Inversions in Split Trees and Conditional Galton–Watson Trees

Xing Shi Cai

Department of Mathematics, Uppsala University, Uppsala, Sweden
xingshi.cai@math.uu.se
https://orcid.org/0000-0002-3768-8078

Cecilia Holmgren

Department of Mathematics, Uppsala University, Uppsala, Sweden
cecilia.holmgren@math.uu.se
https://orcid.org/0000-0003-0717-4671

Svante Janson

Department of Mathematics, Uppsala University, Uppsala, Sweden
svante.janson@math.uu.se
https://orcid.org/0000-0002-9680-2790

Tony Johansson

Department of Mathematics, Uppsala University, Uppsala, Sweden
tony.johansson@math.uu.se

Fiona Skerman

Department of Mathematics, Uppsala University, Uppsala, Sweden
fiona.skerman@math.uu.se
https://orcid.org/0000-0003-4141-7059

—— **Abstract** ——————————————————————————————

We study $I(T)$, the number of inversions in a tree T with its vertices labeled uniformly at random. We first show that the cumulants of $I(T)$ have explicit formulas. 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. For T_n being split trees [6], 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. Our results generalize and extend previous work by Panholzer and Seitz [20].

1   Introduction

1.1   Inversions in a fixed tree

Let $\sigma_1, \ldots, \sigma_n$ be a permutation of $\{1, \ldots, n\}$. 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 [5] (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 ([15, Section 5.1]). Many authors, e.g., Feller [7, pp. 256], have shown that the number of inversions in uniform random permutations has a central limit theorem.

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 $\lambda : V \to \{1, \ldots, |V|\}$ (a *node labeling*), define the number of *inversions*

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

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 $\{1, \ldots, |V|\}$.

The enumeration of trees with a fixed number of inversions has been studied by Mallows and Riordan [16] and Gessel et al. [9] using the so called *inversions polynomial*. While analyzing linear probing hashing, Flajolet et al. [8] noticed that the numbers of inversions in Cayley trees with uniform random labeling converges to an Airy distribution. Panholzer and Seitz [20] 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 $I(T)$ is the basis of most of our results:

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

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

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

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,

$$\mathbb{E}[I(T)] = \sum_{u < v} \mathbb{E}\left[\mathbf{1}_{\lambda(u) > \lambda(v)}\right] = \frac{1}{2}\Upsilon(T), \tag{1.1}$$

where $\Upsilon(T) \overset{\text{def}}{=} \sum_v h(v)$ 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 $\varkappa_2(X) = \mathrm{Var}(X)$. We now define $\Upsilon_k(T)$, *the k-total common ancestors* of T, which allows us to generalize (1.1) to higher cumulants of $I(T)$. For k nodes $v_1, \ldots, v_k$ (not necessarily distinct), let $c(v_1, \ldots, v_k)$ be the number of ancestors that they share, i.e.,

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

We define

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

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. Using Lemma 1, it is easy to show the following:

▶ **Theorem 2.** *Let T be a fixed tree. Let $\varkappa_k(I(T))$ be the k-th cumulant of $I(T)$. Then*

$$\mathbb{E}[I(T)] = \varkappa_1(I(T)) = \frac{1}{2}\Upsilon(T) = \frac{1}{2}(\Upsilon_1(T) - |V|),$$

$$\mathrm{Var}(I(T)) = \varkappa_2(I(T)) = \frac{1}{12}(\Upsilon_2(T) - |V|),$$

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

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

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

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

and for the centralized variable we have the estimate

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

1.2 Inversions in sequences of trees

The total path length $\Upsilon(T)$ has been studied for random trees like split trees [3] and conditional Galton–Watson trees [1, Corollary 9]. This leads us to focus on the deviation

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

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

Fixed trees

The following theorem follows easily from Theorem 2:

▶ **Theorem 3.** *Let T_n be a sequence of fixed trees on n nodes. Let*

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

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

▶ **Example 4.** 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 3, X_n converges to a normal distribution, and we recover the central limit law for inversions in permutations. Also, the vertices have subtree sizes $1, \ldots, n$ and so we also recover from Theorem 2 the moment generating function $\prod_{j=1}^{n}(e^{jt} - 1)/(j(e^t - 1))$ [22, 17].

Other examples where Theorem 2 can be easily applied include complete b-ary trees and stars (a star is a tree containing only a root and leaves).

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 (1.1) 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

$$X_n = Y_n + W_n/2, \tag{1.2}$$

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.

Split trees

The first class of random trees which we study are split trees. They were introduced by Devroye [6] to encompass many families of trees that are frequently used in algorithm analysis, e.g., binary search trees, m-ary search trees, digital search trees, etc.

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* $\mathcal{V} = (V_1, \ldots, V_b)$ 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 2. Any node u such that the subtree rooted at 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

$$\hat{X}_n = \frac{\hat{I}(T_n) - \mathbb{E}\left[\hat{I}(T_n)\right]}{n}, \quad \hat{Y}_n = \frac{\hat{I}(T_n) - s_0 \hat{\Upsilon}(T_n)/2}{n}, \quad \hat{W}_n = \frac{\hat{\Upsilon}(T_n) - \mathbb{E}\left[\hat{\Upsilon}(T_n)\right]}{n}. \quad (1.3)$$

Here s_0 is a fixed integer denoting the number of balls in any internal node, and we have $\hat{X}_n = \hat{Y}_n + s_0 \hat{W}_n/2$ (justified in Section 2). The following theorem gives the limiting distributions of the random vector $(\hat{X}_n, \hat{Y}_n, \hat{W}_n)$. In a longer version of this paper [4], we also have 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 and digital search trees.

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

▶ **Theorem 5.** *Let T_n be a split tree and let $\mathcal{V} = (V_1, \ldots, V_b)$ be a split vector. Define*

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

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 $\mathcal{M}_{0,2}^3$ for the system of fixed-point equations

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{W} \end{bmatrix} \overset{\mathrm{d}}{=} \begin{bmatrix} \displaystyle\sum_{i=1}^b V_i \hat{X}^{(i)} + \sum_{j=1}^{s_0} U_j + \frac{s_0}{2} D(\mathcal{V}) \\ \displaystyle\sum_{i=1}^b V_i \hat{Y}^{(i)} + \sum_{j=1}^{s_0} (U_j - 1/2) \\ \displaystyle\sum_{i=1}^b V_i \hat{W}^{(i)} + 1 + D(\mathcal{V}) \end{bmatrix}. \quad (1.4)$$

Here $(V_1, \ldots, V_b)$, $U_1, \ldots, U_{s_0}$, $(\hat{X}^{(1)}, \hat{Y}^{(1)}, \hat{W}^{(1)}), \ldots, (\hat{X}^{(b)}, \hat{Y}^{(b)}, \hat{W}^{(b)})$ are independent, with $U_j \sim \mathrm{Unif}[0,1]$ for $j = 1, \ldots, s_0$, and $\left(\hat{X}_n^{(i)}, \hat{Y}_n^{(i)}, \hat{W}_n^{(i)}\right) \sim (\hat{X}, \hat{Y}, \hat{W})$ for $i = 1, \ldots, b$. Then the sequence $(\hat{X}_n, \hat{Y}_n, \hat{W}_n)$ defined in (1.3) 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 5 uses the contraction method, introduced by Rösler [21] for finding the total path length of binary search trees. The technique has been applied to d-dimensional quad trees by Neininger and Rüschendorf [19] and to split trees in general by Broutin and Holmgren [3].

Conditional Galton–Watson trees

A conditional Galton–Watson tree T_n is a Galton–Watson tree conditioned on having n nodes, which we define in details in Section 3. 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 [12].

Aldous [1] showed that 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. 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 [14].

In the context of Galton–Watson trees, Aldous [1, 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 (1.2) implies that $n^{-3/2}I(T_n)$ converges to the same Airy distribution, recovering one of the main results of Panholzer and Seitz [20, 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 $C(s,t) \overset{\text{def}}{=} C(t,s) \overset{\text{def}}{=} 2\min_{s \le u \le t} e(u)$ for $0 \le s \le t \le 1$.

We define a random variable, see [11],

$$\eta \overset{\text{def}}{=} \int_{[0,1]^2} C(s,t)\mathrm{d}s\ \mathrm{d}t = 4\int_{0 \le s \le t \le 1} \min_{s \le u \le t} e(u). \tag{1.5}$$

▶ **Theorem 6.** *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 $\mathbb{E}[e^{\alpha\xi}] < \infty$ for some $\alpha > 0$, and define*

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

Then we have

$$Y_n \overset{\mathrm{d}}{\longrightarrow} Y \overset{\text{def}}{=} \frac{1}{\sqrt{12\sigma}}\sqrt{\eta}\ \mathcal{N}, \tag{1.6}$$

where $\mathcal{N}$ is a standard normal random variable, independent from the random variable η defined in (1.5). Moreover, $\mathbb{E}[e^{tY_n}] \to \mathbb{E}[e^{tY}] < \infty$ for all fixed $t \in \mathbb{R}$.

In the rest of the paper, we outline the proofs of our main results, Theorem 5 and 6. The proofs of Theorem 2 and 3 are omitted. The details of the proofs can be found in the longer version of this paper [4].

2 A sequence of split trees

In this section we outline how one can apply the *contraction method* to prove Theorem 5.

We will now define split trees introduced by Devroye [6]. 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

$$2 \le b, \quad 0 < s, \quad 0 \le s_0 \le s, \quad 0 \le bs_1 \le s + 1 - s_0. \tag{2.1}$$

and $\mathcal{V} = (V_1, \ldots, V_b)$ is a random non-negative vector with $\sum_{i=1}^b V_i = 1$. 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 $\mathcal{V}_u = (V_{u,1}, \ldots, V_{u,b})$ be the independent copy of $\mathcal{V}$ assigned to u. Let $u_1, \ldots, u_b$ be the child nodes of u. Conditioning on n_u and $\mathcal{V}_u$, if $n_u \leq s$, then $n_{u_i} = 0$ for all i; if $n_u > s$, then

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

where Mult denotes multinomial distribution, and b, s, s_0, s_1 are integers satisfying (2.1). Note that $\sum_{i=1}^b n_{u_i} \leq n$ (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.

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 $\{1, \ldots, s\}$. We assume, as previous authors, that $\mathbb{P}\{\exists i : V_i = 1\} < 1$. We can assume that $\mathcal{V}$ has a permutation invariant distribution without loss of generality, since a uniform random permutation of subtree order does not change the number of inversions.

2.1 Outline

Recall that in (1.3), we define $\hat{X}_n, \hat{Y}_n$ and $\hat{W}_n$. Let $\bar{n} = (n_1, \ldots, n_b)$ denote the vector of the (random) number of balls in each of the b subtrees of the root. Broutin and Holmgren [3] showed that, conditioning on $\bar{n}$,

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

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

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

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

$$
\hat{X}_n \overset{\mathrm{d}}{=} \sum_{i=1}^b \frac{n_i}{n} \hat{X}_{n_i} + \frac{\hat{Z}_\rho}{n} - \frac{\mathbb{E}\left[\hat{I}(T_n)\right]}{n} + \sum_{i=1}^b \frac{\mathbb{E}\left[\hat{I}(T_{n_i})\right]}{n}
$$

$$
= \sum_{i=1}^b \frac{n_i}{n} \hat{X}_{n_i} + \frac{\hat{Z}_\rho}{n} - \frac{s_0}{2} \frac{\mathbb{E}\left[\hat{\Upsilon}(T_n)\right]}{n} + \frac{s_0}{2} \sum_{i=1}^b \frac{\mathbb{E}\left[\hat{\Upsilon}(T_{n_i})\right]}{n}
$$

$$
= \sum_{i=1}^b \frac{n_i}{n} \hat{X}_{n_i} + \frac{\hat{Z}_\rho}{n} + \frac{s_0}{2} \hat{D}_n(\bar{n}),
$$

where we use that

$$\mathbb{E}\left[\hat{I}(T_n) \mid T_n\right] = \frac{s_0}{2} \hat{\Upsilon}(T_n). \tag{2.2}$$

It follows also from (2.2) that $\hat{X}_n = \hat{Y}_n + \frac{s_0}{2} \hat{W}_n$ and

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

It is not difficult to see 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 [3] have shown that $\hat{D}_n(\overline{n}) \xrightarrow{\text{a.s.}} D(\mathcal{V})$, where

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

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 fixed-point equations (1.4) presented in Theorem 5.

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 [3] proved that $\hat{W}_n \xrightarrow{d_2} \hat{W}$, the unique solution of the last equation of (1.4) 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 (1.4) in $\mathcal{M}_{0,2}^3$. 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 (1.4). Using a coupling argument, we can show that for all $\nu, \lambda \in \mathcal{M}_{0,2}^3$,

$$d_2(F(\nu), F(\lambda)) < cd_2(\nu, \lambda),$$

where $c \in (0, 1)$ is a constant. Thus F is a contraction and by Banach's fixed point theorem, (1.4) 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 $(\hat{X}_n, \hat{Y}_n, \hat{W}_n) \xrightarrow{d_2} (\hat{X}, \hat{Y}, \hat{W})$.

Note that in [4], instead of carrying out the above argument in details, we actually used a result by Neininger [18] which gives us a shortcut.

3   A sequence of conditional Galton–Watson trees

Let ξ be a random variable with $\mathbb{E}[\xi] = 1$, $\text{Var } \xi = \sigma^2 < \infty$, and $\mathbb{E}[e^{\alpha\xi}] < \infty$ 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

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

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 [12, 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 $Q_v \sim \text{Unif}(-1/2, 1/2)$, 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 (1.6) of Theorem 6 will follow from arguing that for a conditional Galton–Watson tree T_n,

$$\frac{J(T_n)}{n^{5/4}} \xrightarrow{\mathrm{d}} Y \stackrel{\text{def}}{=} \frac{1}{\sqrt{12}\sigma} \sqrt{\eta} \mathcal{N}. \tag{3.1}$$

Indeed, under the coupling of Q_v and Z_v above,

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

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 (3.1), once proved, will imply that

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

The quantity $J(T_n)$ and the limiting distribution (3.1) have been considered by several authors. In the interest of keeping this section self-contained, we will now outline the proof of (3.1) 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 [10], and later in more generality by Janson and Marckert [14], 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 $\widetilde{R}_n(t)$ by $\widetilde{R}_n(t) \stackrel{\text{def}}{=} R_n(t)$ when $t \in \{0, 1, \ldots, 2n\}$, and

$$\widetilde{R}_n(t) \stackrel{\text{def}}{=} \begin{cases} R_n(\lfloor t \rfloor), & \text{if} \quad V_n(\lfloor t \rfloor) > V_n(\lceil t \rceil), \\ R_n(\lceil t \rceil), & \text{if} \quad V_n(\lfloor t \rfloor) < V_n(\lceil t \rceil). \end{cases}$$
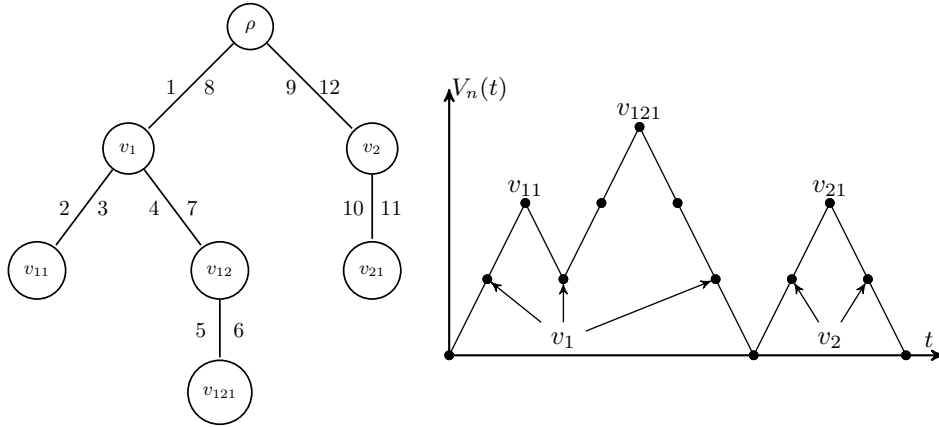
Figure 1 The depth-first walk $V_n(t)$ of a fixed tree.

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 $|R_n(i) - R_n(i-1)| \leq 1/2$ 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 [14] (see also [10]) states that $(r_n, v_n) \stackrel{\mathrm{d}}{\longrightarrow} (r, v)$ 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 $\rho = u_0, u_1 \ldots, u_d = u$, where d is the depth of u. We can define a random walk $\Phi_u(t)$ for $t = 0, \ldots, d$ 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 $t = 0, \ldots, m$, while any subsequent increments are independent. Hence $\text{Cov}(\Phi_{u_1}, \Phi_{u_2}) = cm$ for some constant $c > 0$. Now, for any $i, j \in \{0, \ldots, 2(n-1)\}$, 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 $i \leq k \leq j$. Hence $r(s)$ should be normally distributed with variance given by $v(s)$, and the covariance of $r(s), r(t)$ proportional to $\min_{s \leq u \leq t} v(u)$.

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

Gaussian process on $[0, 1]$ with

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

The constant $1/12$ appears as the variance of the random increments Q_v. Again, Theorem 2 of [14] states that $(r_n, v_n) \xrightarrow{\text{d}} (r, v)$ in $C[0, 1]^2$. We conclude that

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

This integral is the object of study in [13], wherein it is shown that

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

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 [13, 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 ([13, 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 $\beta = 0.981038\dots$. Further analysis of the moments of η and Y, including the moment generating function and tail estimates, can be found in [13].

The last bit of Theorem 6 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. See Section 5.1 of [4] for details.

───── **References** ─────

1 David 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.

2 David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

3 Nicolas Broutin and Cecilia Holmgren. The total path length of split trees. *Ann. Appl. Probab.*, 22(5):1745–1777, 2012.

4 Xing Shi Cai, C. Holmgren, S. Janson, T. Johansson, and F. Skerman. Inversions in split trees and conditional Galton–Watson trees. *ArXiv e-prints*, 2017. `arXiv:1709.00216`.

5 G. Cramer. *Introduction à l'analyse des lignes courbes algébriques.* Cramer, 1750.

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

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

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

9 Ira M. Gessel, Bruce E. Sagan, and Yeong Nan Yeh. Enumeration of trees by inversions. *J. Graph Theory*, 19(4):435–459, 1995.

10 Bernhard 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.

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

12 Svante 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.

13 Svante Janson and Philippe Chassaing. The center of mass of the ISE and the Wiener index of trees. *Electron. Comm. Probab.*, 9:178–187, 2004.

14 Svante Janson and Jean-François Marckert. Convergence of discrete snakes. *J. Theoret. Probab.*, 18(3):615–647, 2005.

15 Donald E. Knuth. *The art of computer programming. Vol. 3.* Addison-Wesley, Reading, MA, 1998.

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

17 Barbara H. Margolius. Permutations with inversions. *J. Integer Seq.*, 4(2):Article 01.2.4, 13, 2001.

18 Ralph 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).

19 Ralph Neininger and Ludger Rüschendorf. On the internal path length of d-dimensional quad trees. *Random Structures Algorithms*, 15(1):25–41, 1999.

20 Alois Panholzer and Georg Seitz. Limiting distributions for the number of inversions in labelled tree families. *Ann. Comb.*, 16(4):847–870, 2012.

21 Uwe Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.

22 Vladimir N. Sachkov. *Probabilistic methods in combinatorial analysis*, volume 56 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1997. Translated from the Russian, Revised by the author.