Inversions in split trees and conditional Galton–Watson trees

Xing Shi Cai, Cecilia Holmgren, Svante Janson, Tony Johansson*
Department of Mathematics, Uppsala University, Sweden
{xingshi.cai, cecilia.holmgren, svante.janson, tony.johansson}@math.uu.se

Fiona Skerman
Heilbronn Institute, Bristol University, UK
f.skerman@bristol.ac.uk

September 1, 2017

Abstract

We study $I(T)$, the number of inversions in a tree T with its vertices labeled uniformly at random, which is a generalization of inversions in permutations. We first show that the cumulants of $I(T)$ have explicit formulas involving the k-total common ancestors of T (an extension of the total path length). Then we consider X_n, the normalized version of $I(T_n)$, for a sequence of trees T_n. For fixed T_n's, we prove a sufficient condition for X_n to converge in distribution. As an application, we identify the limit of X_n for complete b-ary trees. For T_n being split trees [15], we show that X_n converges to the unique solution of a distributional equation. Finally, when T_n's are conditional Galton–Watson trees, we show that X_n converges to a random variable defined in terms of Brownian excursions. By exploiting the connection between inversions and the total path length, we are able to give results that are stronger and much broader compared to previous work by Panholzer and Seitz [45].

1 Introduction

1.1 Inversions in a fixed tree

Let $\sigma_1, \dots, \sigma_n$ be a permutation of $\{1, \dots, n\}$. If $i < j$ and $\sigma_i > \sigma_j$, then the pair (σ_i, σ_j) is called an inversion. The concept of inversions was introduced by Cramer [13] (1750) due to its connection with solving linear equations. More recently, the study of inversions has been motivated by its applications in the analysis of sorting algorithms, see, e.g., [36, Section 5.1]. Many authors, including Feller [20, pp. 256], Sachkov [51, pp. 29], Bender [6], have shown that the number of inversions in uniform random permutations has a central limit theorem. More recently, Margolius [41] and Louchard and Prodinger [38] studied permutations containing a fixed number of inversions.

*This work was partially supported by two grants from the Knut and Alice Wallenberg Foundation and a grant from the Swedish Research Council.

The concept of inversions can be generalized as follows. Consider an unlabeled rooted tree T on node set V. Let ρ denote the root. Write $u < v$ if u is a *proper ancestor* of v, i.e., the unique path from ρ to v passes through u and $u \neq v$. Write $u \leq v$ if u is an ancestor of v, i.e., either $u < v$ or $u = v$. Given a bijection $\lambda : V \to \{1, \dots, |V|\}$ (a *node labeling*), define the number of *inversions*

$$I(T, \lambda) \stackrel{\text{def}}{=} \sum_{u<v} \mathbf{1}_{\lambda(u)>\lambda(v)}.$$

Note that if T is a path, then $I(T, \lambda)$ is nothing but the number of inversions in a permutation. Our main object of study is the random variable $I(T)$, defined by $I(T) = I(T, \lambda)$ where λ is chosen uniformly at random from the set of bijections from V to $\{1, \dots, |V|\}$.

The enumeration of trees with a fixed number of inversions has been studied by Mallows and Riordan [40] and Gessel et al. [24] using the so called *inversions polynomial*. While analyzing linear probing hashing, Flajolet et al. [22] noticed that the numbers of inversions in Cayley trees with uniform random labeling converges to an Airy distribution. Panholzer and Seitz [45] showed that this is true for conditional Galton–Watson trees, which encompasses the case of Cayley trees.

For a node v, let z_v denote the size of the subtree rooted at v. The following representation of $I(T)$, proved in Section 2, is the basis of most of our results:

Lemma 1. *Let T be a fixed tree. Then*

$$I(T) \stackrel{\text{d}}{=} \sum_{v \in V} Z_v, \tag{1.1}$$

where $\{Z_v\}_{v \in V}$ are independent random variables, and $Z_v \sim \mathrm{Unif}\{0, 1, \dots, z_v - 1\}$.

We will generally be concerned with the centralized number of inversions, i.e., $I(T) - \mathbb{E}[I(T)]$. For any $u < v$ we have $\mathbb{P}\{\lambda(u) > \lambda(v)\} = 1/2$. Let $h(v)$ denote the *depth* of v, i.e., the distance from v to the root ρ. It immediately follows that,

$$\mathbb{E}[I(T)] = \sum_{u<v} \mathbb{E}\left[\mathbf{1}_{\lambda(u)>\lambda(v)}\right] = \frac{1}{2}\Upsilon(T), \tag{1.2}$$

where $\Upsilon(T) \stackrel{\text{def}}{=} \sum_v h(v)$ is called the *total path length* (or *internal path length*) of T.

Let $\varkappa_k = \varkappa_k(X)$ denote the k-th cumulant of a random variable X (provided it exists); thus $\varkappa_1(X) = \mathbb{E}[X]$ and $\varkappa_2(X) = \mathrm{Var}(X)$ (see [26, Theorem 4.6.4]). We now define $\Upsilon_k(T)$, *the k-total common ancestors* of T, which allows us to generalize (1.2) to higher cumulants of $I(T)$. For k nodes $v_1, \dots, v_k$ (not necessarily distinct), let $c(v_1, \dots, v_k)$ be the number of ancestors that they share, i.e.,

$$c(v_1, \dots, v_k) \stackrel{\text{def}}{=} |\{u \in V : u \leq v_1, u \leq v_2, \dots, u \leq v_k\}|.$$

We define

$$\Upsilon_k(T) \stackrel{\text{def}}{=} \sum_{v_1, \dots, v_k} c(v_1, \dots, v_k), \tag{1.3}$$

where the sum is over all ordered k-tuples of nodes in the tree. For a single node v, $h(v) = c(v) - 1$, since v itself is counted in $c(v)$. So $\Upsilon(T) = \Upsilon_1(T) - |V|$; i.e., we recover the usual notion of total path length.

Theorem 1. *Let T be a fixed tree. Let $\varkappa_k(I(T))$ be the k-th cumulant of $I(T)$. Then*

$$\mathbb{E}\left[I(T)\right] = \varkappa_1(I(T)) = \frac{1}{2}\Upsilon(T) = \frac{1}{2}(\Upsilon_1(T) - |V|), \tag{1.4}$$
$$\operatorname{Var}\left(I(T)\right) = \varkappa_2(I(T)) = \frac{1}{12}(\Upsilon_2(T) - |V|),$$

and, more generally, for $k \geq 1$,

$$\varkappa_{2k+1}(I(T)) = 0, \qquad \varkappa_{2k}(I(T)) = \frac{B_{2k}}{2k}(\Upsilon_{2k}(T) - |V|), \tag{1.5}$$

where B_k denotes the k-th Bernoulli number. Moreover, $I(T)$ has the moment generating function

$$\mathbb{E}\left[e^{tI(T)}\right] = \prod_{v \in V} \frac{e^{z_v t} - 1}{z_v(e^t - 1)}, \tag{1.6}$$

and for the centralized variable we have the estimate

$$\mathbb{E}\left[e^{t(I(T)-\mathbb{E}[I(T)])}\right] \leq \exp\Bigl(\tfrac{1}{8}t^2 \sum_{v \in T}(z_v - 1)^2\Bigr) \leq \exp\Bigl(\tfrac{1}{8}t^2 \sum_{v \in T} z_v^2\Bigr) = \exp\Bigl(\tfrac{1}{8}t^2 \Upsilon_2(T)\Bigr), \qquad t \in \mathbb{R}. \tag{1.7}$$

Remark 1. *Recalling that $B_1 = -1/2$ and $B_{2k+1} = 0$ for $k \geq 1$, (1.4)–(1.5) can also be written as*

$$\varkappa_k(I(T)) = \frac{B_k}{k}(-1)^k(\Upsilon_k(T) - |V|), \qquad k \geq 1.$$

Remark 2. *Higher moments and central moments can be calculated from the cumulants by standard formulas [52]. (Note that all odd central moments vanish by symmetry.) For example, recalling $B_4 = -1/30$, Theorem 1 implies that*

$$\mathbb{E}\left[(I(T) - \mathbb{E}\left[I(T)\right])^4\right] = 3\varkappa_2(I(T))^2 + \varkappa_4(I(T)) = \frac{1}{48}\Bigl(\Upsilon_2(T) - |V|\Bigr)^2 - \frac{1}{120}(\Upsilon_4(T) - |V|).$$

1.2 Inversions in sequences of trees

The total path length $\Upsilon(T)$ has been studied for random trees like split trees [8] and conditional Galton–Watson trees [3, Corollary 9]. This leads us to focus on the deviation

$$X_n = \frac{I(T_n) - \mathbb{E}\left[I(T_n)\right]}{s(n)},$$

under some appropriate scaling $s(n)$, for a sequence of (random or fixed) trees T_n, where T_n has size n.

Fixed trees

Theorem 2. *Let T_n be a sequence of fixed trees on n nodes. Let*

$$X_n = \frac{I(T_n) - \mathbb{E}\left[I(T_n)\right]}{\sqrt{\Upsilon_2(T_n)}}.$$

Assume that for all $k \geq 1$,

$$\frac{\Upsilon_{2k}(T_n)}{\Upsilon_2(T_n)^k} \to \zeta_{2k},$$

for some sequence (ζ_{2k}). Then there exists a unique distribution X with

$$\varkappa_{2k-1}(X) = 0, \qquad \varkappa_{2k}(X) = \frac{B_{2k}}{2k}\zeta_{2k}, \qquad k \geq 1,$$

such that $X_n \xrightarrow{\mathrm{d}} X$ and, moreover, $\mathbb{E}\left[e^{tX_n}\right] \to \mathbb{E}\left[e^{tX}\right] < \infty$ for every $t \in \mathbb{R}$.

Remark 3. *By Theorem 1, $\operatorname{Var}(X_n) = (\Upsilon_2(T_n) - n)/12s(n)^2$. Thus, it is natural to consider $s(n) = \Theta\big(\sqrt{\Upsilon_2(T_n) - n}\big) = \Theta\big(\sqrt{\Upsilon_2(T_n)}\big)$, where we use $\Upsilon_2(T_n) \stackrel{\text{def}}{=} \sum_{v_1,v_2} c(v_1, v_2) \geq n^2$.*

Remark 4. *The functions $\psi_{X_n}(t) \stackrel{\text{def}}{=} \mathbb{E}\left[e^{tX_n}\right]$ and $\psi_X(t) \stackrel{\text{def}}{=} \mathbb{E}\left[e^{tX}\right]$ are called moment generating functions of X_n and X respectively. The convergence $\psi_{X_n}(t) \to \psi_X(t) < \infty$ in a neighborhood of 0 implies that $X_n \xrightarrow{\mathrm{d}} X$ and $(|X_n|^r)_{n\geq 1}$ is uniformly integrable for all $r > 0$; thus $\mathbb{E}\left[|X_n^r|\right] \to \mathbb{E}\left[|X|^r\right]$ for all $r > 0$ and $\mathbb{E}\left[X_n^r\right] \to \mathbb{E}\left[X^r\right]$ for all integers $r \geq 1$. See, e.g., [26, Theorem 5.9.5].*

As simple examples, we consider two extreme cases.

Example 1. When P_n is a path of n nodes, we have for fixed $k \geq 1$

$$\Upsilon_k(P_n) \sim \frac{1}{k+1} n^{k+1}.$$

Thus $\Upsilon_{2k}(P_n)/\Upsilon_2(P_n)^k \to \varkappa_{2k} = 0$ for $k \geq 2$. So by Theorem 2, X_n converges to a normal distribution, and we recover the central limit law for inversions in permutations. Also, the vertices have subtree sizes $1, \dots, n$ and so we also recover from Theorem 1 the moment generating function $\prod_{j=1}^{n}(e^{jt} - 1)/(j(e^t - 1))$ [51, 41].

Example 2. Let $T_n = S_{n-1}$, a star with $n-1$ leaves, and denote the root by o. We have $z_o = n$ and $z_v = 1$ for $v \neq o$. Hence, by Lemma 1, or directly, $I(S_{n-1}) \sim \operatorname{Unif}\{0, \dots, n-1\}$, and consequently

$$\big(I(T_n) - \mathbb{E}\left[I(T_n)\right]\big)/n \xrightarrow{\mathrm{d}} \operatorname{Unif}[-\tfrac{1}{2}, \tfrac{1}{2}].$$

This follows also by Theorem 2, since $\Upsilon_k(S_{n-1}) \sim n^k$ for $k \geq 2$ (e.g., by Lemma 3 below).

It is straightforward to compute the k-total common ancestors for b-ary trees. Thus our next result follows immediately from Theorem 2.

Theorem 3. *Let $b \geq 2$ and let T_n be the complete b-ary tree of height m with $n = (b^{m+1}-1)/(b-1)$ nodes. Let*

$$X_n = \frac{I(T_n) - \mathbb{E}\left[I(T_n)\right]}{n}, \qquad \text{and} \qquad X = \sum_{d\geq 0}\sum_{j=1}^{b^d} \frac{U_{d,j}}{b^d},$$

where $(U_{d,j})_{d\geq 0, j\geq 1}$ are independent $\operatorname{Unif}[-1/2, 1/2]$. Then $X_n \xrightarrow{\mathrm{d}} X$ and $\mathbb{E}\left[e^{tX_n}\right] \to \mathbb{E}\left[e^{tX}\right] < \infty$, for every $t \in \mathbb{R}$. Moreover X is the unique random variable with

$$\varkappa_{2k-1}(X) = 0, \qquad \varkappa_{2k}(X) = \frac{B_{2k}}{2k}\frac{b^{2k-1}}{b^{2k-1}-1}, \qquad k \geq 1. \tag{1.8}$$

Random trees

We move on to random trees. We consider generating a random tree T_n and, conditioning on T_n, labeling its nodes uniformly at random. The relation (1.2) is maintained for random trees:

$$\mathbb{E}\left[I(T_n)\right] = \mathbb{E}\left[\mathbb{E}\left[I(T_n) \mid T_n\right]\right] = \frac{1}{2}\mathbb{E}\left[\Upsilon(T_n)\right].$$

The deviation of $I(T_n)$ from its mean can be taken to mean two different things. Consider for some scaling function $s(n)$,

$$X_n = \frac{I(T_n) - \mathbb{E}\left[I(T_n)\right]}{s(n)}, \qquad Y_n = \frac{I(T_n) - \mathbb{E}\left[I(T_n) \mid T_n\right]}{s(n)} = \frac{I(T_n) - \frac{1}{2}\Upsilon(T_n)}{s(n)}.$$

Then X_n and Y_n each measure the deviation of $I(T_n)$, unconditionally and conditionally. They are related by the identity

$$X_n = Y_n + W_n/2, \tag{1.9}$$

where

$$W_n = \frac{\Upsilon(T_n) - \mathbb{E}\left[\Upsilon(T_n)\right]}{s(n)}.$$

In the case of fixed trees $W_n = 0$ and $X_n = Y_n$, but for random trees we consider the sequences separately.

We consider two classes of random trees — split trees and conditional Galton–Watson trees.

Split trees

The first class of random trees which we study are split trees. They were introduced by Devroye [15] to encompass many families of trees that are frequently used in algorithm analysis, e.g., binary search trees [27], m-ary search trees [46], quad trees [21], median-of-$(2k+1)$ trees [53], fringe-balanced trees [14], digital search trees [11] and random simplex trees [15, Example 5].

A split tree can be constructed as follows. Consider a rooted infinite b-ary tree where each node is a bucket of finite capacity s. We place n balls at the root, and the balls individually trickle down the tree in a random fashion until no bucket is above capacity. Each node draws a *split vector* $\mathcal{V} = (V_1, \dots, V_b)$ from a common distribution, where V_i describes the probability that a ball passing through the node continues to the ith child. The trickle-down procedure is defined precisely in Section 4. Any node u such that the subtree rooted as u contains no balls is then removed, and we consider the resulting tree T_n.

In the context of split trees we differentiate between $I(T_n)$ (the number of inversions on *nodes*), and $\hat{I}(T_n)$ (the number of inversions on *balls*). In the former case, the nodes (buckets) are given labels, while in the latter the individual balls are given labels. For balls β_1, β_2, write $\beta_1 < \beta_2$ if the node containing β_1 is a proper ancestor of the node containing β_2; if β_1, β_2 are contained in the same node we do not compare their labels. Define

$$\hat{I}(T_n) = \sum_{\beta_1 < \beta_2} \mathbf{1}_{\lambda(\beta_1) > \lambda(\beta_2)}.$$

Similarly define $\hat{\Upsilon}(T_n)$ as the total path length on balls, i.e., the sum of the depth of all balls. And let

$$\hat{X}_n = \frac{\hat{I}(T_n) - \mathbb{E}\left[\hat{I}(T_n)\right]}{n}, \quad \hat{Y}_n = \frac{\hat{I}(T_n) - s_0\hat{\Upsilon}(T_n)/2}{n}, \quad \hat{W}_n = \frac{\hat{\Upsilon}(T_n) - \mathbb{E}\left[\hat{\Upsilon}(T_n)\right]}{n}. \tag{1.10}$$

Here s_0 is a fixed integer denoting the number of balls in any internal node, and we have $\hat{X}_n = \hat{Y}_n + s_0\hat{W}_n/2$ (formally justified in Section 4). The following theorem gives the limiting distributions of the random vector $(\hat{X}_n, \hat{Y}_n, \hat{W}_n)$. In Section 4.4 we state a similar result for (X_n, Y_n, W_n) under stronger assumptions. Note that the concepts are identical for any class of split trees where each node holds exactly one ball, such as binary search trees, quad trees, digital search trees and random simplex trees.

Let d_2 denote the Mallows metric, also called the minimal ℓ_2 metric (defined in Section 4). Let $\mathcal{M}^d_{0,2}$ be the set of probability measures on $\mathbb{R}^d$ with zero mean and finite second moment.

Theorem 4. *Let T_n be a split tree and let $\mathcal{V} = (V_1, \dots, V_b)$ be a split vector. Define*

$$\mu = -\sum_{i=1}^{b} \mathbb{E}\left[V_i \ln V_i\right], \qquad \text{and} \qquad D(\mathcal{V}) = \frac{1}{\mu}\sum_{i=1}^{b} V_i \ln V_i.$$

Assume that $\mathbb{P}\{\exists i : V_i = 1\} < 1$ and $s_0 > 0$. Let $(\hat{X}, \hat{Y}, \hat{W})$ be the unique solution in $\mathcal{M}^3_{0,2}$ for the system of fixed-point equations

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{W} \end{bmatrix} \overset{\mathrm{d}}{=} \begin{bmatrix} \sum_{i=1}^{b} V_i \hat{X}^{(i)} + \sum_{j=1}^{s_0} U_j + \frac{s_0}{2} D(\mathcal{V}) \\ \sum_{i=1}^{b} V_i \hat{Y}^{(i)} + \sum_{j=1}^{s_0} (U_j - 1/2) \\ \sum_{i=1}^{b} V_i \hat{W}^{(i)} + 1 + D(\mathcal{V}) \end{bmatrix}. \tag{1.11}$$

Here $(V_1, \dots, V_b)$, $U_1, \dots, U_{s_0}$, $(\hat{X}^{(1)}, \hat{Y}^{(1)}, \hat{W}^{(1)}), \dots, (\hat{X}^{(b)}, \hat{Y}^{(b)}, \hat{W}^{(b)})$ are independent, with $U_j \sim \mathrm{Unif}[0,1]$ for $j = 1, \dots, s_0$, and $\left(\hat{X}^{(i)}_n, \hat{Y}^{(i)}_n, \hat{W}^{(i)}_n\right) \sim (\hat{X}, \hat{Y}, \hat{W})$ for $i = 1, \dots, b$. Then the sequence $(\hat{X}_n, \hat{Y}_n, \hat{W}_n)$ defined in (1.10) converges to $(\hat{X}, \hat{Y}, \hat{W})$ in d_2 and in moment generating function within a neighborhood of the origin.

The proof of Theorem 4 uses the contraction method, introduced by Rösler [48] for finding the total path length of binary search trees. The technique has been applied to d-dimensional quad trees by Neininger and Rüschendorf [43] and to split trees in general by Broutin and Holmgren [8]. The contraction method also has many other applications in the analysis of recursive algorithms, see, e.g., [49, 50, 44].

Remark 5. *We assume that $s_0 > 0$, for otherwise we trivially have $\hat{X}_n = 0$ and Theorem 4 reduces to Theorem 2.1 in [8].*

Remark 6. *In a recent paper, Janson [33] showed that preferential attachment trees and random recursive trees can be viewed as split trees with infinite-dimensional split vectors. Thus we conjecture that the contraction method should also be applicable for these models and give results similar to Theorem 4.*

Remark 7. *Assume that the constant split vector $\mathcal{V} = (1/b, \dots, 1/b)$ is used and each node holds exactly one ball (a special case of digital search trees, see [14, Example 7]). Then $D(\mathcal{V}) = -1$ and (1.11) has the unique solution $(\hat{X}, \hat{Y}, \hat{W}) = (X, X, 0)$, where X has the limiting distribution for inversions in complete b-ary trees (see Theorem 3). This is as expected, as the shape of a split tree with these parameters is likely to be very similar to a complete b-ary tree.*

Conditional Galton–Watson trees

Finally, we consider conditional Galton–Watson trees (or equivalently, simply generated trees), which were introduced by Bienaymé [7] and Watson and Galton [54] to model the evolution of populations. A Galton–Watson tree starts with a root node. Then recursively, each node in the tree is given a random number of child nodes. The numbers of children are drawn independently from the same distribution ξ called the *offspring distribution*.

A conditional Galton–Watson tree T_n is a Galton–Watson tree conditioned on having n nodes. It generalizes many uniform random tree models, e.g., Cayley trees, Catalan trees, binary trees, b-ary trees, and Motzkin trees. For a comprehensive survey, see Janson [31]. For recent developments, see [32, 37, 16, 9].

In a series of three seminal papers, Aldous showed that T_n converges under re-scaling to a *continuum random tree*, which is a tree-like object constructed from a Brownian excursion [2, 3, 4]. Therefore, many asymptotic properties of conditional Galton–Watson trees, such as the height and the total path length, can be derived from properties of Brownian excursions [3]. Our analysis of inversions follows a similar route. In particular, we relate $I(T_n)$ to the *Brownian snake* studied by e.g., Janson and Marckert [35].

In the context of Galton–Watson trees, Aldous [3, Corollary 9] showed that $n^{-3/2}\Upsilon(T_n)$ converges to an Airy distribution. We will see that the standard deviation of $I(T_n) - \frac{1}{2}\Upsilon(T_n)$ is of order $n^{5/4} \ll n^{3/2}$, which by the decomposition (1.9) implies that $n^{-3/2}I(T_n)$ converges to the same Airy distribution, recovering one of the main results of Panholzer and Seitz [45, Theorem 5.3]. Our contribution for conditional Galton–Watson trees is a detailed analysis of Y_n under the scaling function $s(n) = n^{5/4}$.

Let $e(s), s \in [0,1]$ be the random path of a standard Brownian excursion, and define $C(s,t) \stackrel{\text{def}}{=} C(t,s) \stackrel{\text{def}}{=} 2\min_{s\le u\le t} e(u)$ for $0 \le s \le t \le 1$.

We define a random variable, see [30],

$$\eta \stackrel{\text{def}}{=} \int_{[0,1]^2} C(s,t)\mathrm{d}s\ \mathrm{d}t = 4\int_{0\le s\le t\le 1} \min_{s\le u\le t} e(u). \tag{1.12}$$

Theorem 5. *Suppose T_n is a conditional Galton–Watson tree with offspring distribution ξ such that $\mathbb{E}[\xi] = 1$, $\mathrm{Var}(\xi) = \sigma^2 \in (0,\infty)$, and $\mathbb{E}\left[e^{\alpha\xi}\right] < \infty$ for some $\alpha > 0$, and define*

$$Y_n = \frac{I(T_n) - \frac{1}{2}\Upsilon(T_n)}{n^{5/4}}.$$

Then we have

$$Y_n \xrightarrow{\mathrm{d}} Y \stackrel{\text{def}}{=} \frac{1}{\sqrt{12\sigma}}\sqrt{\eta}\,\mathcal{N}, \tag{1.13}$$

where $\mathcal{N}$ is a standard normal random variable, independent from the random variable η defined in (1.12). Moreover, $\mathbb{E}\left[e^{tY_n}\right] \to \mathbb{E}\left[e^{tY}\right] < \infty$ for all fixed $t \in \mathbb{R}$.

The moments of η and Y are known [34], see Section 5.

The rest of the paper is organized as follows. In Section 2, we prove Lemma 1 and Theorem 1. The results for fixed trees (Theorems 2, 3) are presented in Section 3. Split trees and conditional Galton–Watson trees are considered in Sections 4 and 5 respectively. Sections 4 and 5 are essentially self-contained, and the interested reader may skip ahead.

2 A fixed tree

In this section we study a fixed, non-random tree T. We begin with proving Lemma 1, which shows that $I(T)$ is a sum of independent uniform random variables.

Proof of Lemma 1. We define $Z_u = \sum_{v:v>u} \mathbf{1}_{\lambda(u)>\lambda(v)}$ and note that

$$I(T) \stackrel{\text{def}}{=} \sum_{u<v} \mathbf{1}_{\lambda(u)>\lambda(v)} = \sum_{u\in V} \left(\sum_{v:v>u} \mathbf{1}_{\lambda(u)>\lambda(v)} \right) = \sum_{u\in V} Z_u,$$

showing (1.1). Let $T_u \subseteq T$ denote the subtree rooted at u. It is clear that conditioned on the set $\lambda(T_u)$, λ restricted to T_u is a uniformly random labeling of T_u into $\lambda(T_u)$. Recall that z_u denotes the size of T_u. If the elements of $\lambda(T_u)$ are $\ell_1 < \cdots < \ell_{z_u}$ and if $\lambda(u) = \ell_i$, then $Z_u = i-1$. As $\lambda(u)$ is uniformly distributed, so is Z_u.

We prove independence of the Z_v by induction on V. The base case $|V| = 1$ is trivial. Let $T_1, \dots, T_d$ be the subtrees rooted at the children of the root ρ, and condition on the sets $\lambda(T_1), \dots, \lambda(T_d)$. Given these sets, λ restricted to T_i is a uniformly random labeling of T_i using the given labels $\lambda(T_i)$, and these labelings are independent for different i. Hence, conditioning on $\lambda(T_1), \dots, \lambda(T_d)$, the d families $(Z_v)_{v\in T_i}$ are independent, and each is distributed as the corresponding family for the tree T_i.

Consequently, by induction, still conditioned on $\lambda(T_1), \dots, \lambda(T_d)$, $(Z_v)_{v\neq\rho}$ are independent, with $Z_v \sim \operatorname{Unif}\{0, 1, \dots, z_v - 1\}$. Furthermore, $Z_\rho = \lambda(\rho) - 1$, and $\lambda(\rho)$ is determined by $\lambda(T_1), \dots, \lambda(T_d)$ (as the only label not in $\bigcup_1^d \lambda(T_i)$). Hence the family $(Z_v)_{v\neq\rho}$ of independent random variables is also independent of Z_ρ, and thus $(Z_v)_{v\in V}$ are independent. This completes the induction, and thus the proof. □

Our first use of the representation in Lemma 1 is to prove Theorem 1, which gives both a formula for the moment generating function and explicit formulas for the cumulants of $I(T)$ for a fixed T. The proof begins with a simple lemma giving the cumulants and the moment generating function of Z_v in Lemma 1, from which Theorem 1 will follow immediately.

Recall that the Bernoulli numbers B_k can be defined by their generating function

$$\sum_{k=0}^{\infty} B_k \frac{x^k}{k!} = \frac{x}{e^x - 1} \tag{2.1}$$

(convergent for $|x| < 2\pi$), see, e.g., [17, (24.2.1)]. Recall also $B_0 = 1$, $B_1 = -\frac{1}{2}$ and $B_2 = \frac{1}{6}$, and that $B_{2k+1} = 0$ for $k \geq 1$.

Lemma 2. *Let $N \geq 1$, and let Z_N be uniformly distributed on $\{0, 1, \dots, N-1\}$. Then $\mathbb{E}[Z_N] = (N-1)/2$, $\operatorname{Var}(Z_N) = (N^2-1)/12$ and, more generally,*

$$\varkappa_k(Z_N) = \frac{B_k}{k}(N^k - 1), \qquad k \geq 2, \tag{2.2}$$

where B_k is the k-th Bernoulli number. The moment generating function of Z_N is

$$\mathbb{E}\left[e^{tZ_N}\right] = \frac{e^{Nt}-1}{N(e^t-1)}. \tag{2.3}$$

Proof. This is presumably well-known, but we include a proof for completeness. The moment generating function of Z_N is

$$\mathbb{E}\left[e^{tZ_N}\right] = \frac{1}{N}\sum_{j=0}^{N-1} e^{jt} = \frac{e^{Nt}-1}{N(e^t-1)}, \tag{2.4}$$

verifying (2.3). The function $(e^t-1)/t$ is analytic and non-zero in the disc $|t| < 2\pi$, and thus has there a well-defined analytic logarithm

$$f(t) := \log \frac{e^t - 1}{t}, \tag{2.5}$$

with $f(0) = 0$. By (2.4) and (2.5), the cumulant generating function of Z_N can be written as

$$\log \mathbb{E}\left[e^{tZ_N}\right] = f(Nt) - f(t). \tag{2.6}$$

Differentiating (2.5) yields (for $0 < |t| < 2\pi$)

$$f'(t) = \frac{d}{dt}\bigl(\log(e^t-1) - \log t\bigr) = \frac{e^t}{e^t-1} - \frac{1}{t} = \frac{1}{e^t-1} + 1 - \frac{1}{t},$$

and thus, using (2.1),

$$tf'(t) = \frac{t}{e^t-1} + t - 1 = \sum_{k=0}^{\infty} B_k \frac{t^k}{k!} - 1 + t = \sum_{k=2}^{\infty} B_k \frac{t^k}{k!} + \frac{1}{2}t.$$

Consequently,

$$f(t) = \sum_{k=2}^{\infty} \frac{B_k}{k}\frac{t^k}{k!} + \frac{1}{2}t, \tag{2.7}$$

and thus by (2.6)

$$\log \mathbb{E}\left[e^{tZ_N}\right] = \sum_{k=2}^{\infty} \frac{B_k}{k}(N^k-1)\frac{t^k}{k!} + \frac{N-1}{2}t.$$

The results on cumulants follow. (Of course, $\mathbb{E}[Z_N]$ is more simply calculated directly.) ☐

Remark 8. *Similarly, using (2.7), or by (2.2) and a limiting argument, if $U \sim \mathrm{Unif}[0,1]$ or $U \sim \mathrm{Unif}[-\frac{1}{2}, \frac{1}{2}]$, then $\varkappa_k(U) = B_k/k$, $k \geq 2$.*

Recall that in the introduction, we defined

$$c(v_1, \dots, v_k) \overset{\text{def}}{=} |\{u : u \le v_1, \dots, u \le v_k\}|,$$

i.e., $c(v_1, \dots, v_k)$ is the number of common ancestors of $v_1, \dots, v_k$.

Lemma 3. *Let z_v denote the number of vertices in subtree rooted at v. Then for $k \ge 1$,*

$$\sum_v z_v^k = \Upsilon_k(T) \overset{\text{def}}{=} \sum_{v_1,\dots,v_k} c(v_1, \dots, v_k).$$

Proof. It is easily seen that

$$\sum_u z_u = \sum_{u,v} \mathbf{1}_{\{u \le v\}} = \sum_v c(v).$$

Similarly,

$$\sum_u z_u^2 = \sum_{u,v,w} \mathbf{1}_{\{u \le v, u \le w\}} = \sum_{v,w} c(v, w).$$

More generally,

$$\sum_u z_u^k = \sum_u \prod_{i=1}^{k} \left(\sum_{v_i} \mathbf{1}_{\{u \le v_i\}} \right) = \sum_{v_1,\dots,v_k} c(v_1, \dots, v_k). \qquad \square$$

Remark 9. *Observe that all common ancestors of the k vertices must lie on a path; stretching from the last common ancestor to the root. Define a related parameter $\Upsilon'_k(T)$ to be the sum over all k-tuples of the length of this path (rather than number of vertices in the path). We call this the k-*common path length*. Now $\Upsilon'_1(T) = \Upsilon(T)$ and $\Upsilon'_2(T)$ has appeared in various contexts, see for example [30] (where it is denoted $Q(T)$). Let $v_1 \wedge v_2$ denote the last common ancestor of the vertices v_1 and v_2. It is easy to see that, with $n = |T|$,*

$$\Upsilon'_k(T) \overset{\text{def}}{=} \sum_{v_1,\dots,v_k} h(v_1 \wedge \dots \wedge v_k) = \sum_{v_1,\dots,v_k} (c(v_1, \dots, v_k) - 1) = \Upsilon_k(T) - n^k,$$

and by Lemma 3, $\Upsilon_k(T) = \sum_v z_v^k$, so $\Upsilon'_k(T) = \sum_{v \ne \rho} z_v^k$.

Remark 10. *Let S_k be a star with k leaves $\ell_1, \dots, \ell_k$ and root o. Then $\Upsilon_k(T)$ is the number of embeddings $\phi : V(S_k) \to V(T)$ such that $\phi(o) \le \phi(\ell_i)$ for each i. Similarly the k-common path-length $\Upsilon'_k(T)$ is the number of such embeddings ϕ such that $\phi(o) < \phi(\ell_i)$ for each i.*

Proof of Theorem 1. Since cumulants are additive for sums of independent random variables, an immediate consequence of Lemmas 1 and 2 is that

$$\varkappa_k(I(T)) = \frac{B_k}{k} \sum_{v \in V} \left(z_v^k - 1 \right) = \frac{B_k}{k} (\Upsilon_k(T) - |V|), \qquad k \ge 1.$$

where the last equality follows from Lemma 3. The fact that $\mathbb{E}[I(T)] = \frac{1}{2}\Upsilon(T)$ was noted already in (1.2).

Similarly, (1.6) follows from Lemma 1 and (2.4).

For the estimate (1.7), note first, e.g. by Taylor expansions, that $\cosh x \le e^{x^2/2}$ for every real x. It follows that if U is any symmetric random variable with $|U| \le a$, then

$$\mathbb{E}\left[e^{tU}\right] = \mathbb{E}[\cosh(tU)] \le e^{a^2 t^2/2}. \tag{2.8}$$

(See [28, (4.16)] for a more general result.) Lemma 1 thus implies, applying (2.8) to each $Z_v - \mathbb{E}[Z_v]$,

$$\mathbb{E}\left[e^{t(I(T)-\mathbb{E}[I(T)])}\right] = \prod_v \mathbb{E}\left[e^{t(Z_v - \mathbb{E}[Z_v])}\right] \le \prod_v e^{t^2 (z_v-1)^2/8} = e^{t^2 \sum_v (z_v-1)^2/8},$$

which yields (1.7), using also Lemma 3. □

3 A sequence of fixed trees

In this section, we study

$$X_n = \frac{I(T_n) - \mathbb{E}[I(T_n)]}{s(n)},$$

where T_n is a sequence of fixed trees and $s(n)$ is an appropriate normalization factor. We start by proving Theorem 2, a sufficient condition for X_n to converge in distribution when $s(n) = \sqrt{\Upsilon_2(T_n)}$.

Proof of Theorem 2. First $\varkappa_1(X_n) = \mathbb{E}[X_n] = 0$. For $k \ge 2$, note that shifting a random variable does not change its k-th cumulant. Also note that $\Upsilon_k(T_n) \overset{\text{def}}{=} \sum_{v_1,\dots,v_k} c(v_1,\dots,v_k) \ge n^k$. Therefore, it follows from Theorem 1 that

$$\varkappa_k(X_n) = \frac{\varkappa_k(I(T_n))}{(\Upsilon_2(T_n)-n)^{k/2}} = \frac{B_k}{k} \frac{\Upsilon_k(T_n) - n}{(\Upsilon_2(T_n)-n)^{k/2}} \sim \frac{B_k}{k} \frac{\Upsilon_k(T_n)}{\Upsilon_2(T_n)^{k/2}}, \qquad k \ge 2.$$

Recall that all odd Bernoulli numbers except B_1 are zero. Thus letting $\zeta_k = 0$ for all odd k, the assumption that $\Upsilon_{2k}(T_n)/\Upsilon_2(T_n)^k \to \zeta_{2k}$ for all $k \ge 1$ implies that

$$\varkappa_k(X_n) \to \frac{B_k}{k} \zeta_k, \qquad k \ge 1.$$

Since every moment can be expressed as a polynomial in cumulants, it follows that every moment $\mathbb{E}\left[X_n^k\right]$ converges, $k \ge 1$. Thus to show that there exists an X such that $X_n \xrightarrow{\text{d}} X$, it suffices to show that the moment generating function $\mathbb{E}\left[e^{tX_n}\right]$ stays bounded for all small fixed t; we shall show that this holds for all real t. In fact, using Lemma 3,

$$\sum_v (z_v - 1)^2 \le \sum_v (z_v^2 - 1) = \Upsilon_2(T_n) - n \le \Upsilon_2(T_n).$$

Hence, (1.7) yields

$$\mathbb{E}\left[e^{tX_n}\right] \le \exp\Big(\tfrac{1}{8}\big(t/\sqrt{\Upsilon_2(T_n)}\big)^2 \sum_v (z_v-1)^2\Big) \le \exp\big(\tfrac{1}{8}t^2\big), \qquad t \in \mathbb{R}.$$

This and the moment convergence imply the claims in the theorem. □

3.1 The complete b-ary tree

We prove Theorem 3, which asserts that for complete b-ary trees the limiting variable of X_n is the unique X for which $\varkappa_k(X) = \frac{B_k}{k}\frac{b^{k-1}}{b^{k-1}-1}$ for even $k \ge 2$ and zero for odd k. Fix $b \ge 2$. In the complete b-ary tree of height m, each node v at depth $d \in \{0, 1, \ldots, m\}$ has subtree size $z_v = a_{m,d} \overset{\text{def}}{=} (b^{m-d+1} - 1)/(b - 1)$. Hence Lemma 1 implies that $X_n = \sum_{d=0}^{m} \sum_{j=1}^{b^d} Z_{d,j}/n$, where

$$Z_{d,j} \sim \text{Unif}\left\{-\frac{a_{m,d} - 1}{2}, -\frac{a_{m,d} - 2}{2}, \ldots, \frac{a_{m,d} - 2}{2}, \frac{a_{m,d} - 1}{2}\right\}$$

are independent random variables. Let $U_{d,j}$ be independent $\text{Unif}[-\frac{1}{2}, \frac{1}{2}]$. Approximating $Z_{d,j} \approx U_{d,j} a_{m,d}$ and noticing that $n/a_{m,d} \approx b^d$, intuitively we should have for large n,

$$X_n = \sum_{d=0}^{m} \sum_{j=1}^{b^d} \frac{a_{m,d}}{n} \cdot \frac{Z_{d,j}}{a_{m,d}} \approx \sum_{d \geq 0} \sum_{j=1}^{b^d} \frac{U_{d,j}}{b^d} \overset{\text{def}}{=} X. \tag{3.1}$$

It is not difficult to show this rigorously by truncating the sums. Also, it is not difficult to prove Theorem 3 by showing that $\mathbb{E}\left[e^{tX_n}\right] \to \mathbb{E}\left[e^{tX}\right]$ for all $t \in \mathbb{R}$ and checking the cumulants of X, using Remark 8. But instead we choose the route of computing the k-total common ancestors of b-ary trees and then applying Theorem 2.

Lemma 4. *Assume $b \geq 2$. Let T_n be the complete b-ary tree on $n = (b^{m+1} - 1)/(b - 1)$ nodes. Then*

$$\Upsilon_1(T_n) \sim n \log_b n, \qquad \Upsilon_k(T_n) \sim \frac{b^{k-1}}{b^{k-1} - 1} n^k, \qquad k \geq 2.$$

Proof. The height of T_n is $m \sim \log_b n$. It follows from Lemma 3 that

$$\Upsilon_1(T_n) = \sum_v z_v = \sum_{d=0}^{m} b^d \times a_{m,d} = \frac{b^{m+1}}{b-1} \sum_{d=0}^{m} \left(1 - \frac{1}{b^{m+1-d}}\right) = \frac{b^{m+1}}{b-1} (m + O(1)) \sim n \log_b n.$$

Similarly, for $k \geq 2$,

$$\Upsilon_k(T_n) = \sum_v z_v^k = \sum_{d=0}^{m} b^d \times a_{m,d}^k = \frac{b^{(m+1)k}}{(b-1)^k} \sum_{d=0}^{m} \frac{1}{b^{d(k-1)}} \left(1 - \frac{1}{b^{m+1-d}}\right)^k \sim n^k \frac{b^{k-1}}{b^{k-1} - 1}. \qquad \square$$

Proof of Theorem 3. Let $X_n' = (I(T_n) - \mathbb{E}[I(T_n)])/\sqrt{\Upsilon_2(T_n)}$. By Lemma 4, for fixed $k \geq 1$,

$$\frac{\Upsilon_{2k}(T_n)}{\Upsilon_2(T_n)^k} \sim \frac{n^{2k} \frac{b^{2k-1}}{b^{2k-1}-1}}{\left(n^2 \frac{b}{b-1}\right)^k} = \frac{b^{2k-1}}{b^{2k-1} - 1} \left(\frac{b-1}{b}\right)^k.$$

By Theorem 2, there exists a unique distribution X' such that

$$\varkappa_{2k-1}(X') = 0, \qquad \varkappa_{2k}(X') = \frac{B_{2k}}{2k} \frac{b^{2k-1}}{b^{2k-1} - 1} \left(\frac{b-1}{b}\right)^k, \qquad k \geq 1;$$

moreover, $\mathbb{E}\left[e^{tX_n'}\right] \to \mathbb{E}\left[e^{tX'}\right] < \infty$ for every t. Recall that, using Lemma 4 again,

$$X_n \overset{\text{def}}{=} \frac{I(T_n) - \mathbb{E}[I(T_n)]}{n} = (1 + o(1)) \left(\frac{b}{b-1}\right)^{1/2} X_n'.$$

Let $X'' = \left(b/(b-1)\right)^{1/2} X'$; then $\mathbb{E}\left[e^{tX_n}\right] \to \mathbb{E}\left[e^{tX''}\right]$ for every real t and X'' has cumulants

$$\varkappa_1(X'') = 0, \qquad \varkappa_k(X'') = \frac{B_k}{k} \frac{b^{k-1}}{b^{k-1} - 1}, \qquad k \geq 2,$$

as in (1.8). It is not difficult to show that X'' has the same distribution as X defined in (3.1) by checking the cumulants of X, using Remark 8. $\square$

3.2 Balanced b-ary trees

We call a b-ary tree *balanced* if all but the last level of the tree is full and vertices at the last level take the leftmost positions. A simple example of a balanced binary tree is T_n in which both the left and right subtrees are complete b-ary trees but the left subtree has one more level than the right subtree. Since the left subtree is of size about $2n/3$, and the right subtree is of size about $n/3$, Theorem 3 and Lemma 1 imply that

$$X_n = \frac{I(T_n) - \mathbb{E}\left[I(T_n)\right]}{n} \xrightarrow{\text{d}} U + \frac{2X'}{3} + \frac{X''}{3},$$

where $U \sim \text{Unif}[-\frac{1}{2}, \frac{1}{2}]$ and X', X'' are independent copies of X. The three terms in the limit correspond to inversions involving the root, inversions in the left subtree and inversions in the right subtree.

The above example shows that the limit distribution of X_n in a balanced b-ary tree in which each subtree of the root is complete should be U plus a linear combination of independent copies of X. We formalize this observation in the following corollary.

Corollary 1. *Let T_n be a balanced b-ary tree. Let X_n and X be as in Theorem 3. Let $\{x\} \stackrel{\text{def}}{=} x - \lfloor x \rfloor$. Assume that*

$$\{\log_b((b-1)n)\} = \log_b\left(1 + \frac{b-1}{b} i\right) + o\left(\frac{1}{\log n}\right), \tag{3.2}$$

where $i \in \{0, \dots, b\}$ is a constant. We have

$$X_n \xrightarrow{\text{d}} U + \sum_{j=1}^{i} \frac{b}{b + i(b-1)} X^{(j)} + \sum_{j=i+1}^{b} \frac{1}{b + i(b-1)} X^{(j)} \stackrel{\text{def}}{=} X(b,i),$$

where $U \sim \text{Unif}[-\frac{1}{2}, \frac{1}{2}]$, $X^{(j)} \sim X$ are all independent. Moreover $\mathbb{E}\left[e^{tX_n}\right] \to \mathbb{E}\left[e^{tX(b,i)}\right]$ for all $t \in \mathbb{R}$.

Remark 11. *Condition (3.2) is equivalent of saying that all the b subtrees of the root of T_n except one (either the i-th or the $(i+1)$-th) are complete b-ary trees and the exceptional subtree differs from a complete b-ary tree in size by at most $o(n/\log(n))$.*

4 A sequence of split trees

We will now define split trees introduced by Devroye [15]. The random split tree T_n has parameters $b, s, s_0, s_1, \mathcal{V}$ and n. The integers b, s, s_0, s_1 are required to satisfy the inequalities

$$2 \le b, \quad 0 < s, \quad 0 \le s_0 \le s, \quad 0 \le bs_1 \le s + 1 - s_0. \tag{4.1}$$

and $\mathcal{V} = (V_1, \dots, V_b)$ is a random non-negative vector with $\sum_{i=1}^{b} V_i = 1$. We define T_n algorithmically. Consider the infinite b-ary tree $\mathcal{U}$, and view each node as a bucket with capacity s. Each node u is assigned an independent copy $\mathcal{V}_u$ of the random split vector $\mathcal{V}$. Let $C(u)$ denote the number of balls in node u, initially setting $C(u) = 0$ for all u. Say that u is a *leaf* if $C(u) > 0$ and $C(v) = 0$ for all children v of u, and *internal* if $C(v) > 0$ for some proper descendant v, i.e., $v < u$. We add n balls labeled $\{1, \dots, n\}$ to $\mathcal{U}$ one by one. The j-th ball is added by the following "trickle-down" procedure.

1. Add j to the root.
2. While j is at an internal node u, choose child i with probability $V_{u,i}$, where $(V_{u,1}, \dots, V_{u,b})$ is the split vector at u, and move j to child i.
3. If j is at a leaf u with $C(u) < s$, then j stays at u and we set $C(u) \leftarrow C(u) + 1$.

   If j is at a leaf with $C(u) = s$, then the balls at u are distributed among u and its children as follows. We select $s_0 \le s$ of the balls uniformly at random to stay at u. Among the remaining $s + 1 - s_0$ balls, we uniformly at random distribute s_1 balls to each of the b children of u. Each of the remaining $s + 1 - s_0 - bs_1$ balls is placed at a child node chosen independently at random according to the split vector assigned to u. This splitting process is repeated for any child which receives more than s balls.

For example, if we let $b = 2, s = s_0 = 1, s_1 = 0$ and $\mathcal{V}$ have the distribution of $(U, 1-U)$ where $U \sim \mathrm{Unif}[0,1]$, then we get the well-known binary search tree.

Once all n balls have been placed in $\mathcal{U}$, we obtain T_n by deleting all nodes u such that the subtree rooted at u contains no balls. Note that an internal node of T_n contains exactly s_0 balls, while a leaf contains a random amount in $\{1, \dots, s\}$. We assume, as previous authors, that $\mathbb{P}\{\exists i : V_i = 1\} < 1$. We can assume $\mathcal{V}$ has a symmetric (permutation invariant) distribution without loss of generality, since a uniform random permutation of subtree order does not change the number of inversions.

An equivalent definition of split trees is as follows. Consider an infinite b-ary tree $\mathcal{U}$. The split tree T_n is constructed by distributing n balls (pieces of information) among nodes of $\mathcal{U}$. For a node u, let n_u be the number of balls stored in the subtree rooted at u. Once n_u are all decided, we take T_n to be the largest subtree of $\mathcal{U}$ such that $n_u > 0$ for all $u \in T_n$. Let the split vector $\mathcal{V} \in [0,1]^b$ be as before. Let $\mathcal{V}_u = (V_{u,1}, \dots, V_{u,b})$ be the independent copy of $\mathcal{V}$ assigned to u. Let $u_1, \dots, u_b$ be the child nodes of u. Conditioning on n_u and $\mathcal{V}_u$, if $n_u \le s$, then $n_{u_i} = 0$ for all i; if $n_u > s$, then

$$(n_{u_1}, \dots, n_{u_b}) \sim \mathrm{Mult}(n - s_0 - bs_1, V_{u,1}, \dots, V_{u,b}) + (s_1, s_1, \dots, s_1),$$

where Mult denotes multinomial distribution, and b, s, s_0, s_1 are integers satisfying (4.1). Note that $\sum_{i=1}^{b} n_{u_i} \le n$ (hence the "splitting"). Naturally for the root ρ, $n_\rho = n$. Thus the distribution of $(n_u, \mathcal{V}_u)_{u \in V(\mathcal{U})}$ is completely defined.

4.1 Outline

In this section we outline how one can apply the *contraction method* to prove Theorem 4 but leave the detailed proof to Section 4.2 and Section 4.3. In Section 4.4 we state and outline the proof of the corresponding theorem for inversions on nodes under stronger assumptions.

Recall that in (1.10), we define

$$\hat{X}_n = \frac{\hat{I}(T_n) - \mathbb{E}\left[\hat{I}(T_n)\right]}{n}, \quad \hat{Y}_n = \frac{\hat{I}(T_n) - s_0 \hat{\Upsilon}(T_n)/2}{n}, \quad \hat{W}_n = \frac{\hat{\Upsilon}(T_n) - \mathbb{E}\left[\hat{\Upsilon}(T_n)\right]}{n}.$$

Let $\overline{n} = (n_1, \dots, n_b)$ denote the vector of the (random) number of balls in each of the b subtrees of the root. Broutin and Holmgren [8] showed that, conditioning on $\overline{n}$,

$$\hat{W}_n \overset{\mathrm{d}}{=} \sum_{i=1}^{b} \frac{n_i}{n} \hat{W}_{n_i} + \frac{n - s_0}{n} + \hat{D}_n(\overline{n}), \quad \hat{D}_n(\overline{n}) \overset{\mathrm{def}}{=} -\frac{\mathbb{E}\left[\hat{\Upsilon}(T_n)\right]}{n} + \sum_{i=1}^{b} \frac{\mathbb{E}\left[\hat{\Upsilon}(T_{n_i})\right]}{n}. \tag{4.2}$$

We derive similar recursions for $\hat{X}_n$ and $\hat{Y}_n$. Conditioning on $\overline{n}$, $\hat{I}(T_n)$ satisfies the recursion

$$\hat{I}(T_n) \stackrel{\mathrm{d}}{=} \hat{Z}_\rho + \sum_{i=1}^{b} \hat{I}(T_{n_i}),$$

where $\hat{Z}_\rho$ denotes the number of inversions involving balls contained in the root ρ. Therefore, still conditioning on $\overline{n}$, we have

$$\begin{aligned}\hat{X}_n &\stackrel{\mathrm{d}}{=} \sum_{i=1}^{b} \frac{n_i}{n}\hat{X}_{n_i} + \frac{\hat{Z}_\rho}{n} - \frac{\mathbb{E}\left[\hat{I}(T_n)\right]}{n} + \sum_{i=1}^{b} \frac{\mathbb{E}\left[\hat{I}(T_{n_i})\right]}{n} \\ &= \sum_{i=1}^{b} \frac{n_i}{n}\hat{X}_{n_i} + \frac{\hat{Z}_\rho}{n} - \frac{s_0}{2}\frac{\mathbb{E}\left[\hat{\Upsilon}(T_n)\right]}{n} + \frac{s_0}{2}\sum_{i=1}^{b} \frac{\mathbb{E}\left[\hat{\Upsilon}(T_{n_i})\right]}{n} \\ &= \sum_{i=1}^{b} \frac{n_i}{n}\hat{X}_{n_i} + \frac{\hat{Z}_\rho}{n} + \frac{s_0}{2}\hat{D}_n(\overline{n}),\end{aligned} \tag{4.3}$$

where we use that

$$\mathbb{E}\left[\hat{I}(T_n) \mid T_n\right] = \frac{s_0}{2}\hat{\Upsilon}(T_n). \tag{4.4}$$

(See the proof of Lemma 6.) It follows also from (4.4) that $\hat{X}_n = \hat{Y}_n + \frac{s_0}{2}\hat{W}_n$ and

$$\hat{Y}_n \stackrel{\mathrm{d}}{=} \sum_{i=1}^{b} \frac{n_i}{n}\hat{Y}_{n_i} + \frac{\hat{Z}_\rho}{n} - \frac{s_0}{2}\frac{n - s_0}{n}. \tag{4.5}$$

In Lemma 7 below, we show that

$$\frac{\hat{Z}_\rho}{n} \xrightarrow{L^2} U_1 + \cdots + U_{s_0},$$

where $U_1, \ldots, U_{s_0}$ are independent and uniformly distributed in $[0,1]$. Broutin and Holmgren [8] have shown that $\hat{D}_n(\overline{n}) \xrightarrow{\text{a.s.}} D(\mathcal{V})$, where

$$\mu = -\sum_{i=1}^{b} \mathbb{E}\left[V_i \ln V_i\right], \qquad \text{and} \qquad D(\mathcal{V}) = \frac{1}{\mu}\sum_{i=1}^{b} V_i \ln V_i. \tag{4.6}$$

Together with $(n_1/n, \ldots, n_b/n) \xrightarrow{\text{a.s.}} (V_1, \ldots, V_b)$ (by the law of large number), we arrive at the following fixed-point equations (already presented in Theorem 4)

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{W} \end{bmatrix} \stackrel{\mathrm{d}}{=} \begin{bmatrix} \sum_{i=1}^{b} V_i \hat{X}^{(i)} + \sum_{j=1}^{s_0} U_j + \frac{s_0}{2} D(\mathcal{V}) \\ \sum_{i=1}^{b} V_i \hat{Y}^{(i)} + \sum_{j=1}^{s_0} (U_j - 1/2) \\ \sum_{i=1}^{b} V_i \hat{W}^{(i)} + 1 + D(\mathcal{V}) \end{bmatrix}. \tag{4.7}$$

For a random vector $X \in \mathbb{R}^d$, let $\|X\|$ be the Euclidean norm of X. Let $\|X\|_2 \stackrel{\text{def}}{=} \sqrt{\mathbb{E}\left[\|X\|^2\right]}$. Recall that $\mathcal{M}_{0,2}^d$ denotes the set of probability measures on $\mathbb{R}^d$ with zero mean and finite second moment. The Mallows metric on $\mathcal{M}_{0,2}^d$ is defined by

$$d_2(\nu, \lambda) = \inf\left\{\|X - Y\|_2 : X \sim \lambda, Y \sim \nu\right\}.$$

Using the contraction method, Broutin and Holmgren [8] proved that $\hat{W}_n \xrightarrow{d_2} \hat{W}$, the unique solution of the first equation of (4.7) in $\mathcal{M}_{0,2}^1$.

We can apply the same contraction method to show that the vector $(\hat{X}_n, \hat{Y}_n, \hat{W}_n) \xrightarrow{d_2} (\hat{X}, \hat{Y}, \hat{W})$, the unique solution of (4.7) in $\mathcal{M}_{0,2}^3$. But we only outline the argument here since we will actually use a result by Neininger [42] which gives us a shortcut. Assume that the independent vectors $\left(\hat{X}^{(i)}, \hat{Y}^{(i)}, \hat{W}^{(i)}\right)$, $i = 1, \ldots, b$ share some common distribution $\mu \in \mathcal{M}_{0,2}^3$. Let $F(\mu) \in \mathcal{M}_{0,2}^3$ be the distribution of the random vector given by the right hand side of (4.7). Using a coupling argument, we can show that for all $\nu, \lambda \in \mathcal{M}_{0,2}^3$,

$$d_2(F(\nu), F(\lambda)) < c d_2(\nu, \lambda),$$

where $c \in (0, 1)$ is a constant. Thus F is a contraction and by Banach's fixed point theorem, (4.7) must have a unique solution $(\hat{X}, \hat{Y}, \hat{W}) \in \mathcal{M}_{0,2}^3$. Finally, we can use a similar coupling argument to show that $(\hat{X}_n, \hat{Y}_n, \hat{W}_n) \xrightarrow{d_2} (\hat{X}, \hat{Y}, \hat{W})$.

4.2 Convergence in the Mallows metric

Lemma 5. *Let $(\hat{X}_n, \hat{Y}_n, \hat{W}_n)$ and $(\hat{X}, \hat{Y}, \hat{W})$ be as in Theorem 4. Then*

$$d_2\left((\hat{X}_n, \hat{Y}_n, \hat{W}_n), (\hat{X}, \hat{Y}, \hat{W})\right) \to 0.$$

We will apply Theorem 4.1 of Neininger [42], which summarizes sufficient conditions for the contraction method outlined in the previous section to work. Since the statement of the theorem is rather lengthy, we do not repeat it here and refer the readers to the original paper.

Neininger's theorem implies that $(\hat{X}_n, \hat{Y}_n, \hat{W}_n) \xrightarrow{d_2} (\hat{X}, \hat{Y}, \hat{W})$ if the following three conditions are satisfied:

$$\left(\frac{\hat{Z}_\rho}{n}, \frac{n_1}{n}, \ldots, \frac{n_b}{n}, \hat{D}_n(n)\right) \xrightarrow{d_2} \left(\sum_{j=1}^{s_0} U_j, V_1, \ldots, V_b, D(\mathcal{V})\right), \qquad n \to \infty, \tag{4.8}$$

$$\sum_{i=1}^{b} \mathbb{E}\left[V_i^2\right] < 1, \tag{4.9}$$

$$\mathbb{E}\left[\mathbf{1}_{[n_i \le \ell] \cup [n_i = n]} \left(\frac{n_i}{n}\right)^2\right] \to 0, \qquad n \to \infty, \tag{4.10}$$

for all $\ell \ge 1$ and $i = 1, \ldots, b$. (The three conditions correspond to (11), (12) and (13) in [42].)

Condition (4.9) is satisfied by the assumption that $\mathbb{P}\{\exists i : V_i = 1\} < 1$. Since we assume that $s_0 > 0$, the event $n_i = n$ cannot happen. So the expectation in (4.10) is at most $(\ell/n)^2 \to 0$ and this condition is also satisfied. The last condition (4.8) follows from the following two lemmas.

Lemma 6. *We have* $\hat{D}_n(\overline{n}) \xrightarrow{L^2} D(\mathcal{V})$ *and* $\sup_{n\geq 1} \hat{D}_n(\overline{n})$ *is bounded deterministically.*

Proof. We first derive an expression for the expected number of inversions. Any internal node contains s_0 balls, so any ball at height h has $s_0 \times h$ ancestral balls. Let $B(T_n)$ be the set of balls in T_n. Conditioning on T_n, we have

$$\mathbb{E}\left[\hat{I}(T_n)\,|\,T_n\right] = \mathbb{E}\left[\sum_{\beta\in B(T_n)} |\{\beta' : \beta' < \beta, \lambda(\beta') > \lambda(\beta)\}| \,\middle|\, T_n\right] = \sum_{\beta\in B(T_n)} \frac{s_0}{2} h(\beta) = \frac{s_0}{2}\hat{\Upsilon}(T_n).$$

Thus by Broutin and Holmgren [8, Theorem 3.1],

$$\mathbb{E}\left[\hat{I}(T_n)\right] = \frac{s_0}{2}\mathbb{E}\left[\hat{\Upsilon}(T_n)\right] = \frac{s_0}{2}\left[\frac{1}{\mu} n\ln n + n\varpi(\ln n) + o(n)\right], \tag{4.11}$$

with μ as in (4.6), where ϖ is a continuous function of period $d = \sup\{a \geq 0 : \mathbb{P}\{\ln V_1 \in a\mathbb{Z}\} = 1\}$. In particular, ϖ is constant if $\ln V_1$ is *non-lattice*, meaning that $d = 0$.

The convergence of the toll function can now be deduced from the same result on total path length from [8], but we include the short argument for completeness. Conditioning on the split vector of the root $(V_1, \ldots, V_n)$ and noting that $(n_1/n, \ldots, n_b/n) \xrightarrow{\text{a.s.}} (V_1, \ldots, V_b)$, we have from (4.2), (4.11),

$$\begin{aligned}
\hat{D}_n(\overline{n}) &= -\frac{1}{\mu}\ln n - \varpi(\ln n) + \sum_{i=1}^{b}\left(\frac{1}{\mu}\frac{n_i \ln n_i}{n} + \frac{n_i}{n}\varpi(\ln n_i)\right) + o(1) \\
&= \left(\sum_{i=1}^{b} \frac{1}{\mu}\frac{n_i}{n}\ln\frac{n_i}{n}\right) + \left(\sum_{i=1}^{b}\frac{n_i}{n}\varpi\left(\ln V_i + \ln n\right)\right) - \varpi(\ln n) + o(1) \\
&= \frac{1}{\mu}\sum_{i=1}^{b} V_i \ln V_i + o(1),
\end{aligned}$$

where we use that ϖ is continuous and has the same period as $\ln V_i$. So we have

$$\hat{D}_n(\overline{n}) \xrightarrow{\text{a.s.}} D(\mathcal{V}) \stackrel{\text{def}}{=} \frac{1}{\mu}\sum_{i=1}^{b} V_i \ln V_i,$$

without conditioning on $(V_1, \ldots, V_b)$. Note that since for $x_1, \ldots, x_b \geq 0$ with $\sum_{i=1}^{b} x_i = 1$, we have $\sum_{i=1}^{b} x_i \ln(x_i) \geq -\ln b$ [12, Theorem 3.1], both $\hat{D}_n(\overline{n})$ and $D(\mathcal{V})$ are bounded deterministically. Thus $\hat{D}_n(\overline{n}) \xrightarrow{L^2} D(\mathcal{V})$ by the dominated convergence theorem. □

Lemma 7. *For* $i = 1, \ldots, s_0$*, let* U_i *be a* $\mathrm{Unif}[0,1]$ *random variable independent of all other random variables. Then there exists a coupling such that* $\hat{Z}_\rho/n \xrightarrow{L^2} \sum_{i=1}^{s_0} U_i$.

Proof. We have $\hat{Z}_\rho = \sum_{i=1}^{s_0}(\lambda_i - i)$, where $\lambda_1 < \lambda_2 < \cdots < \lambda_{s_0}$ are the labels for the balls in the root, chosen uniformly at random from $[n]$ without replacement. Indeed, the ball with label λ_i forms an inversion with the balls with labels $\{\lambda : \lambda < \lambda_i, \lambda \neq \lambda_j\ \forall j < i\}$, a set of size $\lambda_i - i$.

Let $\lambda_i' = \lceil nU_i \rceil$ for $i = 1, \ldots, s_0$. Then $\lambda_1', \ldots, \lambda_{s_0}'$ are chosen independently and uniformly at random from $\{1, \ldots, n\}$. Define $\hat{Z}_\rho' = \sum_{i=1}^{s_0}(\lambda_i' - i)$. We couple $\hat{Z}_\rho'$ to $\hat{Z}_\rho$ by setting $\hat{Z}_\rho = \hat{Z}_\rho'$

whenever all λ'_i are distinct, and otherwise setting $\hat{Z}_\rho = \sum_{i=1}^{s_0}(\lambda_i - i)$ for some distinct $\{\lambda_1, \ldots, \lambda_{s_0}\}$ chosen uniformly at random. The probability that $\lambda'_i = \lambda'_j$ for some $i \neq j$ is $O(1/n)$. (See the famous birthday problem [19, Example 3.2.5].) Since $\hat{Z}_\rho \leq s_0 n$ and $\hat{Z}'_\rho \leq s_0 n$,

$$\mathbb{E}\left[\left(\frac{\hat{Z}_\rho}{n} - \frac{\hat{Z}'_\rho}{n}\right)^2\right] \leq \mathbb{P}\left\{\exists i \neq j : \lambda'_i = \lambda'_j\right\} \frac{4s_0^2 n^2}{n^2} = O\left(\frac{1}{n}\right).$$

As $|\lambda'_i/n - U_i| \leq 1/n$, it is clear that $\hat{Z}'_\rho/n = \sum_{i=1}^{s_0}(\lambda'_i - i)/n$ converges in the second moment to $\sum_{j=1}^{s_0} U_j$. By the triangle inequality, this is also true for $\hat{Z}_\rho/n$. ☐

Since $(n_1/n, \ldots, n_b/n) \xrightarrow{\text{a.s.}} (V_1, \ldots, V_b)$ and $n_i/n \leq 1$ for all $i = 1, \ldots, b$, the convergence is also in L^2. This together with Lemma 6 and 7 implies (4.8). Therefore, it follows from Theorem 4.1 of Neininger [42] that $(\hat{X}_n, \hat{Y}_n, \hat{W}_n) \xrightarrow{d_2} (\hat{X}, \hat{Y}, \hat{W})$.

4.3 Convergence in moment generating function

To finish the proof of Theorem 4, it remains to show following lemma:

Lemma 8. *There exists a constant $L \in (0, \infty]$ such that for all fixed $t \in \mathbb{R}^3$ with $\|t\| < L$,*

$$\mathbb{E}\left[\exp\left(t \cdot (\hat{X}_n, \hat{Y}_n, \hat{W}_n)\right)\right] \to \mathbb{E}\left[\exp\left(t \cdot (\hat{X}, \hat{Y}, \hat{W})\right)\right] < \infty,$$

where $\cdot$ denotes the inner product. If we further assume that $\mathbb{P}\{\exists i : V_i = 1\} = 0$, then $L = \infty$.

Remark 12. *The condition $\mathbb{P}\{\exists i : V_i = 1\} = 0$ is necessary for $L = \infty$. Assume the opposite. By (4.7), for all $t \in \mathbb{R}$,*

$$\begin{aligned}\mathbb{E}\left[e^{t\hat{X}}\right] &\geq \mathbb{E}\left[t\left(\sum_{i=1}^b U_i + \sum_{i=1}^b V_i \hat{X}^{(i)} + \frac{s_0}{2} C(\mathcal{V})\right) \middle| \exists i : V_i = 1\right] \mathbb{P}\{\exists i : V_i = 1\} \\ &= \mathbb{E}\left[e^{t\sum_{i=1}^b U_i}\right] \mathbb{P}\{\exists i : V_i = 1\} \mathbb{E}\left[e^{t\hat{X}}\right],\end{aligned}$$

where U_i are independent Unif$[0,1]$. *This implies that $\mathbb{E}\left[e^{t\hat{X}}\right] = \infty$ if we chose t large enough such that $\mathbb{E}\left[e^{t\sum_{i=1}^b U_i}\right] \mathbb{P}\{\exists i : V_i = 1\} > 1$.*

The proofs of the next two lemmas are similar to Lemma 4.1 by Rösler [48], which deals with the total path length of binary search trees. However, we have extended the result to cover general split trees. Moreover, Lemma 10 can be applied not only to inversions and total path length, but also to any properties of split trees that satisfies the assumptions.

Lemma 9. *Let $C_1 > 0$ be a constant. There exists a constant L such that for all $t \in (-L, L)$, there exists $K_t \geq 0$ such that*

$$\mathbb{E}\left[\exp\{C_1|t| + t^2 K_t U_n\}\right] \leq 1, \qquad \text{for all} \quad n \in \mathbb{N}, \tag{4.12}$$

where

$$U_n \stackrel{\text{def}}{=} -1 + \sum_{i=1}^b \left(\frac{n_i}{n}\right)^2.$$

If we further assume that $\mathbb{P}\{\exists i : V_i = 1\} = 0$, then $L = \infty$.

Proof. Let $p = \mathbb{P}\{\exists i : V_i = 1\}$. Recalling the assumption that $p < 1$, we can choose a constant $\delta \in (0, 1-p)$. Then for ε small enough

$$\mathbb{P}\left\{-1 + \sum_{i=1}^{b} V_i^2 \geq -\varepsilon\right\} \leq \mathbb{P}\{\exists i : V_i = 1\} + \frac{\delta}{2} = p + \frac{\delta}{2}.$$

Since $U_n \xrightarrow{\text{a.s.}} -1 + \sum_{i=1}^{b} V_i^2$, there exists $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\{U_n \geq -\varepsilon\} \leq \mathbb{P}\left\{-1 + \sum_{i=1}^{b} V_i^2 \geq -\varepsilon\right\} + \frac{\delta}{2} \leq p + \delta < 1, \qquad \text{for all} \quad n \geq n_0.$$

Together with $U_n \leq 0$, the above inequality implies that for all $n \geq n_0$, $t \in (-L, L)$, and $K_t \in \mathbb{R}$,

$$\mathbb{E}\left[\mathbf{1}_{[U_n \geq -\varepsilon]} \exp\left(C_1|t| + t^2 K_t U_n\right)\right] \leq e^{C_1 L}(p + \delta) < 1, \tag{4.13}$$

if L is small enough. On the other hand, we may assume that $t \neq 0$ and then

$$\mathbb{E}\left[\mathbf{1}_{[U_n < -\varepsilon]} \exp\left(C_1|t| + t^2 K_t U_n\right)\right] \leq \exp\left(C_1|t| - t^2 K_t \varepsilon\right) < 1 - e^{C_1 L}(p + \delta), \tag{4.14}$$

if K_t is large enough. Together (4.13) and (4.14) implies (4.12). Note that if $p = 0$, then L can be arbitrarily large. □

Lemma 10. *Let $(J_n)_{n \geq 1}$ be a sequence of d-dimensional random vectors. Let $(J_n^{(i)})_{n \geq 1}$ for $i = 1, \ldots, b$ be independent copies of (J_n). Let $A_n^{(i)}$ be a diagonal matrix with n_i/n on its diagonal. Let $(B_n)_{n \geq 1}$ be a sequence of random $\mathbb{N}^b \to \mathbb{R}^d$ functions. Assume that conditioning on $\overline{n}$,*

$$J_n \stackrel{\text{d}}{=} \sum_{i=1}^{b} A_n^{(i)} J_{n_i}^{(i)} + B_n(\overline{n}).$$

Further assume that $\sup_{n \geq 1} \|B_n(\overline{n})\| < C_1$ and $\|J_1\| < C_2$ deterministically for some constants C_1, C_2 and that $s_0 > 0$. Then there exists a constant $L \in (0, \infty]$, such that for all $t \in \mathbb{R}^d$ with $\|t\| < L$, there exists $K_t \geq 0$, such that

$$\mathbb{E}[\exp(t \cdot J_n)] \leq \exp(\|t\|^2 K_t), \qquad \textit{for all} \quad n \in \mathbb{N}. \tag{4.15}$$

Moreover, if $J_n \xrightarrow{\text{d}} J^*$, then for all $t \in \mathbb{R}^d$ with $\|t\| < L$,

$$\mathbb{E}[\exp(t \cdot J_n)] \to \mathbb{E}[\exp(t \cdot J^*)] < \infty. \tag{4.16}$$

If we further assume that $\mathbb{P}\{\exists i : V_i = 1\} = 0$, then $L = \infty$.

Proof. It follows from Lemma 9 that there exists an $L \in (0, \infty]$, such that for all t with $\|t\| < L$, there exists $K_t \geq 0$, such that

$$\mathbb{E}\left[\exp\left(C_1 \|t\| + K_t \|t\|^2 U_n\right)\right] \leq 1. \tag{4.17}$$

Now we use induction on n. Since $\|J_1\| < C_2$, we can increase K_t such that (4.15) holds for $n = 1$. Assuming that it holds also for all $J_{n'}$ with $n' < n$, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(t \cdot J_n\right)\right] &= \mathbb{E}\left[\exp\left(t \cdot B_n(\overline{n}) + t \cdot \sum_{i=1}^{b} A_n^{(i)} J_{n_i}^{(i)}\right)\right] \\ &\le e^{C_1\|t\|}\mathbb{E}\left[\sum_{i=1}^{b} K_t\left(\|t\|\frac{n_i}{n}\right)^2\right] \\ &= e^{K_t\|t\|^2}\mathbb{E}\left[\exp\left(C_1\|t\| + K_t\|t\|^2 U_n\right)\right] \le e^{K_t\|t\|^2},\end{aligned}$$

where we use (4.17) and that $n_i < n$ for $i = 1 \ldots, b$ (since $s_0 > 0$). The above inequality implies that $(e^{t \cdot J_n})_{n\ge1}$, are uniformly integrable (see [26, Theorem 5.4.2]). Therefore $J_n \xrightarrow{\mathrm{d}} J^*$ implies (4.16) (see [19, Theorem 5.5.2]). □

Proof of Lemma 8. Let $J_n = (\hat{X}_n, \hat{Y}_n, \hat{W}_n)$. Then (4.3), (4.5), (4.2) can be written as

$$J_n \overset{\mathrm{d}}{=} \sum_{i=1}^{b} A_n^{(i)} J_{n_i}^{(i)} + B_n(\overline{n}),$$

where $A_n^{(i)}$ for $i = 1, \ldots, b$ are as in Lemma 10 and

$$B_n(\overline{n}) = \left[\frac{\hat{Z}_\rho}{n} + \frac{s_0}{2}\hat{D}_n(\overline{n}), \frac{\hat{Z}_\rho}{n} - \frac{s_0}{2}\frac{n - s_0}{n}, \frac{n - s_0}{n} + \hat{D}_n(\overline{n})\right]^{\mathbf{T}},$$

where $\mathbf{T}$ denotes the transposition of a matrix. By Lemma 5, J_n converges in distribution to $(\hat{X}, \hat{Y}, \hat{W})$. Note that $\|B_n(\overline{n})\|$ is bounded. Therefore Lemma 10 implies that there exists an $L \in (0, \infty]$ such that for all $t \in \mathbb{R}^3$ with $\|t\| < L$, $\mathbb{E}\left[e^{t \cdot J_n}\right] \to \mathbb{E}\left[e^{t\cdot(\hat{X},\hat{Y},\hat{W})}\right] < \infty$. □

4.4 Split tree inversions on nodes

We turn to node inversions in a split tree. The main challenge in this context is that the number N of nodes is random in general. Thus we will limit our analysis to split trees satisfying the following two assumptions

$$\frac{N}{n} \xrightarrow{L^2} \alpha, \tag{4.18}$$

and

$$\mathbb{E}\left[\Upsilon(T_n)\right] = \frac{\alpha}{\mu} n \ln n + n\varpi(\ln n) + o(n), \tag{4.19}$$

for some constant $\alpha \in (0, 1]$ and some continuous periodic function ϖ with period $d = \sup\{a \ge 0 : \mathbb{P}\{\ln V \in a\mathbb{Z}\} = 1\}$ (constant if $d = 0$), with $\mu = -\sum \mathbb{E}[V_1 \ln V_1]$.

These two conditions are satisfied for many types of split trees. Holmgren [29] showed that if $\ln V_1$ is non-lattice, i.e., $d = 0$, then $\mathbb{E}[N]/n = \alpha + o(1)$ and furthermore (4.18) holds. However, in the lattice case, Régnier and Jacquet [47] showed that, for tries (split trees with $s_0 = 0$ and $s = 1$) with a fixed split vector $(1/b, \ldots, 1/b)$, $\mathbb{E}[N]/n$ does not converge. Thus (4.18) cannot be true for these trees.

Condition (4.19) has been shown to be true for many types of split trees including m-ary search trees [5, 10, 18, 39]. More specifically, Broutin and Holmgren [8] showed that in the non-lattice case, if $\mathbb{E}[N]/n = \alpha + O(\ln^{-1-\varepsilon} n)$ for some $\varepsilon > 0$, then (4.19) is satisfied. However, Flajolet et al. [23] showed that even in the non-lattice case, there exist tries with some very special parameter values where $\mathbb{E}[n]/n - \alpha$ tends to zero arbitrarily slowly.

We have the following theorem that is similar to Theorem 4:

Theorem 6. *Assume the split tree T_n satisfies (4.18) and (4.19) and define*

$$X_n = \frac{I(T_n) - \mathbb{E}[I(T_n)]}{n}, \quad Y_n = \frac{I(T_n) - \frac{1}{2}\Upsilon(T_n)}{n}, \quad W_n = \frac{\Upsilon(T_n) - \mathbb{E}[\Upsilon(T_n)]}{n}.$$

Assume that $\mathbb{P}\{\exists i : V_i = 1\} < 1$. Let $D(\mathcal{V})$ be as in (4.6). Let (X, Y, W) be the unique solution in $\mathcal{M}_{0,2}^3$ for the system of fixed-point equations

$$\begin{bmatrix} X \\ Y \\ W \end{bmatrix} \overset{\mathrm{d}}{=} \begin{bmatrix} \sum_{i=1}^{b} V_i X^{(i)} + \alpha U_0 + \frac{\alpha}{2} D(\mathcal{V}) \\ \sum_{i=1}^{b} V_i Y^{(i)} + \alpha \left(U_0 - \frac{1}{2} \right) \\ \sum_{i=1}^{b} V_i W^{(i)} + \alpha + \alpha D(\mathcal{V}) \end{bmatrix}.$$

Here $(V_1, \ldots, V_b)$, U_0, $(X^{(1)}, Y^{(1)}, W^{(1)}), \ldots, (X^{(b)}, Y^{(b)}, W^{(b)})$ are independent, with $U_0 \sim \mathrm{Unif}[0,1]$ and $\left(X_n^{(i)}, Y_n^{(i)}, W_n^{(i)}\right) \sim (X, Y, W)$ for $i = 1, \ldots, b$. Then $(X_n, Y_n, W_n) \xrightarrow{d_2} (X, Y, W)$. If $s_0 > 0$, then the convergence is also in moment generating function within a neighborhood of the origin.

The convergence in Mallows metric again follows from Neininger [42, Theorem 4.1]. We leave the details to the reader as it is rather similar to inversions on balls. However, we emphasize that the assumption (4.19) is needed to argue that

$$D_n(\overline{n}) \overset{\mathrm{def}}{=} -\frac{\mathbb{E}[\Upsilon(T_n)]}{n} + \frac{1}{n} \sum_{i=1}^{b} \mathbb{E}[\Upsilon(T_{n_i})] \xrightarrow{L^2} \frac{\alpha}{\mu} \sum_{i=1}^{b} V_i \ln V_i = \alpha D(\mathcal{V}).$$

For convergence in moment generating function, note that $s_0 > 0$ implies $N \le n$ and $Z_\rho/n \le 1$. Therefore, we can again apply Lemma 10 as in Section 4.3.

5 A sequence of conditional Galton–Watson trees

Let ξ be a random variable with $\mathbb{E}[\xi] = 1$, $\mathrm{Var}\, \xi = \sigma^2 < \infty$, and $\mathbb{E}\left[e^{\alpha\xi}\right] < \infty$ for some $\alpha > 0$, (The last condition is used in the proof below, but is presumably not necessary.) Let G^ξ be a (possibly infinite) Galton–Watson tree with offspring distribution ξ. The *conditional Galton–Watson tree* T_n^ξ on n nodes is given by

$$\mathbb{P}\left\{T_n^\xi = T\right\} = \mathbb{P}\left\{G^\xi = T \;\middle|\; G^\xi \text{ has } n \text{ nodes}\right\}$$

for any rooted tree T on n nodes. The assumption $\mathbb{E}[\xi]=1$ is justified by noting that if ζ is such that $\mathbb{P}\{\xi=i\}=c\theta^i\mathbb{P}\{\zeta=i\}$ for all $i\geq 0$ then T_n^ξ and T_n^ζ are identically distributed; hence it is typically possible to replace an offspring distribution ζ by an equivalent one with mean 1, see [31, Sec. 4].

We fix some ξ and drop it from the notation, writing $T_n=T_n^\xi$.

In a fixed tree T with root ρ and n total nodes, for each node $v\neq\rho$ let $Q_v\sim\mathrm{Unif}(-1/2,1/2)$, all independent, and let $Q_\rho=0$. For each node v define

$$\Phi_v \stackrel{\text{def}}{=} \sum_{u\leq v} Q_u, \quad \text{and let} \quad J(T)\stackrel{\text{def}}{=}\sum_{v\in T}\Phi_v.$$

In other words, Φ_u is the sum of Q_v for all v on the path from the root to u. For each $v\neq\rho$ also define $Z_v=\lfloor (Q_v+1/2)z_v\rfloor$, where z_v denotes the size of the subtree rooted at v. Then Z_v is uniform in $\{0,1,\ldots,z_v-1\}$, and by Lemma 1, the quantity

$$I^*(T)\stackrel{\text{def}}{=}\sum_{v\neq\rho}\big(Z_v-\mathbb{E}[Z_v]\big)$$

is equal in distribution to the centralized number of inversions in the tree T, ignoring inversions involving ρ. The main part (1.13) of Theorem 5 will follow from arguing that for a conditional Galton–Watson tree T_n,

$$\frac{J(T_n)}{n^{5/4}} \xrightarrow{\text{d}} Y \stackrel{\text{def}}{=} \frac{1}{\sqrt{12\sigma}}\sqrt{\eta}\mathcal{N}. \tag{5.1}$$

Indeed, under the coupling of Q_v and Z_v above,

$$J(T_n)=\sum_v \Phi_v=\sum_v\sum_{u:u\leq v}Q_u=\sum_u Q_u\sum_{v:u\leq v}1=\sum_u Q_u z_u\leq\sum_{u\neq\rho}\Big(Z_u-\frac{z_u}{2}+1\Big)<n+I^*(T_n),$$

and similarly $J(T_n)>I^*(T_n)-n$. As ρ contributes at most n inversions to $I(T_n)$, it follows from the triangle inequality that $|J(T_n)-(I(T_n)-\Upsilon(T_n)/2)|\leq 2n=o(n^{5/4})$. Thus (5.1), once proved, will imply that

$$Y_n\stackrel{\text{def}}{=}\frac{I(T_n)-\Upsilon(T_n)/2}{n^{5/4}}=o(1)+\frac{J(T_n)}{n^{5/4}}\xrightarrow{\text{d}} Y.$$

The quantity $J(T_n)$ and the limiting distribution (5.1) have been considered by several authors. In the interest of keeping this section self-contained, we will now outline the proof of (5.1) which relies on the concept of a *discrete snake*, a random curve which under proper rescaling converges to a *Brownian snake*, a curve related to a standard Brownian excursion. This convergence was shown by Gittenberger [25], and later in more generality by Janson and Marckert [35], whose notation we use.

Define $f:\{0,\ldots,2(n-1)\}\to V$ by saying that $f(i)$ is the location of a depth-first search (under some fixed ordering of nodes) at stage i, with $f(0)=f(2(n-1))=\rho$. Also define $V_n(i)=d(\rho,f(i))$ where d denotes distance. The process $V_n(i)$ is called the depth-first walk, the Harris walk or the tour of T_n. For non-integer values t, $V_n(t)$ is given by linearly interpolating adjacent values. See Figure 1.

Finally, define $R_n(i)\stackrel{\text{def}}{=}\Phi_{f(i)}$ to be the value at the vertex visited after i steps. For non-integer values t, $R_n(t)$ is defined by linearly interpolating the integer values. Also define $\widetilde{R}_n(t)$

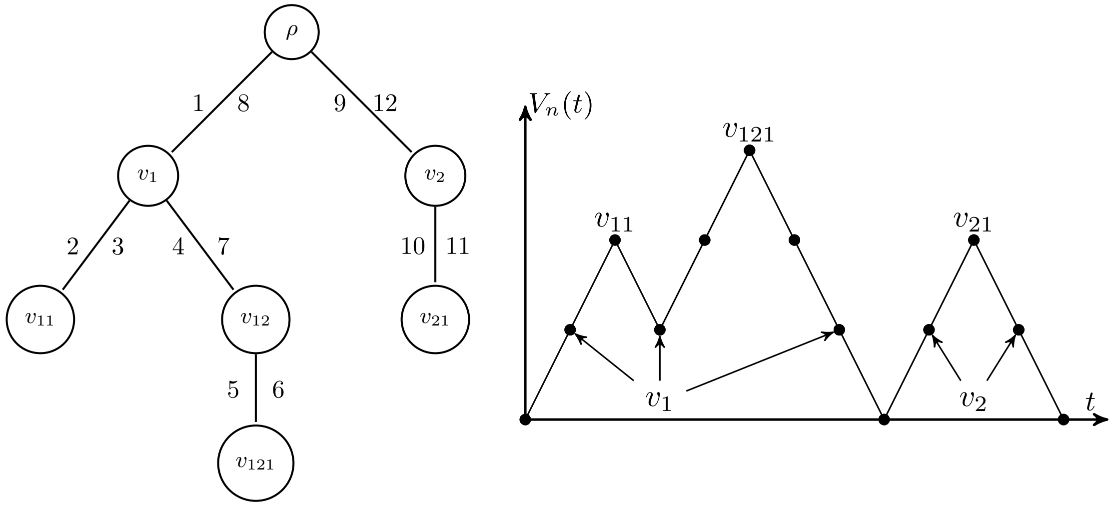

Figure 1: The depth-first walk $V_n(t)$ of a fixed tree.

by $\widetilde{R}_n(t) \stackrel{\text{def}}{=} R_n(t)$ when $t \in \{0, 1, \dots, 2n\}$, and

$$\widetilde{R}_n(t) \stackrel{\text{def}}{=} \begin{cases} R_n(\lfloor t \rfloor), & \text{if} \quad V_n(\lfloor t \rfloor) > V_n(\lceil t \rceil), \\ R_n(\lceil t \rceil), & \text{if} \quad V_n(\lfloor t \rfloor) < V_n(\lceil t \rceil). \end{cases}$$

In other words, $\widetilde{R}_n(t)$ takes the value of node $f(\lfloor t \rfloor)$ or $f(\lceil t \rceil)$, whichever is further from the root. We can recover $J(T_n)$ from $\widetilde{R}_n(t)$ via

$$2J(T_n) = \int_0^{2(n-1)} \widetilde{R}_n(t)dt.$$

Indeed, for each non-root node v there are precisely two unit intervals during which $\widetilde{R}_n(t)$ draws its value from v, namely the two unit intervals during which the parent edge of v is being traversed. Now, since $Q_v \sim \text{Unif}(-1/2, 1/2)$ we have $|R_n(i) - R_n(i-1)| \le 1/2$ for all $i > 0$ and

$$\frac{J(T_n)}{n^{5/4}} = \frac{1}{2n^{5/4}} \int_0^{2(n-1)} \widetilde{R}_n(t)\mathrm{d}t = \frac{1}{2n^{5/4}} \int_0^{2(n-1)} R_n(t)dt + O(n^{-1/4}) = \int_0^1 r_n(s)ds + o(1),$$

where $r_n(s) \stackrel{\text{def}}{=} n^{-1/4} R_n(2(n-1)s)$. Also normalize $v_n(s) \stackrel{\text{def}}{=} n^{-1/2} V_n(2(n-1)s)$. Theorem 2 of [35] (see also [25]) states that $(r_n, v_n) \xrightarrow{\text{d}} (r, v)$ in $C[0,1] \times C[0,1]$, with r, v to be defined shortly.

Before defining r and v, we will briefly motivate what they ought to be. Firstly, as the offspring distribution ξ of T_n satisfies $\mathbb{E}[\xi] = 1$, we expect the tour V_n to be roughly a random walk with zero-mean increments, conditioned to be non-negative and return to the origin at time $2(n-1)$, and the limiting law v ought to be a Brownian excursion (up to a constant scale factor). Secondly, consider a node u and the path $\rho = u_0, u_1 \dots, u_d = u$, where d is the depth of u. We can define a random walk $\Phi_u(t)$ for $t = 0, \dots, d$ by $\Phi_u(0) = 0$ and $\Phi_u(t) = \sum_{i=1}^t Q_{u_i}$ for $t > 0$, noting that $\Phi_u = \Phi_u(d)$. Under rescaling, the random walk $\Phi_u(t)$ will behave like Brownian motion. For any two nodes u_1, u_2 with last common ancestor at depth m, the processes Φ_{u_1}, Φ_{u_2} agree for $t = 0, \dots, m$, while any subsequent increments are independent. Hence $\text{Cov}(\Phi_{u_1}, \Phi_{u_2}) = cm$ for some constant $c > 0$. Now, for any $i, j \in \{0, \dots, 2(n-1)\}$, the nodes $f(i), f(j)$ at depths $V_n(i), V_n(j)$ have last common ancestor $f(k)$, where k is such that $V_n(k)$ is minimal in the range $i \le k \le j$. Hence $r(s)$ should be normally distributed with variance given by $v(s)$, and the covariance of $r(s), r(t)$ proportional to $\min_{s \le u \le t} v(u)$.

We now define r, v precisely. If Var $\xi = \sigma^2$, then $v(s) \overset{\text{def}}{=} 2\sigma^{-1} e(s)$, where $e(s)$ is a standard Brownian excursion, as shown by Aldous [3, 4]. Conditioning on v, we define r as a centered Gaussian process on $[0,1]$ with

$$\mathrm{Cov}(r(s), r(t) \mid v) = \frac{1}{12} \min_{s \le u \le t} v(u) = \frac{1}{12\sigma} C(s,t), \qquad s \le t.$$

The constant $1/12$ appears as the variance of the random increments Q_v. Again, Theorem 2 of [35] states that $(r_n, v_n) \xrightarrow{\text{d}} (r, v)$ in $C[0,1]^2$. We conclude that

$$\lim_{n \to \infty} \frac{J(T_n)}{n^{5/4}} = \int_0^1 r_n(t) dt + o(1) \xrightarrow{\text{d}} \int_0^1 r(t) dt \overset{\text{def}}{=} Y.$$

This integral is the object of study in [34], wherein it is shown that

$$Y \overset{\text{def}}{=} \int_0^1 r(t) dt \overset{\text{d}}{=} \frac{1}{\sqrt{12\sigma}} \sqrt{\eta}\, \mathcal{N},$$

where $\mathcal{N}$ is a standard normal variable, η is given by

$$\eta = \int_{[0,1]^2} C(s,t) \mathrm{d}s \ \mathrm{d}t,$$

and $\eta, \mathcal{N}$ are independent. The odd moments of Y are zero, as this is the case for $\mathcal{N}$, and by [34, Theorem 1.1], for $k \ge 0$

$$\mathbb{E}\left[Y^{2k}\right] = \frac{1}{(12\sigma)^k} \frac{(2k!)\sqrt{\pi}}{2^{(9k-4)/2} \Gamma((5k-1)/2)} a_k,$$

where $a_1 = 1$ and for $k \ge 2$,

$$a_k = 2(5k-4)(5k-6) a_{k-1} + \sum_{i=1}^{k-1} a_i a_{k-i}.$$

In particular ([34, Theorem 1.2]),

$$\mathbb{E}\left[Y^{2k}\right] \sim \frac{1}{(12\sigma)^k} \frac{2\pi^{3/2}\beta}{5} (2k)^{1/2} (10e^3)^{-2k/4} (2k)^{\frac{3}{4}\cdot 2k},$$

as $k \to \infty$, where $\beta = 0.981038\ldots$. Further analysis of the moments of η and Y, including the moment generating function and tail estimates, can be found in [34].

Remark 13. *Conditioning on the value of η, the random variable Y has variance $\eta/(12\sigma)$. The random variable η can be seen as a scaled limit of the second common path length $\Upsilon_2(T_n)$, which appeared in our earlier discussion on cumulants. Indeed, recall that $\Upsilon_2(T_n) \overset{\text{def}}{=} \sum_{u,v \in T_n} c(u,v)$, where $c(u,v)$ denotes the number of common ancestors of u, v.*

5.1 Convergence of the moment generating function

The last bit of Theorem 5 which remains to be proved is that $\mathbb{E}\left[e^{tY_n}\right] \to \mathbb{E}\left[e^{tY}\right]$ for all fixed $t \in \mathbb{R}$. Since we have already shown $Y_n \xrightarrow{\text{d}} Y$, we can apply the Vitali convergence theorem once we have shown that the sequence e^{tY_n} is uniformly integrable. This follows from the following lemma.

Lemma 11. *For all $n \in \mathbb{N}$ and $t \in \mathbb{R}$, there exist positive constants C_1 and c_1 which do not depend on n such that*

$$\mathbb{E}\left[e^{tY_n}\right] \le C_1 e^{c_1 t^4}.$$

Proof. Conditioned on T_n, we have by (1.7)

$$\mathbb{E}\left[e^{tY_n} \mid T_n\right] \le \exp\left(\frac{1}{8}\Big(\frac{t}{n^{5/4}}\Big)^2 \Upsilon_2(T_n)\right) = \exp\left(\frac{t^2}{8} \cdot \frac{\Upsilon_2(T_n)}{n^{5/2}}\right).$$

By (1.3), we have

$$\Upsilon_2(T_n) = \sum_{u,v \in T_n} c(u,v) \le n^2(H_n + 1),$$

where H_n denotes the height of T_n. It follows that

$$\mathbb{E}\left[e^{tY_n}\right] \le \mathbb{E}\left[\exp\left(\frac{\Upsilon_2(T_n)}{n^{5/2}} t^2\right)\right] \le \mathbb{E}\left[\exp\left(\frac{H_n + 1}{\sqrt{n}} t^2\right)\right] \le e^{t^2}\mathbb{E}\left[\exp\left(\frac{H_n}{\sqrt{n}} t^2\right)\right].$$

The random variable H_n has been well-studied. In particular, Addario-Berry et al. [1] showed that there exist positive constants C_2 and c_2 such that

$$\mathbb{P}\left\{H_n > x\right\} \le C_2 \exp\left(-c_2 \frac{x^2}{n}\right),$$

for all $n \in \mathbb{N}$ and $x \ge 0$. Therefore, we have

$$\mathbb{E}\left[\exp\left(\frac{H_n}{\sqrt{n}} t^2\right)\right] = 1 + \int_0^\infty e^x \mathbb{P}\left\{\frac{H_n}{\sqrt{n}} t^2 > x\right\} \mathrm{d}x \le 1 + \int_0^\infty e^x C_2 \exp\left(-c_2 \frac{x^2}{t^4}\right) \mathrm{d}x \le 1 + C_1 t^2 e^{c_3 t^4}$$

for some positive constants c_3 and C_1. (For the equality in the above computation, see [19, pp. 56].) Thus the lemma follows. ☐

References

[1] L. Addario-Berry, L. Devroye, and S. Janson. Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees. *Ann. Probab.*, 41(2):1072–1087, 2013.

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

[3] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

[4] D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[5] R. A. Baeza-Yates. Some average measures in m-ary search trees. *Inform. Process. Lett.*, 25 (6):375–381, 1987.

[6] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *J. Combinatorial Theory Ser. A*, 15:91–111, 1973.

[7] I. J. Bienaymé. De la loi de multiplication et de la durée des familles. *Société Philomatique Paris*, 1845. Reprinted in Kendall (1975).

[8] N. Broutin and C. Holmgren. The total path length of split trees. *Ann. Appl. Probab.*, 22(5): 1745–1777, 2012.

[9] X. S. Cai and L. Devroye. A study of large fringe and non-fringe subtrees in conditional Galton-Watson trees. *Latin American Journal of Probability and Mathematical Statistics*, XIV:579–611, 2017.

[10] B. Chauvin and N. Pouyanne. m-ary search trees when $m \geq 27$: a strong asymptotics for the space requirements. *Random Structures Algorithms*, 24(2):133–154, 2004.

[11] E. G. Coffman, Jr. and J. Eve. File structures using hashing functions. *Commun. ACM*, 13 (7):427–432, July 1970. doi: 10.1145/362686.362693.

[12] K. Conrad. Probability distributions and maximum entropy. http://www.math.uconn.edu/~kconrad/blurbs/analysis/entropypost.pdf. Accessed: 2017-05-26.

[13] G. Cramer. *Introduction à l'analyse des lignes courbes algébriques*. 1750.

[14] L. Devroye. On the expected height of fringe-balanced trees. *Acta Inform.*, 30(5):459–466, 1993.

[15] L. Devroye. Universal limit laws for depths in random trees. *SIAM J. Comput.*, 28(2):409–432, 1999.

[16] L. Devroye, C. Holmgren, and H. Sulzbach. The heavy path approach to Galton-Watson trees with an application to Apollonian networks. *ArXiv e-prints*, Jan. 2017.

[17] DLMF. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.15 of 2017-06-01. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[18] M. Drmota, A. Iksanov, M. Moehle, and U. Roesler. A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures Algorithms*, 34(3):319–336, 2009.

[19] R. Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 4th edition, 2010. ISBN 978-0-521-76539-8.

[20] W. Feller. *An introduction to probability theory and its applications. Vol. I.* John Wiley & Sons, Inc., New York-London-Sydney, 3rd edition, 1968.

[21] R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, 1974. doi: 10.1007/BF00288933.

[22] P. Flajolet, P. Poblete, and A. Viola. On the analysis of linear probing hashing. *Algorithmica*, 22(4):490–515, 1998.

[23] P. Flajolet, M. Roux, and B. Vallée. Digital trees and memoryless sources: from arithmetics to analysis. In *21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10)*, Discrete Math. Theor. Comput. Sci. Proc., AM, pages 233–260. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010.

[24] I. M. Gessel, B. E. Sagan, and Y. N. Yeh. Enumeration of trees by inversions. *J. Graph Theory*, 19(4):435–459, 1995.

[25] B. Gittenberger. A note on: "State spaces of the snake and its tour—convergence of the discrete snake". *J. Theoret. Probab.*, 16(4):1063–1067 (2004), 2003.

[26] A. Gut. *Probability: a graduate course.* Springer Texts in Statistics. Springer, New York, 2nd edition, 2013.

[27] C. A. R. Hoare. Quicksort. *Comput. J.*, 5:10–15, 1962.

[28] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[29] C. Holmgren. Novel characteristic of split trees by use of renewal theory. *Electron. J. Probab.*, 17:no. 5, 27, 2012.

[30] S. Janson. The Wiener index of simply generated random trees. *Random Structures Algorithms*, 22(4):337–358, 2003.

[31] S. Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation: extended abstract. In *23rd Intern. Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA'12)*, Discrete Math. Theor. Comput. Sci. Proc., AQ, pages 479–490. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2012.

[32] S. Janson. Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Structures Algorithms*, 48(1):57–101, 2016.

[33] S. Janson. Random recursive trees and preferential attachment trees are random split trees. *ArXiv e-prints*, June 2017.

[34] S. Janson and P. Chassaing. The center of mass of the ISE and the Wiener index of trees. *Electron. Comm. Probab.*, 9:178–187, 2004.

[35] S. Janson and J.-F. Marckert. Convergence of discrete snakes. *J. Theoret. Probab.*, 18(3): 615–647, 2005.

[36] D. E. Knuth. *The art of computer programming. Vol. 3.* Addison-Wesley, Reading, MA, 1998. ISBN 0-201-89685-0.

[37] I. Kortchemski. Sub-exponential tail bounds for conditioned stable Bienaymé-Galton-Watson trees. *Probab. Theory Related Fields*, 168(1-2):1–40, 2017.

[38] G. Louchard and H. Prodinger. The number of inversions in permutations: a saddle point approach. *J. Integer Seq.*, 6(2):Article 03.2.8, 19, 2003.

[39] H. M. Mahmoud and B. Pittel. Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms*, 10(1):52–75, 1989.

[40] C. L. Mallows and J. Riordan. The inversion enumerator for labeled trees. *Bull. Amer. Math. Soc.*, 74:92–94, 1968.

[41] B. H. Margolius. Permutations with inversions. *J. Integer Seq.*, 4(2):Article 01.2.4, 13, 2001.

[42] R. Neininger. On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms*, 19(3-4):498–524, 2001. Analysis of algorithms (Krynica Morska, 2000).

[43] R. Neininger and L. Rüschendorf. On the internal path length of d-dimensional quad trees. *Random Structures Algorithms*, 15(1):25–41, 1999.

[44] R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14(1):378–418, 2004.

[45] A. Panholzer and G. Seitz. Limiting distributions for the number of inversions in labelled tree families. *Ann. Comb.*, 16(4):847–870, 2012.

[46] R. Pyke. Spacings. (With discussion.). *J. Roy. Statist. Soc. Ser. B*, 27:395–449, 1965.

[47] M. Régnier and P. Jacquet. New results on the size of tries. *IEEE Trans. Inform. Theory*, 35 (1):203–205, 1989.

[48] U. Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.

[49] U. Rösler. On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, 29(1-2): 238–261, 2001.

[50] U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001.

[51] V. N. Sachkov. *Probabilistic methods in combinatorial analysis*, volume 56 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1997. ISBN 0-521-45512-X. Translated from the Russian, Revised by the author.

[52] P. J. Smith. A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *Amer. Statist.*, 49(2):217–218, 1995.

[53] A. Walker and D. Wood. Locally balanced binary trees. *The Computer Journal*, 19(4):322–325, 1976.

[54] H. W. Watson and F. Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.