. (1999). Topological indices and related descriptors in QSAR and QSPR. Amsterdam: Gordon & Breach.

Devroye, L., Neininger, R. (2004). Distances and finger search in random binary search trees. *SIAM Journal on Computing*, *33*, 647–658.

Estrada, E. (2002). The structural interpretation of the Randić index. *Electronic Journal of Molecular Design*, *1*, 360–366.

Gutman, I., Polansky, O. (1986). *Mathematical Concepts in Organic Chemistry.* Berlin: Springer.

Gutman, I., Trinajstić, N. (1972). *Chemical Physics Letters*, *17*, 535–538.

Harary, F. (1969). *Graph theory*. Reading: Addison-Wesley.

Hwang, H. (1998). On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, *19*, 329–343.

Hwang, H., Neininger, R. (2002). Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM Journal on Computing*, *31*, 1687–1722.

Janson, S. (2003). The Wiener index of simply generated random trees. *Random Structures and Algorithms*, *22*, 337–358.

Janson, S., Chassaing, P. (2004). The center of mass of the ISE and the Wiener index of trees. *Electronic Communications of Probability*, *9*, 178–187.

Kemp, R. (1984). *Fundamentals of the average case analysis of particular algorithms*. New York: Wiley.

Kier, L., Hall, L. (1976). Molecular connectivity in chemistry and drug research. New York: Academic.

Kirschenhofer, P., Prodinger, H. (1998). Comparisons in Hoare's Find algorithm. *Combinatorics, Probability, and Computing*, *7*, 111–120.

Kirschenhofer, P., Prodinger, H., Martínez, C. (1997). Analysis of Hoare's FIND algorithm with median-of-three partition. *Random Structures and Algorithms*, *10*, 143–156.

Knuth, D. (1998). *The art of computer programming*, Vol. 3: sorting and searching, 2nd Edn. Reading: Addison-Wesley.

Mahmoud, H. (1992). *Evolution of random search trees*. New York: Wiley.

Mahmoud, H. (2000). *Sorting: A distribution theory*. New York: Wiley.

Mahmoud, M., Neininger, R. (2002). Distribution of distances in random binary search trees. *The Annals of Applied Probability*, *13*, 253–276.

Miličević, A., Nikolić, S. (2004). On variable Zagreb indices. *Croatia Chemica Acta*, *77*, 97–101.

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

Neininger, R. (2002). The Wiener index of random trees. *Combinatorics, Probability and Computing*, *11*, 587–597.

Neininger, R., Rüschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability*, *14*, 378–418.

Panholzer, A. (2003). Analysis of multiple quickselect variants. *Theoretical Computer Science*, *302*, 45–91.

Panholzer, A., Prodinger, H. (1998). A generating functions approach for the analysis of grand averages for Multiple Quickselect. *Random Structures and Algorithms 13*, 189–209.

Panholzer, A., Prodinger, H. (2004). Spanning tree size in random binary search trees. *The Annals of Applied Probability*, *14*, 718–733.

Pittel, B. (1999). Normal convergence problem? Two moments and a recurrence may be the clues. *The Annals of Applied Probability*, *9*, 1260–1302.

Prodinger, H. (1995). Multiple quickselect—Hoare's Find algorithm for several elements. *Information Processing Letters*, *56*, 123–129.

Quintas, L., Szymański J. (1992). Nonuniform random recursive trees with bounded degree. In *Sets, Graphs, and Numbers: Colloquia Mathematica Societas Janos Bolyai*, *60*, 611–620.
Rachev, S., Rüschendorf, L. (1995). Probability metrics and recursive algorithms. *Advances in Applied Probability*, *27*, 770–799.
Randić, M. (1975). On characterization of molecular branching. *Journal of the American Chemical Society*, *97*, 6609–6615.
Rösler, U. (1991). A limit theorem for "Quicksort". *RAIRO Informatiques Théoriques et Applications*, *25*, 85–100.
Rösler, U. (2001). On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, *29*, 238–261.
Rösler, U., Rüschendorf, L. (2001). The contraction method for recursive algorithms. *Algorithmica*, *29*, 3–33.
Trinajstić, N. (1992). *Chemical graph theory*. CRC Press, Boca Raton, Florida.
Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, *69*, 17–20.