Using Stein's Method to Show Poisson and Normal Limit Laws for Fringe Subtrees

Cecilia Holmgren[1†] and Svante Janson[2‡]

[1]*Department of Mathematics, Stockholm University, 114 18 Stockholm, Sweden*
[2]*Department of Mathematics, Uppsala University, SE-75310 Uppsala, Sweden*

Abstract.

We consider sums of functions of fringe subtrees of binary search trees and random recursive trees (of total size n).

The use of Stein's method and certain couplings allow provision of simple proofs showing that in both of these trees, the number of fringe subtrees of size $k < n$, where $k \to \infty$, can be approximated by a Poisson distribution. Combining these results and another version of Stein's method, we can also show that for $k = o(\sqrt{n})$, the number of fringe subtrees in both types of random trees has asymptotically a normal distribution as $n \to \infty$. Furthermore, using the Cramér–Wold device, we show that a random vector with components corresponding to the random number of copies of certain fixed fringe subtrees T_i, has asymptotically a multivariate normal distribution. We can then use these general results on fringe subtrees to obtain simple solutions to a broad range of problems relating to random trees; as an example, we can prove that the number of protected nodes in the binary search tree has asymptotically a normal distribution.

Keywords: Fringe subtrees. Stein's method. Couplings. Limit laws. Binary search trees. Recursive trees.

1 Introduction

In this paper we consider fringe subtrees of the random binary search tree, as well as of the random recursive tree; recall that a fringe subtree is a subtree consisting of some node and all its descendants, see Aldous [1] for a general theory, and note that fringe subtrees typically are "small" compared to the whole tree. This is an extended abstract of [9] where further details are given.

We will use a representation of Devroye [4, 5] for the binary search tree, and a well-known bijection between binary trees and recursive trees, together with different 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 subtrees, and more direct proofs of several earlier results in the field.

†Supported in part by the Swedish Research Council.
‡Supported in part by the Knut and Alice Wallenberg Foundation.

The *binary search tree* is the tree representation of the sorting algorithm Quicksort, see e.g. [11]. Starting with n distinct numbers called keys, we draw one of the keys at random and associate it to the root. Then we draw one of the remaining keys. We compare it with the root, and associate it 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 recursively by drawing new keys 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.

We use the representation of the binary search tree by Devroye [4, 5]. 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 random 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 Λ_n denote a random recursive tree with n nodes.

There is a well-known bijection between ordered trees of size n and binary trees of size $n-1$, see e.g. Knuth [12, 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 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 [4], see also Fuchs, Hwang and Neininger [8], the natural correspondence extends to a coupling between the random recursive tree Λ_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 Λ_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 $1 < k < n$, fringe subtrees of size k in the random recursive tree Λ_n, correspond to left-rooted fringe subtrees of size $k-1$ in the binary search tree $\mathcal{T}_{n-1}$, while fringe subtrees of size 1 (i.e., leaves) correspond to nodes without a left child.

We consider first only the sizes of the fringe subtrees. The results in the following two theorems, except the explicit rate in (3)–(4), were shown by Feng, Mahmoud and Panholzer [6] and Fuchs [7] by using variants of the method of moments. Theorem 1.3 was earlier proved for fixed k by Devroye [4] (using the central limit theorem for m-dependent variables), see also Aldous [1]. The part (5) of Theorem 1.3 was proved for a smaller range of k by Devroye [5] using Stein's method. (The mean (1) is also found in [5].) In the present paper we continue this approach, and use Stein's method for both Poisson and normal approximations to provide simple proofs for the full range.

We recall the definition of the total variation distance between two probability measures.

Definition 1.1 *Let $(\mathcal{X}, \mathcal{A})$ be any measurable space. The total variation distance d_{TV} between two probability measures μ_1 and μ_2 on $\mathcal{X}$ is defined to be*

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

Let $\mathcal{L}(X)$ denote the distribution of a random variable X. $\mathrm{Po}(\mu)$ denotes the Poisson distribution with mean μ, and $\mathcal{N}(0,1)$ the standard normal distribution. Convergence in distribution is denoted by $\xrightarrow{\mathrm{d}}$.

Theorem 1.2 *Let $X_{n,k}$ be the number of fringe subtrees of size k in the random binary search tree $\mathcal{T}_n$ and similarly let $\hat{X}_{n,k}$ be the number of fringe subtrees in the random recursive tree Λ_n. Let $k = k_n$ where $k < n$. Furthermore, let*

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

$$\hat{\mu}_{n,k} := \mathbb{E}(\hat{X}_{n,k}) = \frac{n}{k(k+1)}. \tag{2}$$

Then, for the binary search tree,

$$d_{TV}(\mathcal{L}(X_{n,k}), \mathrm{Po}(\mu_{n,k})) = \frac{1}{2} \sum_{l \geq 0} \Big| \mathbb{P}(X_{n,k} = l) - e^{-\mu_{n,k}} \frac{(\mu_{n,k})^l}{l!} \Big| = O\Big(\frac{1}{k}\Big), \tag{3}$$

and for the random recursive tree,

$$d_{TV}(\mathcal{L}(\hat{X}_{n,k}), \mathrm{Po}(\hat{\mu}_{n,k})) = \frac{1}{2} \sum_{l \geq 0} \Big| \mathbb{P}(\hat{X}_{n,k} = l) - e^{-\hat{\mu}_{n,k}} \frac{(\hat{\mu}_{n,k})^l}{l!} \Big| = O\Big(\frac{1}{k}\Big). \tag{4}$$

Consequently, if $n \to \infty$ and $k \to \infty$, then

$$d_{TV}(\mathcal{L}(X_{n,k}), \mathrm{Po}(\mu_{n,k})) \to 0 \qquad \textit{and} \qquad d_{TV}(\mathcal{L}(\hat{X}_{n,k}), \mathrm{Po}(\hat{\mu}_{n,k})) \to 0.$$

Theorem 1.3 *Let $X_{n,k}$ be the number of fringe subtrees of size k in the binary search tree $\mathcal{T}_n$ and similarly let $\hat{X}_{n,k}$ be the number of fringe subtrees of size k in the random recursive tree Λ_n. Let $k = k_n = o(\sqrt{n})$. Then, as $n \to \infty$, for the binary search tree*

$$\frac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathrm{Var}(X_{n,k})}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \tag{5}$$

and, similarly, for the random recursive tree

$$\frac{\hat{X}_{n,k} - \mathbb{E}(\hat{X}_{n,k})}{\sqrt{\mathrm{Var}(\hat{X}_{n,k})}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{6}$$

Remark 1.4 *If $k/\sqrt{n} \to \infty$, then $\mu_{n,k}, \hat{\mu}_{n,k} \to 0$, and the convergence result in Theorem 1.2 reduces to the trivial $X_{n,k} \xrightarrow{\mathrm{p}} 0$ and $\hat{X}_{n,k} \xrightarrow{\mathrm{p}} 0$; the rate of convergence in* (3)–(4) *is still of interest.*

If $k/\sqrt{n} \to c \in (0,\infty)$, then $\mu_{n,k} \to 2c^{-2}$ and $\hat{\mu}_{n,k} \to c^{-2}$; and we obtain the Poisson distribution limits $X_{n,k} \xrightarrow{\mathrm{d}} \mathrm{Po}(2c^{-2})$ and $\hat{X}_{n,k} \xrightarrow{\mathrm{d}} \mathrm{Po}(c^{-2})$ [6, 7].

We also consider the number of fringe subtrees that are equal to a fixed tree T in the binary search tree $\mathcal{T}_n$, which we denote by X_n^T. Combining the Cramér–Wold device [3, Theorem 7.7] and Stein's method we show that random vectors of fringe subtrees are multivariate normally distributed. These results are also useful for proving general theorems for sums of functions of fringe subtrees.

Theorem 1.5 *Let T be a binary tree of size k and let T' be a binary tree of size m where $m \leq k$. Let X_n^T be the number of fringe subtrees T and let $X_n^{T'}$ be the number of fringe subtrees T' in the binary search tree $\mathcal{T}_n$ with n nodes. 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 fringe subtrees of T that are copies of T'. If $n > k+m+1$, then the covariance between X_n^T and $X_n^{T'}$ is equal to*

$$\mathrm{Cov}(X_n^T, X_n^{T'}) = (n+1)\sigma_{T,T'}, \tag{7}$$

where,

$$\sigma_{T,T'} := \frac{2}{(k+1)(k+2)} q^T_{T'} p_{k,T} - \gamma(k,m) p_{k,T} p_{m,T'}, \tag{8}$$

with

$$\gamma(k,m) := \frac{4(k+m+3)}{(k+1)(k+2)(m+1)(m+2)} - \frac{4(k^2+3km+m^2+4k+4m+3)}{(k+1)\,(m+1)\,(k+m+1)\,(k+m+2)\,(k+m+3)}\,. \tag{9}$$

Theorem 1.6 *Let X_n^T be the number of fringe subtrees T in the random binary search tree $\mathcal{T}_n$. Let $T^1, \dots, T^d$ be a fixed sequence of distinct binary trees and let $\bar{\mathcal{X}}_n^d = (X_n^{T^1}, X_n^{T^2}, \dots, X_n^{T^d})$. Let*

$$\mu_n^d := \Big(\mathbb{E}(X_n^{T^1}), \mathbb{E}(X_n^{T^2}), \dots, \mathbb{E}(X_n^{T^d})\Big)$$

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

$$\gamma_{ij} = \lim_{n\to\infty} \frac{1}{n} \mathrm{Cov}(X_n^{T^i}, X_n^{T^j}) = \sigma_{T^i,T^j}, \tag{10}$$

with notation as in (7)–(8)*. Then Γ is non-singular and*

$$n^{-1/2}(\bar{\mathcal{X}}_n^d - \mu_n^d) \xrightarrow{\mathrm{d}} \mathcal{N}(0,\Gamma). \tag{11}$$

In [9] there are corresponding multivariate-normal distribution results for the number of fringe subtrees Λ in the random recursive tree Λ_n.

In Section 5 we explain how Theorem 1.6 can be used to prove that the number of the so called *protected* nodes in the binary search trees asymptotically has a normal distribution.

2 Representations using uniform random variables

2.1 Devroye's representation for the binary search tree

We use the representation of the binary search tree $\mathcal{T}_n$ by Devroye [4, 5] 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 $1 \le k \le n$ and $1 \le i \le n-k+1$,

$$\sigma(i,k) = \{(i, U_i), \dots, (i+k-1, U_{i+k-1})\}, \tag{12}$$

i.e., the sequence of k labels (j, U_j) starting with $j = i$. For every node $u \in \mathcal{T}_n$, the fringe subtree $\mathcal{T}_n(u)$ 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 fringe subtree*. We define the indicator variable

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

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

$$I_{i,k} = \mathbf{1}\big\{U_{i-1} \text{ and } U_{i+k} \text{ are the two smallest among } U_{i-1}, \dots, U_{i+k}\big\}. \tag{13}$$

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

$$I_{1,k} = \mathbf{1}\big\{U_{k+1} \text{ is the smallest among } U_1, \dots, U_{k+1}\big\}, \tag{14}$$

$$I_{n-k+1,k} = \mathbf{1}\big\{U_{n-k} \text{ is the smallest among } U_{n-k}, \dots, U_n\big\}. \tag{15}$$

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

$$X_n := \sum_{u \in \mathcal{T}_n} f(\mathcal{T}_n(u)), \tag{16}$$

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

For example, one obtains the number of fringe subtrees that are equal (up to labelling) to a given binary tree T' by choosing $f(T) = \mathbf{1}\{T \approx T'\}$ (where $\approx$ denotes equality when we ignore labels), and one obtains the number of fringe subtrees with exactly k nodes by letting $f(T) = \mathbf{1}\{|T| = k\}$. We refer to Devroye [5] for several other examples showing the generality of this representation.

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, $1 \le j \le k$, 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 fringe subtree $\mathcal{T}_n(u)$ of $\mathcal{T}_n$, then, ignoring labels, $\mathcal{T}_n(u)$ is the binary search tree defined by the permutation defined by $\sigma(i,k)$, and thus $f\big(\mathcal{T}_n(u)\big) = f\big(\sigma(i,k)\big)$. Consequently, see [5],

$$X_n := \sum_{u \in \mathcal{T}_n} f(\mathcal{T}_n(u)) = \sum_{k=1}^{n} \sum_{i=1}^{n-k+1} I_{i,k} f(\sigma(i,k)). \tag{17}$$

2.2 The random recursive tree

Consider now instead the random recursive tree Λ_n. Let $f(T)$ be a function from the set of ordered rooted trees to $\mathbb{R}$. In analogy with (16), we define

$$Y_n := \sum_{u\in\Lambda_n} f(\Lambda_n(u)), \tag{18}$$

summing over all fringe subtrees of Λ_n.

As said in the introduction, the natural correspondence yields a coupling between the random recursive tree Λ_n and the binary search tree $\mathcal{T}_{n-1}$. The subtrees in Λ_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 Λ_n. Thus, as noted by [4], 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 Λ_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 $U_{i-1} < U_{i+k-1}$. We define

$$I^{\mathrm{L}}_{i,k-1} := \mathbf{1}\{\sigma(i,k-1) \text{ is a left-rooted subtree in } \mathcal{T}_{n-1}\}. \tag{19}$$

Using (13) it follows that, in analogy with (17),

$$Y_n := \sum_{u\in\Lambda_n} f(\Lambda_n(u)) = \sum_{k=1}^{n}\sum_{i=1}^{n-k+1} I^{\mathrm{L}}_{i,k-1}\bar{f}(\sigma(i,k-1)). \tag{20}$$

This is easily extended to $k=1$ too.

2.3 Cyclic representations

The representation (17) 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 (14)–(15). 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

$$U_i := U_{i \bmod (n+1)}, \qquad i\in\mathbb{Z}, \tag{21}$$

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 (13), but now for all i and k. Similarly, we define $\sigma(i,k)$ by (12) 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* (21). *Then, with notations as above,*

$$X_n := \sum_{u \in \mathcal{T}_n} f(\mathcal{T}_n(u)) \overset{\mathrm{d}}{=} \tilde{X}_n := \sum_{k=1}^{n} \sum_{i=1}^{n+1} I_{i,k} f(\sigma(i,k)). \tag{22}$$

Proof: The double sum in (22) 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,

$$\tilde{X}_n \overset{\mathrm{d}}{=} \bigl(\tilde{X}_n \mid U_0 = \min_i U_i\bigr). \tag{23}$$

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 (13) that $I_{i,k} = 0$ if $i \le n+1 \le i+k-1$; hence the terms in (22) with $n-k+1 < i \le n+1$ vanish. Note also that in the remaining terms, $f(\sigma(i,k))$ does not depend on U_0. Consequently,

$$\tilde{X}_n \overset{\mathrm{d}}{=} \Bigl(\sum_{k=1}^{n} \sum_{i=1}^{n-k+1} I_{i,k} f(\sigma(i,k)) \Bigm| U_0 = 0\Bigr) = X_n, \tag{24}$$

by (17), showing that the cyclic and linear representations in (17) and (22) are equivalent. □

Since U_i is defined for all $i \in \mathbb{Z}$ and has period $n+1$, it is 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

$$|i-j|_{n+1} := \min_{\ell \in \mathbb{Z}} |i - j - \ell \cdot (n+1)|. \tag{25}$$

For the random recursive tree Λ_n we argue in the same way, now using (20); for further details see [9].

The cyclic representations lead to simple exact calculations of means and variances. In [9] we use the cyclic representation to show general results for sums of functions of fringe subtrees.

3 Poisson approximations by Stein's method and couplings

To prove Theorems 1.2 we use Stein's method with couplings as described by Barbour et al. [2].

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 $\mathrm{Po}(\lambda)$, this method uses a coupling for each $\alpha \in \mathcal{A}$ between W and a random variable W_α which is defined on the same probability space as W and has the property

$$\mathcal{L}(W_\alpha) = \mathcal{L}(W - I_\alpha \mid I_\alpha = 1). \tag{26}$$

A common way to construct such a coupling (W, W_α) 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 $\beta \in \mathcal{A}$,

$$\mathcal{L}(J_{\beta\alpha}) = \mathcal{L}(I_\beta \mid I_\alpha = 1). \tag{27}$$

Then $W_\alpha = \sum_{\beta\neq\alpha} J_{\beta\alpha}$ is defined on the same probability space as W and (26) holds.

Suppose that $J_{\beta\alpha}$ are such random variables, and that, for each α, 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

$$J_{\beta\alpha} \le I_\beta \quad \text{if } \beta \in \mathcal{A}_\alpha^-, \tag{28}$$

with no condition if $\beta \in \mathcal{A}_\alpha^0$. We will use the following result from [2] (with a slightly simplified constant). ([2] also contain similar results using a third part $\mathcal{A}_\alpha^+$ of $\mathcal{A}_\alpha$, where (28) 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 3.1 ([2, 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*

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}(\lambda)) \le (1\wedge\lambda^{-1})\Bigl(\lambda - \mathrm{Var}(W) + 2\sum_{\alpha\in\mathcal{A}}\sum_{\beta\in\mathcal{A}_\alpha^0}\mathbb{E}(I_\alpha I_\beta)\Bigr).$$

□

Couplings for proving Theorem 1.2

Returning to the binary search tree, we use the cyclic representation in Section 2.3 to prove the following lemma where we give exact expressions for the expected value and the variance of $X_{n,k}$.

Lemma 3.2 *Let $1 \le k < n$. For the random binary search tree $\mathcal{T}_n$,*

$$\mathbb{E}(X_{n,k}) = \frac{2(n+1)}{(k+1)(k+2)} \tag{29}$$

and

$$\mathrm{Var}(X_{n,k}) = \begin{cases} \mathbb{E}\, X_{n,k} - (n+1)\frac{22k^2+44k+12}{(k+1)(k+2)^2(2k+1)(2k+3)}, & k < \frac{n-1}{2}, \\ \mathbb{E}\, X_{n,k} + \frac{2}{n} - \frac{64}{(n+3)^2}, & k = \frac{n-1}{2}, \\ \mathbb{E}\, X_{n,k} - (\mathbb{E}\, X_{n,k})^2 = \mathbb{E}\, X_{n,k} - \frac{4(n+1)^2}{(k+1)^2(k+2)^2}, & k > \frac{n-1}{2}. \end{cases} \tag{30}$$

Hence,

$$\mathrm{Var}(X_{n,k}) = \mathbb{E}(X_{n,k}) + O\Bigl(\frac{n}{k^3}\Bigr), \tag{31}$$

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

$$\mathrm{Var}(X_{n,k}) = \mathbb{E}(X_{n,k}) + \frac{2}{n} + O\Bigl(\frac{n}{k^3}\Bigr) = \mathbb{E}(X_{n,k}) + O\Bigl(\frac{1}{n}\Bigr). \tag{32}$$

To prove this lemma, we use the cyclic representation (22), which in this case is

$$X_{n,k} \overset{\mathrm{d}}{=} \sum_{i=1}^{n+1} I_{i,k}, \tag{33}$$

where now $I_{i,k}$ are defined by (13) with U_i given by (21). By (13) and symmetry, for any i and $1 \leq k < n$, $\mathbb{E}(I_{i,k}) = \frac{2}{(k+2)(k+1)}$, and thus (29) follows directly from (33). Using the cyclic representation we can also simply prove that (30) holds; however for Poisson approximation we only need the weak asymptotics in (31)–(32).

Recall the construction of $I_{i,k}$ in (13) and the distance $|i-j|_{n+1}$ on $\mathbb{Z}_{n+1}$ given by (25).

Lemma 3.3 *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 $((I_{j,k})_j,(Z^k_{ji})_j)$ such that $\mathcal{L}(Z^k_{ji}) = \mathcal{L}(I_{j,k} \mid I_{i,k}=1)$ jointly for all $j \in 1,\dots,n+1$. Furthermore,*

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

Proof: We define Z^k_{ji} 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. (13),

$$Z^k_{ji} = \mathbf{1}\big\{U'_{j-1} \text{ and } U'_{j+k} \text{ are the two smallest among } U'_{j-1},\dots,U'_{j+k}\big\}. \tag{34}$$

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

Note that $U'_l = U_l$ if $l \notin \{i-1,\dots,i+k\}$ and thus $Z^k_{ji} = I_{j,k}$ if $|j-i|_{n+1} > k+1$. On the other hand, if $0 < j-i < k+1$, then $Z^k_{ji} = 0$ 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 subtrees 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 $Z^k_{ji} \geq I_{j,k}$. 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. □

See Figures 1–2 that illustrate an example for such a coupling in the case $k=3$.

Proof of Theorem 1.2:

We prove the result for the binary search tree, using the representation $X_{n,k} = \sum_{i=1}^{n+1} I_{i,k}$ in (33). (For the random recursive tree the proof is similar and uses the representation $\hat{X}_{n,k} \overset{\mathrm{d}}{=} \sum_{i=1}^{n} I^{\mathrm{L}}_{i,k-1}$; see [9].)

Let $\mathcal{A} := \{1,\dots,n+1\}$. From Lemma 3.3 we see that for each $i \in \mathcal{A}$ we can apply Theorem 3.1 with

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

this yields, using Lemma 3.2 and the fact that $\mathbb{E}(I_{i,k}I_{i+k+1,k}) = O\big(\frac{1}{k^3}\big)$ (which is shown easily),

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

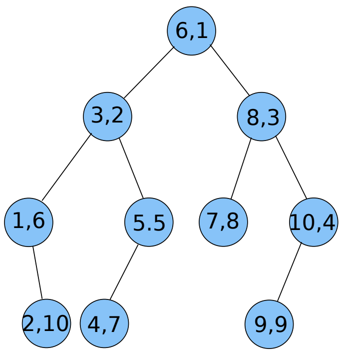

Fig. 1: A binary search tree with no fringe subtree of size three containing the keys $\{4,5,6\}$.

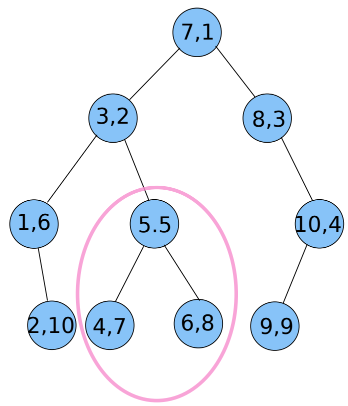

Fig. 2: A coupling forcing a fringe subtree of size three containing the keys $\{4,5,6\}$ in the tree in Fig. 1.

which shows (3).

□

4 Normal approximations by Stein's method

In this section we will prove Theorem 1.6. As in [5, Theorem 5] we will apply the following result, see [10, Theorem 6.33] for a proof, and for the definition of a dependency graph.

Lemma 4.1 *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 $\sum_{\alpha\in V_n} \mathbb{E}(|Z_{n\alpha}|) \le M_n$ and for every $\alpha, \alpha' \in V_n$,*

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

Let $\sigma_n^2 = \operatorname{Var}(S_n)$. If $\lim_{n\to\infty} \frac{M_n Q_n^2}{\sigma_n^3} = 0$, then it holds that $\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\operatorname{Var}(S_n)}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$.

Proof of Theorem 1.6: Recall that $\bar{\mathcal{X}}_n^d = (X_n^{T^1}, X_n^{T^2}, \dots, X_n^{T^d})$ 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 Γ is the matrix with elements $\gamma_{ij} = \lim_{n\to\infty} \frac{1}{n} \operatorname{Cov}(X_n^{T^i}, X_n^{T^j})$, see (10) above.

By the Cramér–Wold device [3, Theorem 7.7], to show that $n^{-\frac{1}{2}}(\bar{\mathcal{X}}_n^d - \mu_n^d)$ 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

$$\frac{\sum_{j=1}^d t_j X_n^{T^j} - \mathbb{E}\left(\sum_{j=1}^d t_j X_n^{T^j}\right)}{\sqrt{n}} \xrightarrow{\mathrm{d}} \sum_{j=1}^d t_j Z_j, \tag{35}$$

where $\sum_{j=1}^{d} t_j Z_j \sim \mathcal{N}(0, \gamma^2)$ with $\gamma^2 := \sum_{j,k=1}^{d} t_j t_k \gamma_{ij}$.

Let $S_n := \sum_{j=1}^{d} t_j X_n^{T^j}$. Theorem 1.5 implies that, as $n \to \infty$,

$$\operatorname{Var}(S_n) \sim n \sum_{j,k=1}^{d} t_j t_k \sigma_{T^i,T^j} = n \sum_{j,k=1}^{d} t_j t_k \gamma_{ij} = n\gamma^2. \tag{36}$$

In particular, if $\gamma^2 = 0$, then (35) is trivial, with the limit 0.

To show that (35) holds when $\gamma^2 > 0$, we will use the same method as was used in [5, Theorem 5] for proving this theorem (in a more general form) in the 1-dimensional case $d = 1$. Let $I_{i,k}^{T} = \mathbf{1}\{\sigma(i,k) \approx T\}$. Let $|T^j| = k_j$, $1 \le j \le d$. We use the cyclic representation (22), which in this case can be written as $X_n^{T^j} = \sum_{i=1}^{n+1} I_i^j$, 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

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

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 4.1 as $M_n = (n+1)dM$ and

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

Since $\sigma_n \sim n^{1/2}$ by (36), it holds that $\lim_{n\to\infty} \frac{M_n Q_n^2}{\sigma_n^3} = 0$, and Lemma 4.1 shows that (35) holds.

It is shown in [9] that the matrix Γ is non-singular. □

The proof of Theorem 1.3 when $k = o(\sqrt{n})$ and k tends to infinity follows directly from Theorem 1.2 and (31), since then $\mathbb{E}(X_{n,k})$ and $\operatorname{Var}(X_{n,k})$ tend to infinity as n tends to infinity. For the case $k = O(1)$ we can apply Lemma 4.1 similarly to the proof of Theorem 1.6 and repeat the arguments used in [5, Theorem 5]; see [9] for details.

5 Protected nodes

We consider the number of protected nodes. A node is *protected* if the shortest distance to a leaf is at least two, i.e., it is neither a leaf or the parent of a leaf. The following theorem was shown by Mahmoud and Ward [13, Theorem 3.1] using generating functions.

Theorem 5.1 *Let X_n denote the number of protected nodes in a binary search tree $\mathcal{T}_n$. Then it holds that*

$$\frac{X_n - \frac{11}{30}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\Big(0, \frac{29}{225}\Big).$$

We can show this result from a simple application of Theorem 1.6. Using the formulation of fringe subtrees, we note that the number of unprotected nodes in the binary search tree equals twice 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). Hence, since any linear combination of the components in a random vector with a multivariate normal distribution is normal, Theorem 5.1 follows from Theorem 1.6.

Moreover, our approach using fringe subtrees also allows us to provide a simple proof of the following result which was conjectured in [13, Conjecture 2.1]. See [9] for detailed proofs of Theorems 5.1–5.2.

Theorem 5.2 *Let X_n denote the number of protected nodes in a binary search tree $\mathcal{T}_n$. 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 $\mathbb{E}(X_n^k) = p_k(n)$ for all $n \geq 4k$.*

References

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

[2] Barbour, A. D., Holst L. and Janson S., *Poisson Approximation.* Oxford University Press, New York, 1992.

[3] Billingsley P., *Convergence of Probability Measures*. John Wiley and Sons, New York, 1968.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Holmgren C. and Janson S., Limit Laws for Functions of Fringe subtrees for Binary Search Trees and Recursive Trees. *In preparation.*

[10] Janson S., Łuczak T. and Ruciński A., *Random Graphs*. John Wiley, New York, 2000.

[11] Knuth D.E., *The Art of Computer Programming, Vol. 3 : Sorting and Searching* Addison-Wesley Publishing Co., Reading, Mass., 1973.

[12] Knuth D.E., *The Art of Computer Programming. Volume 1: Fundamental Algorithms.* Third ed edition. Addison-Wesley Publishing Co., Reading, Mass., 1997.

[13] 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.