MAXIMAL CLADES IN RANDOM BINARY SEARCH TREES

SVANTE JANSON

Abstract. We study maximal clades in random phylogenetic trees with the Yule–Harding model or, equivalently, in binary search trees. We use probabilistic methods to reprove and extend earlier results on moment asymptotics and asymptotic normality. In particular, we give an explanation of the curious phenomenon observed by Drmota, Fuchs and Lee (2014) that asymptotic normality holds, but one should normalize using half the variance.

1. Introduction

Recall that there are two types of binary trees; we fix the notation as follows. A full binary tree is an rooted tree where each node has either 0 or 2 children; in the latter case the two children are designated as left child and right child. A binary tree is a rooted tree where each node has 0, 1 or 2 children; moreover, each child is designated as either left child or right child, and each node has at most one child of each type. (Both versions can be regarded as ordered trees, with the left child before the right when there are two children.) It is convenient to regard also the empty tree $\emptyset$ as a binary tree (but not as a full binary tree). In a full binary tree, the leaves (nodes with no children) are called external nodes; the other nodes (having 2 children) are internal nodes. There is a simple, well-known bijection between full binary trees and binary trees: Given a full binary tree, its internal nodes form a binary tree; this is a bijection, with inverse given by adding, to any given binary tree, external nodes as children at all free places.

Note that a full binary tree with $n$ internal nodes has $n+1$ external nodes, and thus $2n+1$ nodes in total. In particular, the bijection just described yields a bijection between the full binary trees with $2n+1$ nodes and the binary trees with $n$ nodes.

If $T$ is a binary, or full binary, tree, we let $T_L$ and $T_R$ be the subtrees rooted at the left and right child of the root, with $T_L = \emptyset$ [$T_R = \emptyset$] if the root has no left [right] child.

Date: 27 August, 2014.
2010 Mathematics Subject Classification. 60C05; 05C05, 60F05.
Partly supported by the Knut and Alice Wallenberg Foundation.
This research was started during the 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, AofA 2014, in Paris, June 2014. I thank Michael Fuchs for interesting discussions.
A phylogenetic tree is the same as a full binary tree. In this context, the clade of an external node \( v \) is defined to be the set of external nodes that are descendants of the parent of \( v \). (This is called a minimal clade by Blum and François [3] and Chang and Fuchs [6].) Note that two clades are either nested or disjoint; furthermore, each external node belongs to some clade (for example its own). Hence, the set of maximal clades forms a partition of the set of external nodes. We let \( F(T) \) denote the number of maximal clades of a phylogenetic tree \( T \). (Except that for technical reasons, see Section 2, we define \( F(T) = 0 \) for a phylogenetic tree \( T \) with only one external node. Obviously, this does not affect asymptotics.) The maximal clades, and the number of them, were introduced by Durand, Blum and François [11], together with a biological motivation, and further studied by Drmota, Fuchs and Lee [10].

The phylogenetic trees that we consider are random; more precisely, we consider the Yule–Harding model of a random phylogenetic tree \( \tilde{T}_n \) with a given number \( n \) internal, and thus \( n+1 \) external, nodes. These can be defined recursively, with \( \tilde{T}_0 \) the unique phylogenetic tree with 1 node (the root), and \( \tilde{T}_{n+1} \) obtained from \( \tilde{T}_n \) \( (n \geq 0) \) by choosing an external node uniformly at random and converting it to an internal node with two external children. (Alternatively, we obtain the same random model by constructing the tree bottom-up by Kingman’s coalescent [17], see further Aldous [2], Blum and François [3] and Chang and Fuchs [6].) Recall that, for any \( n \geq 1 \), the number of internal nodes in the left subtree \( \tilde{T}_{n,L} \) (or the right subtree \( \tilde{T}_{n,R} \)) is uniformly distributed on \( \{0, \ldots, n-1\} \), and that conditioned on this number being \( m \), \( \tilde{T}_{n,L} \) has the same distribution as \( \tilde{T}_m \); see also Remark 5.1.

Under the bijection above, the Yule–Harding random tree \( \tilde{T}_n \) corresponds to the random binary search tree \( T_n \) with \( n \) nodes, see e.g. Blum, François and Janson [4] and Drmota [9].

The random variable that we study is thus \( X_n := F(\tilde{T}_n) \), the number of maximal clades in the Yule–Harding model. It was proved by Durand and François [12] that the mean number of maximal clades \( \mathbb{E} X_n \sim \alpha n \), where

\[
\alpha = \frac{1 - e^{-2}}{4}.
\] (1.1)

This was reproved by Drmota, Fuchs and Lee [10], in a sharper form:

**Theorem 1.1** ([12; 10]).

\[
\mathbb{E} X_n = \mathbb{E} F(\tilde{T}_n) = \alpha n + O(1),
\] (1.2)

where \( \alpha \) is given by (1.1).

Moreover, Drmota, Fuchs and Lee [10] found also corresponding results for the variance and higher central moments:

**Theorem 1.2** ([10]). As \( n \to \infty \),

\[
\mathbb{E}(X_n - \mathbb{E} X_n)^2 \sim 4\alpha^2 n \log n,
\] (1.3)
and for any fixed integer $k \geq 3$,

$$E(X_n - \mathbb{E} X_n)^k \sim (-1)^k \frac{2k}{k-2} \alpha^k n^{k-1}. \quad (1.4)$$

As a consequence of (1.3)–(1.4), the limit distribution of $F(\bar{T}_n)$ (after centering and normalization) cannot be found by the method of moments. Nevertheless, [10] further proved asymptotic normality, where, unusually, the normalizing uses (the square root of) half the variance:

**Theorem 1.3** ([10]). As $n \to \infty$,

$$\frac{X_n - \mathbb{E} X_n}{\sqrt{2\alpha^2 n \log n}} \xrightarrow{d} N(0,1). \quad (1.5)$$

Here and below, $\xrightarrow{d}$ denotes convergence in distribution; similarly, $\xrightarrow{p}$ will denotes convergence in probability. Unspecified limits (including implicit ones such as $\sim$ and $o(1)$) will be as $n \to \infty$. Furthermore, $Y_p = o_p(a_n)$, for random variables $Y_n$ and positive numbers $a_n$, means $Y_n/a_n \xrightarrow{p} 0$. We let $C, C_1, C_2, \ldots$ denote some unspecified positive constants.

The purpose of the present paper is to use probabilistic methods to reprove these theorems, together with some further results; we hope that this can give additional insight, and it might perhaps also suggest future generalizations to other types of random trees.

In particular, we can explain the appearance of half the variance in Theorem 1.3 as follows:

Fix a sequence of numbers $N = N(n)$, and say that a clade is small if it has at most $N + 1$ elements, and large otherwise. (We use $N + 1$ in the definition only for later notational convenience; the subtree corresponding to a small clade has at most $N$ internal nodes.) Let $X_n^N$ be the number of maximal small clades, i.e., the small clades that are not contained in any other small clade. It turns out that a suitable choice of $N$ is about $\sqrt{n}$; we give two versions in the next theorem.

**Theorem 1.4.** (i) Let $N := \sqrt{n}$. Then $\text{Var}(X_n^N) \sim 2\alpha^2 n \log n$ and

$$\frac{X_n^N - \mathbb{E} X_n^N}{\sqrt{\text{Var} X_n^N}} \xrightarrow{d} N(0,1). \quad (1.6)$$

Furthermore, $X_n - X_n^N = o_p(\sqrt{\text{Var} X_n^N})$ and $\mathbb{E} X_n - \mathbb{E} X_n^N = o(\sqrt{\text{Var} X_n^N})$, so we may replace $X_n^N$ by $X_n$ in the numerator of (1.6). However,

$$\text{Var}(X_n - X_n^N) \sim \text{Var}(X_n^N) \sim 2\alpha^2 n \log n. \quad (1.7)$$

(ii) Let $\sqrt{n} \ll N \ll \sqrt{n \log n}$, for example $N := n \log \log n$. Then the conclusions of (i) still hold; moreover, $\mathbb{P}(X_n \neq X_n^N) \to 0$.

The theorem thus shows that the large clades are rare, and do not contribute to the asymptotic distribution; however, when they appear, the large clades give a large (actually negative) contribution to $X_n$, and as
a result, half the variance of $X_n$ comes from the large clades. (When there is a large clade, there is less room for other clades, so $X_n$ tends to be smaller than usually. See also (2.4) and (2.2) below.)

For higher moments, the large clades play a similar, but even more extreme, role. Note that (for $n \geq 2$) with probability $2/n$, the root of $\bar{T}_n$ has one internal and one external node, and then there is a clade consisting of all external nodes; this is obviously the unique maximal clade, and thus $X_n = 1$. Since $E X_n = \alpha n + O(1)$ by Theorem 1.1, we thus have $X_n - E X_n = -\alpha n + O(1)$ with probability $2/n$, and this single exceptional event gives a contribution $\sim (-1)^k 2\alpha^k n^{k-1}$ to $E(X_n - E X_n)^k$, which explains a fraction $(k - 2)/k$ of the moment (1.4); in particular, this explains why the moment is of order $n^{k-1}$.

We shall see later that, roughly speaking, the moment asymptotic in (1.4) is completely explained by extremely large clades of size $\Theta(n)$, which appear in the $O(1)$ first generations of the tree.

This will also lead to a version of (1.4) for absolute central moments:

**Theorem 1.5.** For any fixed real $p > 2$, as $n \to \infty$,

$$E \left| X_n - E X_n \right|^p \sim \frac{2p}{p - 2} \alpha^p n^{p-1}. \quad (1.8)$$

In Section 2, we transfer the problem from random phylogenetic trees to random binary search tree, which we shall use in the proofs. The theorems above are proved in Sections 3–7.

### 2. Binary trees

We find it technically convenient to work with binary trees instead of full binary trees (phylogenetic trees), so we use the bijection in Section 1 to define $F(T)$ also for binary trees $T$. (We use the same notation $F$; this should not cause any confusion.) With this translation, our problem is thus to study $X_n := F(\bar{T}_n)$, where $\bar{T}_n$ is the binary search tree with $n$ nodes.

The clades in a phylogenetic tree correspond to the internal nodes that have at least one external child, i.e., the nodes in the corresponding binary tree that have outdegree at most 1. We call such nodes green. For a binary tree $T$, the number $F(T)$ is thus the number of maximal green nodes, i.e., the number of green nodes that have no green ancestor. (This holds also for the phylogenetic tree $T$ with a single node, and thus for the empty binary tree, with our definition $F(T) = 0$ in this case.)

It follows that, for any binary tree $T$,

$$F(T) := \begin{cases} 1 & \text{if } T \text{ has a green root}, \\ F(T_L) + F(T_R) & \text{otherwise.} \end{cases} \quad (2.1)$$
Define, for a binary tree $T$,
\[
f(T) := F(T) - F(T_L) - F(T_R) = \begin{cases} 
1 & T_L = \emptyset, T \neq \emptyset, \\
1 & T_R = \emptyset, T \neq \emptyset, \\
0 & \text{otherwise}
\end{cases}
\] (2.2)

Then $F(T)$ is given by the recursion
\[
F(T) = F(T_L) + F(T_R) + f(T),
\] (2.3)
and thus
\[
F(T) = \sum_{v \in T} f(T_v),
\] (2.4)
where $T_v$ is the subtree rooted at $v$, consisting of $v$ and all its descendants.

In another words, $F(T)$ is the additive functional defined by the toll function $f(T)$. The advantage of this point of view is that we have eliminated the maximality condition and now sum over all subtrees $T_v$, and that we can use general results for this type of sums, see Holmgren and Janson [16].

We let $\mathcal{T}$ denote the random binary search tree with a random number of elements such that $P(|\mathcal{T}| = n) = \frac{2}{((n + 1)(n + 2))}$, $n \geq 1$. The random binary tree $\mathcal{T}$ can be constructed by a continuous-time branching process: Let $(\tilde{T}_t)_{t \geq 0}$ be the growing tree that starts with an isolated root at time $t = 0$ and such that each existing node gets a left and a right child after random waiting times that are independent and Exp(1); we stop the process at a random time $\tau \sim \text{Exp}(1)$, independent of everything else, and can take $\mathcal{T} = \tilde{T}_\tau$, see Aldous [1] (where it is also proved that $\mathcal{T}$ is the limit in distribution of a random fringe tree in a binary search tree).

3. The Mean

Recall that $\mathcal{T}_n$ is the random binary search tree with $n$ nodes. Define $\nu_n := E F(\mathcal{T}_n)$ and $\mu_n := E f(\mathcal{T}_n)$, with $F$ and $f$ as in Section 2. (In particular, $\nu_0 = \mu_0 = 0$, while $\nu_1 = \mu_1 = 1$ since $F(\mathcal{T}_1) = f(\mathcal{T}_1) = 1$.) For $n \geq 2$, $\mathcal{T}_{n,L}$ is empty with probability $1/n$, and conditioned on this event, $\mathcal{T}_{n,R}$ has the same distribution as $\mathcal{T}_{n-1}$. The same holds if we interchange $L$ and $R$. Hence, taking the expectation in (2.2),
\[
\mu_n = \frac{2}{n} (1 - E F(\mathcal{T}_{n-1})) = \frac{2}{n} (1 - \nu_{n-1}), \quad n \geq 2.
\] (3.1)

Furthermore, we see that (2.2) implies
\[
P(f(\mathcal{T}_n) \neq 0) \leq \frac{2}{n}.
\] (3.2)

Since obviously $0 \leq F(T) \leq |T|$, we have by (2.2) also $- |T| \leq f(T) \leq 1$ and thus
\[
|f(T)| \leq |T|
\] (3.3)

for any binary tree $T$. In particular, this and (3.2) yield
\[
|\mu_n| \leq E |f(\mathcal{T}_n)| \leq n P(f(\mathcal{T}_n) \neq 0) \leq 2.
\] (3.4)
It is now a simple consequence of general results that $\nu_n := \mathbb{E} F(T_n)$ is asymptotically linear in $n$. Recall the random binary tree $T$ defined in Section 2.

**Lemma 3.1.**

$$\nu_n := \mathbb{E} F(T_n) = n\alpha + O(1), \quad (3.5)$$

where

$$\alpha := \mathbb{E} f(T) = \sum_{n=1}^{\infty} \frac{2}{(n+1)(n+2)} \mathbb{E} f(T_n) = \sum_{n=1}^{\infty} \frac{2}{(n+1)(n+2)} \mu_n$$

$$= \sum_{n=1}^{\infty} \frac{4}{n(n+1)(n+2)}(1 - \nu_{n-1}). \quad (3.6)$$

**Proof.** An instance of Holmgren and Janson [16, Theorem 3.8]. More explicitly, see [16, Theorem 3.4],

$$\mathbb{E} F(T_n) = (n + 1) \sum_{k=1}^{n-1} \frac{2}{(k+1)(k+2)} \mu_k + \mu_n, \quad (3.7)$$

which implies the result by (3.4) and (3.1). \qed

In order to prove Theorem 1.1, it remains to show that $\alpha$ defined in (3.6) equals $(1 - e^{-2})/4$ as asserted in (1.1). In other words, we need the following.

**Lemma 3.2.**

$$\mathbb{E} f(T) = \frac{1 - e^{-2}}{4}. \quad (3.8)$$

We can prove Lemma 3.2 by probabilistic methods, using the construction of $T$ by a branching process in Section 2. However, this proof is considerably longer than the proof of Theorem 1.1 by singularity analysis of generating functions in [12] and [10]; we nevertheless find the probabilistic proof interesting, and perhaps useful for future generalizations, but since the methods in it are not needed for other results in the present paper, we postpone our proof of Lemma 3.2 to Section 7.

4. **Variance**

Let $\gamma_n^2 := \text{Var}(f(T_n))$ and $\sigma_n^2 := \text{Var}(F(T_n))$. Then $\gamma_0^2 = \gamma_1^2 = \sigma_0^2 = \sigma_1^2 = 0$ and, for $n \geq 2$, using (2.2),

$$\gamma_n^2 = \mathbb{E} f(T_n)^2 - \mu_n^2 = \frac{2}{n} \mathbb{E} (F(T_{n-1}) - 1)^2 - \mu_n^2 \leq \frac{2}{n}n^2 = 2n. \quad (4.1)$$

Before proving the variance asymptotics in (1.3), we begin with a weaker estimate.

**Lemma 4.1.** For $n \geq 1$,

$$\sigma_n^2 := \text{Var} F(T_n) = O(n \log^2 n). \quad (4.2)$$
Proof. By [16, Theorem 3.9], where it suffices to sum to \( n \) since we may replace \( f(T) \) by 0 for \( |T| > n \) without changing \( F(T_n) \),

\[
\sigma_n^2 \leq Cn \left( \sum_{k=1}^{n} \frac{\gamma_k}{k^{3/2}} \right)^2 + \sup_k \frac{\gamma_k^2}{k} + \sum_{k=1}^{n} \frac{\mu_k^2}{k^2} = O(n \log^2 n),
\]

(4.3)

using (4.1) and (3.4), provided \( n \geq 2 \). The case \( n = 1 \) is trivial. \( \square \)

Write \( f(T) = g(T) + h(T) \), where

\[
g(T) := \begin{cases} 
1 - \nu_{|T|-1}, & T_L = \emptyset, T \neq \emptyset \text{ or } T_R = \emptyset, T \neq \emptyset, \\
0, & \text{otherwise}.
\end{cases}
\]

(4.4)

and thus, see (2.2),

\[
h(T) := \begin{cases} 
\nu_{|T_L|} - F(T_R), & T_L = \emptyset, \\
\nu_{|T_R|} - F(T_L), & T_R = \emptyset, \\
0, & \text{otherwise}.
\end{cases}
\]

(4.5)

Then \( g(T_1) = 1, h(T_1) = 0 \), and, for \( k \geq 2 \), using (3.1) and (3.4),

\[
\mathbb{E} g(T_k) = \frac{2}{k}(1 - \nu_{k-1}) = \mu_k = O(1),
\]

(4.6)

\[
\mathbb{E} h(T_k) = \frac{2}{k} \mathbb{E}(\nu_{k-1} - F(T_{k-1})) = 0,
\]

(4.7)

and, using Lemma 4.1,

\[
\text{Var} h(T_k) = \frac{2}{k} \mathbb{E}(\nu_{k-1} - F(T_{k-1}))^2 = \frac{2}{k} \sigma_{k-1}^2 = O(\log^2 k).
\]

(4.8)

Let, for an arbitrary binary tree \( T \),

\[
G(T) := \sum_{v \in T} g(T_v) \quad \text{and} \quad H(T) := \sum_{v \in T} h(T_v),
\]

(4.9)

so by (2.4),

\[
F(T) = G(T) + H(T).
\]

(4.10)

**Lemma 4.2.** For \( n \geq 1 \),

\[
\mathbb{E} G(T_n) = \nu_n,
\]

(4.11)

\[
\mathbb{E} H(T_n) = 0,
\]

(4.12)

\[
\text{Var} H(T_n) = O(n).
\]

(4.13)

**Proof.** By [16, Theorem 3.4], cf. (3.7), and (4.7),

\[
\mathbb{E} H(T_n) = (n + 1) \sum_{k=1}^{n-1} \frac{2}{(k+1)(k+2)} \mathbb{E} h(T_k) + \mathbb{E} h(T_n) = 0,
\]

(4.14)

which proves (4.12). This implies (4.11), since by (4.10),

\[
\mathbb{E} G(T_n) = \mathbb{E} F(T_n) - \mathbb{E} H(T_n) = \nu_n.
\]

(4.15)
Similarly, by [16, Theorem 3.9], cf. (4.3), and (4.7)-(4.8),
\[
\text{Var } H(T_n) \leq Cn \left( \left( \sum_{k=1}^{\infty} \frac{\log k}{k^{3/2}} \right)^2 + \sup_{k \geq 1} \frac{\log^2 k}{k} + 0 \right) = O(n). \quad \Box
\]

We shall see that this means that \( H(T_n) \) is asymptotically negligible, and thus it suffices to consider \( G(T_n) \).

Note that \( g(T) \) depends only on the sizes \(|T_L|\) and \(|T_R|\). This enables us to easily estimate the variance of \( G(T_n) \).

**Theorem 4.3.** For all \( n \geq 1 \),
\[
\text{Var } G(T_n) = 4\alpha^2 n \log n + O(n). \quad (4.16)
\]

**Proof.** Write \( g(T) = g(|T|,|T_L|,|T_R|) \). (We only care about \( g(k,j,l) \) when \( j + l = k - 1 \), but use three arguments for emphasis.) Thus \( g(k,0,k-1) = 1 - \nu_{k-1} \) and otherwise \( g(k,j,k-j-1) = 0 \). Let, as in [16, Theorem 1.29], \( I_k \) be uniformly distributed on \( \{0,\ldots,k-1\} \) and
\[
\psi_k := \mathbb{E} \left( \nu_k + \nu_{k-1} - I_k + g(k,I_k,k-1-I_k) - \nu_k \right)^2
\]
\[
= \frac{1}{k} \sum_{j=1}^{k-2} (\nu_j + \nu_{k-1-j} - \nu_k)^2 + \frac{2}{k} (\nu_{k-1} + 1 - \nu_{k-1} - \nu_k)^2
\]
\[
= \frac{1}{k} \sum_{j=1}^{k-2} (\nu_j + \nu_{k-1-j} - \nu_k)^2 + \frac{2}{k} (\nu_k - 1)^2
\]
\[
= O(1) + \frac{2}{k} (\alpha k + O(1))^2 = 2\alpha^2 k + O(1), \quad (4.17)
\]
where we used that \( \nu_j = \alpha j + O(1) \) by Theorem 1.1. By [16, Lemma 7.1], then
\[
\text{Var } G(T_n) = (n+1) \sum_{k=1}^{n-1} \frac{2}{(k+1)(k+2)} \psi_k + \psi_n
\]
\[
= (n+1) \sum_{k=1}^{n-1} \frac{4\alpha^2 k + O(1)}{(k+1)(k+2)} + O(n) = (n+1) \sum_{k=1}^{n-1} \frac{4\alpha^2 k}{k} + O(n)
\]
\[
= 4\alpha^2 n \log n + O(n). \quad (4.18)
\]
\( \Box \)

We can now prove (1.3) in Theorem 1.2. (Higher moments are treated in Section 6.)

**Theorem 4.4.** For all \( n \geq 1 \),
\[
\text{Var } F(T_n) = 4\alpha^2 n \log n + o(n \log n). \quad (4.19)
\]

This follows from (4.10), (4.16) and (4.13) by Minkowski’s inequality (the triangle inequality for \( \sqrt{\text{Var}} \)).
5. Asymptotic normality

We prove the central limit theorem Theorem 1.3 by a martingale central limit theorem for a suitable martingale that we construct in this section.

Consider the infinite binary tree $T_\infty$, where each node has two children, and denote its root by $o$. We may regard any binary tree $T$ as a subtree of $T_\infty$ with the same root $o$. (In the general sense that the node set $V(T)$ is a subset of $V_\infty := V(T_\infty)$, and that the left and right children are the same as in $T_\infty$, when they exist.) In particular we regard the random binary search tree $T_n$ as a subtree of $T_\infty$.

Order the nodes in $T_\infty$ in breadth-first order as $v(1) = o, v(2), \ldots$, and let $V_j := \{v(1), \ldots, v(j)\}$ be the set of the first $j$ nodes. Let $F_j$ be the $\sigma$-field generated by the sizes $|T_{n,v,L}|$ and $|T_{n,v,R}|$ of the two child subtrees of $T_n$ at each node $v \in V_j$. Equivalently, we may regard $V_j$ as the internal nodes in a full binary tree; let $\partial V_j$ be the corresponding set of $j + 1$ external nodes. Then $F_j$ is generated by the subtree sizes $|T_{n,v}|$ for all $v \in \partial V_j$, together with the indicators $\mathbf{1}\{v \in T_n\}$, $v \in V_j$, that describe $T_n \cap V_j$. (We regard the subtree $T_{n,v}$ as defined for all $v \in V_\infty$, with $T_{n,v} = \emptyset$ if $v \notin T_n$.) Then, conditioned on $F_j$, $T_n$ consists of some given subtree of $V_j$ together with attached subtrees $T_{n,v}$ at all nodes $v \in \partial V_j$; these are independent binary search trees of some given orders.

We allow here $j = 0$; $V_0 = \emptyset$ and $F_0$ is the trivial $\sigma$-field.

**Remark 5.1.** As is well-known, see e.g. [9], another construction of the random binary search tree $T_n$ ($n \geq 1$) is to let the random variable $I_n$ be uniformly distributed on $\{0, \ldots, n - 1\}$, and to let $T_n$ be defined recursively such that, given $I_n$, $T_{n,L}$ and $T_{n,R}$ are independent binary search trees with $|T_{n,L}| = I_n$ and $|T_{n,R}| = n - 1 - I_n$. (When the tree is used to sort $n$ keys, $I_n$ tells how many of the keys that are assigned to the left subtree.) The pair $(I_n, n - 1 - I_n)$ thus tells how the tree is split at the root, and there is a similar pair for each node. Then $F_j$ is generated by these pairs (i.e., splits) for the nodes $v_1, \ldots, v_j$.

Recall that $g(T)$ by (4.4) depends only on the sizes $|T_L|$ and $|T_R|$. Hence, $F_j$ specifies the value of $g(T_{n,v})$ for every $v \in V_j$, and it follows that

$$
\mathbb{E}(G(T_n) \mid F_j) = \mathbb{E}\left(\sum_{v \in V_j} g(T_{n,v}) \mid F_j\right) = \sum_{v \in V_j} g(T_{n,v}) + \sum_{v \in \partial V_j} \nu|T_{n,v}|.
$$

Since the sequence of $\sigma$-fields $(F_j)_0^\infty$ is increasing, the sequence $M_{n,j} := \mathbb{E}(G(T_n) \mid F_j)$, $j \geq 0$, is a martingale (for any fixed $n$). It follows from (5.1) that the martingale differences are

$$
\Delta M_{n,j} := M_{n,j} - M_{n,j-1} = g(T_{n,v(j)}) + \nu|T_{n,v(j)}| + \nu|T_{n,v(j)}| - \nu|T_{n,v(j)}|,
$$

where $v(j)_L$ and $v(j)_R$ are the children of $v(j)$. It follows easily that, with $\psi_k$ defined in (4.17),

$$
\mathbb{E}(|\Delta M_{n,j}|^2 \mid F_{j-1}) = \mathbb{E}(|\Delta M_{n,j}|^2 \mid |T_{n,v(j)}|) = \psi|T_{n,v(j)}|.
$$
Consequently, the conditional square function is given by

\[ W_n := \sum_{j=1}^{\infty} \mathbb{E}\left( |\Delta M_{n,j}|^2 \mid \mathcal{F}_{j-1} \right) = \sum_{v \in V_\infty} \psi_{|T_n,v|} = \sum_{v \in T_n} \psi_{|T_n,v|}. \tag{5.4} \]

(It suffices to sum over \( v \in T_n \), since \( \psi_0 = 0 \).) This is again a sum of the same type as (2.4) and (4.9), for the random tree \( T_n \). (Note that the toll function \( \psi_{|T|} \) here depends only on the size of \( T \).) In particular, [16, Theorem 3.4] applies (in this case we can also use [7], [8] or [13]); this yields

\[ \mathbb{E} W_n = (n + 1) \sum_{k=1}^{n-1} \frac{2}{(k + 1)(k + 2)} \psi_k + \psi_n. \tag{5.5} \]

If \( j \) is large enough, say \( j \geq 2^n \), then \( V(T_n) \subseteq V_j \) and thus \( M_{n,j} = G(T_n) \).

In particular, \( G(T_n) = M_{n,\infty} \). Thus, by a standard (and simple) martingale identity, \( \text{Var} G(T_n) = \text{Var} M_{n,\infty} = \mathbb{E} W_n \); hence (5.5) yields the first equality in (4.18). (This is no coincidence; the proof just given of (5.5) is essentially the same as the proof of [16, Lemma 7.1] that was used in (4.18), but stated in martingale formulation.)

We now split the sum \( G(T_n) \) into two parts, roughly corresponding to small and large clades. We fix a cut-off \( N = N(n) \); for definiteness and simplicity we choose \( N = N(n) := \sqrt{n} \), but we note that the arguments below hold with a few minor modifications for any \( N \geq \sqrt{n} \) with \( N = o(\sqrt{n \log n}) \). We then define, for binary trees \( T \),

\[
\begin{align*}
g'(T) &:= g(T) \mathbf{1}\{|T| \leq N\} \tag{5.6} \\
g''(T) &:= g(T) \mathbf{1}\{|T| > N\} = g(T) - g'(T). \tag{5.7}
\end{align*}
\]

In analogy with (2.4) and (4.9), we define further

\[ G'(T) := \sum_{v \in T} g'(T_v) \quad \text{and} \quad G''(T) := \sum_{v \in T} g''(T_v); \tag{5.8} \]

thus \( G(T) = G'(T) + G''(T) \). We shall see that, asymptotically, both \( G'(T_n) \) and \( G''(T) \) contribute to the variance with equal amounts, but nevertheless \( G''(T_n) \) is negligible (in probability).

We begin with the main term \( G'(T_n) \).

**Lemma 5.2.** As \( n \to \infty \),

\[ \text{Var} \left( G'(T_n) \right) = 2\alpha^2 n \log n + O(n), \tag{5.9} \]

\[ \frac{G'(T_n) - \mathbb{E} G'(T_n)}{\sqrt{2\alpha^2 n \log n}} \overset{d}{\to} N(0,1). \tag{5.10} \]

**Proof.** We define \( \nu' := \mathbb{E} G'(T_n) \). Note that \( g'(T) \) depends only on the sizes \( |T_L| \) and \( |T_R| \). Hence we can repeat the argument above and define a martingale \( M'_{n,j} := \mathbb{E}(G'(T_n) \mid \mathcal{F}_j), \ j \geq 0 \), with \( G'(T_n) = M'_{n,\infty} \) and martingale differences

\[ \Delta M'_{n,j} = \varphi'(T_n,v(j)), \tag{5.11} \]

where \( \varphi'(T_n,v(j)) \) are toll functions as in (5.6) and (5.7). By argument above, we have

\[ \text{Var} \left( G'(T_n) \right) = \text{Var} \left( M'_{n,\infty} \right) = \mathbb{E} \left( \left( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \right)^2 \right) - \mathbb{E} \left( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \right)^2. \tag{5.12} \]

By similar arguments to (4.18), the second term is negligible in probability if \( j \) is large enough. Hence

\[ \text{Var} \left( G'(T_n) \right) = \mathbb{E} \left( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \right)^2. \tag{5.13} \]

In particular, \( \mathbb{E} G'(T_n) = \sum_{j=0}^{\infty} \mathbb{E} \varphi'(T_n,v(j)) = \nu' . \)

Therefore, we can think of \( \nu' \) as the expectation of the sum \( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \), and hence the sum also satisfies

\[ \text{Var} \left( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \right) = \text{Var} \left( G'(T_n) \right) \approx \mathbb{E} \left( \left( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \right)^2 \right). \tag{5.14} \]

From (4.9) and (5.10), we have

\[ \text{Var} \left( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \right) = \mathbb{E} \left( \left( \sum_{j=0}^{\infty} \varphi'(T_n,v(j)) \right)^2 \right) = 2\alpha^2 n \log n + O(n). \tag{5.15} \]

This completes the proof of the lemma.
where we define, cf. (5.2),
\[
\varphi'(T) := g'(T) + \nu'_{|T_L|} + \nu'_{|T_R|} - \nu'_{|T|}.
\]  

(5.12)

By [16, Theorem 3.4] again, cf. (3.7) and (5.5), using \( E g(T_k) = \mu_k = O(1) \) by (4.6),
\[
\nu'_m = (m + 1) \sum_{k=1}^{m-1} \frac{2}{(k+1)(k+2)} \left( g(T_k) + g'(T_m) \right) + O(1)
\]  

(5.13)

Hence, (5.12) yields, after cancellations,
\[
\varphi'(T) = g'(T) + O(1) = \begin{cases} 
  g(T) + O(1), & |T| \leq N, \\
  O(1), & |T| > N.
\end{cases}
\]  

(5.14)

Let \( \psi'_k := E |\varphi'(T_k)|^2 \).

(5.15)

Then, by (5.14), (4.4) and (3.5), cf. (4.17),
\[
\psi'_k = \begin{cases} 
  \left( E(g(T_k) + O(1))^2 = 2\alpha^2k + O(1), & k \leq N, \\
  O(1), & k > N.
\end{cases}
\]  

(5.16)

Furthermore, by (5.11) and (5.15),
\[
E(|\Delta M_{n,j}|^2 \mid F_{j-1}) = E(|\varphi'(T_{n,v(j)})|^2 \mid |T_{n,v(j)}|) = \psi'_{|T_{n,v(j)}|}.
\]  

(5.17)

Hence, the conditional square function of \((M'_{n,j})_j\) is
\[
W'_n := \sum_{j=1}^{\infty} E(|\Delta M_{n,j}|^2 \mid F_{j-1}) = \sum_{v \in V_{\infty} \setminus T_n} \psi'_{|T_{n,v}|} = \sum_{v \in T_n} \psi'_{|T_{n,v}|}.
\]  

(5.18)

Yet another application of [16, Theorem 3.4] yields, using (5.16),
\[
E W'_n = (n + 1) \sum_{k=1}^{n-1} \frac{2}{(k+1)(k+2)} \psi'_k + \psi'_n
\]  

\[
= (n + 1) \sum_{k=1}^{N} \frac{4\alpha^2k}{(k+1)(k+2)} + O(n)
\]  

\[
= 4\alpha^2n \log N + O(n) = 2\alpha^2n \log n + O(n).
\]  

(5.19)

Since \( \text{Var} G'(T_n) = \text{Var}(M'_{n,\infty}) = E W'_n \), (5.9) follows from (5.19).
Moreover, the representation (5.18) and [16, Theorem 3.9] (again summing only to $n$, as we may) yield, noting that the toll function $\psi_{|T|}$ depends only on the size of $T$, using (5.16),

$$\text{Var}(W'_n) \leq Cn \sum_{k=1}^{n} \frac{(\psi'_{k})^2}{k^2} \leq C_1n \sum_{k=1}^{N} 1 + C_2n \sum_{k=1}^{n} \frac{1}{k^2} = O(nN) = O(n^2).$$

Hence, $\text{Var}(W'_n/(n \log n)) \to 0$ as $n \to \infty$, which together with (5.19) implies

$$\frac{W'_n}{n \log n} \xrightarrow{p} 2\alpha^2.$$

(5.20)

Note also that $g(T) = O(|T|)$ by (4.4) and (3.5), and thus (5.14) implies $\varphi'(T) = O(N)$ for all trees $T$. Thus (5.11) yields

$$\sup_j |\Delta M_{n,j}| \sqrt{n \log n} = O\left(\frac{N}{\sqrt{n \log n}}\right) = o(1).$$

(5.22)

We now apply the central limit theorem for martingale triangular arrays, in the form in [5, Corollary 1] (see also [15, Theorem 3.1]), which shows that (5.21) and (5.22) together imply

$$\frac{G'(T_n) - \mathbb{E} G'(T_n)}{\sqrt{n \log n}} = \frac{M_{n,\infty} - \mathbb{E} M_{n,\infty}}{\sqrt{n \log n}} \xrightarrow{d} N(0,2\alpha^2).$$

(5.23)

(Actually, [5, Corollary 1] assumes instead of (5.22) only a conditional Lindeberg condition, which is a trivial consequence of the uniform bound (5.22).)

**Remark 5.3.** We used the breadth-first order above as just one convenient order. It is perhaps more natural to consider instead of the sets $V_j$ arbitrary node sets $V$ of (finite) subtrees of $T_\infty$ that include the root $o$. This would give us, instead of $(M_{n,j})$, a martingale indexed by binary trees. However, we have no use for this exotic object here, and use instead the standard martingales above.

**Lemma 5.4.**

$$\mathbb{E}|G''(T_n)| = O(\sqrt{n}),$$

(5.24)

$$\text{Var}(G''(T_n)) = 2\alpha^2 n \log n + O(n).$$

(5.25)

**Proof.** By (5.7), (4.4) and (4.6),

$$\mathbb{E}|g''(T_k)| = |\mathbb{E} g(T_k) \cdot 1\{k > N\} = O(1) \cdot 1\{k > N\}$$

(5.26)

and thus, using the triangle inequality and [16, Theorem 3.4],

$$\mathbb{E}|G''(T_n)| \leq (n + 1) \sum_{N}^{n-1} \frac{2}{(k+1)(k+2)} \mathbb{E}|g''(T_k)| + \mathbb{E}|g''(T_n)| = O\left(\frac{n}{N}\right),$$

yielding (5.24).
For the variance, we use either [16, Theorem 1.29] as in the proof of Theorem 4.4, or the (essentially equivalent) martingale argument in (5.11)–(5.19) and conclude that, with some \( \psi_k'' \) satisfying

\[
\psi_k'' = \begin{cases} 
O(1), & k \leq N, \\
\mathbb{E}(g(T_k) + O(1))^2 = 2\alpha^2 k + O(1), & k > N,
\end{cases}
\]

we have

\[
\text{Var} \, G''(T_n) = (n + 1) \sum_{k=1}^{n-1} \frac{2}{(k + 1)(k + 2)} \psi_k'' + \psi_n''
\]

\[
= (n + 1) \sum_{k=\lfloor N \rfloor + 1}^{n-1} \frac{4\alpha^2 k}{k^2} + O(n)
\]

\[
= 4\alpha^2 n \log(n/N) + O(n) = 2\alpha^2 n \log n + O(n).
\]

\[\square\]

**Proof of Theorem 1.3.** It follows from (5.24) that

\[
\frac{G''(T_n) - \mathbb{E} G''(T_n)}{\sqrt{2\alpha^2 n \log n}} \xrightarrow{p} 0,
\]

which together with (5.10) yields

\[
\frac{G(T_n) - \mathbb{E} G(T_n)}{\sqrt{2\alpha^2 n \log n}} \xrightarrow{d} N(0,1).
\]

Similarly, (4.13) implies

\[
\frac{H(T_n) - \mathbb{E} H(T_n)}{\sqrt{2\alpha^2 n \log n}} \xrightarrow{p} 0,
\]

which together with (5.29) yields (1.5), recalling \( X_n = F(T_n) = G(T_n) + H(T_n) \) by (4.10).

\[\square\]

**Proof of Theorem 1.4.** (i). Define, similarly to (5.6)–(5.7),

\[
f'(T) := f(T) \mathbf{1}\{|T| \leq N\}, \quad f''(T) := f(T) \mathbf{1}\{|T| > N\},
\]

\[
h'(T) := h(T) \mathbf{1}\{|T| \leq N\}, \quad h''(T) := h(T) \mathbf{1}\{|T| > N\},
\]

and corresponding sums \( F'(T) := \sum_{v \in T} f'(T_v) \) and similarly \( F''(T), \, H'(T), \, H''(T) \). The argument in (2.1)–(2.4) is easily modified and shows that

\[
X_n = F'(T_n) = G'(T_n) + H'(T_n).
\]

The same proof as for Lemma 4.2 yields also

\[
\text{Var} \, H'(T_n) = O(n) \quad \text{and} \quad \text{Var} \, H''(T_n) = O(n).
\]

Hence, (1.6) follows from Lemma 5.2 and (5.33).

Furthermore,

\[
X_n - X_n^N = F''(T_n) = G''(T_n) + H''(T_n).
\]
By (5.33) and (5.35), (1.7) follows from (5.9) and (5.25), using (5.34) and Minkowski’s inequality. Similarly,

\[
\mathbb{E}|X_n - X_n^N| \leq \mathbb{E}|G''(T_n)| + \mathbb{E}|H''(T_n)| = O(\sqrt{n}),
\]

(5.36) using (5.24), (5.34) and Hölder’s inequality, together with \(\mathbb{E} H''(T_n) = 0\), which is proved as (4.12).

(ii). The conclusions of (i) hold by the same proofs (with some minor modifications in some estimates).

Moreover, let \(Z_{n,k}\) be the number of clades of size \(k + 1\). Then, for \(n \geq 2\), the expected number is given by

\[
\mathbb{E}Z_{n,k} = \begin{cases} 
\frac{4n}{k(k+1)(k+2)}, & k < n, \\
\frac{2}{n}, & k = n, \\
0, & k > n, 
\end{cases}
\]

(5.37) see [6, Theorem 1]. (This can be seen as another example of [16, Theorem 3.4].) Consequently,

\[
\mathbb{P}(X_n \neq X_n^N) \leq \mathbb{P}\left(\sum_{k>N} Z_{n,k} \geq 1\right)
\]

\[
\leq \mathbb{E}\sum_{k>N} Z_{n,k} = \sum_{[N]+1}^{n-1} \frac{4n}{k(k+1)(k+2)} + \frac{2}{n}
\]

\[
= O\left(\frac{n}{N^2}\right) + O\left(\frac{1}{n}\right) = o(1),
\]

(5.38) which completes the proof. \(\square\)

6. Higher moments

We begin the proof of Theorem 1.5 by proving a weaker estimate. We let \(\|X\|_p := (\mathbb{E} X^p)^{1/p}\) for any random variable \(X\). Recall that \(\nu_n := \mathbb{E} F(T_n)\).

**Lemma 6.1.** For any fixed real \(p > 2\), and all \(n \geq 1\),

\[
\mathbb{E}\|F(T_n) - \nu_n\|^p \leq C(p)n^{p-1}.
\]

(6.1) Equivalently,

\[
\|F(T_n) - \nu_n\|_p = O(n^{1-1/p}).
\]

(6.2)

**Proof.** Fix \(p > 2\) and let \(m \geq 1\) be chosen below. (The constants \(C_i\) below may depend on \(p\) but not on \(m\).) Let \(V_j\) and \(F_j\) be as in Section 5, and write \(V''_m := V_{2m-1}\), \(F''_m := F_{2m-1}\). Thus \(\partial V''_m\) consists of the \(2^m\) nodes in \(T_\infty\) of depth \(m\), and \(V''_m\) consists of the \(2^m - 1\) nodes of smaller depth. It follows from (2.4) that, for any binary tree \(T\),

\[
F(T) = \sum_{v \in V''_m} f(T_v) + \sum_{v \in \partial V''_m} F(T_v).
\]

(6.3)
Furthermore, by (1.2),
\[
\sum_{v \in \partial V_m'} \nu_{|T_v|} = \sum_{v \in \partial V_m'} (\alpha |T_v| + O(1)) = \alpha \sum_{v \in \partial V_m'} |T_v| + O(2^m)
\]
\[
= \alpha |T| + O(2^m) = \nu_{|T|} + O(2^m).
\] (6.4)

Hence, by combining (6.3) and (6.4),
\[
F(T) - \nu_{|T|} = \sum_{v \in V_m'} f(T_v) + \sum_{v \in \partial V_m'} (F(T_v) - \nu_{|T_v|}) + O(2^m).
\] (6.5)

We shall use this decomposition for the binary search tree \(T_n\). Note first that by (3.2)–(3.3),
\[
\mathbb{E} |f(T_n)|^p \leq n^p \mathbb{P}(f(T_n) \neq 0) \leq 2n^{p-1}.
\] (6.6)

(This holds for any \(p > 0\) and generalises (3.4) which is the case \(p = 1\).) Hence, for any \(v \in V_\infty\),
\[
\mathbb{E}(|f(T_{n,v})|^{p} \mid |T_{n,v}|) \leq 2 |T_{n,v}|^{p-1} \leq 2n^{p-1},
\] (6.7)

and thus
\[
\mathbb{E} |f(T_{n,v})|^p \leq 2n^{p-1}.
\] (6.8)

Let \(Y := \sum_{v \in V_m'} f(T_{n,v})\) be the first sum in (6.5) for \(T = T_n\). By Minkowski’s inequality and (6.8),
\[
\|Y\|_p \leq \sum_{v \in V_m'} \|f(T_{n,v})\|_p \leq 2^{m/2} 2^{1/p} n^{(p-1)/p}.
\] (6.9)

Let \(Z := \sum_{v \in \partial V_m'} (F(T_{n,v}) - \nu_{|T_{n,v}|})\) be the second sum in (6.5) for \(T = T_n\). The \(\sigma\)-field \(F'_m\) specifies the sizes of the subtrees \(T_{n,v}\) for \(v \in \partial V_m'\), and conditioned on \(F'_m\), these subtrees are independent and distributed as \(T_{n,v}\) of the given sizes \(n(v)\). Hence, conditionally on \(F'_m\), the terms in the sum \(Z\) are independent and have means zero, so we can apply Rosenthal’s inequality [14, Theorem 3.9.1], which yields
\[
\mathbb{E}(|Z|^p \mid F'_m) \leq C_1 \sum_{v \in \partial V_m'} \mathbb{E}(|F(T_{n,v}) - \nu_{|T_{n,v}|}|^p \mid F'_m)
\]
\[
+ C_1 \left( \sum_{v \in \partial V_m'} \mathbb{E}(|F(T_{n,v}) - \nu_{|T_{n,v}|}|^2 \mid F'_m) \right)^{p/2}.
\] (6.10)

We note first that by (1.3),
\[
\mathbb{E}(|F(T_{n,v}) - \nu_{|T_{n,v}|}|^2 \mid F'_m) \leq C_2 |T_{n,v}| \log |T_{n,v}| \leq C_2 |T_{n,v}| \log n,
\] (6.11)

and thus
\[
\sum_{v \in \partial V_m'} \mathbb{E}(|F(T_{n,v}) - \nu_{|T_{n,v}|}|^2 \mid F'_m) \leq C_2 \sum_{v \in \partial V_m'} |T_{n,v}| \log n \leq C_2 n \log n.
\] (6.12)
Hence the second term on the right-hand side in (6.10) is \( \leq C_3(n \log n)^{p/2} \).
Taking the expectation in (6.10) we thus obtain
\[
\mathbb{E}|Z|^p \leq C_1 \sum_{v \in \partial V_m} \mathbb{E}|F(T_{n,v}) - \nu_{|T_{n,v}|}|^p + C_4(n \log n)^{p/2}.
\] (6.13)

Let \( A_n := \mathbb{E}|F(T_n) - \nu_n|^p \). We can write (6.5) for \( T = T_n \) as
\[
F(T_n) - \nu_n = Y + Z + O(2^m).
\] (6.14)
Thus, by Minkowski’s inequality, (6.9) and (6.13),
\[
A_n = \mathbb{E}|Y + Z + O(2^m)|^p \leq 3^p (\mathbb{E}|Y|^p + \mathbb{E}|Z|^p + O(2^m))
\]
\[
\leq C_5 2^{mp} n^{p-1} + C_6 \mathbb{E}|Z|^p + C_7 2^m \leq C_6 \mathbb{E}|Z|^p + C_8 2^{mp} n^{p-1}.
\] (6.15)
Furthermore, (6.13) can be written
\[
\mathbb{E}|Z|^p \leq C_1 \sum_{v \in \partial V_m} \mathbb{E} A|_{T_{n,v}|} + C_4(n \log n)^{p/2}.
\] (6.16)

We prove the lemma by induction, and assume that \( A_k \leq Ck^{p-1} \) for all \( k < n \). Since \( |T_{n,v}| < n \) for every \( v \in \partial V_m \), (6.16) and the inductive hypothesis yield
\[
\mathbb{E}|Z|^p \leq C_1 C \sum_{v \in \partial V_m} \mathbb{E}|T_{n,v}|^{p-1} + C_4(n \log n)^{p/2}.
\] (6.17)
If \( v \) is a child of the root, then \( |T_{n,v}| \) is uniformly distributed on \( \{0, \ldots, n-1\} \), so \( |T_{n,v}| \overset{d}{=} |nU| \leq nU \), where \( U \sim U(0,1) \) is uniformly distributed on \( [0,1] \).
By induction in \( m \), it follows that for any \( v \in \partial V_m \),
\[
|T_{n,v}| \leq n \prod_{i=1}^m U_i,
\] (6.18)
with \( U_1, \ldots, U_m \) independent and \( U(0,1) \). Consequently,
\[
\mathbb{E}|T_{n,v}|^{p-1} \leq \mathbb{E}\left(n^{p-1} \prod_{i=1}^m U_i^{p-1}\right) = n^{p-1} \prod_{i=1}^m \mathbb{E}U_i^{p-1} = n^{p-1}(1/p)^m,
\] (6.19)
since \( \mathbb{E}U_i^{p-1} = \int_0^1 u^{p-1} du = 1/p \). There are \( 2^m \) nodes in \( \partial V_m \), and thus (6.17) yields
\[
\mathbb{E}|Z|^p \leq C_1 C 2^{m(1/p)^m n^{p-1}} + C_4(n \log n)^{p/2},
\] (6.20)
which together with (6.15) yields, since \( (n \log n)^{p/2} = O(n^{p-1}) \) when \( p > 2 \),
\[
A_n \leq C_6 C_1 C(2/p)^m n^{p-1} + C_6 C_4(n \log n)^{p/2} + C_8 2^{mp} n^{p-1}
\]
\[
\leq C_6 C_1 C(2/p)^m n^{p-1} + C_9 2^{mp} n^{p-1}.
\] (6.21)
Now choose \( m \) such that \( (2/p)^m C_6 C_1 < 1/2 \) (which is possible because \( p > 2 \)). Then choose \( C := 2^{mp+1} C_9 \). With these choices, (6.21) yields
\[
A_n \leq \frac{1}{2} C n^{p-1} + \frac{1}{2} C n^{p-1} = C n^{p-1}.
\] (6.22)
In other words, we have proved the inductive step: \( A_k \leq C k^{p-1} \) for \( k < n \) implies \( A_n \leq C n^{p-1} \). Consequently, this is true for all \( n \geq 0 \), i.e., (6.1) holds. (The initial cases \( n = 0 \) and \( n = 1 \) are trivial, since \( A_0 = A_1 = 0 \).) \( \square \)

**Lemma 6.2.** For any fixed real \( p > 2 \), as \( n \to \infty \),

\[
\| F(T_n) \|_p \sim \alpha n, \tag{6.23}
\]

\[
\| f(T_n) \|_p \sim 2^{1/p} n^{1-1/p}. \tag{6.24}
\]

**Proof.** By Minkowski’s inequality, (6.2) and (1.2),

\[
\| F(T_n) \|_p = |E F(T_n)| + O(n^{1−1/p}) = \alpha n + O(n^{1−1/p}) \sim \alpha n,
\]

which is (6.23).

For \( n \geq 2 \), it follows from (2.2) that

\[
E |f(T_n)|^p = \frac{2}{n} E |1 − F(T_{n−1})|^p = \frac{2}{n} \| F(T_{n−1}) − 1 \|_p^p \sim 2\alpha n^{p−1}, \tag{6.26}
\]

since (6.23) obviously implies also \( \| F(T_n) − 1 \|_p \sim \alpha n \). \( \square \)

The idea in the proof of Theorem 1.5 is to approximate \( E |X_n − E X_n|^p = E |\sum_v (f(T_{n,v}) − E f(T_{n,v}))*|^p \) by \( E \sum_v |f(T_{n,v}) − E f(T_{n,v})|^p \), or simpler by \( E \sum_v |f(T_{n,v})|^p = \sum_v E |f(T_{n,v})|^p \). The heuristic reason for this is that the moment \( E |\sum_v (f(T_{n,v}) − E f(T_{n,v}))*|^p \) is dominated by the event when there is one large term (corresponding to one large clade, cf. the discussion before Theorem 1.5), and then

\[
|\sum_v (f(T_{n,v}) − E f(T_{n,v}))*|^p \approx \sum_v |f(T_{n,v}) − E f(T_{n,v})|^p \approx \sum_v |f(T_{n,v})|^p. \tag{6.27}
\]

We shall justify this in several steps. We begin by finding the expectation of the final sum in (6.27), cf. the sought result (1.8).

**Lemma 6.3.** As \( n \to \infty \),

\[
E \sum_{v \in T_n} |f(T_{n,v})|^p \sim \frac{2^p}{p−2} \alpha n^{p−1}. \tag{6.28}
\]

**Proof.** We apply again [16, Theorem 3.4] and obtain

\[
E \sum_{v \in T_n} |f(T_{n,v})|^p = (n + 1) \sum_{k=1}^{n-1} \frac{2}{(k+1)(k+2)} E |f(T_k)|^p + E |f(T_n)|^p. \tag{6.29}
\]

By (6.26),

\[
\frac{2}{(k+1)(k+2)} E |f(T_k)|^p \sim \frac{2}{k^2} \cdot 2\alpha k^{p−1} = 4\alpha k^{p−3}. \tag{6.30}
\]
as \( k \to \infty \), and it follows that, as \( n \to \infty \), using \( p > 2 \),

\[
\mathbb{E} \sum_{v \in \mathcal{T}_n} |f(\mathcal{T}_{n,v})|^p \sim (n + 1) \sum_{k=1}^{n-1} 4\alpha^p k^{p-3} + 2\alpha^p n^{p-1}
\]

\[
\sim n \frac{4\alpha^p}{p-2} n^{p-2} + 2\alpha^p n^{p-1} = \frac{2p}{p-2} \alpha^p n^{p-1}. \quad \Box
\]

Next we take again some \( m \geq 1 \) and use the notation in the proof of Lemma 6.1. Since we now have proved (6.1), the proof of Lemma 6.1 shows that (6.20) holds for every \( n \), and thus, since \( p > 2 \),

\[
\|Z\|_p \leq C_{10} (2/p)^{m/p} n^{1-1/p} + O((n \log n)^{1/2})
\]

\[
= C_{10} (2/p)^{m/p} n^{1-1/p} + o(n^{1-1/p}). \quad (6.31)
\]

Consequently, by (6.14) and Minkowski’s inequality,

\[
\|F(\mathcal{T}_n) - \nu_n\|_p - \|Y\|_p \leq \|Z\|_p + O(2^m) = C_{10} (2/p)^{m/p} n^{1-1/p} + o(n^{1-1/p}). \quad (6.32)
\]

In particular, (6.32) and (6.2) imply \( \|Y\|_p = O(n^{1-1/p}) \). By the mean value theorem,

\[
|x^p - y^p| \leq p|x - y| \max\{x^{p-1}, y^{p-1}\} \quad (6.33)
\]

for any \( x, y \geq 0 \); hence (6.32) implies, using also (6.2) again,

\[
\mathbb{E}|F(\mathcal{T}_n) - \nu_n|^p - \mathbb{E}|Y|^p = O((2/p)^{m/p} n^{p-1}) + o(n^{p-1}). \quad (6.34)
\]

Let \( \delta > 0 \) be a small positive number to be chosen later, and let \( J_v \) be the indicator of the event that \( v \) is green and \( |\mathcal{T}_{n,v}| \geq \delta n \). (The idea is that the significant contributions only come from nodes \( v \) with \( J_v = 1 \).)

**Lemma 6.4.** For each fixed \( m \geq 1 \) and \( \delta > 0 \), and all \( n \geq 1 \),

\[
\mathbb{P}\left( \sum_{v \in V'_m} J_v \geq 1 \right) \leq 2^{m+1} \delta^{-1} n^{-1} = O(n^{-1}), \quad (6.35)
\]

\[
\mathbb{P}\left( \sum_{v \in V'_m} J_v \geq 2 \right) \leq 2^{2m+1} \delta^{-2} n^{-2} = O(n^{-2}). \quad (6.36)
\]

**Proof.** We use again the \( \sigma \)-fields \( \mathcal{F}_j \) from Section 5. Since \( \mathcal{F}_{j-1} \) specifies \( |\mathcal{T}_{n,v_j}| \), but not how this subtree is split at \( v_j \), we have

\[
\mathbb{P}(J_{v_j} = 1 \mid \mathcal{F}_{j-1}) \leq \frac{2}{|\mathcal{T}_{n,v_j}|} 1\{\mathcal{T}_{n,v_j} \geq \delta n\} \leq \frac{2}{\delta n}, \quad (6.37)
\]

and thus, by taking the expectation, \( \mathbb{P}(J_{v_j} = 1) \leq 2/(\delta n) \). Since there are < \( 2^m \) nodes in \( V'_m \), (6.35) follows.

Furthermore, for any two nodes \( v_i \) and \( v_j \) with \( i < j \), \( J_{v_i} \) is determined by \( \mathcal{F}_{j-1} \), and (6.37) thus gives also

\[
\mathbb{P}(J_{v_i}, J_{v_j} = 1 \mid \mathcal{F}_{j-1}) = \mathbb{E}(J_{v_i}, J_{v_j} \mid \mathcal{F}_{j-1}) = J_{v_i} \mathbb{P}(J_{v_j} = 1 \mid \mathcal{F}_{j-1}) \leq \frac{2}{\delta n} J_{v_i},
\]

\[
(6.38)
\]
Thus, by taking the expectation and using (6.37) again, $P(J_v J_j = 1) \leq 4/(\delta n)^2$. Summing over the less than $2^{2n-1}$ pairs $(v_i, v_j)$ with $v_i, v_j \in V_m'$ yields (6.36). \hfill \Box

**Proof of Theorem 1.5.** We show this in several steps.

**Step 1.** Define

$$Y_1 := \sum_{v \in V_m'} J_v f(T_{n,v}).$$

Since $f(T_{n,v}) = 0$ unless $v$ is green, we have

$$Y - Y_1 = \sum_{v \in V_m'} (1 - J_v) f(T_{n,v}) = \sum_{v \in V_m'} f(T_{n,v}) \mathbf{1}\{|T_{n,v}| < \delta n\}. \quad (6.40)$$

For each $v$, it follows from (6.6) by conditioning on $|T_{n,v}|$ that

$$E|f(T_{n,v})\mathbf{1}\{|T_{n,v}| < \delta n\}|^p \leq 2(\delta n)^{p-1}. \quad (6.41)$$

Hence, (6.40) and Minkowski’s inequality yield

$$||Y||_p - ||Y_1||_p \leq \sum_{v \in V_m'} ||f(T_{n,v})\mathbf{1}\{|T_{n,v}| < \delta n\}||_p \leq 2^{m+1/p}(\delta n)^{1 - 1/p}. \quad (6.42)$$

Thus $||Y_1||_p = O(n^{1-1/p}) + O(2^m \delta^{1-1/p} n^{1-1/p})$, and (6.33) yields

$$E|Y|^p - E|Y_1|^p = O((2^m \delta^{1-1/p} + 2^{mp} \delta^{p-1}) n^{p-1}). \quad (6.43)$$

**Step 2.** Similarly, using (6.41) again,

$$E\left(\sum_{v \in V_m'} |f(T_{n,v})|^p - \sum_{v \in V_m'} J_v |f(T_{n,v})|^p\right) = \sum_{v \in V_m'} E\{|f(T_{n,v})|^p \mathbf{1}\{|T_{n,v}| < \delta n\}\} \leq 2^{m+1}(\delta n)^{p-1}. \quad (6.44)$$

**Step 3.** By (6.39), $|Y_1|^p - \sum_{v \in V_m'} |J_v f(T_{n,v})|^p = 0$ unless $\sum_{v \in V_m'} J_v \geq 2$, and in the latter case we have by (3.3) the trivial bounds $|Y_1|^p \leq (2^m n)^p$ and $\sum_{v \in V_m'} |J_v f(T_{n,v})|^p \leq 2^m n^p$, and thus $||Y_1||_p - \sum_{v \in V_m'} |J_v f(T_{n,v})|^p \leq 2^{mp} n^p$. Consequently, by (6.36),

$$E|Y_1|^p - \sum_{v \in V_m'} |J_v f(T_{n,v})|^p \leq 2^{mp} n^p P\left(\sum_{v \in V_m'} J_v \geq 2\right) = O(n^{p-2}). \quad (6.45)$$

Thus, for fixed $m \geq 1$ and $\delta > 0$,

$$E|Y_1|^p - \sum_{v \in V_m'} E|J_v f(T_{n,v})|^p = O\left(n^{p-2}\right) = o(n^{p-1}). \quad (6.46)$$

**Step 4.** Define $F^{(p)}(T) := \sum_{v \in T} |f(T_v)|^p$. Then, in analogy with (6.3),

$$F^{(p)}(T) = \sum_{v \in V_m'} |f(T_v)|^p + \sum_{v \in \partial V_m'} F^{(p)}(T_v). \quad (6.47)$$
Note that Lemma 6.3 implies \( \mathbb{E} F^{(p)}(\mathcal{T}_n) = O(n^{p-1}) \). Hence, by first conditioning on \( \mathcal{F}'_m \), and using (6.19),

\[
\mathbb{E} \sum_{v \in \partial V'_m} F^{(p)}(\mathcal{T}_{n,v}) \leq C_{11} \mathbb{E} \sum_{v \in \partial V'_m} |\mathcal{T}_{n,v}|^{p-1} = C_{11} (2/p)^m n^{p-1}. \tag{6.48}
\]

Taking \( T = \mathcal{T}_n \) in (6.47) and taking the expectation, we thus find

\[
\mathbb{E} \sum_{v \in \mathcal{T}_n} |f(\mathcal{T}_{n,v})|^p - \mathbb{E} \sum_{v \in V'_m} |f(\mathcal{T}_{n,v})|^p = O\left( (2/p)^m n^{p-1} \right). \tag{6.49}
\]

**Step 5.** Finally, combining (6.34), (6.43), (6.46), (6.44), (6.49) and (6.28), we obtain

\[
\mathbb{E} |F(\mathcal{T}_n) - \nu_n|^p = \frac{2p}{p-2} \alpha_p n^{p-1} + O\left( (2/p)^m n^{p-1} \right) + O\left( 2^m \delta^{1-1/p} n^{p-1} \right) + o(n^{p-1}). \tag{6.50}
\]

For any \( \varepsilon > 0 \), we can make each of the error terms on the right-hand side less than \( \varepsilon n^{p-1} \) by first choosing \( m \) large and then \( \delta \) small, and finally \( n \) large. Consequently, \( \mathbb{E} |F(\mathcal{T}_n) - \nu_n|^p = \frac{2p}{p-2} \alpha_p n^{p-1} + o(n^{p-1}) \). \( \square \)

**Proof of (1.4).** Now \( p = k \) is an integer. If \( k \) is even, then (1.4) is the same as (1.8), so we may assume that \( p = k \geq 3 \) is odd.

In this case, (6.33) holds for all real \( x, y \). Thus for any random variables \( X \) and \( Y \), using also Hölder’s inequality,

\[
\mathbb{E} |X^p - Y^p| \leq p \mathbb{E} \left( |X - Y| |X|^{p-1} + |X - Y| |Y|^{p-1} \right) \\
\leq p \|X - Y\|_p (\|X\|_p^{p-1} + \|Y\|_p^{p-1}). \tag{6.51}
\]

It is now easy to modify the proof of Theorem 1.5 and obtain

\[
\mathbb{E} (F(\mathcal{T}_n) - \nu_n)^p = \mathbb{E} \sum_{v \in \mathcal{T}_n} f(\mathcal{T}_{n,v})^p + o(n^{p-1}). \tag{6.52}
\]

Furthermore, it follows from (2.2) that \( f(T) \leq 0 \) unless \( |T| = 1 \). Hence,

\[
\sum_{v \in \mathcal{T}_n} f(\mathcal{T}_{n,v})^p = - \sum_{v \in \mathcal{T}_n} |f(\mathcal{T}_{n,v})|^p + O(n). \tag{6.53}
\]

The estimate (1.4) now follows from (6.52), (6.53) and (6.28). \( \square \)

**7. Proof of Lemma 3.2**

Define a *chain* of length \( k \) in a (binary) tree \( T \) to be a sequence of \( k \) nodes \( v_1 \cdots v_k \) such that \( v_{i+1} \) is a (strict) descendant of \( v_i \) for each \( i = 1, \ldots, k-1 \). In other words, \( v_1, \ldots, v_k \) are some nodes (in order) on some path from the root. We say that the chain \( v_1 \cdots v_k \) is *green* if all nodes \( v_1, \ldots, v_k \) are green. (The nodes between the \( v_i \)'s may have any colour.)
For a binary tree $T$ and $k \geq 1$, let $F_k(T)$ be the number of green chains $v_1 \cdots v_k$ in $T$, and let $f_k(T)$ be the number of such chains where $v_1$ is the root. Obviously, cf. (2.4),

$$F_k(T) = \sum_{v \in T} f_k(T_v). \quad (7.1)$$

These functionals are useful to us because of the following simple relations, that are cases of inclusion-exclusion.

**Lemma 7.1.** For any binary tree $T$,

$$f(T) = \sum_{k=1}^{\infty} (-1)^{k-1} f_k(T), \quad (7.2)$$

$$F(T) = \sum_{k=1}^{\infty} (-1)^{k-1} F_k(T). \quad (7.3)$$

**Proof.** Let $v$ be a node in $T$ and consider the contribution to the sum in (7.3) of all chains with final node $v_k = v$. This is clearly 0 if 1 if $v$ is not green, and it is 1 if $v$ is a maximal green node; furthermore, if $v$ is green but has $j \geq 1$ green ancestors, then the contribution is easily seen to be $\sum_{i=0}^{j} \binom{j}{i} (-1)^i = (1 - 1)^j = 0$. Hence the right-hand side of (7.3) is the number of maximal green nodes, i.e., $F(T)$.

For (7.2) we can argue similarly: Both sides are 0 unless the root $o$ is green. If it is, the chain $o$ gives contribution 1, and by inclusion-exclusion, the chains with a given final node $v \neq o$ yield together a contribution $-1$ if $v$ is green and there are no green nodes between $v$ and $o$, and 0 otherwise. Hence the sum equals $f(T)$ by (2.2). (Alternatively, (7.2) follows by induction from (7.3), (2.4) and (7.1).) \qed

**Lemma 7.2.** For every $k \geq 1$,

$$\mathbb{E} f_k(T) = \frac{k(k+3)}{(k+1)(k+2)} \cdot \frac{2^{k-1}}{k!} = \frac{2^{k-1}}{k!} - \frac{2^k}{(k+2)!}. \quad (7.4)$$

**Proof.** We use the construction of $T = \tilde{T}_\tau$ in Section 2, which we formulate as follows. Consider again the infinite binary tree $T_\infty$, and grow $\tilde{T}_t$ as a subtree of $T_\infty$, cf. Section 5. To do this, we equip each node $v$ in $T_\infty$ with two clocks $C_L(v)$ and $C_R(v)$. These are started when $v$ is added to the growing tree $\tilde{T}_t$, and each chimes after a random time with an exponential distribution with mean 1; when the clock chimes we add a left or right child, respectively, to $v$. There is also a *doomsday clock* $C_0$, started at 0 and with the same Exp(1) distribution; when it chimes (at time $\tau$), the process is stopped and the tree $\tilde{T}_\tau$ is output. All clocks are independent of each other.

Fix a chain $v_1 \cdots v_k$ in the infinite tree $T_\infty$, with $v_1 = o$, the root. Let $l_i \geq 0$ be the number of nodes between $v_i$ and $v_{i+1}$. We compute the probability that $v_1 \cdots v_k$ is a green chain in $T = \tilde{T}_\tau$ by following the construction of $\tilde{T}_t$ as time progresses, checking in several steps whether still $v_1 \cdots v_k$ is a
candidate for a green chain, and computing the probability of this. (We use throughout the proof the Markov property and the memoryless property of the exponential distribution.) We assume for notational convenience that the path from \( v_1 \) to \( v_k \) always uses the left child of each node. (By symmetry, this does not affect the result.)

1. If \( k > 1 \), we first need that \( v_1 = o \) has a left child but no right child (in order to be green); in particular, of the three clocks \( C_L(v_1), C_R(v_1), C_0 \) that run from the beginning, \( C_L(v_1) \) has to chime first. This has probability \( \frac{1}{3} \).

2. Given that Step 1 succeeds, \( v_1 \) gets a left child \( w_1 \). If \( \ell_1 > 0 \), we need a left child of \( w_1 \), and still no right child at \( v_1 \). (But we do not care whether we get a right child at \( w_1 \) or not.) Hence we need that \( C_L(w_1) \) chimes first among the three clocks \( C_L(w_1), C_R(v_1), C_0 \) (ignoring all other clocks). This has probability \( \frac{1}{3} \).

This is repeated for \( \ell_1 \) nodes; thus, the total probability that steps 1 and 2 succeed is \( 3^{-(\ell_1+1)} \).

3. This takes us to \( v_2 \). If \( k > 2 \), we need a left child but no right child at \( v_2 \), and still no right child at \( v_1 \). Hence, the next chime from the four clocks \( C_L(v_2), C_R(v_2), C_R(v_1), C_0 \) has to come from \( C_L(v_2) \). This has probability \( \frac{1}{4} \).

4. Similarly for each of the \( \ell_2 \) nodes between \( v_2 \) and \( v_3 \); again the probability of success at each of these nodes is \( \frac{1}{4} \). Hence the probability that Steps 3 and 4 succeed is \( 4^{-(\ell_2+1)} \).

5. Steps 3 and 4 are repeated for \( v_i \) for each \( i < k \), yielding a probability \( (i + 2)^{-(\ell_i+1)} \) of success for each \( i \).

6. Finally, we have obtained \( v_k \), and wait for the doomsday clock. Until it chimes, we must not get any right child at \( v_1, \ldots, v_{k-1} \), and we must get at most one child at \( v_k \). Hence, among the \( k + 2 \) clocks \( C_R(v_1), \ldots, C_R(v_k), C_L(v_k), C_0 \), the next chime must be either from \( C_0 \) (probability \( \frac{1}{k+2} \)), or from \( C_L(v_k) \) or \( C_R(v_k) \), followed by \( C_0 \) (probability \( \frac{2}{k+2} \cdot \frac{1}{k+1} \)). The probability of success in this step is thus

\[
\frac{1}{k+2} + \frac{2}{k+2} \cdot \frac{1}{k+1} = \frac{k+3}{(k+1)(k+2)}. \quad (7.5)
\]

Combining the six steps above, we see that the probability that \( v_1 \cdots v_k \) is a green chain in \( \mathcal{T}_\tau \) is

\[
\frac{k+3}{(k+1)(k+2)} \prod_{i=1}^{k-1} \left( \frac{1}{i+2} \right)^{\ell_i+1}. \quad (7.6)
\]
Given $\ell_1, \ldots, \ell_{k-1}$, there are $\prod_{i=1}^{k-1} 2^{\ell_i+1}$ choices of the chain $v_1 \cdots v_k$, all with the same probability, so summing over all $\ell_1, \ldots, \ell_{k-1} \geq 0$, we obtain

$$E f_k(T) = \frac{k+3}{(k+1)(k+2)} \prod_{i=1}^{k-1} \sum_{\ell_i=0}^{\infty} \left( \frac{2}{i+2} \right)^{\ell_i+1} = \frac{k+3}{(k+1)(k+2)} \prod_{i=1}^{k-1} \frac{2}{i}.$$ 

$$= \frac{k+3}{(k+1)(k+2)} \cdot \frac{2^{k-1}}{(k-1)!} = \frac{k(k+3)}{(k+1)(k+2)} \cdot \frac{2^{k-1}}{k!}. \quad \Box$$

**Proof of Lemma 3.2.** By Lemmas 7.1 and 7.2, and a simple calculation,

$$E f(T) = \sum_{k=1}^{\infty} (-1)^{k-1} E f_k(T) = \sum_{k=1}^{\infty} \left( \frac{(-2)^{k-1}}{k!} + \frac{(-2)^k}{(k+2)!} \right) = \frac{1-e^{-2}}{4},$$

noting that we may take the expectation inside the sum since it also follows from Lemma 7.2 that $\sum_{k=1}^{\infty} E |f_k(T)| = \sum_{k=1}^{\infty} E f_k(T) < \infty$. \quad \Box

Recall that this, together with Lemma 3.1, completes our probabilistic proof of Theorem 1.1.

**Remark 7.3.** If we in the proof above change the doomsday clock and let it have an arbitrary rate $\lambda > 0$, and denote the resulting random binary tree by $T^{(\lambda)}$, then the same argument yields

$$E f_k(T^{(\lambda)}) = \frac{k+\lambda+2}{(k+\lambda)(k+\lambda+1)} \prod_{i=1}^{k-1} \sum_{\ell_i=0}^{\infty} \left( \frac{2}{i+\lambda+1} \right)^{\ell_i+1}$$

$$= \frac{k+\lambda+2}{(k+\lambda)(k+\lambda+1)} \prod_{i=1}^{k-1} \frac{2}{i+\lambda-1}$$

$$= \frac{(k+\lambda-1)(k+\lambda+2)}{(k+\lambda)(k+\lambda+1)} \cdot \frac{2^{k-1}}{\lambda^{k+1}}$$

$$= \frac{2^{k-1}}{\lambda^{k+2}}.$$  

Thus by Lemma 7.1, letting $\,_1F_1$ denote the confluent hypergeometric function, see e.g. [18, §§13.1–13.2 and 16.1–16.2],

$$E f(T^{(\lambda)}) = \sum_{k=1}^{\infty} (-1)^{k-1} E f_k(T^{(\lambda)}) = \sum_{k=1}^{\infty} \left( \frac{(-2)^{k-1}}{\lambda^k} + \frac{(-2)^k}{\lambda^{k+2}} \right)$$

$$= -\frac{1}{2} \left( \,_1F_1(1; \lambda; -2) - 1 \right) + \frac{1}{4} \left( \,_1F_1(1; \lambda; -2) - \left( 1 - \frac{2}{\lambda} + \frac{2 \cdot 2}{\lambda(\lambda+1)} \right) \right)$$

$$= \frac{1}{4} + \frac{\lambda-1}{2\lambda(\lambda+1)} - \frac{1}{4} \,_1F_1(1; \lambda; -2). \quad (7.8)$$

Furthermore, if $\lambda > 1$ we can compute $E F(T^{(\lambda)})$ by the same method; the only difference is that we also allow a path of length $\ell_0 \geq 0$ from the root to
$v_1$, which gives an additional factor $(1 + \lambda)^{-\ell_0}$ for each $v_1 \cdots v_k$, leading to

$$\mathbb{E} F_k(T^{(\lambda)}) = \sum_{\ell_0=0}^{\infty} \left( \frac{2}{\lambda + 1} \right)^{\ell_0} \mathbb{E} f_k(T^{(\lambda)}) = \frac{\lambda + 1}{\lambda - 1} \mathbb{E} f_k(T^{(\lambda)}), \quad (7.9)$$

and hence, using both parts of Lemma 7.1,

$$\mathbb{E} F(T^{(\lambda)}) = \sum_{k=1}^{\infty} (-1)^{k-1} \mathbb{E} F_k(T^{(\lambda)}) = \frac{\lambda + 1}{\lambda - 1} \mathbb{E} f(T^{(\lambda)}). \quad (7.10)$$

Moreover, a simple argument shows that, for any $n \geq 1$,

$$\mathbb{P}(|T^{(\lambda)}| = n) = \prod_{i=2}^{n} \frac{i}{i + \lambda} \cdot \frac{\lambda}{n + 1 + \lambda} = \frac{\lambda n!}{(2 + \lambda)^n}, \quad (7.11)$$

and conditioned on $|T^{(\lambda)}| = n$, $T^{(\lambda)}$ has the same distribution as $T_n$, i.e., $(T^{(\lambda)} | |T^{(\lambda)}| = n) \overset{d}{=} T_n$. Hence,

$$\mathbb{E} F(T^{(\lambda)}) = \sum_{n=1}^{\infty} \frac{\lambda n!}{(2 + \lambda)^n} \nu_n, \quad (7.12)$$

which can be interpreted as an unusual type of generating function for the sequence $(\nu_n)$; note that (7.10) and (7.8) yield an explicit expression for it.

REFERENCES

[1] David Aldous, Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* 1 (1991), no. 2, 228–266.

[2] David Aldous, Probability distributions on cladograms. In *Random Discrete Structures (Minneapolis, MN, 1993)*, 1–18, IMA Vol. Math. Appl., 76, Springer, New York, 1996.

[3] Michael G. B. Blum and Olivier François, Minimal clade size and external branch length under the neutral coalescent. *Adv. in Appl. Probab.* 37 (2005), no. 3, 647–662.

[4] Michael G. B. Blum, Olivier François and Svante Janson, The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance. *Ann. Appl. Probab.* 16 (2006), no. 4, 2195–2214.

[5] B. M. Brown and G. K. Eagleson, Martingale convergence to infinitely divisible laws with finite variances. *Trans. Amer. Math. Soc.* 162 (1971), 449–453.

[6] Huilan Chang and Michael Fuchs, Limit theorems for patterns in phylogenetic trees. *J. Math. Biol.* 60 (2010), no. 4, 481–512.

[7] Luc Devroye, Limit laws for local counters in random binary search trees. *Random Structures Algorithms* 2 (1991), no. 3, 303–315.

[8] Luc Devroye, Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.* 32 (2002/03), no. 1, 152–171.

[9] Michael Drmota, *Random Trees*. Springer, Vienna, 2009.
[10] Michael Drmota, Michael Fuchs and Yi-Wen Lee, Limit laws for the number of groups formed by social animals under the extra clustering model. (Extended abstract.) Proceedings, 2014 Conference on Analysis of Algorithms, AofA ’14 (Paris, 2014), DMTCS Proceedings, 2014.

[11] Eric Durand, Michael G. B. Blum and Olivier François, Prediction of group patterns in social mammals based on a coalescent model. *J. Theoret. Biol.* **249** (2007), no. 2, 262–270.

[12] Eric Durand and Olivier François, Probabilistic analysis of a genealogical model of animal group patterns. *J. Math. Biol.* **60** (2010), no. 3, 451–468.

[13] Philippe Flajolet, Xavier Gourdon and Conrado Martínez, Patterns in random binary search trees. *Random Structures Algorithms* **11** (1997), no. 3, 223–244.

[14] Allan Gut, *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[15] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*. Academic Press, New York, 1980.

[16] Cecilia Holmgren and Svante Janson, Limit laws for functions of fringe trees for binary search trees and recursive trees. Preprint, 2014. arXiv: 1406.6883v1

[17] J. F. C. Kingman, The coalescent. *Stochastic Process. Appl.* **13** (1982), no. 3, 235–248.

[18] *NIST Handbook of Mathematical Functions*. Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert and Charles W. Clark. Cambridge Univ. Press, 2010. Also available as *NIST Digital Library of Mathematical Functions*, http://dlmf.nist.gov/