

Electron. J. Probab. **20** (2015), no. 4, 1–51.  
ISSN: 1083-6489 DOI: 10.1214/EJP.v20-3627

## Limit laws for functions of fringe trees for binary search trees and random recursive trees

Cecilia Holmgren\*      Svante Janson†

### Abstract

We prove general limit theorems for sums of functions of subtrees of (random) binary search trees and random recursive trees. The proofs use a new version of a representation by Devroye, and Stein's method for both normal and Poisson approximation together with certain couplings.

As a consequence, we give simple new proofs of the fact that the number of fringe trees of size  $k = k_n$  in the binary search tree or in the random recursive tree (of total size  $n$ ) has an asymptotical Poisson distribution if  $k \rightarrow \infty$ , and that the distribution is asymptotically normal for  $k = o(\sqrt{n})$ . Furthermore, we prove similar results for the number of subtrees of size  $k$  with some required property  $P$ , e.g., the number of copies of a certain fixed subtree  $T$ . Using the Cramér–Wold device, we show also that these random numbers for different fixed subtrees converge jointly to a multivariate normal distribution.

We complete the paper by giving examples of applications of the general results, e.g., we obtain a normal limit law for the number of  $\ell$ -protected nodes in a binary search tree or in a random recursive tree.

**Keywords:** Fringe trees; Stein's method; Couplings; Limit laws; Binary search trees; Random recursive trees.

**AMS MSC 2010:** Primary 60C05, Secondary 05C05; 60F05.

Submitted to EJP on June 26, 2014, final version accepted on January 3, 2015.

Supersedes arXiv:1406.6883.

## 1 Introduction

In this paper we consider fringe trees of the random binary search tree as well as of the random recursive tree; recall that a *fringe tree* is a subtree consisting of some node and all its descendants, see Aldous [1] for a general theory, and note that fringe trees typically are "small" compared to the whole tree. (All subtrees considered in the present paper are of this type, and we will use 'subtree' and 'fringe tree' as synonyms.) We will use a representation of Devroye [10, 11] for the random binary search tree, and a well-known bijection between binary trees and recursive trees, together with different

---

\*Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden.

E-mail: cecilia@math.su.se Supported by the Swedish Research Council.

†Department of Mathematics, Uppsala University, SE-75310 Uppsala, Sweden.

E-mail: svante.janson@math.uu.se Supported in part by the Knut and Alice Wallenberg Foundation.

applications of Stein’s method for both normal and Poisson approximation to give both new general results on the asymptotic distributions for random variables depending on fringe trees, and more direct proofs of several earlier results in the field. We state the main results in this section: More precisely, we identify two main results that we present in Section 1.3, one on Poisson approximation and one on normal approximation, for general functions of fringe trees for the random binary search tree, as well as for the random recursive tree. In Section 1.4 we then give some important consequences and extensions of these results.

In Section 2 we describe the representation of the binary search tree by Devroye, we explain how it can be extended to the random recursive tree, and we also extend it to a cyclic version. Furthermore, in Section 3 we state and prove general results on means and variances of functions of fringe trees. Proofs of the results in Section 1 are given in Sections 4–7. We also give some examples of applications of the general results on functions of fringe trees in Section 8, for example to the number of protected nodes in the random binary search tree or in the random recursive tree studied by Mahmoud and Ward [35, 36] and to the shape functionals for the random binary search tree or the random recursive tree.

### 1.1 Binary search trees and random recursive trees

A *binary search tree* is the tree representation of the sorting algorithm Quicksort, see e.g. [33] or [14]. Starting with  $n$  distinct ordered numbers called keys, we associate the first key to the root. The second key is then compared with the root, and is associated to the left child if it is smaller than the key at the root, and to the right child if it is larger. We continue to add the remaining keys to the tree until the set is exhausted. The comparison for each new key starts at the root, and at each node the key visits, it proceeds to the left/right child if it is smaller/larger than the key associated to that node; eventually, the new key is associated to the first empty node it visits. In the final tree, all the  $n$  ordered numbers are sorted by size, so that smaller numbers are in left subtrees, and larger numbers are in right subtrees. A binary tree with such a labelling of the nodes is called a binary search tree.

The *random binary search tree* with  $n$  nodes, which we denote by  $\mathcal{T}_n$ , is the binary search tree constructed from a (uniformly) random permutation of  $n$  keys. For convenience, we slightly abuse the notation and follow the custom of writing just “binary search tree” also for this random version.

We use the representation of the binary search tree by Devroye [10, 11]. We may clearly assume that the keys are  $1, \dots, n$ . We assign, independently, each key  $k$  a uniform random variable  $U_k$  in  $(0, 1)$  which we regard as a time stamp indicating the time when the key is drawn. (We may and will assume that the  $U_k$  are distinct.) The binary search tree constructed by drawing the keys in this order, i.e., in order of increasing  $U_k$ , then is the unique binary tree with nodes labelled by  $(1, U_1), \dots, (n, U_n)$  with the property that it is a binary search tree with respect to the first coordinates in the pairs, and along every path down from the root the values  $U_i$  are increasing. We will also use a cyclic version of this representation described in Section 2.3.

Recall that the *random recursive tree* is constructed recursively, by starting with a root with label 1, and at stage  $i$  ( $i = 2, \dots, n$ ) a new node with label  $i$  is attached uniformly at random to one of the previous nodes  $1, \dots, i - 1$ . We let  $\Lambda_n$  denote a random recursive tree with  $n$  nodes. We may regard the random recursive tree as an ordered tree by ordering the children of each node by their labels, from left to right.

There is a well-known bijection between ordered trees of size  $n$  and binary trees of size  $n - 1$ , see e.g. Knuth [32, Section 2.3.2] who calls this the *natural correspondence* (the same bijection is also called the *rotation correspondence*): Given an ordered tree with  $n$  nodes, eliminate first the root, and arrange all its children in a path from left to right, as right children of each other. Continue recursively, with the children of each

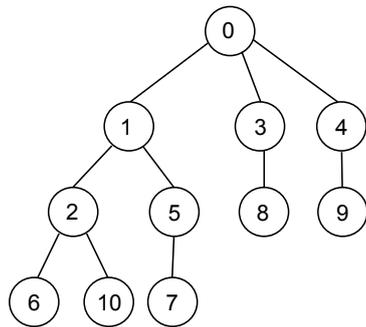


Figure 1: A recursive tree. The root has label 0 instead of 1, to better illustrate the bijection.

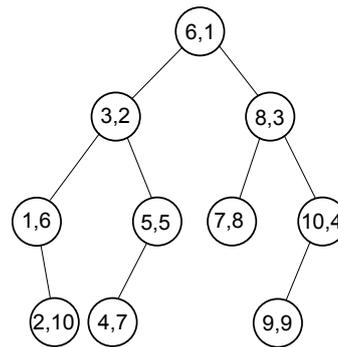


Figure 2: The corresponding binary search tree. The first and second labels are the keys and time stamps, respectively, using the time stamps  $1, \dots, 10$  for convenience.

node arranged in a path from left to right, with the first child attached to its parent as the left child. This yields a binary tree with  $n - 1$  nodes, and the transformation is invertible. As noted by Devroye [10], see also Fuchs, Hwang and Neininger [24], the natural correspondence extends to a coupling between the random recursive tree  $\Lambda_n$  and the binary search tree  $\mathcal{T}_{n-1}$ ; the probability distributions are equal by induction because the  $n$  possible places to add a new node to  $\Lambda_n$  correspond to the  $n$  possible places (external leaves) to add a new node to  $\mathcal{T}_{n-1}$ , and these places have equal probabilities for both models.

Note that a left child in the binary search tree corresponds to an eldest child in the random recursive tree, while a right child corresponds to a sibling. We say that a proper subtree in a binary tree is left-rooted [right-rooted] if its root is a left [right] child. Thus, for  $1 < k < n$ , subtrees of size  $k$  in the random recursive tree  $\Lambda_n$ , correspond to left-rooted subtrees of size  $k - 1$  in the binary search tree  $\mathcal{T}_{n-1}$ , while subtrees of size 1 (i.e., leaves) correspond to nodes without left child. (Alternatively, we can say that subtrees of size 1 in the recursive tree correspond to empty left subtrees in the binary tree.)

An example of a bijection obtained from the natural correspondence is illustrated in Figures 1–2. Note that the labels in the random recursive tree correspond to the time stamps in the binary search tree (replaced by  $1, 2, \dots$  in increasing order), while the keys in the binary search tree are determined by the tree structure and thus redundant.

**Remark 1.1.** The binary search tree with its time stamps and the random recursive tree with its labels are both increasing trees, i.e., labelled trees where the label of a node is greater than the label of its parent. We allow the labels in an increasing tree to be arbitrary real numbers, but we are only interested in the order relations between them and consider two increasing trees that are isomorphic as trees and with labels in the same order to be the same; hence we may freely relabel (preserving the order), for example by  $1, 2, \dots$

Note that the binary search tree yields a uniformly distributed increasing binary tree, and the random recursive tree a uniformly distributed (unordered) increasing tree, see e.g. [14, Sections 1.3–1.4]. Note also that the natural correspondence extends to a bijection between increasing binary trees and increasing ordered trees that have the children of each node ordered according to their labels.

**Remark 1.2.** We may consider a subtree of the binary search tree in two different ways; either we regard it as an (unlabelled) binary tree by ignoring the time stamps (and keys),

or we may regard it as an increasing binary tree by keeping the time stamps (perhaps replacing them, in order, by  $1, 2, \dots$ , see Remark 1.1).

Similarly, there are three different ways to look at a subtree of a random recursive tree: as an increasing tree, as an unlabelled ordered tree (ignoring the labels but keeping the order defined by them), or as an unlabelled unordered rooted tree (by ignoring both labels and ordering).

The theorems and other results below, and their proofs, apply (unless explicitly stated otherwise) to all these interpretations. (The different interpretations may be useful in different applications.) For convenience, we state most results for the unlabelled versions (which seem to be more common in applications), and leave the versions with increasing trees to the reader. Recall also that an ordered tree can be regarded as unordered by ignoring the orderings.

### 1.2 Properties and functionals of fringe trees

Our main theorems are stated in terms of a property  $P$  or a functional  $f$  of trees.

By a *property*  $P$ , in the binary tree case we formally mean any set of binary trees; we let  $P_k$  be the set of binary trees of size  $k$  in  $P$  and we let  $p_{k,P} := \mathbb{P}(\mathcal{T}_k \in P)$ . Similarly, in the random recursive tree case, a property  $P$  is any set of ordered (or unordered) trees, and  $\hat{p}_{k,P} := \mathbb{P}(\Lambda_k \in P)$ . (As said in Remark 1.2, we may also, more generally, let  $P$  be a set of increasing binary or unordered trees.)

Let  $X_{n,k}^P$  be the number of subtrees of size  $k$  with some given property  $P$  in the binary search tree  $\mathcal{T}_n$ , and similarly let  $\hat{X}_{n,k}^P$  be the number of subtrees of size  $k$  with some given property  $P$  in the random recursive tree  $\Lambda_n$ . As a special case (with  $P$  the set of all trees), we let  $X_{n,k}$  be the number of subtrees of size  $k$  in  $\mathcal{T}_n$ , and  $\hat{X}_{n,k}$  be the number of subtrees of size  $k$  in  $\Lambda_n$ . Another important example of  $X_{n,k}^P$  is the number of subtrees of the binary search tree  $\mathcal{T}_n$  that are copies of a fixed binary tree  $T$ , which we denote by  $X_n^T$ ; similarly we denote by  $\hat{X}_n^\Lambda$  the number of copies of an ordered (or unordered) tree  $\Lambda$  in the random recursive tree  $\Lambda_n$ .

**Example 1.3.** Further examples come from evolutionary biology, where it is important to study such *tree patterns* in phylogenetic trees. A *phylogenetic tree*, more precisely a *cladogram*, is a (non-ordered) tree where every node has outdegree 2 (internal nodes) or 0 (external nodes). A binary tree of size  $n$  yields a phylogenetic tree with  $n$  internal nodes by adding  $n + 1$  external nodes. An important model for a random phylogenetic tree is the *Yule model*, which gives the same distribution as the correspondence just described applied to a binary search tree, see e.g. [2] and [5]. Hence, fringe trees in random phylogenetic trees under the Yule model correspond to fringe trees in binary search trees, and our results can be translated. (Note that the size of a phylogenetic tree usually is defined as the number of external nodes; a phylogenetic tree of size  $k$  thus corresponds to a binary search tree with  $k - 1$  nodes.) Some important examples of tree patterns that have been studied are *k-pronged nodes* (trees of size  $k$ ), *k-caterpillars* (trees of size  $k$  such that the internal nodes form a path) and *minimal clade size k* (trees of size  $k$  with either left or right subtree of the root empty); see e.g., [7] and the references there.

**Remark 1.4.** If  $A_k^P$  is the set of permutations of length  $k$  that give rise to binary search trees of size  $k$  with the property  $P$ , then  $p_{k,P} = |A_k^P|/k!$ . In particular, if  $P_k$  is nonempty, then  $1 \geq p_{k,P} \geq 1/k!$ . Similarly,  $1 \geq \hat{p}_{k,P} \geq 1/(k - 1)!$  unless  $\hat{p}_{k,P} = 0$ .

A *functional* is any real-valued function, of (binary, ordered or unordered) rooted trees. (Again, we may also, more generally, consider functionals of increasing trees.) Given a functional  $f$  and a rooted tree  $T$ , let  $T(v)$  be the fringe tree rooted at the node  $v \in T$ , and define the sum over all fringe trees

$$F(T) = F(T; f) := \sum_{v \in T} f(T(v)). \tag{1.1}$$

**Remark 1.5.** Functionals  $F$  that can be written as (1.1) for some  $f$  are called *additive functionals*. They can also be defined recursively by

$$F(T) = f(T) + F(T_1) + \dots + F(T_d), \tag{1.2}$$

where  $T_1, \dots, T_d$  are the subtrees rooted at the children of the root of  $T$ . In this context,  $f(T)$  is often called a *toll function*.

We consider the random variables  $F(\mathcal{T}_n)$  and  $F(\Lambda_n)$ , where as above  $\mathcal{T}_n$  and  $\Lambda_n$  are the binary search tree and the random recursive tree, respectively.

Note that the formulations using properties and functionals are closely related. In fact, if  $P$  is a property and we choose  $f(T)$  as the indicator function  $\mathbf{1}\{T \in P_k\}$ , then  $X_{n,k}^P = F(\mathcal{T}_n)$ ; in particular,  $X_{n,k} = F(\mathcal{T}_n)$  with  $f(T) = \mathbf{1}\{|T| = k\}$  and  $X_n^T = F(\mathcal{T}_n)$  with  $f(T') = \mathbf{1}\{T' = T\}$ . Conversely, for any  $f$ ,

$$F(\mathcal{T}_n) = \sum_T f(T) X_n^T, \tag{1.3}$$

summing over all binary trees  $T$ . The recursive tree case is similar.

Generally speaking, the formulation with properties is more convenient for results with an asymptotic Poisson distribution, while the formulation with functionals is more general and flexible for cases with asymptotic normality. For example, by the Cramér–Wold device, it implies immediately also multivariate results as Theorem 1.22 below, see Remark 1.27.

We refer to Devroye [11] for several other examples showing the generality of the functionals  $F(\mathcal{T}_n)$ .

### 1.3 Main results

In this subsection we state two main results: Theorem 1.7 on Poisson approximation for  $X_{n,k}^P$  and  $\hat{X}_{n,k}^P$  for a property  $P$ , and Theorem 1.14 on normal limits for  $F(\mathcal{T}_n)$  and  $F(\Lambda_n)$  for a functional  $f$  (under some conditions).

We let  $\mathcal{L}(X)$  denote the distribution of a random variable  $X$ . Let  $\text{Po}(\mu)$  denote the Poisson distribution with mean  $\mu$ , and let  $\mathcal{N}(0, 1)$  denote the standard normal distribution. Convergence in distribution is denoted by  $\xrightarrow{d}$ . We recall also the definition of the total variation distance between two probability measures.

**Definition 1.6.** Let  $(\mathcal{X}, \mathcal{A})$  be any measurable space. The total variation distance  $d_{TV}$  between two probability measures  $\mu_1$  and  $\mu_2$  on  $\mathcal{X}$  is defined to be

$$d_{TV}(\mu_1, \mu_2) := \sup_{A \in \mathcal{A}} |\mu_1(A) - \mu_2(A)|.$$

Chang and Fuchs [7] studied fringe trees in random phylogenetic trees. The following theorem (the binary search tree case) improves the convergence rate in their Theorem 9 and yields in a simple way the rate stated in their Remark 1.

**Theorem 1.7.** Let  $X_{n,k}^P$  be the number of subtrees of size  $k$  with some given property  $P$  in the binary search tree  $\mathcal{T}_n$ , and similarly let  $\hat{X}_{n,k}^P$  be the number of subtrees of size  $k$  with some given property  $P$  in the random recursive tree  $\Lambda_n$ . Let  $p_{k,P}$  be the probability that a binary search tree of size  $k$  has property  $P$ , and similarly let  $\hat{p}_{k,P}$  be the probability that a random recursive tree of size  $k$  has property  $P$ . Let  $k = k_n$  where  $k < n$ . Furthermore, let  $\mu_{n,k}^P := \mathbb{E}(X_{n,k}^P)$  and  $\hat{\mu}_{n,k}^P := \mathbb{E}(\hat{X}_{n,k}^P)$ . Then

$$\mu_{n,k}^P := \mathbb{E}(X_{n,k}^P) = \frac{2(n+1)p_{k,P}}{(k+1)(k+2)}, \tag{1.4}$$

$$\hat{\mu}_{n,k}^P := \mathbb{E}(\hat{X}_{n,k}^P) = \frac{n\hat{p}_{k,P}}{k(k+1)}. \tag{1.5}$$

Then, for the binary search tree, if  $k \neq (n - 1)/2$ ,

$$d_{TV}(\mathcal{L}(X_{n,k}^P), \text{Po}(\mu_{n,k}^P)) = \frac{1}{2} \sum_{l \geq 0} \left| \mathbb{P}(X_{n,k}^P = l) - e^{-\mu_{n,k}^P} \frac{(\mu_{n,k}^P)^l}{l!} \right|$$

$$= \begin{cases} O\left(\frac{p_{k,P}}{k}\right) & \text{if } \mu_{n,k}^P \geq 1 \\ O\left(\frac{p_{k,P}}{k} \cdot \mu_{n,k}^P\right) & \text{if } \mu_{n,k}^P < 1, \end{cases} \quad (1.6)$$

and if  $k = (n - 1)/2$ ,

$$d_{TV}(\mathcal{L}(X_{n,k}^P), \text{Po}(\mu_{n,k}^P)) = O\left(\frac{p_{k,P}^2}{k}\right). \quad (1.7)$$

For the random recursive tree,

$$d_{TV}(\mathcal{L}(\hat{X}_{n,k}^P), \text{Po}(\hat{\mu}_{n,k}^P)) = \frac{1}{2} \sum_{l \geq 0} \left| \mathbb{P}(\hat{X}_{n,k}^P = l) - e^{-\hat{\mu}_{n,k}^P} \frac{(\hat{\mu}_{n,k}^P)^l}{l!} \right|$$

$$= \begin{cases} O\left(\frac{\hat{p}_{k,P}}{k}\right) & \text{if } \hat{\mu}_{n,k}^P \geq 1 \\ O\left(\frac{\hat{p}_{k,P}}{k} \cdot \hat{\mu}_{n,k}^P\right) & \text{if } \hat{\mu}_{n,k}^P < 1. \end{cases} \quad (1.8)$$

Consequently, if  $n \rightarrow \infty$  and  $k \rightarrow \infty$ , then  $d_{TV}(\mathcal{L}(X_{n,k}^P), \text{Po}(\mu_{n,k}^P)) \rightarrow 0$ , and similarly  $d_{TV}(\mathcal{L}(\hat{X}_{n,k}^P), \text{Po}(\hat{\mu}_{n,k}^P)) \rightarrow 0$ .

**Remark 1.8.** Theorem 1.7 implies asymptotic normality in all cases when  $k \rightarrow \infty$  and  $\mu_{n,k}^P \rightarrow \infty$  or  $\hat{\mu}_{n,k}^P \rightarrow \infty$ . Asymptotic normality holds for  $k = O(1)$  too, see Examples 1.24 and 1.28 below. For the binary tree, the asymptotic normality in these cases was proved by Devroye [10, Theorem 1] ( $k$  fixed) and [11, Theorem 5] (at least for  $k = o(\log n / \log \log n)$  which implies that  $\mu_{n,k}^P \rightarrow \infty$  for every  $P$  unless  $P_k$  is empty).

Before stating Theorem 1.14, we give (exact) expressions for the covariances between  $X_n^T$  and  $X_n^{T'}$  (i.e., the number of subtrees that are copies of a fixed tree  $T$  respectively  $T'$  in the binary search tree), and similar expression for the random recursive tree. (The variances, i.e. the special case  $T = T'$ , in the binary case were found by [21].) As said in Remark 1.2, these results hold also if  $T$  or  $\Lambda$  is a given increasing tree, and for the random recursive tree we may let  $\Lambda$  be either an ordered or unordered tree; the results are valid for all cases, but note that e.g.  $\hat{p}_{k,\Lambda}$  and  $\hat{q}_{\Lambda}^{\Lambda}$  depend on the version. (In particular, for increasing trees  $T$  and  $\Lambda$ ,  $p_{k,T} = 1/k!$  and  $\hat{p}_{k,\Lambda} = 1/(k - 1)!.$ )

**Proposition 1.9.** Let  $T$  be a binary tree of size  $k$  and let  $T'$  be a binary tree of size  $m$  where  $m \leq k$ . Let  $p_{k,T} := \mathbb{P}(\mathcal{T}_k = T)$  and  $p_{m,T'} := \mathbb{P}(\mathcal{T}_m = T')$ , and let  $q_{T'}^T$  be the number of subtrees of  $T$  that are copies of  $T'$ ; further, let

$$\beta(k, m) := \frac{4(k + m + 3)}{(k + 1)(k + 2)(m + 1)(m + 2)} - \frac{4(k^2 + 3km + m^2 + 4k + 4m + 3)}{(k + 1)(m + 1)(k + m + 1)(k + m + 2)(k + m + 3)}. \quad (1.9)$$

If  $n > k + m + 1$ , then the covariance between  $X_n^T$  and  $X_n^{T'}$  is equal to

$$\text{Cov}(X_n^T, X_n^{T'}) = (n + 1)\sigma_{T,T'}, \quad (1.10)$$

where

$$\sigma_{T,T'} := \frac{2}{(k + 1)(k + 2)} q_{T'}^T p_{k,T} - \beta(k, m) p_{k,T} p_{m,T'}. \quad (1.11)$$

We note also the corresponding result for  $X_{n,k}$ , combining all subtrees of the same size. (The variances  $\sigma_{k,k}$  are given by Feng, Mahmoud and Panholzer [16], as well as higher moments, and the covariances are given by Dennert and Grübel [9].)

**Proposition 1.10** (Dennert and Grübel [9]). *Let  $k, m \geq 1$  and suppose  $n > k + m + 1$ . The covariance between  $X_{n,k}$  and  $X_{n,m}$  is equal to*

$$\text{Cov}(X_{n,k}, X_{n,m}) = (n + 1)\sigma_{k,m} \tag{1.12}$$

where  $\sigma_{k,m} = \sigma_{m,k}$  and

$$\sigma_{k,m} = -\frac{4m(2k + m + 3)}{(k + 1)(k + 2)(k + m + 1)(k + m + 2)(k + m + 3)}, \quad m < k, \tag{1.13}$$

$$\sigma_{k,k} = \frac{2k(4k^2 + 5k - 3)}{(k + 1)(k + 2)^2(2k + 1)(2k + 3)}. \tag{1.14}$$

For the random recursive tree we have similar results.

**Proposition 1.11.** *Let  $\Lambda$  be an ordered [or unordered] tree of size  $k$ , and let  $\Lambda'$  be an ordered [or unordered] tree of size  $m$  where  $m \leq k$ .*

*Let  $\hat{p}_{k,\Lambda} := \mathbb{P}(\Lambda_k = \Lambda)$  and  $\hat{p}_{m,\Lambda'} := \mathbb{P}(\Lambda_m = \Lambda')$ , and let  $\hat{q}_{\Lambda}^{\Lambda'}$  be the number of subtrees of  $\Lambda'$  that are copies of  $\Lambda$ ; further, let*

$$\hat{\beta}(k, m) := \frac{k^2 + km + m^2 + k + m}{k(k + 1)m(m + 1)(k + m + 1)}. \tag{1.15}$$

*If  $n > k + m$ , then the covariance between  $\hat{X}_n^{\Lambda}$  and  $\hat{X}_n^{\Lambda'}$  is equal to*

$$\text{Cov}(\hat{X}_n^{\Lambda}, \hat{X}_n^{\Lambda'}) = n\hat{\sigma}_{\Lambda,\Lambda'}, \tag{1.16}$$

where

$$\hat{\sigma}_{\Lambda,\Lambda'} := \frac{1}{k(k + 1)}\hat{q}_{\Lambda}^{\Lambda'}\hat{p}_{k,\Lambda} - \hat{\beta}(k, m)\hat{p}_{k,\Lambda}\hat{p}_{m,\Lambda'}. \tag{1.17}$$

**Remark 1.12.** Note that using the natural correspondence between ordered trees and binary trees, if  $\Lambda$  corresponds to the binary tree  $T$  of size  $k - 1$ , then  $\hat{p}_{k,\Lambda} = p_{k-1,T}$ .

**Proposition 1.13.** *Let  $k, m \geq 1$  and suppose  $n > k + m$ . The covariance between  $\hat{X}_{n,k}$  and  $\hat{X}_{n,m}$  is equal to*

$$\text{Cov}(\hat{X}_{n,k}, \hat{X}_{n,m}) = n\hat{\sigma}_{k,m} \tag{1.18}$$

where  $\hat{\sigma}_{k,m} = \hat{\sigma}_{m,k}$  and

$$\hat{\sigma}_{k,m} = -\frac{1}{k(k + 1)(k + m + 1)}, \quad m < k, \tag{1.19}$$

$$\hat{\sigma}_{k,k} = \frac{2k^2 - 1}{k(k + 1)^2(2k + 1)}. \tag{1.20}$$

We are now ready to state our general limit theorem for  $F(\mathcal{T}_n)$  and  $F(\Lambda_n)$ .

**Theorem 1.14.** *Let  $F$  be given by (1.1) for some functional  $f$ .*

(i) *For the binary search tree, assume that*

$$\sum_{k=1}^{\infty} \frac{(\text{Var } f(\mathcal{T}_k))^{1/2}}{k^{3/2}} < \infty, \tag{1.21}$$

$$\lim_{k \rightarrow \infty} \frac{\text{Var } f(\mathcal{T}_k)}{k} = 0, \tag{1.22}$$

$$\sum_{k=1}^{\infty} \frac{(\mathbb{E} f(\mathcal{T}_k))^2}{k^2} < \infty. \tag{1.23}$$

Then, as  $n \rightarrow \infty$ ,

$$\mathbb{E}(F(\mathcal{T}_n))/n \rightarrow \mu_F := \sum_{k=1}^{\infty} \frac{2}{(k+1)(k+2)} \mathbb{E} f(\mathcal{T}_k), \tag{1.24}$$

$$\text{Var}(F(\mathcal{T}_n))/n \rightarrow \sigma_F^2 := \lim_{N \rightarrow \infty} \sum_{|T|, |T'| \leq N} f(T)f(T')\sigma_{T,T'} < \infty \tag{1.25}$$

and

$$\frac{F(\mathcal{T}_n) - \mathbb{E} F(\mathcal{T}_n)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_F^2). \tag{1.26}$$

(ii) For the random recursive tree, assume that

$$\sum_{k=1}^{\infty} \frac{(\text{Var} f(\Lambda_k))^{1/2}}{k^{3/2}} < \infty, \tag{1.27}$$

$$\lim_{k \rightarrow \infty} \frac{\text{Var} f(\Lambda_k)}{k} = 0, \tag{1.28}$$

$$\sum_{k=1}^{\infty} \frac{(\mathbb{E} f(\Lambda_k))^2}{k^2} < \infty. \tag{1.29}$$

Then, as  $n \rightarrow \infty$ ,

$$\mathbb{E}(F(\Lambda_n))/n \rightarrow \hat{\mu}_F := \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \mathbb{E} f(\Lambda_k), \tag{1.30}$$

$$\text{Var}(F(\mathcal{T}_n))/n \rightarrow \hat{\sigma}_F^2 := \lim_{N \rightarrow \infty} \sum_{|\Lambda|, |\Lambda'| \leq N} f(\Lambda)f(\Lambda')\hat{\sigma}_{\Lambda, \Lambda'} < \infty \tag{1.31}$$

and

$$\frac{F(\Lambda_n) - \mathbb{E} F(\Lambda_n)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}_F^2). \tag{1.32}$$

**Corollary 1.15.** Let  $F$  be given by (1.1) for some functional  $f$  such that  $f(T) = O(|T|^\alpha)$  for some  $\alpha < 1/2$ . Then the conclusions (1.24)–(1.26) and (1.30)–(1.32) hold. Furthermore, the asymptotic normality (1.26) can be written as

$$\frac{F(\mathcal{T}_n) - n\mu_F}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_F^2) \tag{1.33}$$

and similarly, (1.32) can be written

$$\frac{F(\Lambda_n) - n\hat{\mu}_F}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}_F^2). \tag{1.34}$$

**Remark 1.16.** For the binary search tree and  $f(T)$  depending on the size  $|T|$  only, Corollary 1.15 was proved by methods similar to the ones used here by Devroye [11, Theorem 6] (under somewhat stronger hypotheses), and also by Hwang and Neininger [28] using the contraction method (somewhat more generally, and with a somewhat different expression for  $\sigma_F^2$ ); Hwang and Neininger [28] also show that, for example,  $f(T) = |T|^\alpha$  with  $\alpha > 1/2$  yields different limit behaviour. See further Fill and Kapur [20] for similar results (extended to general  $m$ -ary search trees). Cf. also Fill, Flajolet and Kapur [18, Theorem 13] for related results for the mean.

A well-known case when  $f$  grows too rapidly for the results above to hold is  $f(T) = |T|$ , when  $F(T)$  is the total path length in the tree. In this case, for the binary search tree, the expectation grows like  $2n \log n$  and the limit is non-normal, see Régnier [39], Rösler [40], Fill and Janson [19].

**Remark 1.17.** Of course, (1.25) means that (summing over all binary trees)

$$\sigma_F^2 = \sum_{T, T'} f(T)f(T')\sigma_{T, T'}, \tag{1.35}$$

provided this sum is absolutely convergent. However, this fails in general, even if  $f$  is bounded, since, as is shown in the appendix,

$$\sum_{T, T'} |\sigma_{T, T'}| = \infty. \tag{1.36}$$

Similarly, for the random recursive tree in (1.31),

$$\sum_{\Lambda, \Lambda'} |\hat{\sigma}_{\Lambda, \Lambda'}| = \infty. \tag{1.37}$$

Hence, in general, we need the less elegant expressions in (1.25) and (1.31). The same applies to the special cases in (1.52) and (1.55) below.

Note that if  $f(T)$  depends on the size  $|T|$  only (a case considered in [28] and [20]), so  $f(T) = \mu_{|T|}$  for some sequence  $\mu_k, k \geq 1$ , then (1.25) implies

$$\sigma_F^2 = \sum_{k, m \geq 1} \mu_k \mu_m \sigma_{k, m}, \tag{1.38}$$

where it is easily shown that the sum is absolutely convergent as a consequence of (1.13)–(1.14) and the assumption (1.23), i.e.  $\sum_k \mu_k^2/k^2 < \infty$ . The analogous result for the random recursive tree holds too for such  $f$ , now using (1.19)–(1.20).

The asymptotic means  $\mu_F$  and  $\hat{\mu}_F$  in (1.24) and (1.30) can also be written as follows. Let  $\mathcal{T}$  be the random binary search tree  $\mathcal{T}_N$  with random size  $N$  such that  $\mathbb{P}(|\mathcal{T}| = k) = \mathbb{P}(N = k) = \frac{2}{(k+1)(k+2)}, k \geq 1$ . Similarly, let  $\Lambda$  be the random recursive tree  $\Lambda_N$  with random size  $N$  such that  $\mathbb{P}(|\Lambda| = k) = \mathbb{P}(N = k) = \frac{1}{k(k+1)}, k \geq 1$ . Then, by definition,

$$\mu_F = \mathbb{E} f(\mathcal{T}), \tag{1.39}$$

$$\hat{\mu}_F = \mathbb{E} f(\Lambda). \tag{1.40}$$

Moreover, as shown by Aldous [1],  $\mathcal{T}$  is the limit in distribution of a uniformly random fringe tree of  $\mathcal{T}_n$  as  $n \rightarrow \infty$ , and similarly  $\Lambda$  is the limit in distribution of a uniformly random fringe tree of  $\Lambda_n$  as  $n \rightarrow \infty$ , see also [10] and [12]. (In fact, this is an immediate consequence of (1.4) and (1.5).)

Aldous [1] gave also direct constructions of  $\mathcal{T}$  and  $\Lambda$  using branching processes. For  $\Lambda$  we consider a tree  $\Lambda_t$  growing randomly in continuous time, starting with an isolated root at time  $t = 0$  and such that each existing node gets children according to a Poisson process with rate 1. For  $\mathcal{T}$  we similarly grow a random binary tree  $\mathcal{T}_t$  by letting each node get a left and a right child after waiting times that are independent and  $\text{Exp}(1)$ . In both cases, we stop the process at a random time  $\tau \sim \text{Exp}(1)$ , independent of everything else; this gives  $\Lambda$  and  $\mathcal{T}$ , see [1]. This construction often simplifies the calculation of  $\mu_F$  and  $\hat{\mu}_F$ , see [12] and examples in Section 8. ( $\Lambda$  and  $\mathcal{T}$  can be regarded as increasing trees, using the birth times of the nodes as labels.)

#### 1.4 Some consequences and extensions

An important special case of Theorem 1.7 is to consider all trees of size  $k$ , i.e.,  $X_{n,k}$  and  $\hat{X}_{n,k}$ . This yields Theorem 1.18 below. We give also an accompanying Theorem 1.19 on normal convergence. Note that in both theorems,  $k = k_n$  may depend on  $n$ . The results in the Theorems 1.18–1.19, except the explicit rate in (1.43)–(1.44), were shown

by Feng, Mahmoud and Panholzer [16] and Fuchs [22] by using variants of the method of moments. Theorem 1.19 was earlier proved for fixed  $k$  by Devroye [10] (using the central limit theorem for  $m$ -dependent variables), and the means (1.41)–(1.42) are implicit in [10], see also [11] and Flajolet, Gourdon and Martínez [21]. (The corresponding, weaker, laws of large numbers were also given by Aldous [1] by another method.) The part (1.45) for binary search trees was extended to  $k = k_n$  (for a smaller range than here) by Devroye [11] using Stein’s method for normal approximation. In the present paper we continue and extend this approach, and use Stein’s method for both normal and Poisson approximations (these are two completely different methods) to provide simple proofs for the full range.

**Theorem 1.18.** *Let  $X_{n,k}$  be the number of subtrees of size  $k$  in the binary search tree  $\mathcal{T}_n$  and similarly let  $\hat{X}_{n,k}$  be the number of subtrees in the random recursive tree  $\Lambda_n$ . Let  $k = k_n$  where  $k < n$ . Furthermore, let  $\mu_{n,k} := \mathbb{E}(X_{n,k})$  and  $\hat{\mu}_{n,k} := \mathbb{E}(\hat{X}_{n,k})$ . Then*

$$\mu_{n,k} := \mathbb{E}(X_{n,k}) = \frac{2(n+1)}{(k+1)(k+2)}, \tag{1.41}$$

$$\hat{\mu}_{n,k} := \mathbb{E}(\hat{X}_{n,k}) = \frac{n}{k(k+1)}. \tag{1.42}$$

Then, for the binary search tree,

$$d_{TV}(\mathcal{L}(X_{n,k}), \text{Po}(\mu_{n,k})) = \frac{1}{2} \sum_{l \geq 0} \left| \mathbb{P}(X_{n,k} = l) - e^{-\mu_{n,k}} \frac{(\mu_{n,k})^l}{l!} \right| = O\left(\frac{1}{k}\right), \tag{1.43}$$

and for the random recursive tree,

$$d_{TV}(\mathcal{L}(\hat{X}_{n,k}), \text{Po}(\hat{\mu}_{n,k})) = \frac{1}{2} \sum_{l \geq 0} \left| \mathbb{P}(\hat{X}_{n,k} = l) - e^{-\hat{\mu}_{n,k}} \frac{(\hat{\mu}_{n,k})^l}{l!} \right| = O\left(\frac{1}{k}\right). \tag{1.44}$$

Consequently, if  $n \rightarrow \infty$  and  $k \rightarrow \infty$ , then  $d_{TV}(\mathcal{L}(X_{n,k}), \text{Po}(\mu_{n,k})) \rightarrow 0$ , and similarly  $d_{TV}(\mathcal{L}(\hat{X}_{n,k}), \text{Po}(\hat{\mu}_{n,k})) \rightarrow 0$ .

**Theorem 1.19.** *Let  $X_{n,k}$  be the number of subtrees of size  $k$  in the binary search tree  $\mathcal{T}_n$  and similarly let  $\hat{X}_{n,k}$  be the number of subtrees of size  $k$  in the random recursive tree  $\Lambda_n$ . Let  $k = k_n = o(\sqrt{n})$ . Then, as  $n \rightarrow \infty$ , for the binary search tree*

$$\frac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\text{Var}(X_{n,k})}} \xrightarrow{d} \mathcal{N}(0, 1), \tag{1.45}$$

and similarly, for the random recursive tree

$$\frac{\hat{X}_{n,k} - \mathbb{E}(\hat{X}_{n,k})}{\sqrt{\text{Var}(\hat{X}_{n,k})}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{1.46}$$

**Remark 1.20.** If  $k/\sqrt{n} \rightarrow \infty$ , then  $\mu_{n,k}, \hat{\mu}_{n,k} \rightarrow 0$ , and the convergence result in Theorem 1.18 reduces to the trivial  $X_{n,k} \xrightarrow{P} 0$  and  $\hat{X}_{n,k} \xrightarrow{P} 0$ ; the rate of convergence in (1.43)–(1.44) is still of interest. Dennert and Grübel [9] considered instead the sum  $\sum_{k \geq (1-t)n} X_{n,k}$  and obtained a functional central limit theorem.

If  $k/\sqrt{n} \rightarrow c \in (0, \infty)$ , then  $\mu_{n,k} \rightarrow 2c^{-2}$  and  $\hat{\mu}_{n,k} \rightarrow c^{-2}$ ; and we obtain the Poisson distribution limits  $X_{n,k} \xrightarrow{d} \text{Po}(2c^{-2})$  and  $\hat{X}_{n,k} \xrightarrow{d} \text{Po}(c^{-2})$  [16, 22].

**Remark 1.21.** The proofs yield immediately, using the classical Berry–Esseen estimate for Poisson distributions, also the estimate  $O(k/\sqrt{n} + 1/k)$  of the convergence rates in (1.45) and (1.46) for the Kolmogorov distance; however, for slowly growing  $k$  this is inferior to the bound  $O(k/\sqrt{n})$  given by Fuchs [22]. (We have not investigated whether this bound can be shown by a more careful application of Stein’s method.) Other distances might also be studied, but we have not done so.

Next we consider the subtree counts  $X_n^T$  and  $\hat{X}_n^\Lambda$  for subtrees of different sizes together, giving result on joint asymptotic normality for several sizes and properties; however, now we do not allow the sizes to depend on  $n$ . Theorem 1.22 below shows that these numbers are asymptotically normal, and moreover, jointly so for different trees  $T$  or  $\Lambda$ . (For a single binary tree  $T$  this was shown by Devroye [10, 11], and by another method by Flajolet, Gourdon and Martínez [21]; for a single unordered tree  $\Lambda$  this was shown by Feng and Mahmoud [15].)

**Theorem 1.22.** (i) Let  $T^1, \dots, T^d$  be a fixed sequence of distinct binary trees and let  $\mathbf{X}_n = (X_n^{T^1}, X_n^{T^2}, \dots, X_n^{T^d})$ . Let

$$\boldsymbol{\mu}_n := \mathbb{E} \mathbf{X}_n = \left( \mathbb{E}(X_n^{T^1}), \mathbb{E}(X_n^{T^2}), \dots, \mathbb{E}(X_n^{T^d}) \right)$$

and let  $\Gamma = (\gamma_{ij})_{i,j=1}^d$  denote the matrix with elements

$$\gamma_{ij} = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Cov}(X_n^{T^i}, X_n^{T^j}) = \sigma_{T^i, T^j}, \tag{1.47}$$

with notation as in (1.10)–(1.11). Then  $\Gamma$  is non-singular and

$$n^{-1/2}(\mathbf{X}_n - \boldsymbol{\mu}_n) \xrightarrow{d} \mathcal{N}(0, \Gamma). \tag{1.48}$$

(ii) Similarly, let  $\Lambda^1, \dots, \Lambda^d$  be a fixed sequence of distinct ordered (or unordered) trees and let  $\hat{\mathbf{X}}_n = (\hat{X}_n^{\Lambda^1}, \hat{X}_n^{\Lambda^2}, \dots, \hat{X}_n^{\Lambda^d})$ . Let

$$\hat{\boldsymbol{\mu}}_n := \mathbb{E} \hat{\mathbf{X}}_n = \left( \mathbb{E}(\hat{X}_n^{\Lambda^1}), \mathbb{E}(\hat{X}_n^{\Lambda^2}), \dots, \mathbb{E}(\hat{X}_n^{\Lambda^d}) \right)$$

and let  $\hat{\Gamma} = (\hat{\gamma}_{ij})_{i,j=1}^d$  denote the matrix with elements

$$\hat{\gamma}_{ij} = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Cov}(\hat{X}_n^{\Lambda^i}, \hat{X}_n^{\Lambda^j}) = \hat{\sigma}_{\Lambda^i, \Lambda^j} \tag{1.49}$$

with notation as in (1.16)–(1.17). Then  $\hat{\Gamma}$  is non-singular and

$$n^{-1/2}(\hat{\mathbf{X}}_n - \hat{\boldsymbol{\mu}}_n) \xrightarrow{d} \mathcal{N}(0, \hat{\Gamma}). \tag{1.50}$$

For binary search trees, (1.48) can be proved as the univariate case in Devroye [10], but the formula (1.11) for the covariances seems to be new. For random recursive trees, as said above, Feng and Mahmoud [15] showed the univariate case  $d = 1$  of (1.50) (for unordered  $\Lambda$ ), together with formulas for the mean and variance.

**Remark 1.23.** Since the covariance matrices  $\Gamma$  and  $\hat{\Gamma}$  in Theorem 1.22 are non-singular, the limiting multivariate normal distributions  $\mathcal{N}(0, \Gamma)$  and  $\mathcal{N}(0, \hat{\Gamma})$  are non-degenerate. Furthermore, let  $\text{Cov}(\mathbf{X}_n)$  denote the covariance matrix of  $\mathbf{X}_n$ . Since  $n^{-1} \text{Cov}(\mathbf{X}_n) \rightarrow \Gamma$  as  $n \rightarrow \infty$ ,  $\text{Cov}(\mathbf{X}_n)$  is non-singular for large enough  $n$  and thus  $\text{Cov}(\mathbf{X}_n)^{-1/2}$  exists and the conclusion (1.48) is equivalent to  $\text{Cov}(\mathbf{X}_n)^{-1/2}(\mathbf{X}_n - \boldsymbol{\mu}_n) \xrightarrow{d} \mathcal{N}(0, I_d)$ , where  $I_d$  is the  $d \times d$  identity matrix and  $\mathcal{N}(0, I_d)$  is the  $d$ -dimensional standard normal distribution with  $d$  i.i.d.  $\mathcal{N}(0, 1)$  components. Similarly, (1.50) is equivalent to  $\text{Cov}(\hat{\mathbf{X}}_n)^{-1/2}(\hat{\mathbf{X}}_n - \hat{\boldsymbol{\mu}}_n) \xrightarrow{d} \mathcal{N}(0, I_d)$ .

**Example 1.24.** For any property  $P$  of binary trees and any fixed  $k$ ,  $X_{n,k}^P = \sum_{T \in P_k} X_n^T$ , summing over all trees  $T \in P_k$ ; hence the joint asymptotic normality in Theorem 1.22 implies asymptotic normality of  $X_{n,k}^P$ , as asserted in Remark 1.8. Moreover, this also yields joint asymptotic normality for several properties  $P$  and several (fixed)  $k$ ; in particular, we obtain joint asymptotic normality of  $X_{n,k}$  for any finite set of  $k$ , as earlier shown by Dennert and Grübel [9]. The random recursive tree case is similar.

Example 1.24 generalizes immediately to any finite linear combination of subtree counts  $X_n^T$  or  $\hat{X}_n^T$ ; in fact, this is equivalent to the joint asymptotic normality.

Next, consider again functionals. Corollary 1.15 shows, in particular, that  $F(\mathcal{T}_n)$  or  $F(\Lambda_n)$  is asymptotically normal for any bounded  $f$ , unless  $\sigma_F^2 = 0$  or  $\hat{\sigma}_F^2 = 0$ . Letting  $f$  be the indicator function of a set of trees, we obtain the following general result. (In the binary case, Devroye [11, Theorem 2] showed (1.51) and the corresponding weak law of large numbers, which is a consequence of (1.53). See also Devroye [11, Lemma 4] for a result similar to (1.53).)

**Corollary 1.25.** *Let  $P$  be any property of binary trees and let  $X_n^P$  be the number of subtrees of  $T_n$  with this property. Then, as  $n \rightarrow \infty$ ,*

$$\mathbb{E} X_n^P / n \rightarrow \mu_P := \mathbb{P}(\mathcal{T} \in P), \tag{1.51}$$

$$\text{Var} X_n^P / n \rightarrow \sigma_P^2 := \lim_{N \rightarrow \infty} \sum_{T, T' \in P: |T|, |T'| \leq N} \sigma_{T, T'} < \infty, \tag{1.52}$$

and

$$\frac{X_n^P - \mathbb{E} X_n^P}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_P^2). \tag{1.53}$$

Similarly, if  $P$  is any property of ordered (or unordered) trees and  $\hat{X}_n^P$  is the number of subtrees of  $\Lambda_n$  with this property, then, as  $n \rightarrow \infty$ ,

$$\mathbb{E} \hat{X}_n^P / n \rightarrow \hat{\mu}_P := \mathbb{P}(\Lambda \in P), \tag{1.54}$$

$$\text{Var} \hat{X}_n^P / n \rightarrow \hat{\sigma}_P^2 := \lim_{N \rightarrow \infty} \sum_{\Lambda, \Lambda' \in P: |\Lambda|, |\Lambda'| \leq N} \hat{\sigma}_{\Lambda, \Lambda'} < \infty, \tag{1.55}$$

and

$$\frac{\hat{X}_n^P - \mathbb{E} \hat{X}_n^P}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}_P^2). \tag{1.56}$$

Furthermore, we can replace  $\mathbb{E} X_n^P$  in (1.53) and  $\mathbb{E} \hat{X}_n^P$  in (1.56) by  $n\mu_P$  and  $n\hat{\mu}_P$ , respectively.  $\square$

Note that, unlike Theorem 1.7, we do not count subtrees of a given size only; we have  $X_n^P = \sum_k X_{n,k}^P$  and similarly for  $\hat{X}_n^P$ .

**Problem 1.26.** Is the asymptotic variance  $\sigma_P^2$  or  $\hat{\sigma}_P^2$  in Theorem 1.14 always non-zero except in trivial cases when  $F(\mathcal{T}_n)$  or  $F(\Lambda_n)$  is deterministic? (We conjecture so, but have no general proof.) Note that by (1.3) and the non-singularity of the finite covariance matrices in Theorem 1.22, this holds for any  $f$  such that  $f(T)$  is non-zero only for finitely many  $T$ . Another special case where this holds is given in Theorem 1.30 below.

In particular, can  $\sigma_P^2 = 0$  or  $\hat{\sigma}_P^2 = 0$  occur in Corollary 1.25 except in trivial cases when  $\text{Var} X_n^P = 0$  or  $\text{Var} \hat{X}_n^P = 0$ , respectively, for every  $n$ ?

Note that  $F$  may be deterministic also when  $f$  is not; for example, if  $f(T)$  equals the degree of the root of  $T$  minus 1, then  $F(T) = -1$  for any rooted tree  $T$ . (See also Remark 8.9 for a related example where different functionals  $f$  yield the same  $F$  for binary trees.)

**Remark 1.27.** Theorem 1.14 extends immediately to joint asymptotic normality for several functionals  $f$  and  $F$  by the Cramér–Wold device. Hence Corollaries 1.15 and 1.25 too extend to joint asymptotic normality.

**Example 1.28.** For any property  $P$ , Corollary 1.25 applied to  $P_k$ , or taking  $f(T) = \mathbf{1}\{T \in P_k\}$  in Corollary 1.15 or in Theorem 1.14, yields again the asymptotic normality of  $X_{n,k}^P$  and  $\hat{X}_{n,k}^P$  for fixed  $k$ , obtained more directly in Example 1.24.

**Remark 1.29.** Similar results for conditioned Galton–Watson trees are given in [30]. Note, however, that for the result corresponding to Theorem 1.14 there, stronger conditions on the size of  $f$  are required than for the results above; in particular, Corollary 1.15 holds in that setting only for  $\alpha < 0$ . We believe that, similarly, the analogue of Corollary 1.25 does not hold for conditioned Galton–Watson trees for arbitrary properties, although we do not know any counter example.

We note a special case where we can give an alternative formula for the asymptotic variance  $\sigma_F^2$  or  $\hat{\sigma}_F^2$  and prove the conjecture in Problem 1.26. (Theorem 1.30, for the binary search tree, is essentially the same as the case treated by Hwang and Neininger [28, Theorem 2’], with an equivalent formula for the variance, except for the extra randomization allowed there. It includes the case when  $F(T)$  only depends on the size  $|T|$ , where the formula is the case  $m = 2$  of Fill and Kapur [20, (5.3)]. In this case, a very similar result was also proved by Devroye [11, Lemma 5]. Another example where Theorems 1.30–1.31 apply is provided by the 2-protected nodes in Section 8.2.)

For a rooted tree  $T$ , let  $v_1, \dots, v_d$  be children of the root (in order if  $T$  is an ordered tree), where  $d = d(T)$  is the degree of the root. We call the subtrees  $T(v_1), \dots, T(v_d)$  *principal subtrees* of  $T$ . In the case of a binary tree  $T$ , we let  $T_L$  and  $T_R$  be the subtrees rooted at the left and right child of the root, and call these the *left* and *right subtree*; these are thus the principal subtrees, except that  $T_L$  and  $T_R$  may be the empty tree  $\emptyset$ . (We define  $\mathcal{T}_0 = \emptyset$  and  $F(\emptyset) = 0$ .)

**Theorem 1.30.** *Suppose, in addition to the hypotheses of Theorem 1.14(i), that the functional  $f(T) = f(|T|, |T_L|, |T_R|)$  depends only on the sizes of  $T$  and of its left and right subtrees. Let  $\nu_k := \mathbb{E} F(\mathcal{T}_k)$ , let  $I_k$  be uniformly distributed on  $\{0, \dots, k - 1\}$  and let*

$$\begin{aligned} \psi_k &:= \text{Var}(\nu_{I_k} + \nu_{k-1-I_k} + f(k, I_k, k - 1 - I_k)) \\ &= \mathbb{E}(\nu_{I_k} + \nu_{k-1-I_k} + f(k, I_k, k - 1 - I_k) - \nu_k)^2. \end{aligned} \tag{1.57}$$

Then

$$\sigma_F^2 = \sum_{k=1}^{\infty} \frac{2}{(k+1)(k+2)} \psi_k < \infty. \tag{1.58}$$

Moreover,  $\sigma_F^2 > 0$  unless  $\text{Var} F(\mathcal{T}_n) = 0$  for every  $n \geq 1$ ; this happens if and only if  $f(n, k, n - 1 - k) = a_n - a_k - a_{n-1-k}$  for some real numbers  $a_n, n \geq 0$ .

Note that  $|T| = |T_L| + |T_R| + 1$ , so two of  $|T|, |T_L|, |T_R|$  determine the third; nevertheless we write  $f(|T|, |T_L|, |T_R|)$  for emphasis.

**Theorem 1.31.** *Suppose, in addition to the hypotheses of Theorem 1.14(ii), that the functional  $f(\Lambda) = f(|\Lambda|, d(\Lambda), |\Lambda_{v_1}|, \dots, |\Lambda_{v_{d(\Lambda)}}|)$  depends only on the size  $|\Lambda|$  and the number and sizes of the principal subtrees. Let  $\nu_k := \mathbb{E} F(\Lambda_k)$ , and let*

$$\begin{aligned} \psi_k &:= \text{Var} \left( f(k, d(\Lambda_k), |\Lambda_{k,1}|, \dots) + \sum_{i=1}^{d(\Lambda_k)} \nu_{|\Lambda_{k,i}|} \right) \\ &= \mathbb{E} \left( f(k, d(\Lambda_k), |\Lambda_{k,1}|, \dots) + \sum_{i=1}^{d(\Lambda_k)} \nu_{|\Lambda_{k,i}|} - \nu_k \right)^2. \end{aligned} \tag{1.59}$$

Then

$$\hat{\sigma}_F^2 = \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \psi_k < \infty. \tag{1.60}$$

Moreover,  $\hat{\sigma}_F^2 > 0$  unless  $\text{Var} F(\Lambda_n) = 0$  for every  $n \geq 1$ ; this happens if and only if  $f(n, d, n_1, \dots, n_d) = a_n - \sum_{i=1}^d a_{n_i}$  for some real numbers  $a_n, n \geq 0$ .

The distribution of  $(d(\Lambda_k), |\Lambda_{k,v_1}|, \dots)$  in (1.59) is the same as the distribution of the number of cycles in a random permutation of length  $k - 1$  and their lengths (taken in the order of their minimal elements), see Drmota [14, Section 6.1.1].

## 2 Representations using uniform random variables

In this section we describe the representation of the binary search tree by Devroye [10, 11], and extend it to the random recursive tree. Finally, we give a new cyclic representation for both of the random trees, which leads to simpler calculations in later sections.

### 2.1 Devroye's representation for the binary search tree

We use the representation of the binary search tree  $\mathcal{T}_n$  by Devroye [10, 11] described in Section 1, using i.i.d. random time stamps  $U_i \sim U(0, 1)$  assigned to the keys  $i = 1, \dots, n$ . Write, for  $1 \leq k \leq n$  and  $1 \leq i \leq n - k + 1$ ,

$$\sigma(i, k) = \{(i, U_i), \dots, (i + k - 1, U_{i+k-1})\}, \quad (2.1)$$

i.e., the sequence of  $k$  labels  $(j, U_j)$  starting with  $j = i$ . For every node  $u \in \mathcal{T}_n$ , the fringe tree  $\mathcal{T}_n(u)$  rooted at  $u$  consists of the nodes with labels in a set  $\sigma(i, k)$  for some such  $i$  and  $k$ , where  $k = |\mathcal{T}_n(u)|$ , but note that not every set  $\sigma(i, k)$  is the set of labels of the nodes of a fringe subtree; if it is, we say simply that  $\sigma(i, k)$  is a *subtree*. We define the indicator variable

$$I_{i,k} := \mathbf{1}\{\sigma(i, k) \text{ is a subtree in } \mathcal{T}_n\}.$$

It is easy to see that, for convenience defining  $U_0 = U_{n+1} = 0$ ,

$$I_{i,k} = \mathbf{1}\{U_{i-1} \text{ and } U_{i+k} \text{ are the two smallest among } U_{i-1}, \dots, U_{i+k}\}. \quad (2.2)$$

Note that if  $i = 1$  or  $i = n - k + 1$ , this reduces to

$$I_{1,k} = \mathbf{1}\{U_{k+1} \text{ is the smallest among } U_1, \dots, U_{k+1}\}, \quad (2.3)$$

$$I_{n-k+1,k} = \mathbf{1}\{U_{n-k} \text{ is the smallest among } U_{n-k}, \dots, U_n\}. \quad (2.4)$$

For  $k = n$ , when we only consider  $i = 1$ , we have  $I_{1,n} = 1$ .

Let  $f(T)$  be a function from the set of (unlabelled) binary trees to  $\mathbb{R}$ . We are interested in the functional, see (1.1),

$$X_n := F(\mathcal{T}_n) = \sum_{u \in \mathcal{T}_n} f(\mathcal{T}_n(u)), \quad (2.5)$$

summing over all fringe trees of  $\mathcal{T}_n$ .

Since a permutation  $(\sigma_1, \dots, \sigma_k)$  defines a binary search tree (by drawing the keys in order  $\sigma_1, \dots, \sigma_k$ ), we can also regard  $f$  as a function of permutations (of arbitrary length). Moreover, any set  $\sigma(i, k)$  defines a permutation  $(\sigma_1, \sigma_2, \dots, \sigma_k)$  where the values  $j$ ,  $1 \leq j \leq k$ , are ordered according to the order of  $U_{i+j-1}$ . We can thus also regard  $f$  as a mapping from the collection of all sets  $\sigma(i, k)$ . Note that if  $\sigma(i, k)$  corresponds to a subtree  $\mathcal{T}_n(u)$  of  $\mathcal{T}_n$ , then  $\mathcal{T}_n(u)$  is the binary search tree defined by the permutation defined by  $\sigma(i, k)$ , and thus  $f(\mathcal{T}_n(u)) = f(\sigma(i, k))$ . Consequently, see [11],

$$X_n := \sum_{u \in \mathcal{T}_n} f(\mathcal{T}_n(u)) = \sum_{k=1}^n \sum_{i=1}^{n-k+1} I_{i,k} f(\sigma(i, k)). \quad (2.6)$$

### 2.2 The random recursive tree

Consider now instead the random recursive tree  $\Lambda_n$ . Let  $f(T)$  be a function from the set of ordered rooted trees to  $\mathbb{R}$ . (The case when  $f$  is a functional of unordered trees is a special case, and the case when  $f$  is a functional of increasing trees is similar.) In analogy with (2.5), we define

$$Y_n := F(\Lambda_n) = \sum_{u \in \Lambda_n} f(\Lambda_n(u)), \quad (2.7)$$

summing over all fringe trees of  $\Lambda_n$ .

As said in the introduction, the natural correspondence yields a coupling between the random recursive tree  $\Lambda_n$  and the binary search tree  $\mathcal{T}_{n-1}$ , where the subtrees in  $\Lambda_n$  correspond to the left subtrees at the nodes in  $\mathcal{T}_{n-1}$  together with the whole tree, including an empty left subtree  $\emptyset$  at every node in  $\mathcal{T}_{n-1}$  without a left child, corresponding to a subtree of size 1 (a leaf) in  $\Lambda_n$ . Thus, as noted by [10], the representation in Section 2.1 yields a similar representation for the random recursive tree, which can be described as follows.

Define  $\bar{f}$  as the functional on binary trees corresponding to  $f$  by  $\bar{f}(T) := f(T')$ , where  $T'$  is the ordered tree corresponding to the binary tree  $T$  by the natural correspondence. (Thus  $|T'| = |T| + 1$ .) We regard the empty binary tree  $\emptyset$  as corresponding to the (unique) ordered tree  $\bullet$  with only one vertex, and thus we define  $\bar{f}(\emptyset) := f(\bullet)$ .

Assume first  $1 < k < n$  and recall that subtrees of size  $k$  in the random recursive tree  $\Lambda_n$  correspond to left-rooted subtrees of size  $k - 1$  in the binary search tree  $\mathcal{T}_{n-1}$ . As said in Section 2.1, a subtree of size  $k - 1$  in  $\mathcal{T}_{n-1}$  corresponds to a set  $\sigma(i, k - 1)$  for some  $i \in \{1, \dots, n - k + 1\}$ . The parent of the root of this subtree is either  $i - 1$  or  $i + k - 1$ ; it is  $i - 1$ , and the subtree is right-rooted, if  $U_{i-1} > U_{i+k-1}$  and it is  $i + k - 1$ , and the subtree is left-rooted, if  $U_{i-1} < U_{i+k-1}$ . Thus, if we define

$$I_{i,k-1}^L := \mathbf{1}\{\sigma(i, k - 1) \text{ is a left-rooted subtree in } \mathcal{T}_{n-1}\}, \tag{2.8}$$

then, using (2.2),

$$I_{i,k-1}^L = \mathbf{1}\{U_{i-1} \leq U_{i+k-1} < \min_{i \leq j \leq i+k-2} U_j\}. \tag{2.9}$$

Note that, since we consider  $\mathcal{T}_{n-1}$ , we have defined  $U_0 = U_n = 0$ , and the argument above holds also in the boundary cases  $i = 1$  and  $i = n - k + 1$ . Furthermore, in the case  $k = n$ , we define the whole binary tree as left-rooted, so  $I_{1,n-1}^L = 1$  and (2.9) holds also for  $k = n$  (and thus  $i = 1$ ). (This is the reason for using a weak inequality  $U_{i-1} \leq U_{i+k-1}$  in (2.9); for  $k < n$  we might as well require  $U_{i-1} < U_{i+k-1}$  since  $U_0, \dots, U_{n-1}$  are assumed to be distinct.)

Finally, consider the case  $k = 1$ . Subtrees of size 1 in  $\Lambda_n$  correspond to nodes without left child in  $\mathcal{T}_{n-1}$ , and it is easily seen that a node  $i$  lacks a left child if and only if  $U_i \geq U_{i-1}$ . Hence, defining  $I_{i,0}^L := \mathbf{1}\{i \text{ has no left child}\}$ , (2.9) holds also for  $k = 1$  (with the empty minimum interpreted as  $+\infty$ ).

Consequently, (2.9) holds for all  $k$ , and the fringe trees in  $\Lambda_n$  correspond to the sets  $\sigma(i, k - 1)$  with  $1 \leq k \leq n$  and  $1 \leq i \leq n - k + 1$  such that  $I_{i,k-1}^L = 1$ . It follows that, in analogy with (2.6),

$$Y_n := \sum_{u \in \Lambda_n} f(\Lambda_n(u)) = \sum_{k=1}^n \sum_{i=1}^{n-k+1} I_{i,k-1}^L \bar{f}(\sigma(i, k - 1)). \tag{2.10}$$

Note that (for  $k = 1$ )  $\sigma(i, 0) = \emptyset$ , the empty set corresponding to the empty subtree  $\emptyset$ , and thus  $f(\sigma(i, 0)) = f(\emptyset) = f(\bullet)$ . Note also the boundary cases, because  $U_0 = U_n = 0$ ,

$$I_{1,k-1}^L = \mathbf{1}\{U_k \text{ is the smallest among } U_1, \dots, U_k\}, \tag{2.11}$$

and

$$I_{n-k+1,k-1}^L = \begin{cases} 0, & 1 \leq k < n, \\ 1, & k = n. \end{cases} \tag{2.12}$$

### 2.3 Cyclic representations

The representation (2.6) of  $X_n$  using a linear sequence  $U_1, \dots, U_n$  of i.i.d. random variables is natural and useful, but it has the (minor) disadvantage that terms with  $i = 1$

or  $i = n - k + 1$  have to be treated specially because of boundary effects, as seen in (2.3)–(2.4). It will be convenient to use a related cyclic representation, where we take  $n + 1$  i.i.d. uniform variables  $U_0, \dots, U_n \sim U(0, 1)$  and extend them to an infinite periodic sequence of random variables by

$$U_i := U_{i \bmod (n+1)}, \quad i \in \mathbb{Z}, \tag{2.13}$$

where  $i \bmod (n + 1)$  is the remainder when  $i$  is divided by  $n + 1$ , i.e., the integer  $\ell \in [0, n]$  such that  $i \equiv \ell \pmod{n + 1}$ . (We may and will assume that  $U_0, \dots, U_n$  are distinct.) We define further  $I_{i,k}$  as in (2.2), but now for all  $i$  and  $k$ . Similarly, we define  $\sigma(i, k)$  by (2.1) for all  $i$  and  $k$ . We then have the following cyclic representation of  $X_n$ . (We are indebted to Allan Gut for suggesting a cyclic representation.)

**Lemma 2.1.** *Let  $U_0, \dots, U_n \sim U(0, 1)$  be independent and extend this sequence periodically by (2.13). Then, with notations as above,*

$$X_n := \sum_{u \in \mathcal{T}_n} f(\mathcal{T}_n(u)) \stackrel{d}{=} \tilde{X}_n := \sum_{k=1}^n \sum_{i=1}^{n+1} I_{i,k} f(\sigma(i, k)). \tag{2.14}$$

*Proof.* The double sum in (2.14) is invariant under a cyclic shift of  $U_0, \dots, U_n$ . If we shift these values so that  $U_0$  becomes the smallest, we obtain the same distribution of  $(U_0, \dots, U_n)$  as if we instead condition on the event that  $U_0$  is the smallest  $U_i$ , i.e., on  $\{U_0 = \min_i U_i\}$ . Hence,

$$\tilde{X}_n \stackrel{d}{=} (\tilde{X}_n \mid U_0 = \min_i U_i). \tag{2.15}$$

Furthermore, the variables  $I_{i,k}$  depend only on the order relations among  $\{U_i\}$ , so if  $U_0$  is minimal, they remain the same if we put  $U_0 = 0$ . Moreover, in this case also  $U_{n+1} = U_0 = 0$  and it follows from (2.2) that  $I_{i,k} = 0$  if  $i \leq n + 1 \leq i + k - 1$ ; hence the terms in (2.14) with  $n - k + 1 < i \leq n + 1$  vanish. Note also that in the remaining terms,  $f(\sigma(i, k))$  does not depend on  $U_0$ . Consequently,

$$\tilde{X}_n \stackrel{d}{=} \left( \sum_{k=1}^n \sum_{i=1}^{n-k+1} I_{i,k} f(\sigma(i, k)) \mid U_0 = 0 \right) = X_n, \tag{2.16}$$

by (2.6), showing that the cyclic and linear representations in (2.6) and (2.14) are equivalent.  $\square$

**Remark 2.2.** In terms of the tree  $\mathcal{T}_n$ , the construction above means that we find  $i_0 \in \{0, \dots, n + 1\}$  such that  $U_{i_0}$  is minimal and then construct the tree  $\mathcal{T}_n$  from the pairs  $(1, U_{i_0+1}), \dots, (n, U_{i_0+n})$  by Devroye’s construction.

For the random recursive tree  $\Lambda_n$  we argue in the same way, now using (2.10). We start with  $n$  i.i.d. uniform random variables  $U_0, \dots, U_{n-1}$  and extend them to a sequence with period  $n$ ; we then define  $\sigma(i, k - 1)$  and  $I_{i,k-1}^L$  by (2.1) and (2.9) for all  $i$  and  $k$ . This yields the following; we omit the details.

**Lemma 2.3.** *Let  $U_0, \dots, U_{n-1} \sim U(0, 1)$  be independent and extend this sequence periodically by  $U_i := U_{i \bmod n}$ . Then, with notations as above,*

$$Y_n := \sum_{u \in \Lambda_n} f(\Lambda_n(u)) \stackrel{d}{=} \tilde{Y}_n := \sum_{k=1}^n \sum_{i=1}^n I_{i,k-1}^L \bar{f}(\sigma(i, k - 1)). \tag{2.17}$$

$\square$

We may (and will) assume that the equalities in distribution in the lemmas above are equalities.

### 3 Means and variances

In this section we prove general results on means and variances for functions on fringe trees for binary search trees and random recursive trees. The cyclic representations in Section 2.3 lead to simple calculations.

#### 3.1 Binary search tree

We begin by computing the mean and variance of  $X_{n,k}$ , the number of subtrees of size  $k$  in the random binary search tree  $\mathcal{T}_n$ . This has earlier been done using the linear representation in Section 2.1 by Devroye [10] (implicitly) and [11] (explicitly); our proof is very similar but the cyclic representation avoids the (asymptotically insignificant) boundary terms. Explicit expressions have also been derived by other (analytic) methods, see Feng, Mahmoud and Panholzer [16], Chang and Fuchs [7], Fuchs [22, 23]. We give a detailed proof for completeness, and as an introduction to later proofs. (The lemma is a special case of later results, but we find it convenient to start with the simplest case.) For completeness, note also that  $X_{n,k} = 1$  when  $k = n$  and  $X_{n,k} = 0$  when  $k > n$ .

Note that  $X_{n,k}$  is given by (2.5) with  $f(T) = \mathbf{1}\{|T| = k\}$ , and thus by (2.6) with  $f(\sigma(i, \ell)) = \mathbf{1}\{\ell = k\}$ , i.e.,  $X_{n,k} = \sum_{i=1}^{n-k+1} I_{i,k}$ . However, we prefer to instead use the cyclic representation (2.14), which in this case is

$$X_{n,k} = \sum_{i=1}^{n+1} I_{i,k}, \tag{3.1}$$

where now  $I_{i,k}$  are defined by (2.2) with  $U_i$  given by (2.13). Recall that  $U_i$  thus is defined for all  $i \in \mathbb{Z}$  and has period  $n + 1$ ; it is thus natural to regard the index  $i$  as an element of  $\mathbb{Z}_{n+1}$ ; similarly,  $I_{i,k}$  is defined for all  $i \in \mathbb{Z}$  with period  $n + 1$  in  $i$ , so we can regard it as defined for  $i \in \mathbb{Z}_{n+1}$ . When discussing these variables, we will use the natural metric on  $\mathbb{Z}_{n+1}$  defined by

$$|i - j|_{n+1} := \min_{\ell \in \mathbb{Z}} |i - j - \ell(n + 1)|. \tag{3.2}$$

**Lemma 3.1** (Cf. Devroye [10, 11] and Feng, Mahmoud and Panholzer [16]). *Let  $1 \leq k < n$ . For the binary search tree  $\mathcal{T}_n$ ,*

$$\mathbb{E}(X_{n,k}) = \frac{2(n + 1)}{(k + 1)(k + 2)} \tag{3.3}$$

and

$$\text{Var}(X_{n,k}) = \begin{cases} \mathbb{E} X_{n,k} - (n + 1) \frac{22k^2 + 44k + 12}{(k+1)(k+2)^2(2k+1)(2k+3)}, & k < \frac{n-1}{2}, \\ \mathbb{E} X_{n,k} + \frac{2}{n} - \frac{64}{(n+3)^2}, & k = \frac{n-1}{2}, \\ \mathbb{E} X_{n,k} - (\mathbb{E} X_{n,k})^2 = \mathbb{E} X_{n,k} - \frac{4(n+1)^2}{(k+1)^2(k+2)^2}, & k > \frac{n-1}{2}. \end{cases} \tag{3.4}$$

Hence,

$$\text{Var}(X_{n,k}) = \mathbb{E}(X_{n,k}) + O\left(\frac{n}{k^3}\right), \tag{3.5}$$

except when  $k = (n - 1)/2$ ; in this case

$$\text{Var}(X_{n,k}) = \mathbb{E}(X_{n,k}) + \frac{2}{n} + O\left(\frac{n}{k^3}\right) = \mathbb{E}(X_{n,k}) + O\left(\frac{1}{n}\right). \tag{3.6}$$

Another, equivalent, expression for the variance in the case  $k < (n - 1)/2$  is given in Proposition 1.10 with  $m = k$ . (It is easily checked that when  $n > 2k + 1$ , (3.4) and (1.12) with (1.14) are equivalent.)

*Proof.* We use (3.1). By (2.2) and symmetry, for any  $i$  and  $1 \leq k < n$ ,

$$\mathbb{E}(I_{i,k}) = \frac{2}{(k+2)(k+1)} \tag{3.7}$$

and thus (3.3) follows directly from (3.1).

We now consider the variance. Note that by (2.2),  $I_{i,k}$  and  $I_{j,k}$  are independent unless the sets  $i-1, \dots, i+k$  and  $j-1, \dots, j+k$  overlap modulo  $n+1$ , i.e., unless  $|i-j|_{n+1} \leq k+1$ . Furthermore, if  $0 < |i-j|_{n+1} \leq k$ , then (2.2) implies  $I_{i,k}I_{j,k} = 0$  (this says that two distinct subtrees of size  $k$  are disjoint and, moreover, have their corresponding intervals of  $k$  indices non-adjacent, which is obvious). Hence, by (3.1) and symmetry, if  $k < (n-1)/2$ ,

$$\begin{aligned} \text{Var}(X_{n,k}) &= \sum_{i=0}^n \sum_{j=0}^n \text{Cov}(I_{i,k}, I_{j,k}) \\ &= (n+1) \text{Var}(I_{0,k}) + 2(n+1) \sum_{j=1}^{k+1} \text{Cov}(I_{0,k}, I_{j,k}) \\ &= (n+1) \left( \mathbb{E} I_{0,k} + 2 \mathbb{E}(I_{0,k}I_{k+1,k}) - (2k+3)(\mathbb{E} I_{0,k})^2 \right). \end{aligned} \tag{3.8}$$

If  $k = (n-1)/2$  (and thus  $n$  is odd) this has to be modified since  $-(k+1) \equiv k+1 \pmod{n+1}$ , so the terms for  $j-i = \pm(k+1)$  coincide and should only be counted once; thus

$$\text{Var}(X_{n,k}) = (n+1) \left( \mathbb{E} I_{0,k} + \mathbb{E}(I_{0,k}I_{k+1,k}) - (2k+2)(\mathbb{E} I_{0,k})^2 \right). \tag{3.9}$$

Finally, if  $k > (n-1)/2$ , then always  $I_{i,k}I_{j,k} = 0$  unless  $i = j$  (there is no room for two distinct subtrees of size  $k \geq n/2$ ) and

$$\text{Var}(X_{n,k}) = (n+1) \left( \mathbb{E} I_{0,k} - (n+1)(\mathbb{E} I_{0,k})^2 \right). \tag{3.10}$$

This can also be seen directly, since in this case  $X_{n,k} \leq 1$ , so  $X_{n,k} \sim \text{Be}(\mu_{n,k})$  with  $\mu_{n,k} = \mathbb{E} X_{n,k} = (n+1) \mathbb{E} I_{0,k}$ .

It remains to compute  $\mathbb{E}(I_{0,k}I_{k+1,k}) = \mathbb{E}(I_{1,k}I_{k+2,k})$ . By (2.2),  $I_{1,k}I_{k+2,k} = 1$  when  $U_0$  and  $U_{k+1}$  are smaller than  $U_1, \dots, U_k$  and  $U_{k+1}$  and  $U_{2k+2}$  are smaller than  $U_{k+2}, \dots, U_{2k+1}$ . Consider first  $k < (n-1)/2$  and condition on  $U_{k+1} = u$ . Then the first condition is satisfied if either  $U_0 < u$  and  $U_1, \dots, U_k > u$ , which has probability  $u(1-u)^k$ , or if  $U_0, \dots, U_k > u$  and  $U_0$  is the smallest among them, which by symmetry has the probability  $\frac{1}{k+1} \mathbb{P}(U_0, \dots, U_k > u) = \frac{1}{k+1}(1-u)^{k+1}$ . The second condition has the same probability, and by independence we obtain, letting  $x = 1-u$ ,

$$\begin{aligned} \mathbb{E}(I_{1,k}I_{k+2,k}) &= \int_0^1 \left( u(1-u)^k + \frac{1}{k+1}(1-u)^{k+1} \right)^2 du \\ &= \int_0^1 \left( x^k - \frac{k}{k+1}x^{k+1} \right)^2 dx = \int_0^1 \left( x^{2k} - \frac{2k}{k+1}x^{2k+1} + \frac{k^2}{(k+1)^2}x^{2k+2} \right) dx \\ &= \frac{1}{2k+1} - \frac{2k}{(k+1)(2k+2)} + \frac{k^2}{(k+1)^2(2k+3)} \\ &= \frac{5k+3}{(k+1)^2(2k+1)(2k+3)}. \end{aligned} \tag{3.11}$$

(This can alternatively be obtained by a combinatorial argument, considering the 6 possible orderings of  $U_0, U_{k+1}, U_{2k+2}$  separately.)

In the case  $k = (n - 1)/2$ ,  $U_{2k+2} = U_{n+1} = U_0$ , and thus  $I_{1,k}I_{k+2,k} = 1$  if and only if  $U_0$  and  $U_{k+1}$  are the two smallest among  $U_0, \dots, U_n$ ; hence

$$\mathbb{E}(I_{1,k}I_{k+2,k}) = \frac{2}{n(n+1)}. \tag{3.12}$$

The result (3.4) now follows from (3.7)–(3.10) by elementary calculations. Finally, (3.5)–(3.6) follow.  $\square$

Lemma 3.1 is easily extended to  $X_{n,k}^P$ , the number of subtrees of size  $k$  with some property  $P$ . (The mean and estimates of the variance are given by Devroye [11]. The special case when we count copies of a given tree  $T$  was given by Flajolet, Gourdon and Martínez [21].)

**Lemma 3.2.** *Let  $P$  be some property of binary trees. Let  $1 \leq k < n$  and let  $p_{k,P} := \mathbb{P}(\mathcal{T}_k \in P)$ . For the binary search tree  $\mathcal{T}_n$ ,*

$$\mathbb{E}(X_{n,k}^P) = \frac{2(n+1)p_{k,P}}{(k+1)(k+2)} \tag{3.13}$$

and

$$\text{Var}(X_{n,k}^P) = \begin{cases} \mathbb{E} X_{n,k}^P - (n+1) \frac{22k^2+44k+12}{(k+1)(k+2)^2(2k+1)(2k+3)} p_{k,P}^2, & k < \frac{n-1}{2}, \\ \mathbb{E} X_{n,k}^P + \left(\frac{2}{n} - \frac{64}{(n+3)^2}\right) p_{k,P}^2, & k = \frac{n-1}{2}, \\ \mathbb{E} X_{n,k}^P - (\mathbb{E} X_{n,k}^P)^2 = \mathbb{E} X_{n,k}^P - \frac{4(n+1)^2}{(k+1)^2(k+2)^2} p_{k,P}^2, & k > \frac{n-1}{2}. \end{cases} \tag{3.14}$$

Hence,

$$\text{Var}(X_{n,k}^P) = \mathbb{E}(X_{n,k}^P) + O\left(\frac{n}{k^3} p_{k,P}^2\right), \tag{3.15}$$

except when  $k = (n - 1)/2$ ; in this case

$$\text{Var}(X_{n,k}^P) = \mathbb{E}(X_{n,k}^P) + O\left(\frac{1}{n} p_{k,P}^2\right). \tag{3.16}$$

*Proof.* Let  $I_{i,k}^P$  be the indicator of the event that the binary search tree defined by the permutation defined by  $\sigma(i, k)$  belongs to  $P$ . Then the cyclic representation Lemma 2.1 with  $f(T) = \mathbf{1}\{T \in P_k\}$  yields

$$X_{n,k}^P = \sum_{i=1}^{n+1} I_{i,k} I_{i,k}^P. \tag{3.17}$$

By (2.2), conditioning on  $I_{i,k} = 1$  says nothing about the relative order of  $U_i, \dots, U_{i+k-1}$ ; hence  $I_{i,k}$  and  $I_{i,k}^P$  are independent. Consequently, by (3.7),

$$\mathbb{E}(I_{i,k} I_{i,k}^P) = \mathbb{E}(I_{i,k}) \mathbb{E}(I_{i,k}^P) = \frac{2}{(k+1)(k+2)} \mathbb{P}(\mathcal{T}_k \in P) = \frac{2}{(k+1)(k+2)} p_{k,P}, \tag{3.18}$$

and (3.13) follows immediately.

Similarly, for the variance we use (3.17), (3.18) and the argument in the proof of Lemma 3.1. Note that  $I_{0,k}^P$  and  $I_{k+1,k}^P$  are independent of  $I_{0,k}I_{k+1,k}$  and of each other; thus

$$\mathbb{E}(I_{0,k} I_{0,k}^P I_{k+1,k} I_{k+1,k}^P) = \mathbb{E}(I_{0,k} I_{k+1,k}) p_{k,P}^2.$$

The result follows by simple calculations.  $\square$

To further extend this, we consider a real-valued functional  $f(T)$  of binary trees and the sum  $F(T)$  defined by (1.1). We begin with two such functionals of a special type.

**Lemma 3.3.** *Let  $1 \leq m \leq k$ . Suppose that  $f(T)$  and  $g(T)$  are two functionals of binary trees such that  $f(T) = 0$  unless  $|T| = k$  and  $g(T) = 0$  unless  $|T| = m$ , and let  $F(T)$  and  $G(T)$  be the corresponding sums (1.1) over subtrees. Let*

$$\mu_f := \mathbb{E} f(\mathcal{T}_k) \quad \text{and} \quad \mu_g := \mathbb{E} g(\mathcal{T}_m). \tag{3.19}$$

(i) *The means of  $F(\mathcal{T}_n)$  and  $G(\mathcal{T}_n)$  are given by*

$$\mathbb{E} F(\mathcal{T}_n) = \begin{cases} \frac{2(n+1)}{(k+1)(k+2)} \mu_f, & n > k, \\ \mu_f, & n = k, \\ 0, & n < k, \end{cases} \tag{3.20}$$

*and similarly for  $\mathbb{E} G(\mathcal{T}_n)$ .*

(ii) *If  $n > k + m + 1$ , then*

$$\text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) = (n + 1) \left( \frac{2}{(k + 1)(k + 2)} \mathbb{E}(f(\mathcal{T}_k)G(\mathcal{T}_k)) - \beta(k, m)\mu_f\mu_g \right)$$

*where  $\beta(k, m)$  is given by (1.9).*

(iii) *If  $n = k + m + 1$ , then*

$$\text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) = (n + 1) \left( \frac{2}{(k + 1)(k + 2)} \mathbb{E}(f(\mathcal{T}_k)G(\mathcal{T}_k)) - \beta_1(k, m)\mu_f\mu_g \right)$$

*where*

$$\beta_1(k, m) := \frac{4(k + m + 2)}{(k + 1)(k + 2)(m + 1)(m + 2)} - \frac{2}{n(n + 1)}. \tag{3.21}$$

(iv) *If  $k < n < k + m + 1$ , then*

$$\text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) = (n + 1) \left( \frac{2}{(k + 1)(k + 2)} \mathbb{E}(f(\mathcal{T}_k)G(\mathcal{T}_k)) - \beta_2(k, m)\mu_f\mu_g \right)$$

*where*

$$\beta_2(k, m) := \frac{4(n + 1)}{(k + 1)(k + 2)(m + 1)(m + 2)}. \tag{3.22}$$

(v) *If  $n = k$ , then*

$$\text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) = \mathbb{E}(f(\mathcal{T}_k)G(\mathcal{T}_k)) - (n + 1)\beta_3(k, m)\mu_f\mu_g$$

*where*

$$\beta_3(k, m) := \begin{cases} \frac{2}{(m+1)(m+2)}, & m < k, \\ \frac{1}{k+1}, & m = k. \end{cases} \tag{3.23}$$

(vi) *If  $n < k$ , then  $F(\mathcal{T}_n) = 0$  and thus  $\text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) = 0$ .*

*Proof.* (i): The result is trivial for  $k \geq n$  since  $F(\mathcal{T}_n) = F(\mathcal{T}_k) = f(\mathcal{T}_k)$  if  $k = n$  and  $F(\mathcal{T}_n) = 0$  if  $k > n$ .

Hence, assume  $k < n$ . Using the cyclic representation (2.14), we find

$$\mathbb{E} F(\mathcal{T}_n) = \sum_{i=0}^n \mathbb{E}(I_{i,k} f(\sigma(i, k))) = (n + 1) \mathbb{E}(I_{i,k} f(\sigma(i, k))). \tag{3.24}$$

Recalling (2.2) and noting that  $f(\sigma(i, k))$  depends only on the relative order of  $U_i, \dots, U_{i+k-1}$ , we see that  $I_{i,k}$  and  $f(\sigma(i, k))$  are independent. Thus, using (3.7),

$$\mathbb{E}(I_{i,k} f(\sigma(i, k))) = \mathbb{E}(I_{i,k}) \mathbb{E}(f(\sigma(i, k))) = \mathbb{E}(I_{i,k}) \mathbb{E}(f(\mathcal{T}_k)) = \frac{2}{(k + 1)(k + 2)} \mu_f \tag{3.25}$$

and thus

$$\mathbb{E} F(\mathcal{T}_n) = (n + 1) \mathbb{E}(I_{i,k}) \mathbb{E}(f(\mathcal{T}_k)) = (n + 1) \frac{2}{(k + 1)(k + 2)} \mu_f, \quad (3.26)$$

showing (3.20) in the case  $k < n$ .

(ii)–(iv): The cyclic representation (2.14) similarly yields

$$\text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) = \sum_{i=0}^n \sum_{j=0}^n \text{Cov}(I_{i,k} f(\sigma(i, k)), I_{j,m} g(\sigma(j, m))), \quad (3.27)$$

where  $I_{i,k} f(\sigma(i, k))$  and  $I_{j,m} g(\sigma(j, m))$  are independent unless the sets  $\{i - 1, \dots, i + k\}$  and  $\{j - 1, \dots, j + m\}$  overlap (as subsets of  $\mathbb{Z}_{n+1}$ ). Furthermore, as a consequence of (2.2), if these sets overlap by more than one element but none of the sets is a subset of the other, then  $I_{i,k} I_{j,m} = 0$ , except in the case  $k + m = n - 1$  and  $j - 1 \equiv i + k, i - 1 \equiv j + m \pmod{n + 1}$  (again, this says that two subtrees cannot overlap or be adjacent unless one is contained in the other).

(ii): We now assume  $k + m < n - 1$  and  $k \geq m$ . Then (3.27), symmetry and the observations just made yield

$$\begin{aligned} \text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) &= (n + 1) \left( \mathbb{E}(I_{0,k} f(\sigma(0, k)) I_{-m-1,m} g(\sigma(-m - 1, m))) \right. \\ &\quad + \sum_{j=0}^{k-m} \mathbb{E}(I_{0,k} f(\sigma(0, k)) I_{j,m} g(\sigma(j, m))) \\ &\quad + \mathbb{E}(I_{0,k} f(\sigma(0, k)) I_{k+1,m} g(\sigma(k + 1, m))) \\ &\quad \left. - (k + m + 3) \mathbb{E}(I_{0,k} f(\sigma(0, k))) \mathbb{E}(I_{0,m} g(\sigma(0, m))) \right). \end{aligned}$$

As seen in the proof of (i),  $I_{i,k}$  is independent of  $f(\sigma(i, k))$ , and thus (3.25) holds. Similarly,

$$\mathbb{E}(I_{j,m} g(\sigma(j, m))) = \frac{2}{(m + 1)(m + 2)} \mu_g, \quad (3.28)$$

and

$$\mathbb{E}(I_{0,k} f(\sigma(0, k)) I_{k+1,m} g(\sigma(k + 1, m))) = \mathbb{E}(I_{0,k} I_{k+1,m}) \mu_f \mu_g. \quad (3.29)$$

Furthermore, the argument for (3.11) generalizes to

$$\begin{aligned} \mathbb{E}(I_{0,k} I_{k+1,m}) &= \int_0^1 \left( u(1 - u)^k + \frac{1}{k+1} (1 - u)^{k+1} \right) \left( u(1 - u)^m + \frac{1}{m+1} (1 - u)^{m+1} \right) du \\ &= \int_0^1 \left( x^k - \frac{k}{k+1} x^{k+1} \right) \left( x^m - \frac{m}{m+1} x^{m+1} \right) dx \\ &= \frac{1}{k + m + 1} - \frac{k}{(k + 1)(k + m + 2)} - \frac{m}{(m + 1)(k + m + 2)} \\ &\quad + \frac{km}{(k + 1)(m + 1)(k + m + 3)} \\ &= \frac{2(k^2 + 3km + m^2 + 4k + 4m + 3)}{(k + 1)(m + 1)(k + m + 1)(k + m + 2)(k + m + 3)}. \end{aligned} \quad (3.30)$$

(Again, this can also be obtain by a combinatorial argument.)

The term  $\mathbb{E}(I_{0,k} f(\sigma(0, k)) I_{-m-1,m} g(\sigma(-m - 1, m)))$  is calculated in the same way, and yields the same result.

Finally, for convenience shifting the indices,

$$\begin{aligned} & \sum_{j=0}^{k-m} \mathbb{E}(I_{0,k} f(\sigma(0, k)) I_{j,m} g(\sigma(j, m))) \\ &= \mathbb{E}(I_{1,k}) \mathbb{E}\left(f(\sigma(1, k)) \sum_{j=1}^{k-m+1} I_{j,m} g(\sigma(j, m)) \mid I_{1,k} = 1\right) \\ &= \frac{2}{(k+1)(k+2)} \mathbb{E}(f(\mathcal{T}_k) G(\mathcal{T}_k)), \end{aligned} \tag{3.31}$$

where the last equality follows because the conditioning on  $I_{1,k} = 1$  yields the same result as conditioning on  $U_0 = U_{k+1} = 0$ , and the linear representation (2.6) shows that then the sum is  $G(\mathcal{T}_k)$ . The result follows by collecting the terms above.

(iii): In the case  $k + m = n - 1$ , we argue in the same way, but as in the case  $k = (n - 1)/2$  of Lemma 3.1 (a special case of the present lemma), there are only  $k + m + 2 = n + 1$  terms to subtract and (3.30) is replaced by the simple

$$\mathbb{E}(I_{0,k} I_{k+1,m}) = \frac{2}{n(n+1)}, \tag{3.32}$$

cf. (3.9) and (3.12).

(iv): In the case  $k + m > n - 1$ , there cannot be two disjoint subtrees of sizes  $k$  and  $m$ . Hence the arguments above yield

$$\begin{aligned} \text{Cov}(F(\mathcal{T}_n), G(\mathcal{T}_n)) &= (n+1) \left( \sum_{j=0}^{k-m} \mathbb{E}(I_{0,k} f(\sigma(0, k)) I_{j,m} g(\sigma(j, m))) \right. \\ &\quad \left. - (n+1) \mathbb{E}(I_{0,k} f(\sigma(0, k))) \mathbb{E}(I_{0,m} g(\sigma(0, m))) \right) \end{aligned}$$

and the result follows from (3.31) and (3.25), (3.28).

(v): In the case  $k = n$  we have  $F(\mathcal{T}_n) = F(\mathcal{T}_k) = f(\mathcal{T}_k)$ , and the result follows from (3.20).

(vi): Trivial. □

This leads to the following formulas for a general functional  $f$ . (Note that Lemmas 3.1–3.3 treat special cases. The mean (3.35) is computed by Devroye [11].)

**Theorem 3.4.** *Let  $f(T)$  be a functional of binary trees, and let  $F(T)$  be the sum (1.1). Further, let*

$$\mu_k := \mathbb{E} f(\mathcal{T}_k) \tag{3.33}$$

and

$$\pi_{k,n} := \begin{cases} \frac{2}{(k+1)(k+2)}, & k < n, \\ \frac{1}{n+1}, & k = n, \\ 0, & k > n. \end{cases} \tag{3.34}$$

Then, for the random binary search tree,

$$\mathbb{E} F(\mathcal{T}_n) = (n+1) \sum_{k=1}^n \pi_{k,n} \mu_k \tag{3.35}$$

and

$$\text{Var}(F(\mathcal{T}_n)) = (n+1) \left( \sum_{k=1}^n \pi_{k,n} \mathbb{E}\left(f(\mathcal{T}_k)(2F(\mathcal{T}_k) - f(\mathcal{T}_k))\right) - \sum_{k=1}^n \sum_{m=1}^n \beta^*(k, m) \mu_k \mu_m \right) \tag{3.36}$$

where, using (1.9) and (3.21)–(3.23),

$$\beta^*(k, m) := \begin{cases} \beta(k, m), & k + m + 1 < n, \\ \beta_1(k, m), & k + m + 1 = n, \\ \beta_2(k, m), & \max\{k, m\} < n < k + m + 1, \\ \beta_3(k, m), & k = n \geq m, \\ \beta_3(m, k), & m = n \geq k. \end{cases} \quad (3.37)$$

*Proof.* Let  $f_k(T) := f(T)\mathbf{1}\{|T| = k\}$ , and let  $F_k$  be the corresponding sum (1.1). Then  $f(T) = \sum_k f_k(T)$  and  $F(T) = \sum_k F_k(T)$ . Hence, using Lemma 3.3(i),

$$\mathbb{E} F(\mathcal{T}_n) = \sum_{k=1}^n \mathbb{E} F_k(\mathcal{T}_n) = \sum_{k=1}^n (n+1)\pi_{k,n}\mu_k, \quad (3.38)$$

which shows (3.35).

Similarly, using symmetry and Lemma 3.3(ii)–(v), noting  $\mathbb{E} f_k(\mathcal{T}_k) = \mathbb{E} f(\mathcal{T}_k) = \mu_k$ ,

$$\begin{aligned} \text{Var}(F(\mathcal{T}_n)) &= \sum_{k=1}^n \sum_{m=1}^k (2 - \delta_{km}) \text{Cov}(F_k(\mathcal{T}_n), F_m(\mathcal{T}_n)) \\ &= \sum_{k=1}^n \sum_{m=1}^k (2 - \delta_{km})(n+1) \left( \pi_{k,n} \mathbb{E}(f_k(\mathcal{T}_k)F_m(\mathcal{T}_k)) - \beta^*(k, m)\mu_k\mu_m \right) \end{aligned}$$

(where  $\delta_{km}$  denotes the Kronecker delta). Furthermore,  $F_m(\mathcal{T}_k) = 0$  for  $m > k$ , and  $F_k(\mathcal{T}_k) = f_k(\mathcal{T}_k) = f(\mathcal{T}_k)$ , and thus

$$\sum_{m=1}^k (2 - \delta_{km})F_m(\mathcal{T}_k) = 2 \sum_{m=1}^{\infty} F_m(\mathcal{T}_k) - F_k(\mathcal{T}_k) = 2F(\mathcal{T}_k) - f(\mathcal{T}_k)$$

and (3.36) follows, noting that  $\beta^*(k, m)$  by definition is symmetric in  $k$  and  $m$ . □

The formula (3.35) for the expectation is also easily obtained by induction, using a simple recurrence, see Hwang and Neininger [28, Lemma 1].

The notation above is a little cheating, since not only  $\pi_{k,n}$  but also  $\beta^*(k, m)$  depends on  $n$ ; however, if  $n > k + m + 1$ , neither depends on  $n$ , and we obtain the following. Define

$$\pi_k := \frac{2}{(k+1)(k+2)} \quad (3.39)$$

and recall that  $\mathcal{T}$  is the random binary search tree  $\mathcal{T}_N$  with random size  $N$  such that  $\mathbb{P}(|\mathcal{T}| = k) = \mathbb{P}(N = k) = \pi_k$ .

**Corollary 3.5.** *In the notation above, assume further that  $f(T) = 0$  when  $|T| > K$ , for some  $K < \infty$ . If  $n > 2K + 1$ , then*

$$\mathbb{E} F(\mathcal{T}_n) = (n+1) \mathbb{E} f(\mathcal{T}) \quad (3.40)$$

and

$$\text{Var}(F(\mathcal{T}_n)) = (n+1) \left( \mathbb{E} \left( f(\mathcal{T})(2F(\mathcal{T}) - f(\mathcal{T})) \right) - \sum_{k=1}^K \sum_{m=1}^K \beta(k, m)\mu_k\mu_m \right). \quad (3.41)$$

□

We can now prove Propositions 1.9 and 1.10 as two special cases of the results above.

*Proof of Proposition 1.9.* Apply Lemma 3.3(ii) with  $f(T_1) := \mathbf{1}\{T_1 = T\}$  and  $g(T_1) := \mathbf{1}\{T_1 = T'\}$ . Then  $X_n^T = F(\mathcal{T}_n)$  and  $X_n^{T'} = G(\mathcal{T}_n)$ . We have  $\mu_f = p_{k,T}$  and  $\mu_g = p_{m,T'}$ . Furthermore, if  $f(\mathcal{T}_k) \neq 0$ , then  $\mathcal{T}_k = T$  and  $G(\mathcal{T}_k) = G(T) = q_{T'}^T$ . Hence,

$$\mathbb{E}(f(\mathcal{T}_k)G(\mathcal{T}_k)) = q_{T'}^T \mathbb{E} f(\mathcal{T}_k) = q_{T'}^T p_{k,T}. \quad \square$$

*Proof of Proposition 1.10.* In principle, this follows from Proposition 1.9 by summing over all trees of sizes  $k$  and  $m$ , and evaluating the resulting sum; however, it is easier to give a direct proof. By symmetry we may assume  $k \geq m$ . We apply Lemma 3.3(ii) with  $f(T) := \mathbf{1}\{|T| = k\}$  and  $g(T) := \mathbf{1}\{|T| = m\}$ . Then  $X_{n,k} = F(\mathcal{T}_n)$  and  $X_{n,m} = G(\mathcal{T}_n)$ . Furthermore,  $f(\mathcal{T}_k) = 1$ ,  $g(\mathcal{T}_m) = 1$  and  $G(\mathcal{T}_k) = X_{k,m}$ . Hence  $\mu_f = \mu_g = 1$ , and, using (3.3),

$$\mathbb{E}(f(\mathcal{T}_k)G(\mathcal{T}_k)) = \mathbb{E} X_{k,m} = \begin{cases} \frac{2(k+1)}{(m+1)(m+2)}, & m < k, \\ 1, & m = k. \end{cases} \quad (3.42)$$

Hence, Lemma 3.3(ii) yields (1.12) with

$$\sigma_{k,m} = \begin{cases} \frac{4}{(k+2)(m+1)(m+2)} - \beta(k,m), & m < k, \\ \frac{2}{(k+1)(k+2)} - \beta(k,k), & m = k, \end{cases} \quad (3.43)$$

which yields (1.13)–(1.14) by elementary calculations.  $\square$

**Lemma 3.6.** *Let  $T_1, \dots, T_N$  be a finite sequence of distinct binary trees. Then the matrix  $(\sigma_{T_i, T_j})_{i,j=1}^N$  in Proposition 1.9 is non-singular and thus positive definite.*

*Proof.* Let  $K := \max_i |T_i|$ . For any real numbers  $a_1, \dots, a_N$  and any  $n > 2K + 1$ , Proposition 1.9 yields

$$\text{Var}\left(\sum_{i=1}^N a_i X_n^{T_i}\right) = \sum_{i,j=1}^N a_i a_j \text{Cov}(X_n^{T_i}, X_n^{T_j}) = (n+1) \sum_{i,j=1}^N a_i a_j \sigma_{T_i, T_j}. \quad (3.44)$$

Since a variance always is nonnegative, it follows that the matrix  $(\sigma_{T_i, T_j})_{i,j=1}^N$  is positive semi-definite.

Suppose that the matrix is singular. Then, using (3.44), there exist  $a_1, \dots, a_N$ , not all 0, such that if  $Z_n := \sum_{i=1}^N a_i X_n^{T_i}$ , then  $\text{Var}(Z_n) = 0$  for every  $n > 2K + 1$ . Hence  $Z_n$  is a constant, i.e., it takes the same value (possibly depending on  $n$ ) for every realization of  $\mathcal{T}_n$ . We shall see that this leads to a contradiction.

We may assume that  $a_i \neq 0$  for every  $i$  (otherwise we just ignore the remaining trees  $T_i$ ). We may further assume that  $T_1, \dots, T_N$  are ordered with  $k := |T_1| = \min_i |T_i|$ . For  $n > K + k + 1$ , let  $T_{0,n}$  be the tree consisting of a path to the right from the root with  $n$  nodes, and let  $T_{1,n}$  consist of a path to the right from the root with  $n - k$  nodes together with a left subtree  $T_1$  at the root. The subtrees of  $T_{1,n}$  with size in  $[k, K]$  are paths to the right, one each of each length  $l \in [k, K]$ , and in addition one copy of  $T_1$ ;  $T_{0,n}$  have the same paths as subtrees but no other subtrees of these sizes. Thus, denoting the values of  $Z_n$  for a realization  $T$  of  $\mathcal{T}_n$  by  $Z_n(T)$ , and similarly for  $X_n^{T_i}$ , we have  $X_n^{T_1}(T_{1,n}) = X_n^{T_1}(T_{0,n}) + 1$  and  $X_n^{T_i}(T_{1,n}) = X_n^{T_i}(T_{0,n})$  for  $i > 1$ , and hence  $Z_n(T_{1,n}) = Z_n(T_{0,n}) + a_1$ . This exhibits two possible realizations of  $\mathcal{T}_n$  with different values of  $Z_n$ . Hence  $\text{Var}(Z_n) > 0$ , a contradiction which completes the proof.  $\square$

**Lemma 3.7.** *For every  $N \geq 1$ , the matrix  $(\sigma_{k,m})_{k,m=1}^N$  of the values defined in Proposition 1.10 is non-singular and thus positive definite.*

*Proof.* This can be proved in exactly the same way as Lemma 3.6. Alternatively, it is an easy corollary of Lemma 3.6, since  $X_{n,k} = \sum_{|T|=k} X_n^T$  for every  $k$ .  $\square$

In the finitely supported case in Corollary 3.5, both  $\mathbb{E} F(\mathcal{T}_n)$  and  $\text{Var} F(\mathcal{T}_n)$  grow linearly in  $n + 1$ . Asymptotically, this is true under much weaker assumptions. We begin with the mean. (The binary tree case (3.45) was shown by Devroye [11, Lemma 1].)

**Theorem 3.8.** *Under the assumptions in Theorem 3.4, assume further that  $\mathbb{E} |f(\mathcal{T})| < \infty$  and  $\mu_n = o(n)$  as  $n \rightarrow \infty$ . Then*

$$\mathbb{E} F(\mathcal{T}_n) = n \mathbb{E} f(\mathcal{T}) + o(n). \tag{3.45}$$

More generally, if  $\mathbb{E} |f(\mathcal{T})| < \infty$  and  $\mu_n = o(n^\alpha)$  for some  $\alpha \in (0, 1]$ , then

$$\mathbb{E} F(\mathcal{T}_n) = n \mathbb{E} f(\mathcal{T}) + o(n^\alpha), \tag{3.46}$$

and if  $\mathbb{E} |f(\mathcal{T})| < \infty$  and  $\mu_n = O(n^\alpha)$  for some  $\alpha \in [0, 1)$ , then

$$\mathbb{E} F(\mathcal{T}_n) = n \mathbb{E} f(\mathcal{T}) + O(n^\alpha). \tag{3.47}$$

*Proof.* We have

$$\sum_{k=1}^{\infty} \pi_k |\mu_k| \leq \sum_{k=1}^{\infty} \pi_k \mathbb{E} |f(\mathcal{T}_k)| = \mathbb{E} |f(\mathcal{T})| < \infty \tag{3.48}$$

and similarly

$$\mathbb{E} f(\mathcal{T}) = \sum_{k=1}^{\infty} \pi_k \mathbb{E} f(\mathcal{T}_k) = \sum_{k=1}^{\infty} \pi_k \mu_k, \tag{3.49}$$

where the sum converges absolutely by (3.48). Thus (3.35) implies

$$\left| \frac{1}{n+1} \mathbb{E} F(\mathcal{T}_n) - \mathbb{E} f(\mathcal{T}) \right| \leq \sum_{k=1}^{\infty} |\pi_{k,n} - \pi_k| |\mu_k| \leq \frac{|\mu_n|}{n} + \sum_{k=n+1}^{\infty} \pi_k |\mu_k|, \tag{3.50}$$

which tends to 0 by the assumption  $\mu_n = o(n)$  and (3.48). This implies (3.45).

This is the case  $\alpha = 1$  of (3.46). For  $\alpha < 1$ , (3.50) similarly implies (3.46) and (3.47) under the stated assumptions.  $\square$

For the variance we begin with an upper bound that is uniform in  $n$  and  $f$ .

**Theorem 3.9.** *There exists a universal constant  $C$  such that, under the assumptions and notations of Theorem 3.4, for all  $n \geq 1$ ,*

$$\text{Var}(F(\mathcal{T}_n)) \leq Cn \left( \left( \sum_{k=1}^{\infty} \frac{(\text{Var} f(\mathcal{T}_k))^{1/2}}{k^{3/2}} \right)^2 + \sup_k \frac{\text{Var} f(\mathcal{T}_k)}{k} + \sum_{k=1}^{\infty} \frac{\mu_k^2}{k^2} \right). \tag{3.51}$$

*Proof.* We split  $f(T) = f^{(1)}(T) + f^{(2)}(T)$ , where for a tree  $T$  with  $|T| = k$  we define  $f^{(1)}(T) := \mathbb{E} f(\mathcal{T}_k) = \mu_k$  and  $f^{(2)}(T) := f(T) - \mu_k$ ; thus  $\mathbb{E} f^{(2)}(\mathcal{T}_k) = 0$ . This yields a corresponding decomposition  $F(\mathcal{T}_n) = F^{(1)}(\mathcal{T}_n) + F^{(2)}(\mathcal{T}_n)$ , and it suffices to estimate the variance of each term separately. For convenience, we drop the superscripts, and note that the two terms correspond to the two special cases  $f(T) = \mu_k$  when  $|T| = k$  (i.e.,  $f(T)$  depends on  $|T|$  only), and  $\mu_k = \mathbb{E} f(\mathcal{T}_k) = 0$ , respectively.

*Case 1:*  $f(T) = \mu_{|T|}$ . In this case,  $f(T) = \sum_{k=1}^{\infty} \mu_k \mathbf{1}\{|T| = k\}$  and  $F(\mathcal{T}_n) = \sum_{k=1}^n \mu_k X_{n,k}$ ; furthermore,  $X_{n,n} = 1$  is deterministic. Hence,

$$\text{Var}(F(\mathcal{T}_n)) = \sum_{k=1}^{n-1} \sum_{m=1}^{n-1} \text{Cov}(X_{n,k}, X_{n,m}) \mu_k \mu_m. \tag{3.52}$$

These covariances are evaluated by Lemma 3.3(ii)–(iv), as in the special case  $n > k + m + 1$  treated in Proposition 1.10; this yields, assuming  $m \leq k < n$  and recalling (3.42),

$$\frac{1}{n+1} \text{Cov}(X_{n,k}, X_{n,m}) = \frac{2}{(k+1)(k+2)} \mathbb{E} X_{k,m} - \beta^*(k, m). \tag{3.53}$$

Suppose first that  $m < k < n$ . If  $n > k + m + 1$ , then  $\text{Cov}(X_{n,k}, X_{n,m}) < 0$  by (1.13). If  $n = k + m + 1$ , then, similar calculations as in the proof of Proposition 1.10, now using (3.42) and (3.21), yield

$$\frac{1}{n+1} \text{Cov}(X_{n,k}, X_{n,m}) = -\frac{4}{(k+1)(k+2)(m+2)} + \frac{2}{n(n+1)} \leq \frac{2}{n(n+1)}, \tag{3.54}$$

and when  $k < n < k + m + 1$ , (3.22) similarly implies,

$$\frac{1}{n+1} \text{Cov}(X_{n,k}, X_{n,m}) = -\frac{4(n-k)}{(k+1)(k+2)(m+1)(m+2)} < 0. \tag{3.55}$$

In the case  $m = k < n$  we obtain similarly, or simpler from (3.5)–(3.6),

$$\frac{1}{n+1} \text{Var}(X_{n,k}) = O\left(\frac{1}{k^2}\right). \tag{3.56}$$

Suppose now that all  $\mu_k \geq 0$ . The (3.52), (1.13) and (3.54)–(3.56) yield, for some  $C_1$ , using the Cauchy–Schwarz inequality,

$$\begin{aligned} \frac{1}{n+1} \text{Var}(F(\mathcal{T}_n)) &\leq C_1 \sum_{k=1}^{n-1} \frac{\mu_k^2}{k^2} + 2 \sum_{k=1}^{n-2} \frac{\mu_k \mu_{n-1-k}}{n^2} \leq C_1 \sum_{k=1}^{n-1} \frac{\mu_k^2}{k^2} + \frac{2}{n^2} \sum_{k=1}^{n-2} \mu_k^2 \\ &\leq (C_1 + 2) \sum_{k=1}^{\infty} \frac{\mu_k^2}{k^2}. \end{aligned} \tag{3.57}$$

This proves (3.51) in the case  $f(T) = \mu_{|T|}$ , if we further assume  $\mu_k \geq 0$ , i.e.,  $f(T) \geq 0$ . For a general sequence  $\mu_k$ , we split  $f$  (and thus  $\mu_k$ ) into its positive and negative parts, and apply the estimate just obtained to each part. This yields (3.51) in general for Case 1.

Case 2:  $\mu_k = 0, k \geq 1$ . Let

$$a_k^2 := \text{Var}(f(\mathcal{T}_k)) \quad \text{and} \quad b_n^2 := \frac{1}{n+1} \text{Var}(F(\mathcal{T}_n)). \tag{3.58}$$

Then (3.36) implies, since we assume  $\mu_k = 0$ , using the Cauchy–Schwarz inequality and recalling (3.34),

$$\begin{aligned} b_n^2 &= \sum_{k=1}^n \pi_{k,n} \mathbb{E}(f(\mathcal{T}_k)(2F(\mathcal{T}_k) - f(\mathcal{T}_k))) \leq 2 \sum_{k=1}^n \pi_{k,n} \mathbb{E}(f(\mathcal{T}_k)F(\mathcal{T}_k)) \\ &\leq 2 \sum_{k=1}^n \pi_{k,n} a_k (k+1)^{1/2} b_k \leq 4 \sum_{k=1}^{n-1} k^{-3/2} a_k b_k + 2n^{-1/2} a_n b_n. \end{aligned} \tag{3.59}$$

Now let  $A := \max\{\sum_{k=1}^{\infty} a_k k^{-3/2}, \sup_k k^{-1/2} a_k\}$ . We find from (3.59)

$$b_n^2 \leq 4A \max_{k < n} b_k + 2Ab_n \tag{3.60}$$

and thus  $(b_n - A)^2 \leq 4A \max_{k < n} b_k + A^2$ , which by induction implies  $b_n \leq 6A$ .

In other words,

$$\text{Var}(F(\mathcal{T}_n)) \leq 36A^2(n+1) \leq 36(n+1) \left( \left( \sum_{k=1}^{\infty} a_k k^{-3/2} \right)^2 + \sup_k \frac{a_k^2}{k} \right), \tag{3.61}$$

which proves (3.51) in Case 2. □

**Remark 3.10.** In the proof of Case 1, it was convenient to reduce to the case  $\mu_k \geq 0$  in order to require only upper bounds for  $\text{Cov}(X_{n,k}, X_{n,m})$ . This is not necessary, however. An alternative is to note that by (1.13) and (3.54)–(3.55), whenever  $m < k < n$ ,

$$\frac{1}{n+1} \text{Cov}(X_{n,k}, X_{n,m}) > -\frac{8}{k^2 m}. \tag{3.62}$$

Hence, the proof can be concluded (for general  $\mu_k$ ) by the additional estimate

$$\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{|\mu_k \mu_m|}{km \max\{k, m\}} \leq 2 \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{|\mu_k \mu_m|}{(k+m)km} \leq C_2 \sum_{k=1}^{\infty} \frac{\mu_k^2}{k^2}, \tag{3.63}$$

which (by the substitution  $x_k = |\mu_k|/k$ ) is an application of Hilbert’s inequality saying that the infinite matrix  $(1/(k+m))_{k,m=1}^{\infty}$  defines a bounded operator on  $\ell^2$ , see [26, Chapter IX].

**Remark 3.11.** In order for the estimate in Theorem 3.9 to be useful, the three terms in the right-hand side of (3.51) have to be finite. These conditions are the best possible that imply  $\text{Var}(F(\mathcal{T}_n)) = O(n)$  in general, as is seen in the following examples. (We do not claim that these terms have to be finite in all cases for  $\text{Var}(F(\mathcal{T}_n)) = O(n)$  to hold, but at least in some examples they have to.)

Consider first the case  $f(T) = \mu_{|T|}$ . By (3.52) and the estimates above,

$$\frac{\text{Var } F(\mathcal{T}_n)}{n+1} = \sum_{k=1}^{n-1} \left( \frac{2}{k^2} + O(k^{-3}) \right) \mu_k^2 + \sum_{k=1}^{n-2} \frac{2}{n^2} \mu_k \mu_{n-1-k} + \sum_{m < k < n} O\left( \frac{1}{k^2 m} \right) \mu_k \mu_m. \tag{3.64}$$

Note that the factor  $2/k^2 + O(k^{-3}) \geq 1/k^2$  unless  $k \leq k_0$ , for some  $k_0$ , and suppose for simplicity that  $\mu_k = 0$  for  $k \leq k_0$ . Then the first sum is at least  $\sum_{k=1}^{n-1} \mu_k^2/k^2$ .

Now consider instead of the sequence  $(\mu_k)$  a random thinning  $(\mu'_k)$  obtained by letting  $\mu'_k = \mu_k$  with some small fixed probability  $p > 0$ , and  $\mu_k = 0$  otherwise, independently for all  $k$ . Replacing  $\mu_k$  by  $\mu'_k$  in (3.64) and taking the expectation over the thinnings yields, using the Cauchy–Schwarz inequality and (3.63),

$$\begin{aligned} \mathbb{E} \frac{\text{Var } F(\mathcal{T}_n)}{n+1} &\geq \sum_{k=1}^{n-1} \frac{p \mu_k^2}{k^2} + \sum_{k=1}^{n-2} \frac{2}{n^2} p^2 \mu_k \mu_{n-1-k} + \sum_{m < k < n} O\left( \frac{1}{k^2 m} \right) p^2 \mu_k \mu_m \\ &\geq p \sum_{k=1}^{n-1} \frac{\mu_k^2}{k^2} - p^2 C_3 \sum_{k=1}^{n-1} \frac{\mu_k^2}{k^2}. \end{aligned} \tag{3.65}$$

Choose  $p \leq 1/2C_3$ ; then the right-hand side of (3.65) is at least  $\frac{p}{2} \sum_{k=1}^{n-1} \mu_k^2/k^2$ . Suppose now that  $\sum_{k=1}^{\infty} \mu_k^2/k^2 = \infty$ . By choosing  $n$  large we then can make  $\mathbb{E} \text{Var } F(\mathcal{T}_n)/n$  arbitrarily large, so there exists an  $n$  and a thinning with  $\text{Var } F(\mathcal{T}_n)/n$  arbitrarily large. This holds also if we fix a finite number of the elements  $\mu'_k$  of the thinning, and it follows by using this argument recursively that there exists a (deterministic) thinning  $(\mu'_k)$  and a sequence  $n_\nu \rightarrow \infty$  such that  $\text{Var } F(\mathcal{T}_{n_\nu})/n_\nu \rightarrow \infty$  as  $n \rightarrow \infty$  along this sequence.

For the case  $\mu_k = 0$ , suppose that  $(a_k)_{k=3}^{\infty}$  is a given sequence of positive numbers. Define  $f_1 = f_2 := 0$  and let  $g_3(T) := \#\{\text{leaves in } T\} - 4/3$  when  $|T| = 3$ , where the constant  $4/3$  is chosen such that  $\mathbb{E} g_3(\mathcal{T}_3) = 0$ , cf. (1.41). Let  $f_3(T) = c_3 g_3(T)$  for a constant  $c_3 > 0$  such that  $\text{Var } f_3(\mathcal{T}_3) = a_3^2$ . Continue recursively as follows: If we have chosen  $f_1, \dots, f_{k-1}$ , let for a tree  $T$  with  $|T| = k$ ,  $g_k(T) := \sum'_{v \in T} f_{|T(v)|}(T(v))$ , where  $\sum'$  denotes summation over all nodes except the root. Define  $f_k(T) = c_k g_k(T)$  for a constant  $c_k > 0$  such that  $\text{Var } f_k(\mathcal{T}_k) = a_k^2$ . Note that, by induction using (3.35),  $\mathbb{E} f_k(\mathcal{T}_k) = \mathbb{E} g_k(\mathcal{T}_k) = 0$  for every  $k$ .

Consider  $f := \sum_k f_k$  and the corresponding  $F$ . By construction, for  $k > 3$ ,  $F(T) = f_k(T) + g_k(T) = (1 + c_k)g_k(T)$  for every tree  $T$  with  $|T| = k$ . If we let  $d_k^2 := \text{Var } g_k(\mathcal{T}_k)$ , we have  $a_k^2 = c_k^2 d_k^2$  and

$$\mathbb{E}(f(\mathcal{T}_k)(2F(\mathcal{T}_k) - f(\mathcal{T}_k))) = \mathbb{E}(c_k g_k(\mathcal{T}_k)(c_k + 2)g_k(\mathcal{T}_k)) = c_k(c_k + 2)d_k^2 = a_k(2d_k + a_k). \tag{3.66}$$

Similarly,

$$\text{Var } F(\mathcal{T}_k) = (1 + c_k)^2 \text{Var } g_k(\mathcal{T}_k) = (1 + c_k)^2 d_k^2 = (a_k + d_k)^2. \tag{3.67}$$

(For  $k = 3$ ,  $F(\mathcal{T}_3) = f_3(\mathcal{T}_3)$ , and (3.66)–(3.67) hold if we redefine  $d_3 := 0$ .)

Note first that we have  $\text{Var } F(\mathcal{T}_k) = (a_k + d_k)^2 \geq a_k^2 = \text{Var } f(\mathcal{T}_k)$ ; hence, if  $\text{Var } F(\mathcal{T}_n) = O(n)$ , then  $a_n^2 = \text{Var } f(\mathcal{T}_n) = O(n)$ , i.e.,  $\sup_k \text{Var } f(\mathcal{T}_k)/k < \infty$ .

Next, (3.36) and (3.66)–(3.67) yield

$$\begin{aligned} (d_n + a_n)^2 &= \text{Var } F(\mathcal{T}_n) = (n + 1) \sum_{k=3}^n \pi_{k,n} a_k (2d_k + a_k) \\ &= 2a_n d_n + a_n^2 + \sum_{k=3}^{n-1} \frac{2(n+1)}{(k+1)(k+2)} a_k (d_k + 2a_k) \end{aligned} \tag{3.68}$$

and thus

$$d_n^2 = \sum_{k=3}^{n-1} \frac{2(n+1)}{(k+1)(k+2)} a_k (2d_k + a_k). \tag{3.69}$$

It follows that, for  $n \geq 4$ ,  $d_n^2 > na_3^2/10$ , and thus  $d_n \geq c_1 n^{1/2}$  for some  $c_1 > 0$ . Using this in (3.69) we obtain

$$d_n^2 > c_2 n \sum_{k=4}^{n-1} k^{-3/2} a_k. \tag{3.70}$$

Hence, if  $\sum_{k=1}^\infty k^{-3/2} a_k = \infty$ , then  $\text{Var } F(\mathcal{T}_n)/n \rightarrow \infty$  as  $n \rightarrow \infty$ .

### 3.2 Random recursive tree

For the random recursive tree we similarly compute mean and variance using the cyclic representation (2.17). Again, these have earlier been computed using the linear representation (see Section 2.2) by Devroye [10], and also by other (analytic) methods, see Feng, Mahmoud and Panholzer [16], Fuchs [22].

The representation (2.17) gives, recalling that subtrees of size  $k$  correspond to subtrees of size  $k - 1$  in the corresponding binary tree,

$$\hat{X}_{n,k} = \sum_{i=1}^n I_{i,k-1}^L, \quad 1 \leq k \leq n. \tag{3.71}$$

**Lemma 3.12.** *Let  $1 \leq k < n$ . For the random recursive tree,*

$$\mathbb{E}(\hat{X}_{n,k}) = \frac{n}{k(k+1)} \tag{3.72}$$

and

$$\text{Var}(\hat{X}_{n,k}) = \begin{cases} \mathbb{E} \hat{X}_{n,k} - n \frac{3k+2}{k(k+1)^2(2k+1)}, & k < \frac{n}{2}, \\ \mathbb{E} \hat{X}_{n,k} - (\mathbb{E} \hat{X}_{n,k})^2 = \mathbb{E} \hat{X}_{n,k} - \frac{n^2}{k^2(k+1)^2}, & k \geq \frac{n}{2}. \end{cases} \tag{3.73}$$

Hence, for  $1 \leq k < n$ ,

$$\text{Var}(\hat{X}_{n,k}) = \mathbb{E}(\hat{X}_{n,k}) + O\left(\frac{n}{k^3}\right). \tag{3.74}$$

*Proof.* We use (3.71) and argue as in the proof of Lemma 3.1, replacing  $k$  and  $n$  by  $k - 1$  and  $n - 1$ . By (2.9) and symmetry, for any  $i$  and  $1 \leq k < n$ ,

$$\mathbb{E}(I_{i,k-1}^L) = \frac{1}{k(k+1)} \tag{3.75}$$

and thus (3.72) follows from (3.71).

For the variance, we obtain as in the proof of Lemma 3.1, if  $k < n/2$ ,

$$\text{Var}(\hat{X}_{n,k}) = n \left( \mathbb{E} I_{0,k-1}^L + 2 \mathbb{E}(I_{0,k-1}^L I_{k,k-1}^L) - (2k+1)(\mathbb{E} I_{0,k-1}^L)^2 \right). \tag{3.76}$$

If  $k \geq n/2$ , then  $I_{i,k}^L I_{j,k}^L = 0$  unless  $i = j$  and thus, or because  $\hat{X}_{n,k} \leq 1$ ,

$$\text{Var}(\hat{X}_{n,k}) = n \left( \mathbb{E} I_{0,k-1}^L - n(\mathbb{E} I_{0,k-1}^L)^2 \right) = \mathbb{E} \hat{X}_{n,k} - (\mathbb{E} \hat{X}_{n,k})^2. \tag{3.77}$$

(There is no exceptional case when  $k = n/2$ , since  $I_{0,k-1}^L I_{k,k-1}^L = 0$  in this case.)

It remains to compute  $\mathbb{E}(I_{0,k-1}^L I_{k,k-1}^L) = \mathbb{E}(I_{1,k-1}^L I_{k+1,k-1}^L)$ . We can argue as in the proof of Lemma 3.1, recalling also the condition  $U_{i-1} \leq U_{i+k-1}$  in (2.9), which yields

$$\begin{aligned} \mathbb{E}(I_{1,k-1}^L I_{k+1,k-1}^L) &= \int_0^1 u(1-u)^{k-1} \cdot \frac{1}{k}(1-u)^k \, du \\ &= \int_0^1 \frac{1}{k}(1-x)x^{2k-1} \, dx = \frac{1}{2k^2(2k+1)}. \end{aligned} \tag{3.78}$$

Alternatively, it is this time easy to use a combinatorial argument;  $I_{1,k-1}^L I_{k+1,k-1}^L = 1$  if  $U_0$  is the smallest of  $U_0, \dots, U_{2k}$ ,  $U_k$  is the smallest of the rest, and  $U_{2k}$  is the smallest of  $U_{k+1}, \dots, U_{2k}$ ; these events are independent and have probabilities  $1/(2k+1)$ ,  $1/(2k)$  and  $1/k$ .

Finally, (3.73)–(3.74) follow by simple calculations from (3.75)–(3.78). □

**Lemma 3.13.** *Let  $P$  be some property of ordered rooted trees. Let  $1 \leq k < n$  and let  $\hat{p}_{k,P} := \mathbb{P}(\Lambda_k \in P)$ . For the random recursive tree  $\Lambda_n$ ,*

$$\mathbb{E}(\hat{X}_{n,k}^P) = \frac{n\hat{p}_{k,P}}{k(k+1)}. \tag{3.79}$$

Furthermore,

$$\text{Var}(\hat{X}_{n,k}^P) = \begin{cases} \mathbb{E} \hat{X}_{n,k}^P - n \frac{3k+2}{k(k+1)^2(2k+1)} \hat{p}_{k,P}^2, & k < \frac{n}{2}, \\ \mathbb{E} \hat{X}_{n,k}^P - (\mathbb{E} \hat{X}_{n,k}^P)^2 = \mathbb{E} \hat{X}_{n,k}^P - \frac{n^2}{k^2(k+1)^2} \hat{p}_{k,P}^2, & k \geq \frac{n}{2}, \end{cases} \tag{3.80}$$

and hence

$$\text{Var}(\hat{X}_{n,k}^P) = \mathbb{E}(\hat{X}_{n,k}^P) + O\left(\frac{n\hat{p}_{k,P}^2}{k^3}\right). \tag{3.81}$$

*Proof.* Let  $I_{i,k-1}^{\bar{P}}$  be the indicator of the event that the binary search tree defined by the permutation defined by  $\sigma(i, k - 1)$  belongs to  $\bar{P}$ , where  $\bar{P}$  is the property of binary trees corresponding to (by the natural correspondence) the property  $P$  of ordered rooted trees. Then the cyclic representation Lemma 2.3 with  $f(\Lambda) = \mathbf{1}\{\Lambda \in P_k\}$  and thus  $\bar{f}(T) = \mathbf{1}\{T \in \bar{P}_{k-1}\}$  yields

$$\hat{X}_{n,k}^P = \sum_{i=1}^n I_{i,k-1}^L I_{i,k-1}^{\bar{P}}. \tag{3.82}$$

The rest of the proof is analogous to the proof of Lemma 3.2. □

**Lemma 3.14.** *Let  $1 \leq m \leq k$ . Suppose that  $f(\Lambda)$  and  $g(\Lambda)$  are two functionals of ordered rooted trees such that  $f(\Lambda) = 0$  unless  $|\Lambda| = k$  and  $g(\Lambda) = 0$  unless  $|\Lambda| = m$ , and let  $F(\Lambda)$  and  $G(\Lambda)$  be the corresponding sums (1.1) over subtrees. Let*

$$\lambda_f := \mathbb{E} f(\Lambda_k) \quad \text{and} \quad \lambda_g := \mathbb{E} g(\Lambda_m). \quad (3.83)$$

(i) *The means of  $F(\Lambda_n)$  and  $G(\Lambda_n)$  are given by*

$$\mathbb{E} F(\Lambda_n) = \begin{cases} \frac{n}{k(k+1)} \lambda_f, & n > k, \\ \lambda_f, & n = k, \\ 0, & n < k, \end{cases} \quad (3.84)$$

*and similarly for  $\mathbb{E} G(\Lambda_n)$ .*

(ii) *If  $n > k + m$ , then*

$$\text{Cov}(F(\Lambda_n), G(\Lambda_n)) = n \left( \frac{1}{k(k+1)} \mathbb{E}(f(\Lambda_k)G(\Lambda_k)) - \hat{\beta}(k, m)\lambda_f\lambda_g \right)$$

*where  $\hat{\beta}(k, m)$  is given by (1.15).*

(iii) *If  $k < n \leq k + m$ , then*

$$\text{Cov}(F(\Lambda_n), G(\Lambda_n)) = n \left( \frac{1}{k(k+1)} \mathbb{E}(f(\Lambda_k)G(\Lambda_k)) - \hat{\beta}_2(k, m)\lambda_f\lambda_g \right)$$

*where*

$$\beta_2(k, m) := \frac{n}{k(k+1)m(m+1)}. \quad (3.85)$$

(iv) *If  $n = k$ , then*

$$\text{Cov}(F(\Lambda_n), G(\Lambda_n)) = \mathbb{E}(f(\Lambda_k)G(\Lambda_k)) - n\hat{\beta}_3(k, m)\lambda_f\lambda_g$$

*where*

$$\hat{\beta}_3(k, m) := \begin{cases} \frac{1}{m(m+1)}, & m < k, \\ \frac{1}{k}, & m = k. \end{cases} \quad (3.86)$$

(v) *If  $n < k$ , then  $F(\Lambda_n) = 0$  and thus  $\text{Cov}(F(\Lambda_n), G(\Lambda_n)) = 0$ .*

*Proof.* The proof is similar to the proof of Lemma 3.3.

(i): The result is trivial for  $k \geq n$  since  $F(\Lambda_n) = F(\Lambda_k) = f(\Lambda_k)$  if  $k = n$  and  $F(\Lambda_n) = 0$  if  $k > n$ . Hence, assume  $k < n$ . Using the cyclic representation (2.17), we find by similar calculations as in (3.25), using (3.75),

$$\begin{aligned} \mathbb{E} F(\Lambda_n) &= \sum_{i=1}^n \mathbb{E}(I_{i,k-1}^L \bar{f}(\sigma(i, k-1))) = n \mathbb{E}(I_{i,k-1}^L \bar{f}(\sigma(i, k-1))) \\ &= n \mathbb{E}(I_{i,k-1}^L) \mathbb{E}(f(\Lambda_k)) = \frac{n}{k(k+1)} \lambda_f, \end{aligned} \quad (3.87)$$

showing (3.84) in the case  $k < n$ .

(ii)–(iii): The cyclic representation (2.17) similarly yields

$$\text{Cov}(F(\Lambda_n), G(\Lambda_n)) = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(I_{i,k-1}^L \bar{f}(\sigma(i, k-1)), I_{j,m-1}^L \bar{g}(\sigma(j, m-1))), \quad (3.88)$$

where  $I_{i,k-1}^L \bar{f}(\sigma(i, k-1))$  and  $I_{j,m-1}^L \bar{g}(\sigma(j, m-1))$  are independent unless the sets  $\{i-1, \dots, i+k-1\}$  and  $\{j-1, \dots, j+m-1\}$  overlap (as subsets of  $\mathbb{Z}_n$ ). Furthermore, as a

consequence of (2.2), if these sets overlap by more than one element but none of the sets is a subset of the other, then  $I_{i,k-1}^L I_{j,m-1}^L = 0$ . (Note that there is no exception with  $k + m = n$ ; there is not room for two disjoint subtrees of sizes  $k$  and  $m$ .)

(ii): We now assume  $k+m < n$  and  $k \geq m$ . Then (3.88), symmetry and the observations just made yield

$$\begin{aligned} \text{Cov}(F(\Lambda_n), G(\Lambda_n)) &= n \left( \mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1)) I_{-m,m-1}^L \bar{g}(\sigma(-m, m-1))) \right. \\ &\quad + \sum_{j=0}^{k-m} \mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1)) I_{j,m-1}^L \bar{g}(\sigma(j, m-1))) \\ &\quad + \mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1)) I_{k,m-1}^L \bar{g}(\sigma(k, m-1))) \\ &\quad \left. - (k+m+1) \mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1))) \mathbb{E}(I_{0,m-1}^L \bar{g}(\sigma(0, m-1))) \right). \end{aligned}$$

As seen in the proof of (i),  $I_{0,k-1}^L$  is independent of  $\bar{f}(\sigma(i, k-1))$ , and thus, cf. (3.87),

$$\mathbb{E}(I_{i,k-1}^L \bar{f}(\sigma(i, k-1))) = \frac{1}{k(k+1)} \lambda_f; \tag{3.89}$$

similarly,

$$\mathbb{E}(I_{j,m-1}^L \bar{g}(\sigma(j, m-1))) = \frac{1}{m(m+1)} \lambda_g, \tag{3.90}$$

$$\mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1)) I_{k,m-1}^L \bar{g}(\sigma(k, m-1))) = \mathbb{E}(I_{0,k-1}^L I_{k,m-1}^L) \lambda_f \lambda_g. \tag{3.91}$$

Furthermore, the argument for (3.78) generalizes to

$$\mathbb{E}(I_{0,k-1}^L I_{k,m-1}^L) = \int_0^1 x(1-x)^{k-1} \frac{1}{m} (1-x)^m dx = \frac{1}{m(k+m)(1+k+m)}. \tag{3.92}$$

(Again, this can also be obtain by a combinatorial argument.) By analogous calculations we obtain

$$\mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1)) I_{-m,m-1}^L \bar{g}(\sigma(-m, m-1))) = \frac{1}{k(k+m)(1+k+m)} \lambda_f \lambda_g.$$

(Note that this differs from (3.91)–(3.92), unlike the corresponding terms for the binary search tree case where (3.30) is symmetric in  $k$  and  $m$ .) Finally, for convenience shifting the indices,

$$\begin{aligned} &\sum_{j=0}^{k-m} \mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1)) I_{j,m-1}^L \bar{g}(\sigma(j, m-1))) \\ &= \mathbb{E}(I_{1,k-1}^L) \mathbb{E}\left(\bar{f}(\sigma(1, k-1)) \sum_{j=1}^{k-m+1} I_{j,m-1}^L \bar{g}(\sigma(j, m-1)) \mid I_{1,k-1}^L = 1\right) \\ &= \frac{1}{k(k+1)} \mathbb{E}(f(\Lambda_k) G(\Lambda_k)), \end{aligned} \tag{3.93}$$

where the last equality follows from the linear representation in (2.10). The result follows by collecting the terms above.

(iii): In the case  $k+m \geq n$ , there cannot be two disjoint subtrees of sizes  $k$  and  $m$ . Hence the arguments above yield

$$\begin{aligned} \text{Cov}(F(\Lambda_n), G(\Lambda_n)) &= n \left( \sum_{j=0}^{k-m} \mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1)) I_{j,m-1}^L \bar{g}(\sigma(j, m-1))) \right. \\ &\quad \left. - n \mathbb{E}(I_{0,k-1}^L \bar{f}(\sigma(0, k-1))) \mathbb{E}(I_{0,m-1}^L \bar{g}(\sigma(0, m-1))) \right) \end{aligned}$$

and the result follows from (3.93) and (3.89), (3.90).

(iv): In the case  $k = n$  we have  $F(\Lambda_n) = F(\Lambda_k) = f(\Lambda_k)$ , and the result follows from (3.84).

(v): Trivial. □

**Theorem 3.15.** *Let  $f$  be a functional of ordered rooted trees, and let  $F$  be the sum (1.1). Further, let*

$$\lambda_k := \mathbb{E} f(\Lambda_k) \tag{3.94}$$

and

$$\hat{\pi}_{k,n} := \begin{cases} \frac{1}{k(k+1)}, & k < n, \\ \frac{1}{n}, & k = n, \\ 0, & k > n. \end{cases} \tag{3.95}$$

Then, for the random recursive tree,

$$\mathbb{E} F(\Lambda_n) = n \sum_{k=1}^n \hat{\pi}_{k,n} \lambda_k \tag{3.96}$$

and

$$\text{Var}(F(\Lambda_n)) = n \left( \sum_{k=1}^n \hat{\pi}_{k,n} \mathbb{E} \left( f(\Lambda_k) (2F(\Lambda_k) - f(\Lambda_k)) \right) - \sum_{k=1}^n \sum_{m=1}^n \hat{\beta}^*(k, m) \lambda_k \lambda_m \right) \tag{3.97}$$

where, using (1.15) and (3.85)–(3.86),

$$\hat{\beta}^*(k, m) := \begin{cases} \hat{\beta}(k, m), & k + m < n, \\ \hat{\beta}_2(k, m), & \max\{k, m\} < n \leq k + m, \\ \hat{\beta}_3(k, m), & k = n \geq m, \\ \hat{\beta}_3(m, k), & m = n \geq k. \end{cases} \tag{3.98}$$

*Proof.* Analogous to the proof of Theorem 3.4, using Lemma 3.14. □

Recall that  $\Lambda$  is the random recursive tree  $\Lambda_N$  with random size  $N$  such that  $\mathbb{P}(|\Lambda| = k) = \mathbb{P}(N = k) = \hat{\pi}_k := 1/(k(k + 1))$ .

**Corollary 3.16.** *In the notation above, assume further that  $f(\Lambda) = 0$  when  $|\Lambda| > K$ , for some  $K < \infty$ . If  $n > 2K$ , then*

$$\mathbb{E} F(\Lambda_n) = n \mathbb{E} f(\Lambda) \tag{3.99}$$

and

$$\text{Var}(F(\Lambda_n)) = n \left( \mathbb{E} \left( f(\Lambda) (2F(\Lambda) - f(\Lambda)) \right) - \sum_{k=1}^K \sum_{m=1}^K \hat{\beta}(k, m) \lambda_k \lambda_m \right). \tag{3.100}$$

□

We can now prove Propositions 1.11 and 1.13 as two special cases of the results above. The proofs are analogous to the proofs of Propositions 1.9 and 1.10, but we include them for completeness.

*Proof of Proposition 1.11.* Apply Lemma 3.14(ii) with the functionals  $f(\Lambda_n(u)) := \mathbf{1}\{\Lambda_n(u) = \Lambda\}$  and  $g(\Lambda_n(u)) := \mathbf{1}\{\Lambda_n(u) = \Lambda'\}$ . Then  $\hat{X}_n^\Lambda = F(\Lambda_n)$  and  $\hat{X}_n^{\Lambda'} = G(\Lambda_n)$ . We have  $\lambda_f = \hat{p}_{k,\Lambda}$  and  $\lambda_g = \hat{p}_{m,\Lambda'}$ . Furthermore, if  $f(\Lambda_k) \neq 0$ , then  $\Lambda_k = \Lambda$  and  $G(\Lambda_k) = G(\Lambda) = \hat{q}_{\Lambda'}^\Lambda$ . Hence,

$$\mathbb{E}(f(\Lambda_k)G(\Lambda_k)) = \hat{q}_{\Lambda'}^\Lambda \mathbb{E} f(\Lambda_k) = \hat{q}_{\Lambda'}^\Lambda \hat{p}_{k,\Lambda}. \tag{3.101} \quad \square$$

*Proof of Proposition 1.13.* In principle, this follows from Proposition 1.11 by summing over all trees of sizes  $k$  and  $m$ , and evaluating the resulting sum; however as noted for the binary search tree, it is easier to give a direct proof. By symmetry we may assume  $k \geq m$ . We apply Lemma 3.14(ii) with  $f(\Lambda) := \mathbf{1}\{|\Lambda| = k\}$  and  $g(\Lambda) := \mathbf{1}\{|\Lambda| = m\}$ . Then  $\hat{X}_{n,k} = F(\Lambda_n)$  and  $\hat{X}_{n,m} = G(\Lambda_n)$ . Furthermore,  $f(\Lambda_k) = 1$ ,  $g(\Lambda_m) = 1$  and  $G(\Lambda_k) = \hat{X}_{k,m}$ . Hence  $\lambda_f = \lambda_g = 1$ , and, using (3.72),

$$\mathbb{E}(f(\Lambda_k)G(\Lambda_k)) = \mathbb{E} \hat{X}_{k,m} = \begin{cases} \frac{k}{m(m+1)}, & m < k, \\ 1, & m = k. \end{cases} \quad (3.101)$$

Hence, Lemma 3.14(ii) yields (1.18) with

$$\sigma_{k,m} = \begin{cases} \frac{1}{(k+1)m(m+1)} - \hat{\beta}(k, m), & m < k, \\ \frac{1}{k(k+1)} - \hat{\beta}(k, k), & m = k, \end{cases} \quad (3.102)$$

which yields (1.19)–(1.20) by elementary calculations.  $\square$

**Lemma 3.17.** *Let  $\Lambda_1, \dots, \Lambda_N$  be a finite sequence of distinct ordered or unordered rooted trees. Then the matrix  $(\hat{\sigma}_{\Lambda_i, \Lambda_j})_{i,j=1}^N$  in Proposition 1.11 is non-singular and thus positive definite.*

*Proof.* The proof is analogous to the proof of Lemma 3.6.  $\square$

**Lemma 3.18.** *For every  $N \geq 1$ , the matrix  $(\hat{\sigma}_{k,m})_{k,m=1}^N$  of the values defined in Proposition 1.13 is non-singular and thus positive definite.*

*Proof.* The proof is analogous to the proof of Lemma 3.7.  $\square$

In the finitely supported case in Corollary 3.16, both  $\mathbb{E} F(\Lambda_n)$  and  $\text{Var} F(\Lambda_n)$  grow linearly in  $n$ . Asymptotically, this is true under much weaker assumptions.

**Theorem 3.19.** *Under the assumptions in Theorem 3.15, assume further that  $\mathbb{E}|f(\Lambda)| < \infty$  and  $\lambda_n = o(n)$  as  $n \rightarrow \infty$ . Then*

$$\mathbb{E} F(\Lambda_n) = n \mathbb{E} f(\Lambda) + o(n). \quad (3.103)$$

*More generally, if  $\mathbb{E}|f(\Lambda)| < \infty$  and  $\lambda_n = o(n^\alpha)$  for some  $\alpha \leq 1$ , then*

$$\mathbb{E} F(\Lambda_n) = n \mathbb{E} f(\Lambda) + o(n^\alpha), \quad (3.104)$$

*and if  $\mathbb{E}|f(\Lambda)| < \infty$  and  $\lambda_n = O(n^\alpha)$  for some  $\alpha < 1$ , then*

$$\mathbb{E} F(\Lambda_n) = n \mathbb{E} f(\Lambda) + O(n^\alpha). \quad (3.105)$$

*Proof.* The proof is analogous to the proof of Theorem 3.8.  $\square$

**Theorem 3.20.** *There exists a universal constant  $C$  such that, under the assumptions and notations of Theorem 3.15, for all  $n \geq 1$ ,*

$$\text{Var}(F(\Lambda_n)) \leq Cn \left( \left( \sum_{k=1}^{\infty} \frac{(\text{Var} f(\Lambda_k))^{1/2}}{k^{3/2}} \right)^2 + \sup_k \frac{\text{Var} f(\Lambda_k)}{k} + \sum_{k=1}^{\infty} \frac{\lambda_k^2}{k^2} \right). \quad (3.106)$$

*Proof.* The proof is analogous to the proof of Theorem 3.20.  $\square$

### 4 Poisson approximation by Stein’s method and couplings

To prove Theorems 1.7 and 1.18 we use Stein’s method with couplings as described by Barbour, Holst and Janson [3]. In general, let  $\mathcal{A}$  be a finite index set and let  $(I_\alpha, \alpha \in \mathcal{A})$  be indicator random variables. We write  $W := \sum_{\alpha \in \mathcal{A}} I_\alpha$  and  $\lambda := \mathbb{E}(W)$ . To approximate  $W$  with a Poisson distribution  $\text{Po}(\lambda)$ , this method uses a coupling for each  $\alpha \in \mathcal{A}$  between  $W$  and a random variable  $W_\alpha$  which is defined on the same probability space as  $W$  and has the property

$$\mathcal{L}(W_\alpha) = \mathcal{L}(W - I_\alpha \mid I_\alpha = 1). \tag{4.1}$$

A common way to construct such a coupling  $(W, W_\alpha)$  is to find random variables  $(J_{\beta\alpha}, \beta \in \mathcal{A})$  defined on the same probability space as  $(I_\alpha, \alpha \in \mathcal{A})$  in such a way that for each  $\alpha \in \mathcal{A}$ , and jointly for all  $\beta \in \mathcal{A}$ ,

$$\mathcal{L}(J_{\beta\alpha}) = \mathcal{L}(I_\beta \mid I_\alpha = 1). \tag{4.2}$$

Then  $W_\alpha = \sum_{\beta \neq \alpha} J_{\beta\alpha}$  is defined on the same probability space as  $W$  and (4.1) holds.

Suppose that  $J_{\beta\alpha}$  are such random variables, and that, for each  $\alpha$ , the set  $\mathcal{A}_\alpha := \mathcal{A} \setminus \{\alpha\}$  is partitioned into  $\mathcal{A}_\alpha^-$  and  $\mathcal{A}_\alpha^0$  in such a way that

$$J_{\beta\alpha} \leq I_\beta \quad \text{if } \beta \in \mathcal{A}_\alpha^-, \tag{4.3}$$

with no condition if  $\beta \in \mathcal{A}_\alpha^0$ . We will use the following result from [3] (with a slightly simplified constant). ([3] also contain similar results using a third part  $\mathcal{A}_\alpha^+$  of  $\mathcal{A}_\alpha$ , where (4.3) holds in the opposite direction; we will not need them and note that it is always possible to include  $\mathcal{A}_\alpha^+$  in  $\mathcal{A}_\alpha^0$  and then use the following result.)

**Theorem 4.1** ([3, Corollary 2.C.1]). *Let  $W = \sum_{\alpha \in \mathcal{A}} I_\alpha$  and  $\lambda = \mathbb{E}(W)$ . Let  $\mathcal{A}_\alpha = \mathcal{A} \setminus \{\alpha\}$  and  $\mathcal{A}_\alpha^-, \mathcal{A}_\alpha^0$  be defined as above. Then*

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq (1 \wedge \lambda^{-1}) \left( \lambda - \text{Var}(W) + 2 \sum_{\alpha \in \mathcal{A}} \sum_{\beta \in \mathcal{A}_\alpha^0} \mathbb{E}(I_\alpha I_\beta) \right). \quad \square$$

#### 4.1 Couplings for proving Theorems 1.7 and 1.18

Returning to the binary search tree, we use the cyclic representation  $X_{n,k} = \sum_{i=1}^{n+1} I_{i,k}$  in (3.1). Recall the construction of  $I_{i,k}$  in (2.2) and the distance  $|i - j|_{n+1}$  on  $\mathbb{Z}_{n+1}$  given by (3.2).

**Lemma 4.2.** *Let  $k \in \{1, \dots, n - 1\}$  and let  $I_{i,k}$  be as in Section 2.3. Then for each  $i \in \{1, \dots, n + 1\}$ , there exists a coupling  $((I_{j,k})_j, (Z_{ji}^k)_j)$  such that  $\mathcal{L}(Z_{ji}^k) = \mathcal{L}(I_{j,k} \mid I_{i,k} = 1)$  jointly for all  $j \in \{1, \dots, n + 1\}$ . Furthermore,*

$$\begin{cases} Z_{ji}^k = I_{j,k} & \text{if } |j - i|_{n+1} > k + 1, \\ Z_{ji}^k \geq I_{j,k} & \text{if } |j - i|_{n+1} = k + 1, \\ Z_{ji}^k = 0 \leq I_{j,k} & \text{if } 0 < |j - i|_{n+1} \leq k. \end{cases}$$

*Proof.* We define  $Z_{ji}^k$  as follows. (Indices are taken modulo  $n + 1$ .) Let  $m$  and  $m'$  be the indices in  $i - 1, \dots, i + k$  such that  $U_m$  and  $U_{m'}$  are the two smallest of  $U_{i-1}, \dots, U_{i+k}$ ; if one of these is  $i - 1$  we choose  $m = i - 1$ , and if one of them is  $i + k$  we choose  $m' = i + k$ , otherwise, we randomize the choice of  $m$  among these two indices so that  $\mathbb{P}(m < m') = \frac{1}{2}$ , independently of everything else. Now exchange  $U_{i-1} \leftrightarrow U_m$  and  $U_{i+k} \leftrightarrow U_{m'}$ , i.e., let  $U'_{i-1} := U_m, U'_m := U_{i-1}, U'_{i+k} := U_{m'}, U'_{m'} := U_{i+k}$ , and  $U'_l := U_l$  for all other indices  $l$ . Finally, let, cf. (2.2),

$$Z_{ji}^k = \mathbf{1}\{U'_{j-1} \text{ and } U'_{j+k} \text{ are the two smallest among } U'_{j-1}, \dots, U'_{j+k}\}. \tag{4.4}$$

Then,  $\mathcal{L}(U'_1, \dots, U'_n) = \mathcal{L}((U_1, \dots, U_n) \mid I_{i,k} = 1)$  and thus  $\mathcal{L}(Z_{ji}^k) = \mathcal{L}(I_{j,k} \mid I_{i,k} = 1)$  jointly for all  $j$ .

Note that  $U'_l = U_l$  if  $l \notin \{i-1, \dots, i+k\}$  and thus  $Z_{ji}^k = I_{j,k}$  if  $|j-i|_{n+1} > k+1$ . On the other hand, if  $0 < j-i < k+1$ , then  $Z_{ji}^k = 0$  since  $i+k$  lies in  $\{j, \dots, j+k-1\}$  and  $U'_{i+k}$  is smaller than  $U'_{j-1}$  by construction; the case  $-k-1 < j-i < 0$  is similar. (This says simply that two different fringe trees of the same size cannot overlap, which is obvious.)

Finally, if  $j = i+k+1$  with  $j+k+1 < i+n+1$  (i.e.,  $k+1 < (n+1)/2$ ), then  $j-1 = i+k$  and thus  $U'_{j-1} \leq U_{j-1}$  while  $U'_l = U_l$  for  $l \in \{j, \dots, j+k\}$ ; hence  $Z_{ji}^k \geq I_{j,k}$ . The cases  $j = i+k+1$  with  $j+k+1 = i+n+1$  and  $j = i-k-1$  with  $j-k-1 > i-n-1$  are similar.  $\square$

Figures 3–4 show an example of this coupling, illustrated by the corresponding binary search trees; in this example  $i = 4$ ,  $k = 3$ ,  $m = i - 1 = 3$ ,  $m' = 6$  and  $U_0 = 0$ .

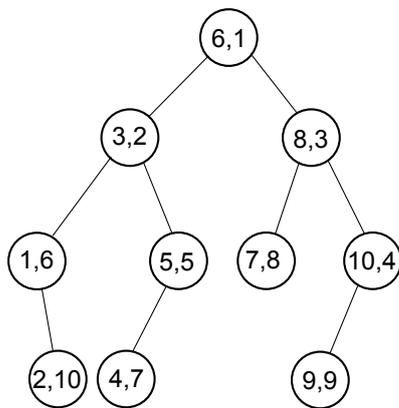


Figure 3: A binary search tree with no fringe subtree of size three containing the keys  $\{4, 5, 6\}$ .

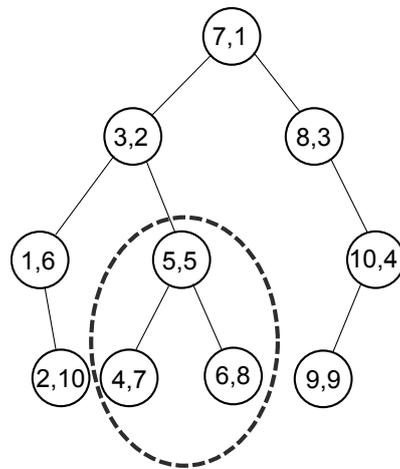


Figure 4: A coupling forcing a fringe subtree of size three containing the keys  $\{4, 5, 6\}$  in the tree in Fig. 3.

For proving the Poisson approximation result in (1.44) for the random recursive tree there is a similar coupling using the representation (3.71) where  $I_{i,k-1}^L$  is defined by (2.9) and the indicators  $U_i$  have period  $n$ :  $U_i := U_{i \bmod n}$ .

**Lemma 4.3.** *Let  $k \in \{1, \dots, n-1\}$  and let  $I_{i,k-1}^L$  be as in Section 2.3. Then for each  $i \in \{1, \dots, n\}$ , there exists a coupling  $((I_{j,k-1}^L)_j, (Z_{ji}^{k-1})_j)$  such that  $\mathcal{L}(Z_{ji}^{k-1}) = \mathcal{L}(I_{j,k-1}^L \mid I_{i,k-1}^L = 1)$  jointly for all  $j \in \{1, \dots, n\}$ . Furthermore,*

$$\begin{cases} \hat{Z}_{ji}^{k-1} = I_{j,k-1}^L & \text{if } |j-i|_n > k, \\ \hat{Z}_{ji}^{k-1} = 0 \leq I_{j,k-1}^L & \text{if } 0 < |j-i|_n < k. \end{cases}$$

In contrast to Lemma 4.2, there is no monotonicity (in any direction) between  $\hat{Z}_{ji}^{k-1}$  and  $I_{j,k-1}^L$  when  $|j-i|_n = k$ , as easily is seen by simple examples.

*Proof.* We use the same construction as in the proof of Lemma 4.2 except that if  $U'_{i-1} > U'_{i+k}$  then we make a final additional interchange  $U'_{i-1} \leftrightarrow U'_{i+k}$ . Denote the result by  $U''_1, \dots, U''_{n-1}$ . The rest of the argument is as above, now defining

$$\hat{Z}_{ji}^{k-1} = \mathbf{1}\{U''_{j-1} < U''_{j+k-1} < \min_{j \leq l \leq j+k-2} U''_l\}. \quad (4.5)$$

$\square$

Theorem 1.18 is a special case of Theorem 1.7. However, we prefer to prove Theorem 1.18 first, to show the main ideas of the proof more clearly, and then give the modifications needed for Theorem 1.7.

*Proof of Theorem 1.18.* The means are given in Lemmas 3.1 and 3.12.

We prove the Poisson approximation result first for the binary search tree, using the representation  $X_{n,k} = \sum_{i=1}^{n+1} I_{i,k}$  in (3.1). Let  $\mathcal{A} := \{1, \dots, n+1\}$ . From Lemma 4.2 we see that for each  $i \in \mathcal{A}$  we can apply Theorem 4.1 with

$$\mathcal{A}_i^- := \mathcal{A} \setminus \{i, i \pm (k+1)\}, \quad \mathcal{A}_i^0 := \{i \pm (k+1)\};$$

this yields, using Lemma 3.1 and (3.11), provided  $k \neq (n-1)/2$ ,

$$\begin{aligned} d_{TV}(\mathcal{L}(X_{n,k}), \text{Po}(\mu_{n,k})) &\leq (1 \wedge \mu_{n,k}^{-1}) \left( \mu_{n,k} - \text{Var}(X_{n,k}) + 4 \sum_{1 \leq i \leq n+1} \mathbb{E}(I_{i,k} I_{i+k+1,k}) \right) \\ &= O\left(\frac{1}{\mu_{n,k}} \cdot \frac{n}{k^3}\right) = O\left(\frac{1}{k}\right), \end{aligned}$$

which shows (1.43); the case  $k = (n-1)/2$  follows similarly from Lemma 3.1 and (3.12).

For the random recursive tree, we argue similarly, using the representation  $\hat{X}_{n,k} = \sum_{i=1}^n I_{i,k-1}^L$  in (3.71) and Theorem 4.1 together with Lemmas 4.3 and 3.12, and (3.78).  $\square$

Lemmas 4.2 and 4.3 can be extended to include a property  $P$ . We state only the binary search tree case, and leave the random recursive tree to the reader. Recall that  $I_{i,k}^P$  is the indicator of the event that the binary search tree defined by the permutation defined by  $\sigma(i, k)$  belongs to  $P$ .

**Lemma 4.4.** *Let  $k \in \{1, \dots, n-1\}$ , and let  $\tilde{I}_{i,k}^P := I_{i,k} I_{i,k}^P$ . Then for each  $i \in \{1, \dots, n+1\}$ , there exists a coupling  $((\tilde{I}_{j,k}^P)_j, (W_{j,i}^k)_j)$  such that  $\mathcal{L}(W_{j,i}^k) = \mathcal{L}(\tilde{I}_{j,k}^P \mid \tilde{I}_{i,k}^P = 1)$  jointly for all  $j \in \{1, \dots, n+1\}$ . Furthermore,*

$$\begin{cases} W_{j,i}^k = \tilde{I}_{j,k}^P & \text{if } |j-i|_{n+1} > k+1, \\ W_{j,i}^k \geq \tilde{I}_{j,k}^P & \text{if } |j-i|_{n+1} = k+1, \\ W_{j,i}^k = 0 \leq \tilde{I}_{j,k}^P & \text{if } 0 < |j-i|_{n+1} \leq k. \end{cases} \quad (4.6)$$

*Proof.* We use the same notations as in the proof of Lemma 4.2. (In particular, indices are taken modulo  $n+1$ .) Let  $m$  and  $m'$  be the indices in  $i-1, \dots, i+k$  defined in proof of Lemma 4.2, and exchange  $U_{i-1} \leftrightarrow U_m$  and  $U_{i+k} \leftrightarrow U_{m'}$ . So far we have used exactly the same coupling as in Lemma 4.2. However, since we want  $\sigma(i, k)$  to have the property  $P$ , we also exchange the values  $U'_i, \dots, U'_{i+k-1}$  with each other so that this property is fulfilled (choosing uniformly at random between the orderings satisfying  $P$ ). We abuse notation and write  $U'_i, \dots, U'_{i+k-1}$  for the new values after this exchange. Write

$$\sigma^i(j, k) = \{(j, U'_j), \dots, (j+k-1, U'_{j+k-1})\}$$

and note that  $\sigma^i(j, k) = \sigma(j, k)$  if  $|j-i| \geq k+1$ . Finally, let

$$W_{j,i}^k := Z_{j,i}^k \cdot \mathbf{1}\{\sigma^i(j, k) \text{ has property } P\}, \quad (4.7)$$

where  $Z_{j,i}^k$  is defined by (4.4). Then,  $\mathcal{L}(U'_1, \dots, U'_n) = \mathcal{L}((U_1, \dots, U_n) \mid \tilde{I}_{i,k}^P = 1)$  and thus  $\mathcal{L}(W_{j,i}^k) = \mathcal{L}(\tilde{I}_{j,k}^P \mid \tilde{I}_{i,k}^P = 1)$  jointly for all  $j$ . To see that (4.6) holds, we argue as in the proof of Lemma 4.2.  $\square$

*Proof of Theorem 1.7.* We prove the result for  $X_{n,k}^P$ , the result for  $\hat{X}_{n,k}^P$  follows by similar calculations.

The mean  $\mu_{n,k}^P := \mathbb{E}(X_{n,k}^P)$  is given by Lemma 3.2. By Theorem 4.1 together with Lemma 4.4, Lemma 3.2 and (3.11)–(3.12), we deduce that for  $k \neq (n - 1)/2$ ,

$$\begin{aligned} d_{TV}(\mathcal{L}(X_{n,k}^P), \text{Po}(\mu_{n,k}^P)) &\leq (1 \wedge (\mu_{n,k}^P)^{-1}) \left( \mu_{n,k}^P - \text{Var}(X_{n,k}^P) + 4 \sum_{1 \leq i \leq n+1} (\mathbb{E}(\tilde{I}_{i,k}^P \tilde{I}_{i+k+1,k}^P)) \right) \\ &= \begin{cases} O\left(\frac{p_{k,P}}{k}\right) & \text{if } \mu_{n,k}^P \geq 1 \\ O\left(\frac{p_{k,P}}{k}\right) \cdot \mu_{n,k}^P & \text{if } \mu_{n,k}^P < 1 \end{cases} \end{aligned}$$

and for  $k = (n - 1)/2$ ,

$$d_{TV}(\mathcal{L}(X_{n,k}^P), \text{Po}(\mu_{n,k}^P)) = O\left(\frac{p_{k,P}^2}{k}\right),$$

which shows Theorem 1.7 in the binary tree case. □

### 5 Normal approximation by Stein’s method

In this section we will prove Theorem 1.19 and Theorem 1.22. As in [11, Theorem 5] we use Stein’s method in the following form, see e.g. [31, Theorem 6.33] for a proof, and for the definition of a dependency graph.

**Lemma 5.1.** *Suppose that  $(S_n)_1^\infty$  is a sequence of random variables such that  $S_n = \sum_{\alpha \in V_n} Z_{n\alpha}$ , where for each  $n$ ,  $\{Z_{n\alpha}\}_\alpha$  is a family of random variables with dependency graph  $(V_n, E_n)$ . Let  $N(\cdot)$  denote the closed neighborhood of a node or set of nodes in this graph. Suppose further that there exist numbers  $M_n$  and  $Q_n$  such that*

$$\sum_{\alpha \in V_n} \mathbb{E}(|Z_{n\alpha}|) \leq M_n$$

and for every  $\alpha, \alpha' \in V_n$ ,

$$\sum_{\beta \in N(\alpha, \alpha')} \mathbb{E}(|Z_{n\beta}| \mid Z_{n\alpha}, Z_{n\alpha'}) \leq Q_n.$$

Let  $\sigma_n^2 = \text{Var}(S_n)$ . If

$$\lim_{n \rightarrow \infty} \frac{M_n Q_n^2}{\sigma_n^3} = 0, \tag{5.1}$$

then

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

*Proof of Theorem 1.19.* We consider the binary search tree. The random recursive tree is similar.

From Lemma 3.1 we have

$$\mathbb{E}(X_{n,k}) = \frac{2(n+1)}{(k+2)(k+1)} \tag{5.2}$$

and

$$\text{Var}(X_{n,k}) = \mathbb{E}(X_{n,k}) + O\left(\frac{n}{k^3}\right). \tag{5.3}$$

By the usual argument with subsequences, it suffices to consider the two cases  $k \rightarrow \infty$  and  $k = O(1)$ .

If  $k \rightarrow \infty$  and  $k = o(\sqrt{n})$ , then Theorem 1.18 shows that  $X_{n,k}$  can be approximated by a random variable with a  $\text{Po}(\mathbb{E}(X_{n,k}))$  distribution, where by (5.2)–(5.3),  $\text{Var}(X_{n,k}) \sim \mathbb{E}(X_{n,k}) \rightarrow \infty$  as  $n \rightarrow \infty$ . Thus, from Theorem 1.18 and the central limit theorem for Poisson distributions, it follows that  $\frac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\text{Var}(X_{n,k})}} \xrightarrow{d} \mathcal{N}(0, 1)$  as  $n \rightarrow \infty$ .

Thus, it remains to only show Theorem 1.19 for  $k = O(1)$ . We repeat the arguments used in [11, Theorem 5], but using the representation (3.1). (In fact, it suffices to consider a fixed  $k$  and then the result follows by Theorem 1.22. However, we prefer to give a direct, and somewhat more general, proof.)

We define the dependency graph  $(V_n, E_n)$  for the collection of random variables  $\{I_{i,k}, 1 \leq i \leq n + 1\}$  by taking

$$V_n = \{1, \dots, n + 1\}$$

and  $E_n := \{(i, j) : 0 < |i - j|_{n+1} \leq k + 1\}$ . Then  $|N(\alpha, \alpha')| \leq 2(2k + 3)$  for all  $\alpha, \alpha' \in V_n$ , and thus we may take  $Q_n = 4k + 6$  in Lemma 5.1. We further take  $M_n = \mathbb{E} X_{n,k} = O(n/k^2)$ . Thus,  $M_n Q_n^2 = O(n)$ , and to show (5.1) and thus Theorem 1.19 for the binary search tree it is enough to show that

$$\frac{n}{\text{Var}(X_{n,k})^{3/2}} \xrightarrow{n \rightarrow \infty} 0. \tag{5.4}$$

For  $k = O(1)$ , Proposition 1.10 shows that  $\text{Var}(X_{n,k}) \geq cn$ , and (5.4) follows, which completes the proof.

More generally, Proposition 1.10 shows that  $\text{Var}(X_{n,k}) \geq cn/k^2$  for all  $k < (n - 1)/2$ . Thus  $n/\text{Var}(X_{n,k})^{3/2} = O(k^3/n^{1/2})$ , and it follows that (5.4) holds if  $k = o(n^{1/6})$ .  $\square$

*Proof of Theorem 1.22.* We show the result for the binary search tree, for the random recursive tree the proof follows by analogous calculations. Recall that  $\mathbf{X}_n = (X_n^{T^1}, X_n^{T^2}, \dots, X_n^{T^d})$  and let  $\mathcal{Z}_d = (Z_1, \dots, Z_d)$ , where  $\mathcal{Z}_d$  is multivariate normal with the distribution  $\mathcal{N}(0, \Gamma)$ , where  $\Gamma$  is the matrix with elements  $\gamma_{ij} = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Cov}(X_n^{T^i}, X_n^{T^j})$ , see (1.47). Note that  $\Gamma$  is non-singular by Lemma 3.6.

By the Cramér–Wold device [4, Theorem 7.7], to show that  $n^{-\frac{1}{2}}(\mathbf{X}_n - \boldsymbol{\mu}_n)$  converges in distribution to  $\mathcal{Z}_d$ , it is enough to show that for every fixed vector  $(t_1, \dots, t_d) \in \mathbb{R}^d$  we have

$$\frac{\sum_{j=1}^d t_j X_n^{T^j} - \mathbb{E} \left( \sum_{j=1}^d t_j X_n^{T^j} \right)}{\sqrt{n}} \xrightarrow{d} \sum_{j=1}^d t_j Z_j, \tag{5.5}$$

where  $\sum_{j=1}^d t_j Z_j \sim \mathcal{N}(0, \gamma^2)$  with

$$\gamma^2 := \sum_{j,k=1}^d t_j t_k \gamma_{jk}. \tag{5.6}$$

Let  $S_n := \sum_{j=1}^d t_j X_n^{T^j}$ . Proposition 1.9 implies that, as  $n \rightarrow \infty$ ,

$$\text{Var}(S_n) \sim n \sum_{j,k=1}^d t_j t_k \sigma_{T^i, T^j} = n \sum_{j,k=1}^d t_j t_k \gamma_{ij} = n\gamma^2. \tag{5.7}$$

In particular, if  $\gamma^2 = 0$ , then (5.5) is trivial, with the limit 0.

To show that (5.5) holds when  $\gamma^2 > 0$ , we will use the same method as was used in [11, Theorem 5] for proving this theorem (in a more general form) in the 1-dimensional case  $d = 1$ . Let  $|T^j| = k_j$ ,  $1 \leq j \leq d$ . We use the cyclic representation (2.14), which in this case can be written as, see (3.17),

$$X_n^{T^j} = \sum_{i=1}^{n+1} I_i^j$$

for some indicator variable  $I_i^j = I_{i,k_j} I_{i,k_j}^{T^j}$  depending only on  $U_{i-1}, \dots, U_{i+k_j}$ . We define

$$V_n := \{(i, j) : 1 \leq i \leq n + 1, 1 \leq j \leq d\}$$

and let for each  $(i, j) \in V_n$ ,  $A_{i,j}$  be the set  $\{i - 1, \dots, i + k_j\}$ , regarded as a subset of  $\mathbb{Z}_{n+1}$ . Thus  $I_i^j$  depends only on  $\{U_k : k \in A_{i,j}\}$ , and thus we can define a dependency graph  $L_n$  with vertex set  $V_n$  by connecting  $(i, j)$  and  $(i', j')$  when  $A_{i,j} \cap A_{i',j'} \neq \emptyset$ .

Let  $K := \max\{k_1, k_2, \dots, k_d\}$  and  $M := \max\{t_1, t_2, \dots, t_d\}$ . It is easy to see that for the sum

$$S_n := \sum_{j=1}^d t_j X_n^{T^j} = \sum_{i=1}^{n+1} \sum_{j=1}^d t_j I_i^j = \sum_{(i,j) \in V_n} t_j I_i^j,$$

we can choose the numbers  $M_n$  and  $Q_n$  in Lemma 5.1 as  $M_n = (n + 1)dM$  and

$$Q_n = 2M \sup_{(i,j) \in V_n} |N((i, j))| \leq 2Md(2K + 3).$$

Since  $\sigma_n \sim n^{1/2}$  by (5.7), (5.1) holds and Lemma 5.1 shows that (5.5) holds. □

### 6 Proof of Theorem 1.14

We prove Theorem 1.14 by combining Theorem 1.22 with a truncation argument to deal with more general additive functionals.

*Proof of Theorem 1.14.* We consider again the binary search tree. The random recursive tree is similar.

Note first that (1.21)–(1.22) imply

$$\sum_{k=1}^{\infty} \frac{\text{Var } f(\mathcal{T}_k)}{k^2} \leq \left( \sup_k \frac{\text{Var } f(\mathcal{T}_k)}{k} \right)^{1/2} \sum_{k=1}^{\infty} \frac{(\text{Var } f(\mathcal{T}_k))^{1/2}}{k^{3/2}} < \infty, \tag{6.1}$$

and thus, using also (1.23),

$$\sum_{k=1}^{\infty} \frac{\mathbb{E} |f(\mathcal{T}_k)|^2}{k^2} = \sum_{k=1}^{\infty} \frac{\text{Var } f(\mathcal{T}_k)}{k^2} + \sum_{k=1}^{\infty} \frac{(\mathbb{E} f(\mathcal{T}_k))^2}{k^2} < \infty. \tag{6.2}$$

It follows that  $\sum_{k=1}^{\infty} \frac{\mathbb{E} |f(\mathcal{T}_k)|}{k^2} < \infty$ , and thus, see (3.48) and (3.39), that  $\mathbb{E} |f(\mathcal{T})| < \infty$ . Since (1.23) also implies  $\mathbb{E} f(\mathcal{T}_k)/k \rightarrow 0$  as  $k \rightarrow \infty$ , (1.24) follows by Theorem 3.8 and (3.49).

Next, define the truncations  $f^N(T) := f(T)\mathbf{1}\{|T| \leq N\}$ , and the corresponding sums  $F^N(T)$ . Then

$$F^N(\mathcal{T}_n) = \sum_{|T| \leq N} f(T) X_n^T, \tag{6.3}$$

and thus Proposition 1.9 yields, as  $n \rightarrow \infty$ ,

$$\text{Var } F^N(\mathcal{T}_n)/n \rightarrow \sigma_{F,N}^2 := \sum_{|T|, |T'| \leq N} f(T)f(T')\sigma_{T,T'}. \tag{6.4}$$

Moreover, Theorem 3.9 applied to  $f - f^N$  yields

$$\frac{1}{n} \text{Var}(F(\mathcal{T}_n) - F^N(\mathcal{T}_n)) \leq C\delta_N \tag{6.5}$$

where

$$\delta_N := \left( \sum_{k>N} \frac{(\text{Var } f(\mathcal{T}_k))^{1/2}}{k^{3/2}} \right)^2 + \sup_{k>N} \frac{\text{Var } f(\mathcal{T}_k)}{k} + \sum_{k>N} \frac{\mu_k^2}{k^2}. \tag{6.6}$$

Note that  $\delta_N$  is independent of  $n$ , and by the assumptions (1.21)–(1.23),  $\delta_N \rightarrow 0$  as  $N \rightarrow \infty$ . It follows by Minkowski's inequality that the sequences  $(\text{Var}(F^N(\mathcal{T}_n))/n)_{n \geq 1}$  converge uniformly to  $(\text{Var}(F(\mathcal{T}_n))/n)_{n \geq 1}$ . This and (6.4) imply (1.25) (including the existence of the limit in (1.25)).

For the convergence in distribution (1.26), we use again the truncation  $f^N$  and  $F^N$ , and note that Theorem 1.22 implies

$$\frac{F^N(\mathcal{T}_n) - \mathbb{E} F^N(\mathcal{T}_n)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_{F,N}^2) \tag{6.7}$$

as  $n \rightarrow \infty$ , for each fixed  $N$ . This together with the uniform bound (6.5) where  $\delta_N \rightarrow 0$  as  $N \rightarrow \infty$  implies (1.26), see e.g. [4, Theorem 4.2].  $\square$

*Proof of Corollary 1.15.* The assumption  $f(T) = O(|T|^\alpha)$  with  $\alpha < 1/2$  implies (1.21)–(1.23) and (1.27)–(1.29). The result follows by Theorem 1.14. The version (1.33) of the asymptotic normality (1.26) follows by (3.47) and (1.39), and similarly for the random recursive case (1.34).  $\square$

## 7 Proofs of Theorems 1.30–1.31

Finally, we prove Theorems 1.30 and 1.31, beginning with exact formulas for finite  $n$ .

**Lemma 7.1.** *Let  $F(T)$  be given for binary trees  $T$  by (1.1), and assume that the functional  $f(T) = f(|T|, |T_L|, |T_R|)$  depends only on the sizes of  $T$  and of its left and right subtrees. Let  $\psi_k$  be as in Theorem 1.30. Then*

$$\text{Var}(F(\mathcal{T}_n)) = (n + 1) \sum_{k=1}^{n-1} \frac{2}{(k + 1)(k + 2)} \psi_k + \psi_n. \tag{7.1}$$

*Proof.* We use the notation in Theorem 1.30, and let  $\sigma_n^2 := \text{Var}(F(\mathcal{T}_n))$ .

Condition the random tree  $\mathcal{T}_n$  on having a left subtree of size  $|\mathcal{T}_{n,L}| = k$ ; then the two subtrees  $\mathcal{T}_{n,L}$  and  $\mathcal{T}_{n,R}$  are independent random trees with the distributions  $\mathcal{T}_{n,L} \stackrel{d}{=} \mathcal{T}_k$  and  $\mathcal{T}_{n,R} \stackrel{d}{=} \mathcal{T}_{n-1-k}$ . Hence, (1.2) implies that the conditional distribution of  $F(\mathcal{T}_n)$  is given by

$$(F(\mathcal{T}_n) \mid |\mathcal{T}_{n,L}| = k) \stackrel{d}{=} f(n, k, n - 1 - k) + F(\mathcal{T}_k) + F(\mathcal{T}'_{n-k-1}), \tag{7.2}$$

where  $\mathcal{T}'_{n-k-1} \stackrel{d}{=} \mathcal{T}_{n-k-1}$  is independent of  $\mathcal{T}_k$ .

Taking the expectation in (7.2) we obtain the conditional expectation of  $F(\mathcal{T}_n)$  as

$$\mathbb{E}(F(\mathcal{T}_n) \mid |\mathcal{T}_{n,L}| = k) = g(k) := f(n, k, n - 1 - k) + \nu_k + \nu_{n-1-k}. \tag{7.3}$$

Since  $|\mathcal{T}_{n,L}| \stackrel{d}{=} I_n$ , it follows that

$$\mathbb{E}(F(\mathcal{T}_n) \mid |\mathcal{T}_{n,L}|) \stackrel{d}{=} g(I_n). \tag{7.4}$$

Consequently,

$$\text{Var}(\mathbb{E}(F(\mathcal{T}_n) \mid |\mathcal{T}_{n,L}|)) = \text{Var}(g(I_n)) = \psi_n \tag{7.5}$$

by (1.57); the last equality in (1.57) follows because taking the expectation in (7.4) yields

$$\mathbb{E} g(I_n) = \mathbb{E} F(\mathcal{T}_n) = \nu_n. \tag{7.6}$$

Furthermore, taking the variance in (7.2) we obtain the conditional variance

$$\text{Var}(F(\mathcal{T}_n) \mid |\mathcal{T}_{n,L}| = k) = \text{Var}(F(\mathcal{T}_k)) + \text{Var}(F(\mathcal{T}'_{n-1-k})) = \sigma_k^2 + \sigma_{n-1-k}^2. \tag{7.7}$$

Consequently, by a standard variance decomposition formula (“the law of total variance”), see, e.g., [25, Exercise 10.17-2], together with (7.5) and (7.7),

$$\begin{aligned} \sigma_n^2 &= \text{Var}(F(\mathcal{T}_n)) = \mathbb{E}(\text{Var}(F(\mathcal{T}_n) \mid |T_{n,L}|)) + \text{Var}(\mathbb{E}(F(\mathcal{T}_n) \mid |T_{n,L}|)) \\ &= \mathbb{E}(\sigma_{I_n}^2 + \sigma_{n-1-I_n}^2) + \psi_n. \end{aligned} \tag{7.8}$$

If we define  $\Psi(T)$  by (1.2) using the toll function  $\psi(T) := \psi_{|T|}$ , it follows from (7.8) and induction that  $\mathbb{E}(\Psi(\mathcal{T}_n)) = \sigma_n^2$ , and thus (7.1) follows from (3.35).  $\square$

**Lemma 7.2.** *Let  $F(\Lambda)$  be given for rooted trees  $T$  by (1.1), and assume that the functional  $f(\Lambda) = f(|\Lambda|, d(\Lambda), \Lambda_{v_1}, \dots, \Lambda_{v_{d(\Lambda)}})$  depends only on the size  $|\Lambda|$  and the number and sizes of the principal subtrees. Let  $\psi_k$  be as in Theorem 1.31. Then*

$$\text{Var}(F(\Lambda_n)) = n \sum_{k=1}^{n-1} \frac{1}{k(k+1)} \psi_k + \psi_n. \tag{7.9}$$

*Proof.* Similar to the proof of Lemma 7.1 with mainly notational changes, now conditioning on the degree  $d = d(\Lambda_n)$  and the sizes of the principal subtrees  $\Lambda_{n,v_1}, \dots, \Lambda_{n,v_d}$ , and using (3.96).  $\square$

*Proof of Theorem 1.30.* By Theorem 1.14,  $\text{Var}(F(\mathcal{T}_n))/(n+1) \rightarrow \sigma_F^2 < \infty$ . Since  $\psi_n \geq 0$ , this and (7.1) imply that

$$\sum_{k=1}^{\infty} \frac{2}{(k+1)(k+2)} \psi_k < \infty \tag{7.10}$$

and

$$\sigma_F^2 = \sum_{k=1}^{\infty} \frac{2}{(k+1)(k+2)} \psi_k + \lim_{n \rightarrow \infty} \frac{\psi_n}{n+1}, \tag{7.11}$$

where the limit has to exist. However, if  $\lim_{n \rightarrow \infty} \psi_n/(n+1) \neq 0$ , then (7.10) cannot hold. Hence  $\lim_{n \rightarrow \infty} \psi_n/(n+1) = 0$  and (1.58) follows from (7.11).

It follows from (1.58) and (7.1) that  $\sigma_F^2 = 0 \iff \psi_k = 0 \forall k \iff \text{Var}(F(\mathcal{T}_n)) = 0 \forall n$ . The final conclusion follows by (1.57). (If  $f(n, k, n-1-k) = a_n - a_k - a_{n-1-k}$ , then  $F(T) = a_{|T|} - (|T|+1)a_0$  is deterministic.)  $\square$

*Proof of Theorem 1.31.* Similar.  $\square$

## 8 Applications

In this section we give some simple examples of applications of the results above.

### 8.1 Outdegrees

First we consider the number of nodes in  $\mathcal{T}_n$  or  $\Lambda_n$  of a certain outdegree (number of children)  $d \geq 0$ ; we denote these numbers by  $D_{n,d}$  and  $\hat{D}_{n,d}$ , respectively. These equal  $X_n^P$  and  $\hat{X}_n^P$ , where  $P$  is the property that the root has degree  $d$ . Consequently, Corollary 1.25 immediately yields convergence of the expectation and variance divided by  $n$ , and asymptotic normality provided the asymptotic variance does not vanish. By Remark 1.27, this extends to joint convergence for several outdegrees  $d$ .

The case  $d = 0$  is simple; the vertices with outdegree 0 are the leaves, and thus  $D_{n,0} = X_{n,1}$  and  $\hat{D}_{n,0} = \hat{X}_{n,1}$  with means given by (1.41)–(1.42) and variances given in Propositions 1.10 and 1.13. (In this case, the asymptotic normality also follows by Theorem 1.19 or 1.22.) To find the asymptotic variances for  $d > 0$  (and covariances) directly from Corollary 1.25 seems much more difficult. However, as noted already

by Devroye [10], for the binary search tree, when the only outdegrees are 0, 1, 2, it is possible to reduce to the case  $d = 0$ , because

$$D_{n,0} + D_{n,1} + D_{n,2} = n \quad \text{and} \quad D_{n,1} + 2D_{n,2} = n - 1, \quad (8.1)$$

and hence

$$D_{n,2} = D_{n,0} - 1 \quad \text{and} \quad D_{n,1} = n + 1 - 2D_{n,0}. \quad (8.2)$$

Hence we recover the result by Devroye [10, Theorem 2]:

**Example 8.1** (Devroye [10]).  $D_{n,d}$ , the number of vertices with outdegree  $d$  in the binary search tree,  $d = 0, 1, 2$ , has expectation (for  $n > 1$ )

$$\mathbb{E}(D_{n,0}) = \mathbb{E}(D_{n,1}) = \frac{n+1}{3}, \quad \mathbb{E}(D_{n,2}) = \frac{n-2}{3} \quad (8.3)$$

and variance (for  $n > 3$ )

$$\text{Var } D_{n,0} = \text{Var } D_{n,2} = \frac{2}{45}(n+1), \quad \text{Var } D_{n,1} = \frac{8}{45}(n+1) \quad (8.4)$$

and for each  $d \in \{0, 1, 2\}$ , as  $n \rightarrow \infty$ ,

$$\frac{D_{n,d} - \mathbb{E}(D_{n,d})}{\sqrt{\text{Var}(D_{n,d})}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (8.5)$$

**Remark 8.2.** The asymptotic means  $\mu_{D,d} := \lim_{n \rightarrow \infty} \mathbb{E} D_{n,d}/n$  can also be calculated by (1.51) or (1.24). For (1.51), we note that the growing binary tree  $\mathcal{T}_t$  has root degree distributed as  $\text{Bin}(2, 1 - e^{-t})$ , and thus, by the definition of  $\mathcal{T} := \mathcal{T}_\tau$ ,

$$\mu_{D,d} := \int_0^\infty \binom{2}{d} (1 - e^{-t})^d (e^{-t})^{2-d} e^{-t} dt = \binom{2}{d} \int_0^1 (1-x)^d x^{2-d} dx = \frac{1}{3}, \quad (8.6)$$

for each  $d = 0, 1, 2$ , see Aldous [1]. If we instead use (1.24), we obtain

$$\mu_{D,d} = \sum_{k=1}^\infty \frac{2}{(k+1)(k+2)} p_{k,d},$$

where  $p_{k,d}$  is the probability that the root of  $\mathcal{T}_k$  has degree  $d$ . For  $d = 0$  we have  $p_{1,0} = 1$  and  $p_{k,0} = 0$  for  $k > 1$ ; hence  $\mu_{D,0} = \frac{2}{2 \cdot 3} = \frac{1}{3}$ . For  $d = 1$  we have  $p_{1,1} = 0$  and  $p_{k,1} = 2/k$  for  $k \geq 2$ , since the binary search tree generated by a sequence of keys has root degree 1 if and only if the first key is either the largest or the smallest. Hence (1.24) yields

$$\mu_{D,1} = \sum_{k=2}^\infty \frac{2}{(k+2)(k+1)} \cdot \frac{2}{k} = \frac{1}{3}. \quad (8.7)$$

We can similarly show  $\mu_{D,2} = \frac{1}{3}$  too by (1.24).

For the random recursive tree, Corollary 1.25 yields the following, which was proved (using an urn model) by Janson [29], extending earlier results by Mahmoud and Smythe [34]. In fact, [29] gave also a generating function for the variances  $\sigma_{\hat{D},d}^2$  (and the covariances), enabling us to calculate them; as said above, it seems difficult to obtain  $\sigma_{\hat{D},d}^2$  by the methods of this paper except for  $d = 0$ , when  $\sigma_{\hat{D},0}^2 = \hat{\sigma}_{1,1} = \frac{1}{12}$  by (1.20). (The asymptotic formula (8.8) for the expectation was shown earlier by Na and Rapoport [38]. The convergence in probability  $\hat{D}_{n,d}/n \xrightarrow{p} 2^{-d-1}$ , which follows from (8.9), was shown by Meir and Moon [37].)

**Theorem 8.3.** For  $\hat{D}_{n,d}$ , the number of vertices with outdegree  $d \geq 0$  in the random recursive tree, it holds that, as  $n \rightarrow \infty$ ,

$$\frac{\mathbb{E} \hat{D}_{n,d}}{n} \rightarrow 2^{-d-1} \tag{8.8}$$

and furthermore

$$\frac{\hat{D}_{n,d} - 2^{-d-1}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \sigma_{\hat{D},d}^2\right) \tag{8.9}$$

for some constant  $\sigma_{\hat{D},d}^2 \geq 0$ .

*Proof.* By Corollary 1.25, it remains only to calculate  $\mu_{\hat{D},d} := \lim_{n \rightarrow \infty} \mathbb{E} \hat{D}_{n,d}/n$ . We use (1.54) and note that the growing random tree  $\Lambda_t$  has root degree with the Poisson distribution  $\text{Po}(t)$ . Since we stop the process at a random time  $\tau \sim \text{Exp}(1)$  it follows that

$$\mu_{\hat{D},d} = \int_0^\infty \frac{t^d e^{-t}}{d!} \cdot e^{-t} dt = 2^{-d-1},$$

as calculated by Aldous [1]. □

**Remark 8.4.** An alternative approach for finding  $\mu_{\hat{D},d}$  is to use the the natural correspondence between the recursive tree and the binary search tree. A node of outdegree  $d$  in the recursive tree (except the root of the whole tree) corresponds to a left-rooted subtree in the binary search tree with a rightmost path of length  $d - 1$ , and thus to a left-rooted right path of length  $d - 1$ , considering here only paths that cannot be continued further to the right. By symmetry, the expected number of such paths equals the expected number of rightrouted right paths of length  $d - 1$ , but these paths are the right paths of length  $d$ . By symmetry again, on the average half of these paths (except paths from the root) are left-rooted, and thus  $\mathbb{E} \hat{D}_{n,d+1} = \frac{1}{2} \mathbb{E} \hat{D}_{n,d} + O(1)$ . Hence,  $\mu_{\hat{D},d} = 2^{-d-1}$  follows by induction since  $\mu_{\hat{D},0} = \frac{1}{2}$  (see (1.42)).

### 8.2 Protected nodes

We proceed to use fringe trees to study the so-called protected nodes that recently have been studied in several types of random trees, see, e.g., [6, 8, 12, 35, 36] and the references there. A node is  $\ell$ -protected if the shortest distance to a descendant that is a leaf is at least  $\ell$ . The most studied case is  $\ell = 2$ : a node is *two-protected* if it is neither a leaf nor the parent of a leaf.

**Remark 8.5.** The case  $\ell = 1$  is a bit trivial, at least for the random trees studied here: a node is 1-protected if and only if it is a non-leaf. Hence, for binary search trees and random recursive trees, where the number of nodes is given, it is equivalent to study the number of leaves, which was done in Section 8.1. (However, for random trees with a random number of nodes, for example the ternary search tree studied in [27], this case too is interesting.)

Corollary 1.25 implies immediately that for any  $\ell$ , the number of  $\ell$ -protected nodes is asymptotically normal in both the binary search tree and the random recursive tree, at least provided the asymptotic variances below are non-zero, which is an obvious conjecture although we have no rigorous proof for  $\ell \geq 3$ , cf. Problem 1.26.

**Theorem 8.6.** Let  $\ell \geq 1$  and let  $Y_{\ell,n}$  denote the number of  $\ell$ -protected nodes in a binary search tree  $\mathcal{T}_n$ . Then, for some constants  $\mu_{Y,\ell} = \mathbb{P}(\text{the root of } \mathcal{T} \text{ is } \ell\text{-protected}) > 0$  and  $\sigma_{Y,\ell}^2 \geq 0$ , with at least  $\sigma_{Y,2}^2 > 0$ ,

$$\frac{\mathbb{E}(Y_{\ell,n})}{n} \rightarrow \mu_{Y,\ell}, \tag{8.10}$$

$$\frac{\text{Var}(Y_{\ell,n})}{n} \rightarrow \sigma_{Y,\ell}^2, \tag{8.11}$$

and

$$\frac{Y_{\ell,n} - n\mu_{Y,\ell}}{\sqrt{n}}, \frac{Y_{\ell,n} - \mathbb{E}(Y_{\ell,n})}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_{Y,\ell}^2). \quad (8.12)$$

Similarly, let  $Z_{\ell,n}$  denote the number of  $\ell$ -protected nodes in a random recursive tree  $\Lambda_n$ . Then, for some constants  $\mu_{Z,\ell} = \mathbb{P}(\text{the root of } \Lambda \text{ is } \ell\text{-protected}) > 0$  and  $\sigma_{Z,\ell}^2 \geq 0$ , with at least  $\sigma_{Z,2}^2 > 0$ ,

$$\frac{\mathbb{E}(Z_{\ell,n})}{n} \rightarrow \mu_{Z,\ell}, \quad (8.13)$$

$$\frac{\text{Var}(Z_{\ell,n})}{n} \rightarrow \sigma_{Z,\ell}^2 \quad (8.14)$$

and

$$\frac{Z_{\ell,n} - \mathbb{E}(Z_{\ell,n})}{\sqrt{n}}, \frac{Z_{\ell,n} - n\mu_{Z,\ell}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_{Z,\ell}^2). \quad (8.15)$$

*Proof.* Let  $P$  be the class of trees such that the root is  $\ell$ -protected and apply Corollary 1.25, noting that  $Y_{\ell,n} = X_n^P$  and  $Z_{\ell,n} = \hat{X}_n^P$ .

That  $\sigma_{Y,2}^2 > 0$  and  $\sigma_{Z,2}^2 > 0$  follows from Theorems 1.30 and 1.31.  $\square$

By Remark 1.27, we also obtain joint normality for several  $\ell$ .

Theorem 8.6 includes several earlier results, proved by several different methods: (8.10) was shown by Mahmoud and Ward [35] for  $\ell = 2$  and by Bóna [6] and Devroye and Janson [12] in general; [35] also shows (8.11)–(8.12) for  $\ell = 2$ ; (8.13) was shown by Mahmoud and Ward [36] for  $\ell = 2$  and by Devroye and Janson [12] in general; [36] also shows for  $\ell = 2$  the weaker version of (8.14) that  $\text{Var}(Z_{2,n}) = O(1/n)$ .

To calculate the asymptotic means and variances is more complicated, however. For the binary search tree, the asymptotic means  $\mu_{Y,\ell}$  were calculated for  $\ell \leq 4$  by Bóna [6] (using generating functions) and Devroye and Janson [12] (using the formula (1.51) as here) to be  $\mu_{Y,1} = \frac{2}{3}$ ,  $\mu_{Y,2} = \frac{11}{30}$ ,  $\mu_{Y,3} = \frac{1249}{8100}$ ,  $\mu_{Y,4} = \frac{103365591157608217}{2294809143026400000}$ ; the methods in these papers apply to arbitrary  $\ell$  (and yield rational numbers) but explicit calculations quickly become cumbersome.

For the random recursive tree,  $\mu_{Z,1} = \frac{1}{2}$  as a consequence of Theorem 8.3 (with  $d = 0$ ) and  $\mu_{Z,2} = \frac{1}{2} - e^{-1}$  by Mahmoud and Ward [36] and Devroye and Janson [12]; the method in [12] is based on (1.54) as here and yields (recursively) a complicated integral expression for every  $\ell$ , but we do not know any closed form for  $\ell \geq 3$ .

For the asymptotic variances, the formulas (1.52) and (1.55) do not seem to easily yield explicit formulas (although they might be useful for numerical approximations). The only value that we know, except for  $\ell = 1$  when  $\sigma_{Y,1}^2 = \sigma_{1,1} = \frac{2}{45}$  (cf. (8.4)) and  $\sigma_{Z,1}^2 = \hat{\sigma}_{1,1} = \frac{1}{12}$ , is  $\sigma_{Y,2}^2 = \frac{29}{225}$ . In fact, for the binary search tree and  $\ell = 2$  we can compute the mean and variance of  $Y_{2,n}$  exactly by a special trick; the result is stated in the following theorem earlier shown by Mahmoud and Ward [35, Theorems 2.1, 2.2 and 3.1] (using generating functions and recurrences), which is a more precise version of the special case  $\ell = 2$  of Theorem 8.6 for the binary search tree. (See also [27, Theorem 1.2] for a different proof of the asymptotic normality using Pólya urns.)

**Theorem 8.7** (Mahmoud and Ward [35]). *Let  $Y_{2,n}$  denote the number of two-protected nodes in a binary search tree  $\mathcal{T}_n$ . Then*

$$\mathbb{E}(Y_{2,n}) = \frac{11}{30}n - \frac{19}{30}, \quad \text{for } n \geq 4, \quad (8.16)$$

and

$$\text{Var}(Y_{2,n}) = \frac{29}{225}(n + 1), \quad \text{for } n \geq 8. \quad (8.17)$$

Furthermore, as  $n \rightarrow \infty$ ,

$$\frac{Y_{2,n} - \frac{11}{30}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \frac{29}{225}\right). \tag{8.18}$$

We provide a simple proof of this theorem using our results on fringe trees. Moreover, our approach using fringe trees also allows us to provide a simple proof of the following result which was conjectured in [35, Conjecture 2.1].

**Theorem 8.8.** *For each fixed integer  $k \geq 1$ , there exists a polynomial  $p_k(n)$  of degree  $k$ , the leading term of which is  $(\frac{11}{30})^k$ , such that  $\mathbb{E}(Y_{2,n}^k) = p_k(n)$  for all  $n \geq 4k$ .*

*Proof of Theorem 8.7.* In a binary tree (with at least 2 nodes), the number of nodes that are not two-protected equals two times the number of leaves (counting all the leaves and all the parents of the leaves) minus the number of cherry subtrees, i.e., subtrees consisting of a root with one left and one right child that both are leaves (since these are the only cases when a parent is counted twice). Thus, writing  $L$  for a tree that is a single leaf and  $C$  for a tree that is a cherry,

$$Y_{2,n} = n - 2X_n^L + X_n^C. \tag{8.19}$$

Hence,

$$\mathbb{E}(Y_{2,n}) = n - 2\mathbb{E}(X_n^L) + \mathbb{E}(X_n^C) \tag{8.20}$$

and

$$\text{Var}(Y_{2,n}) = 4\text{Var}(X_n^L) + \text{Var}(X_n^C) - 4\text{Cov}(X_n^L, X_n^C). \tag{8.21}$$

By (1.4), the expected number of subtrees of  $\mathcal{T}_n$  isomorphic to a tree  $T$  of size  $|T| = k$  is

$$\mathbb{E}(X_n^T) = \frac{2(n+1)}{(k+2)(k+1)} p_{k,T}, \tag{8.22}$$

where  $p_{k,T} = \mathbb{P}(\mathcal{T}_k = T) = |A_k^T|/k!$  where  $A_k^T$  is the set of permutations of length  $k$  that give rise to the binary search tree  $T$ . Evidently  $p_{1,L} = 1$ , and for the cherry  $C$  we have  $|C| = 3$  and  $p_{3,C} = \frac{1}{3}$ . Thus  $\mathbb{E}(X_n^L) = (n+1)/3$  (for  $n \geq 2$ ), as already seen in (8.3), and  $\mathbb{E}(X_n^C) = (n+1)/30$  (for  $n \geq 4$ ). Hence (8.20) yields (8.16).

To calculate  $\text{Var}(Y_{2,n})$  we use Proposition 1.9. Using the notations there  $q_L^L = q_C^C = 1$  and  $q_L^C = 2$ , and simple calculations yield (for  $n \geq 8$ )  $\text{Var}(X_n^L) = \frac{2}{45}(n+1)$  (as shown in (8.4)),  $\text{Cov}(X_n^L, X_n^C) = \frac{2}{105}(n+1)$  and  $\text{Var}(X_n^C) = \frac{43}{1575}(n+1)$ , which together with (8.21) yield (8.17).

Since any linear combination of the components in a random vector with a multivariate normal distribution is normal, the asymptotic normality (8.18) follows from (8.19) and Theorem 1.22.  $\square$

**Remark 8.9.** Alternatively, (8.19) shows that  $Y_{2,n} = F(\mathcal{T}_n)$  for the functional

$$f(T) := 1 - 2 \cdot \mathbf{1}\{T = L\} + \mathbf{1}\{T = C\} \tag{8.23}$$

and the results follow by Theorem 1.14 (with the same calculations as above).

*Proof of Theorem 8.8.* We use again (8.19) and the cyclic representation (2.14), which show that

$$Y_{2,n} = n + \sum_{i=1}^{n+1} g(\sigma(i-1, 4)) \tag{8.24}$$

for some functional  $g$  defined by  $g(\sigma(i-1, 5)) := -2I_{i,1} + I_{i,3}f_C(\sigma(i, 3))$  where  $f_C$  is the indicator that the permutation defines a cherry. Thus  $\mathbb{E}Y_{2,n}^k$  can be calculated by substituting (8.24) and expanding, and the result follows easily by collecting terms that are equal since the random variables  $g(\sigma(i-1, 5))$  are i.i.d. and 4-dependent.  $\square$

**Remark 8.10.** The asymptotic mean  $\mu_{Y,2}$  in (8.10) can also be derived directly from (1.24) in Theorem 1.14. We give this alternative calculation to illustrate our results, although in this case (1.39) (see [12]) or (8.20) yield simpler calculations. Let  $p_k$  be the probability that the root of  $\mathcal{T}_k$  is two-protected. Since the complement of the two-protected nodes consists of the leaves and the parents of the leaves we obtain (for  $k \geq 2$ ), that the root is not two-protected if and only if it has a child that is a leaf, which going back to the construction of the binary search tree by a sequence of keys means that the first key is either the second smallest or the second largest key. Hence,  $p_k = 1 - 2/k$  for  $k \geq 4$ . Furthermore,  $p_1 = p_2 = 0$  and  $p_3 = 2/3$ . Consequently, (1.24) yields

$$\mu_{Y,2} = \frac{2}{4 \cdot 5} \cdot \frac{2}{3} + \sum_{k=4}^{\infty} \frac{2}{(k+2)(k+1)} \cdot \left(1 - \frac{2}{k}\right) = \frac{11}{30}. \tag{8.25}$$

Mahmoud and Ward [35] also discuss the two-protected nodes in the extended binary search tree. Recall that an extended binary search tree is a binary search tree where the  $n + 1$  external children are added. The leaves in the extended binary tree are the external vertices; hence the two-protected nodes are those that have at least distance two to an external vertex, i.e., the internal vertices that have no external children. In other words, the two-protected nodes are precisely the nodes in the binary search tree that have outdegree 2. Thus Example 8.1 directly implies the following theorem in [35].

**Theorem 8.11** (Mahmoud and Ward [35]). *Let  $Z_n$  denote the number of two-protected nodes in an extended binary search tree. Then*

$$\mathbb{E}(Z_n) = \frac{n}{3} - \frac{2}{3}, \quad \text{for } n \geq 2, \tag{8.26}$$

and

$$\text{Var}(Z_n) = \frac{2}{45}(n + 1), \quad \text{for } n \geq 4. \tag{8.27}$$

Furthermore, as  $n \rightarrow \infty$ ,

$$\frac{Z_n - \frac{n}{3}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \frac{2}{45}\right). \tag{8.28}$$

We can also show the following result which was conjectured in [35, Conjecture 4.1].

**Theorem 8.12.** *Let  $Z_n$  denote the number of two-protected nodes in the extended binary search tree. For each fixed integer  $k \geq 1$ , there exists a polynomial  $p_k(n)$  of degree  $k$ , the leading term of which is  $\frac{1}{3^k}$ , such that  $\mathbb{E}(Z_n^k) = p_k(n)$  for all  $n \geq 2k$ .  $\square$*

*Proof of Theorem 8.12.* By the comments above, (8.2) and (3.1),

$$Z_n = D_{n,2} = D_{n,0} - 1 = X_{n,1} - 1 = \sum_{i=1}^{n+1} I_{i,1} - 1. \tag{8.29}$$

The result follows from the fact that the indicator functions  $I_{i,1}$  are 2-dependent, cf. the proof of Theorem 8.8.  $\square$

### 8.3 Shape functionals

#### 8.3.1 Binary search trees

Consider first a binary tree  $T$  with  $|T| = n$ , and define  $P(T) := p_{n,T} = \mathbb{P}(\mathcal{T}_n = T)$ . It is easy to see that

$$P(T) := \mathbb{P}(\mathcal{T}_n = T) = \prod_{v \in T} |T(v)|^{-1}, \tag{8.30}$$

see e.g. (more generally for  $m$ -ary search trees) Dobrow and Fill [13]. The functional  $P(T)$  is known as the *shape functional* for binary trees.

By (8.30), the functional  $F(T) := -\log P(T)$  is given by (1.1) with  $f(T) = \log |T|$ .

**Example 8.13** (Fill [17]). Theorems 1.14 and 1.30 apply to  $F(T) = -\log P(T)$ , and it follows immediately that, as shown by Fill [17] (with some further details), see also Fill and Kapur [20] for  $m$ -ary search trees, as  $n \rightarrow \infty$ ,

$$-\mathbb{E} \log P(\mathcal{T}_n) \sim n \sum_{k=2}^{\infty} \frac{2 \log k}{(k+1)(k+2)} \tag{8.31}$$

and

$$\frac{\log P(\mathcal{T}_n) - \mathbb{E} \log P(\mathcal{T}_n)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{8.32}$$

for some  $\sigma^2 > 0$  which can be computed from (1.58). (We have  $\sigma^2 > 0$  by Theorem 1.30, since e.g.  $P(\mathcal{T}_3)$  is not deterministic.)

### 8.3.2 Unordered random recursive trees

For an unordered rooted tree  $\Lambda$  with  $|\Lambda| = n$ , we similarly define  $\hat{P}(\Lambda) := \hat{p}_{n,\Lambda} = \mathbb{P}(\Lambda_n = \Lambda)$  (regarding  $\Lambda_n$  as an unordered tree).. Then, see Feng and Mahmoud [15],

$$\hat{P}(\Lambda) := \mathbb{P}(\Lambda_n = \Lambda) = n \prod_{v \in \Lambda} s(\Lambda, v)^{-1} |\Lambda(v)|^{-1}, \tag{8.33}$$

where  $s(\Lambda, v)$  is the number of permutations of the children of  $v$  that can be extended to automorphisms of the tree  $\Lambda$ , i.e., if  $v$  has  $\nu_1$  children  $v_{1i}$  such that  $\Lambda(v_{1i}) \cong \Lambda_1$  for some rooted tree  $\Lambda_1$ ,  $\nu_2$  children  $v_{2i}$  such that  $\Lambda(v_{2i}) \cong \Lambda_2$  for some different rooted tree  $\Lambda_2, \dots$ , then  $s(\Lambda, v) = \prod_j \nu_j!$ . This functional  $\hat{P}(T)$  is the shape functional for unordered rooted trees.

By (8.33), the functional  $-\log \hat{P}(\Lambda) = F(\Lambda) - \log |\Lambda|$ , where  $F(\Lambda)$  is given by (1.1) with  $f(\Lambda) = \log |\Lambda| + \log s(\Lambda, o)$ , where  $o$  is the root. Note that this functional is more complicated than the corresponding one for binary trees, and that  $f(\Lambda)$  no longer depends only on the size  $|\Lambda|$  (nor only on the size of  $\Lambda$  and of the principal subtrees as in Theorem 1.31). Nevertheless, Theorem 1.14 applies and yields the following. (It seems obvious that  $\hat{\sigma}^2 > 0$ , but we have no rigorous proof. We have not attempted any numerical estimate.)

**Theorem 8.14.** As  $n \rightarrow \infty$ ,

$$-\mathbb{E} \log \hat{P}(\Lambda_n) \sim n \hat{\mu} \tag{8.34}$$

and

$$\frac{\log \hat{P}(\Lambda_n) - \mathbb{E} \log \hat{P}(\Lambda_n)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}^2) \tag{8.35}$$

for some  $\hat{\mu} > 0$  and  $\hat{\sigma}^2 \geq 0$  which in principle can be computed from (1.30)–(1.31).

*Proof.* Let  $d(\Lambda)$  be the degree of the root  $o$  of  $\Lambda$ . Then, crudely,  $s(o, \Lambda) \leq d(\Lambda)!$  and thus

$$\log s(o, \Lambda) \leq d(\Lambda) \log d(\Lambda) \leq d(\Lambda) \log |\Lambda|. \tag{8.36}$$

From the definition of the random recursive tree,  $d(\Lambda_k) \stackrel{d}{=} \sum_{i=1}^{k-1} I_i$ , where  $I_i \sim \text{Be}(1/i)$  are i.i.d., and a simple calculation shows that

$$\mathbb{E} d(\Lambda_k)^2 = O(\log^2 k). \tag{8.37}$$

By (8.36) and (8.37) we have, rather crudely,

$$\begin{aligned} \mathbb{E}(f(\Lambda_k)^2) &= \mathbb{E}((\log s(o, \Lambda_k) + \log k)^2) \leq 2 \mathbb{E}(d(\Lambda_k)^2 \log^2 k) + 2 \log^2 k \\ &= O(\log^4 k). \end{aligned} \tag{8.38}$$

Hence, (1.27)–(1.29) hold, and Theorem 1.14(ii) applies. □

### 8.3.3 Ordered random recursive trees

Now consider the random recursive tree  $\Lambda_n$  as an ordered tree. For an ordered rooted tree  $\Lambda$  with  $|\Lambda| = n$ , we define  $\hat{P}(\Lambda) := \hat{p}_{n,\Lambda} = \mathbb{P}(\Lambda_n = \Lambda)$ . It is easily seen that if we denote the children of a node  $v$  by  $v_1, \dots, v_{d(v)}$ , then

$$\hat{P}(\Lambda) := \mathbb{P}(\Lambda_n = \Lambda) = \prod_{v \in \Lambda} \prod_{i=1}^{d(v)} \left( \sum_{j=i}^{d(v)} |\Lambda(v_j)| \right)^{-1} \tag{8.39}$$

This functional  $\hat{P}(T)$  is the shape functional for ordered rooted trees.

By (8.39), the functional  $-\log \hat{P}(\Lambda) = F(\Lambda)$ , where  $F(\Lambda)$  is given by (1.1) with

$$f(\Lambda) = \sum_{i=1}^d \log \sum_{j=i}^d |\Lambda_j| \tag{8.40}$$

where  $d$  is the degree of the root and  $\Lambda_1, \dots, \Lambda_d$  are the principal subtrees. This functional is of the type in Theorem 1.31, and we obtain the following.

**Theorem 8.15.** *As  $n \rightarrow \infty$ ,*

$$-\mathbb{E} \log \hat{P}(\Lambda_n) \sim n\hat{\mu} \tag{8.41}$$

and

$$\frac{\log \hat{P}(\Lambda_n) - \mathbb{E} \log \hat{P}(\Lambda_n)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}^2) \tag{8.42}$$

for some  $\hat{\mu} > 0$  and  $\hat{\sigma}^2 > 0$  which in principle can be computed from (1.30) and (1.60).

*Proof.* Let again  $d(\Lambda)$  be the degree of the root of  $\Lambda$ . Then,  $f(\Lambda) \leq d(\Lambda) \log |\Lambda|$ , and (8.37) shows that (8.38) holds in the ordered case too; thus the result follows by Theorems 1.14 and 1.31.  $\square$

## A Appendix: proof of (1.36)–(1.37)

Let  $|T| = k$  and  $|T'| = m < k$ . By (1.9),

$$\beta(k, m) = \frac{4}{(k+2)(m+1)(m+2)} + O\left(\frac{1}{k^2 m}\right) \tag{A.1}$$

and thus (1.11) yields

$$\sigma_{T, T'} = \frac{2}{(k+1)(k+2)} p_{k, T} \left( q_{T'}^T - \frac{2(k+1)}{(m+1)(m+2)} p_{m, T'} \right) + O\left(\frac{p_{k, T} p_{m, T'}}{k^2 m}\right). \tag{A.2}$$

Note first that we may ignore the  $O$  term in (A.2), since

$$\sum_{m=1}^{\infty} \sum_{k>m} \sum_{|T|=k} \sum_{|T'|=m} \frac{p_{k, T} p_{m, T'}}{k^2 m} = \sum_{m=1}^{\infty} \sum_{k>m} \frac{1}{k^2 m} < \infty; \tag{A.3}$$

hence it suffices to show that

$$S_1 := \sum_{m=1}^{\infty} \sum_{k>m} \sum_{|T|=k} \sum_{|T'|=m} \frac{2}{(k+1)(k+2)} p_{k, T} \left| q_{T'}^T - \frac{2(k+1)}{(m+1)(m+2)} p_{m, T'} \right| = \infty. \tag{A.4}$$

Note that  $q_{T'}^T$  is an integer. Thus, if  $\frac{2(k+1)}{(m+1)(m+2)} p_{m, T'} \leq \frac{1}{2}$ , then

$$\left| q_{T'}^T - \frac{2(k+1)}{(m+1)(m+2)} p_{m, T'} \right| \geq \frac{2(k+1)}{(m+1)(m+2)} p_{m, T'}. \tag{A.5}$$

Hence,

$$\begin{aligned}
 S_1 &\geq \sum_{m=1}^{\infty} \sum_{|T'|=m} \sum_{k=m+1}^{m^2/(4p_{m,T'})} \sum_{|T|=k} \frac{2}{(k+1)(k+2)} p_{k,T} \frac{2(k+1)}{(m+1)(m+2)} p_{m,T'} \\
 &= \sum_{m=1}^{\infty} \sum_{|T'|=m} \frac{4}{(m+1)(m+2)} p_{m,T'} \sum_{k=m+1}^{m^2/(4p_{m,T'})} \frac{1}{k+2} \\
 &\geq \sum_{m=1}^{\infty} \sum_{|T'|=m} \frac{4}{(m+1)(m+2)} p_{m,T'} \log \frac{m^2}{4p_{m,T'}(m+3)} \\
 &\geq \sum_{m=6}^{\infty} \frac{1}{m^2} \sum_{|T'|=m} p_{m,T'} \log \frac{1}{p_{m,T'}} = \sum_{m=6}^{\infty} \frac{1}{m^2} \mathbb{E} \log \frac{1}{p_{m,\mathcal{T}_m}}.
 \end{aligned} \tag{A.6}$$

However, we saw in (8.31) that  $\mathbb{E} \log p_{m,\mathcal{T}_m}^{-1} \sim \mu m$  as  $m \rightarrow \infty$  for some constant  $\mu > 0$ , and thus the sum in (A.6) diverges, which as said above implies (1.36) by (A.2)–(A.3).

The proof of the random recursive case (1.37) is similar, using (1.17), (1.15) and (8.34). (It suffices to consider the versions with unlabelled binary trees and unordered rooted trees, since the versions with increasing trees or ordered trees have larger  $\mathbb{E} \log p_{m,\mathcal{T}_m}^{-1}$  and  $\mathbb{E} \log \hat{p}_{m,\Lambda_m}^{-1}$ .)  $\square$

**Remark A.1.** The proof above needs only lower bounds for  $\mathbb{E} \log p_{m,\mathcal{T}_m}^{-1}$  and  $\mathbb{E} \log \hat{p}_{m,\Lambda_m}^{-1}$ . We can use the simple bounds  $\log p_{m,\mathcal{T}_m}^{-1} \geq X_{m,2} \log 2$  and  $\log \hat{p}_{m,\Lambda_m}^{-1} \geq \hat{X}_{m,2} \log 2 - \log m$ , see (8.30) and (8.33), together with (1.41)–(1.42) instead of (8.31) and (8.34).

**Remark A.2.** Recalling the notation in Section 1,  $q_{T'}^T$  is the value of  $X_k^{T'}$  when  $\mathcal{T}_k = T$ , and it follows, using (1.4), that

$$\sum_{|T|=k} p_{k,T} \left| q_{T'}^T - \frac{2(k+1)}{(m+1)(m+2)} p_{m,T'} \right| = \mathbb{E} \left| X_k^{T'} - \mathbb{E} X_k^{T'} \right| \tag{A.7}$$

and thus, dropping the prime,

$$S_1 = \sum_T \sum_{k>|T|} \frac{2}{(k+1)(k+2)} \mathbb{E} |X_k^T - \mathbb{E} X_k^T|. \tag{A.8}$$

A similar sum appears in the proof for the random recursive tree.

## References

- [1] Aldous D.: Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1**, (1991), no. 2, 228–266. MR-1102319
- [2] Aldous D.: Probability distributions on cladograms. *Random Discrete Structures (Minneapolis, MN, 1993)*, 1–18, IMA Vol. Math. Appl., 76, Springer, New York, 1996. MR-1395604
- [3] Barbour A. D., Holst L. and Janson S.: *Poisson Approximation*. Oxford University Press, New York, 1992. MR-1163825
- [4] Billingsley P., *Convergence of Probability Measures*. John Wiley and Sons, New York, 1968. MR-0233396
- [5] Blum M. G. B., François O. and Janson S.: The mean, variance and limiting distributions of two statistics sensitive to phylogenetic tree balance. *Ann. Appl. Prob.* **16**, (2006), 2195–2214. MR-2288718

- [6] Bóna M.:  $k$ -protected nodes in binary search trees. *Adv. Appl. Math.* **53**, (2014), 1–11. MR-3149690
- [7] Chang H. and Fuchs M.: Limit theorems for patterns in phylogenetic trees. *J. Math. Biol.* **60**, (2010), 481–512. MR-2587587
- [8] Cheon G.-S. and Shapiro L. W.: Protected points in ordered trees. *Applied Mathematics Letters* **21**, (2008), no. 5, 516–520. MR-2402845
- [9] Dennert F. and Grübel R.: On the subtree size profile of binary search trees, *Combin. Probab. Comput.* **19**, (2010), no. 4, 561–578. MR-2647493
- [10] Devroye L.: Limit laws for local counters in random binary search trees. *Random Structures Algorithms* **2**, (1991), no. 3, 303–315. MR-1109697
- [11] Devroye L.: Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.* **32**, (2002/03), no. 1, 152–171. MR-1954858
- [12] Devroye L. and Janson S.: Protected nodes and fringe subtrees in some random trees. *Electronic Communications in Probability* **19**, (2014), no. 6, 1–10. MR-3164753
- [13] Dobrow R. P. and Fill J. A.: Multiway trees of maximum and minimum probability under the random permutation model. *Combin. Probab. Comput.* **5**, (1996), no. 4, 351–371. MR-1426429
- [14] Drmota M.: *Random Trees*. Springer, Vienna, 2009. MR-2484382
- [15] Feng Q. and Mahmoud H. M.: On the variety of shapes on the fringe of a random recursive tree. *J. Appl. Probab.* **47**, (2010), no. 1, 191–200. MR-2654767
- [16] Feng Q., Mahmoud H. M. and Panholzer A.: Phase changes in subtree varieties in random recursive and binary search trees. *SIAM J. Discrete Math.* **22**, (2008), no. 1, 160–184. MR-2383234
- [17] Fill J. A.: On the distribution of binary search trees under the random permutation model. *Random Structures Algorithms* **8**, (1996), no. 1, 1–25. MR-1368848
- [18] Fill J. A., Flajolet P. and Kapur N.: Singularity analysis, Hadamard products, and tree recurrences. *J. Comput. Appl. Math.* **174**, (2005), no. 2, 271–313. MR-2106441
- [19] Fill J. A. and Janson S.: Smoothness and decay properties of the limiting Quicksort density function. *Mathematics and Computer Science (Versailles, 2000)*, 53–64, D. Gardy and A. Mekkadem, eds., Birkhäuser, Basel, 2000. MR-1798287
- [20] Fill J. A. and Kapur N.: Transfer theorems and asymptotic distributional results for  $m$ -ary search trees. *Random Structures Algorithms* **26**, (2005), no. 4, 359–391. MR-2139868
- [21] Flajolet P., Gourdon X. and Martínez C.: Patterns in random binary search trees. *Random Structures Algorithms* **11**, (1997), no. 3, 223–244. MR-1609509
- [22] Fuchs M.: Subtree sizes in recursive trees and binary search trees: Berry–Esseen bounds and Poisson approximations. *Combin. Probab. Comput.* **17**, (2008), no. 5, 661–680. MR-2454562
- [23] Fuchs M.: Limit theorems for subtree size profiles of increasing trees. *Combin. Probab. Comput.* **21**, (2012), no. 3, 412–441. MR-2912789

- [24] Fuchs M., Hwang H.-K. and Neininger R.: Profiles of random trees: limit theorems for random recursive trees and binary search trees. *Algorithmica* **46**, (2006), no. 3-4, 367–407. MR-2291961
- [25] Gut A.: *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013. MR-2977961
- [26] Hardy G. H., Littlewood J. E. and Pólya G.: *Inequalities*. 2nd ed., Cambridge University Press, Cambridge, 1952. MR-0046395
- [27] Holmgren C. and Janson S.: Asymptotic distribution of two-protected nodes in ternary search trees. Preprint, 2014. arXiv:1403.5573
- [28] Hwang H. K. and Neininger R.: Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.* **31**, (2002), no. 6, 1687–1722. MR-1954876
- [29] Janson S.: Asymptotic degree distribution in random recursive trees. *Random Structures Algorithms* **26**, (2005), no. 1-2, 69–83. MR-2116576
- [30] Janson S.: Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. *Random Structures Algorithms*, to appear. MR-2116576
- [31] Janson S., Łuczak T., and Ruciński A.: *Random Graphs*. John Wiley, New York, 2000. MR-1782847
- [32] Knuth D. E.: *The Art of Computer Programming. Volume 1: Fundamental Algorithms*. Third ed., Addison-Wesley, Reading, Mass., 1997. MR-3077152
- [33] Knuth D. E.: *The Art of Computer Programming. Volume 3: Sorting and Searching*. Second ed., Addison-Wesley, Reading, Mass., 1998. MR-3077154
- [34] Mahmoud H. M. and Smythe R. T.: Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms* **3**, (1992), no. 3, 255–266. MR-1164839
- [35] Mahmoud H. M. and Ward M. D.: Asymptotic distribution of two-protected nodes in random binary search trees. *Appl. Math. Lett.* **25**, (2012), no. 12, 2218–2222. MR-2967819
- [36] Mahmoud H. M. and Ward M. D.: Asymptotic properties of protected nodes in random recursive trees. Preprint, 2014.
- [37] Meir A. and Moon J. W.: Recursive trees with no nodes of out-degree one. *Congr. Numer.* **66**, (1988), 49–62. MR-0992887
- [38] Na N. S. and Rapoport A.: Distribution of nodes of a tree by degree. *Math. Biosci.* **6**, (1970), 313–329. MR-0278985
- [39] Régnier M.: A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* **23**, (1989), 335–343. MR-1020478
- [40] Rösler U.: A limit theorem for ‘Quicksort’. *RAIRO Inform. Théor. Appl.* **25**, (1991), 85–100. MR-1104413

---

# Electronic Journal of Probability

## Electronic Communications in Probability

---

### Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)

### Economical model of EJP-ECP

- Low cost, based on free software (OJS<sup>1</sup>)
- Non profit, sponsored by IMS<sup>2</sup>, BS<sup>3</sup>, PKP<sup>4</sup>
- Purely electronic and secure (LOCKSS<sup>5</sup>)

### Help keep the journal free and vigorous

- Donate to the IMS open access fund<sup>6</sup> (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

---

<sup>1</sup>OJS: Open Journal Systems <http://pkp.sfu.ca/ojs/>

<sup>2</sup>IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

<sup>3</sup>BS: Bernoulli Society <http://www.bernoulli-society.org/>

<sup>4</sup>PK: Public Knowledge Project <http://pkp.sfu.ca/>

<sup>5</sup>LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

<sup>6</sup>IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>