

ASYMPTOTIC RESULTS ON HOPPE TREES AND THEIR VARIATIONS

ELLA HIESMAYR,* *University of California, Berkeley*
ÜMIT IŞLAK,** *Boğaziçi University*

Abstract

A uniform recursive tree on n vertices is a random tree where each possible $(n - 1)!$ labelled recursive rooted tree is selected with equal probability. We introduce and study weighted trees, a non-uniform recursive tree model departing from the recently introduced Hoppe trees. This class generalizes both uniform recursive trees and Hoppe trees, providing diversity among the nodes and making the model more flexible for applications. We analyse the number of leaves, the height, the depth, the number of branches, and the size of the largest branch in these weighted trees.

Keywords: Uniform recursive trees; Hoppe trees; Ewens sampling formula; random permutations; coupling; random tree statistics

2010 Mathematics Subject Classification: Primary 05C80
Secondary 60C05

1. Introduction

A uniform recursive tree (URT) on n vertices is a rooted random recursive tree where each possible $(n - 1)!$ distinct tree has the same probability of appearing. Another way of looking at uniform recursive trees is that one starts constructing the tree with only the root (node 1) and node 2 attached to the root. At each subsequent step, $k = 3, \dots, n$, node k connects to one of the prior nodes with equal probability $1/(k - 1)$. A detailed survey on URTs can be found in [11], and a book length treatment of random trees can be found in [5].

URT's have been used in various applications. These include, but are not restricted to, the spread of epidemics [6], determining the genealogy of ancient and medieval texts [13], analysing pyramid schemes [8], and the spread of a fire in a tree [12]. Though these investigations use uniformity in their models, having different distributions would provide more flexibility to the researcher. Using URTs implies that in the model all nodes are assumed to be identical in a certain sense. In the spread of epidemics, this means, for instance, that every infected person is equally likely to infect the next one. Similarly, it means that in the study of medieval texts every book is equally likely to be copied. This is obviously not the case in real-world applications.

Parallel to the development of the theory of uniform recursive trees, various other recursive tree structures have already been studied. Among the most well known are binary recursive trees, which are described in [3]. Other non-uniform recursive tree models include plane-oriented recursive trees [16], scaled attachment random recursive trees [4], and biased recursive trees [1].

Received 15 December 2017; revision received 9 December 2017.

* Postal address: University of California, Berkeley, U.S. Email: ella.hiesmayr@berkeley.edu

** Postal address: Boğaziçi University, Istanbul, Turkey. Email: umit.islak1@boun.edu.tr

Our interest here is on another natural generalization, Hoppe trees, as recently considered in [10]. There, the root is assigned a weight θ ; all other nodes get weight 1. Node i then attaches to the root with probability $\frac{\theta}{\theta+i-2}$ and to any other node with probability $\frac{1}{\theta+i-2}$. This model is associated with Hoppe's urn, which has an application in modelling the alleles of a gene with mutation rate $\theta > 0$. Concerning many properties like the number of leaves, the height, and the depth of node n , Hoppe trees behave similarly to uniform recursive trees in an asymptotic sense.

The model in this paper generalizes the idea of Hoppe trees: we assign every node a weight w_i . Node j then attaches to node $1 \leq i < j$ with probability $\frac{w_i}{w_1 + \dots + w_{j-1}}$. We call the resulting tree construction a weighted recursive tree (WRT).

Using weights is interesting from the point of view of applications since it allows one to introduce diversity among the nodes. In the other non-uniform models discussed above, all nodes have the same behaviour, or in other words they attract nodes according to the same rule. When a recursive process does not satisfy such conditions, weighted recursive trees can be used to model it more precisely. Moreover, the properties of weighted recursive trees and how much they differ from the uniform model can be interpreted as an indicator for the stability of a process. In various applications it is reasonable to assume that it is more probable for some nodes to get children than others. For example, some persons might be more likely to infect others, some copies of ancient texts are more likely to have been copied again, and some people might be more likely to recruit new individuals. Thus it is interesting to see how much fluctuation in the attachment probabilities can be tolerated when using the uniform model.

The rest of the paper is organized as follows. The next section introduces a coupling used to construct a specific WRT from a Hoppe tree. Section 3 applies the coupling construction for an analysis of the number of leaves and the height of the resulting random tree. In Section 4 we relate Hoppe trees to Hoppe permutations, and use the coupling construction of Section 2 to study the size of the largest branch. In Sections 5 and 6 we study the depth and the number of branches of a general WRT.

2. A useful coupling construction

2.1. Constructing WRTs from URTs

We will first introduce a coupling allowing us to construct a special kind of WRT from a URT. We will not use this coupling in the analysis of random tree statistics because the second coupling construction introduced below can be applied to a more general class of WRTs. We still wanted to discuss this version because it has URTs rather than Hoppe trees as a starting point, and URTs are a much better studied structure.

We will call trees that have a weight sequence such that the first k nodes have a constant weight equal to θ and all other nodes have weight 1, θ^k -RTs. It is possible to construct a θ^k -RT from a URT by a coupling construction when $\theta \in \mathbb{N}^+$. To emphasize this assumption we will use m instead of θ in this part. In particular, we can construct Hoppe trees for which the weight of the root is a natural number by this coupling. To avoid confusion let us denote the nodes in the URT by i and the nodes in the m^k -RT obtained by the coupling by i^* .

The coupling construction in that case goes as follows: First construct a URT with $mk + n - k$ nodes. We write \mathcal{T}_{mk+n-k} for this URT and \mathcal{T}_n^{mk} for the tree that we obtain at the end of the coupling. Since we want the weight of the first k nodes to be m , we need to join several

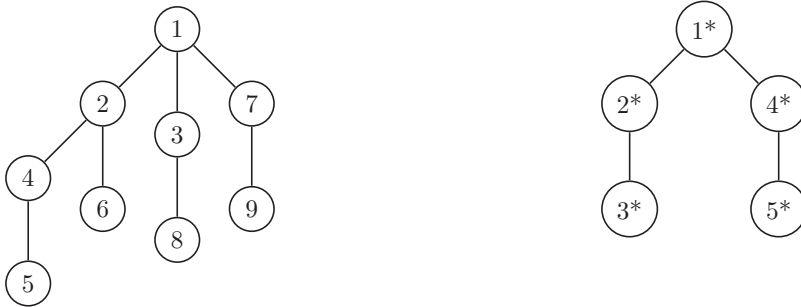


FIGURE 1: A WRT $\mathcal{T}_5^{2^4}$ is constructed from a URT \mathcal{T}_5 .

nodes into one. More precisely, we combine m nodes of the URT to get each of the first k nodes of the m^k -RT in the following way:

- Nodes $1, \dots, m$ become node 1^* ;
- $m + 1, \dots, 2m$ become node 2^* ; \dots
- $(k - 1)m + 1, \dots, km$ become node k^* .

All nodes after km in \mathcal{T}_{mk+n-k} correspond only to a single node in $\mathcal{T}_n^{m^k}$; we just need to ‘translate’ the names of the nodes to take into account that we used mk nodes instead of k for the first k nodes in the reconstructed tree. More precisely, for $j > km$, node j in the URT becomes node $(j - k(m - 1))^*$ in the WRT.

The parent–child relations in the WRT follow from those in the URT as follows: If, for $j > km$ and $i \leq k$, j is a child of one of the nodes $(i - 1)m + 1, \dots, im$ in the URT, $(j - k(m - 1))^*$ becomes a child of i^* in the WRT. If, for $j > km$, the parent of node j is h with $h > km$, the parent of node $(j - k(m - 1))^*$ is node $(h - k(m - 1))^*$.

For the nodes before or equal to km , we need to be more careful, since we joined several nodes into one and the nodes $(j - 1)m + 1, \dots, jm$ might have different parents in the URT. To settle this ambiguity, for $1 < j \leq k$ we set the parent of j^* as the parent of $(j - 1)m + 1$, i.e. of the node with the smallest label among those that become j^* . Thus, if in the URT the parent of $(j - 1)m + 1$ is any of the nodes $(i - 1)m + 1, \dots, im$, the parent of j^* is i^* . It can be easily verified that the attachment probabilities obtained correspond to those of an m^k -tree; for details, see [9].

In Figure 1 we present a small example of the coupling construction. In the example, nodes 1 and 2 combine to form node 1^* , and nodes 3 and 4 become node 2^* . Then nodes 5 and 6 become node 3^* , and 3^* is a child of 2^* , since node 5 is a child of node 4 in the URT. Nodes 7 and 8 become node 4^* , and 4^* is a child of 1^* since node 7 is a child of node 1 in the URT. Finally, node 9 becomes node 5^* , and 5^* is a child of 4^* since node 9 is a child of node 7 in the URT.

We now show, as an example, how the coupling can be used to study the number of leaves, i.e. the number of nodes without children, of an m^k -RT. Let \mathcal{L}_{mk+n-k} denote the number of leaves of \mathcal{T}_{mk+n-k} , and $\mathcal{L}_n^{m^k}$ denote the number of leaves of $\mathcal{T}_n^{m^k}$. Then \mathcal{L}_{mk+n-k} can be used to bound $\mathcal{L}_n^{m^k}$. First, if node $i > km$ is a leaf in \mathcal{T}_{mk+n-k} , the corresponding node in $\mathcal{T}_n^{m^k}$, which

is $(i - k(m - 1))^*$, is also a leaf. The reconstruction process thus only affects the children of the nodes i^* with $1 \leq i \leq k$, and the root cannot be a leaf, so we can have at most $k - 1$ additional leaves. Moreover, for each $1 \leq i \leq k$ we can ‘lose’ at most $m - 1$ leaves since if all of $(i - 1)m + 1, \dots, im$ are leaves in \mathcal{T}_{mk+n-k} , i^* will be a leaf in \mathcal{T}_n^{mk} too. Hence, we can conclude that

$$\mathcal{L}_{mk+n-k} - k(m - 1) < \mathcal{L}_n^{mk} < \mathcal{L}_{mk+n-k} + k - 1.$$

Together with results about the expected number of leaves of URTs, this implies, after some simple manipulations, that

$$\left| \mathbb{E}[\mathcal{L}_n^{mk}] - \mathbb{E}[\mathcal{L}_n] \right| \leq \frac{k(m + 1)}{2},$$

where \mathcal{L}_n denotes the number of leaves in a URT on n nodes. It is possible to derive other results from this coupling, but since the coupling we present below is more general, we will not go further into it here.

2.2. Constructing WRTs from Hoppe trees

We will now introduce a coupling construction for WRTs whose nodes have constant weight after a certain index, a class similar to, but more general, than θ^k -RTs. Let \mathcal{T}_n^w be a WRT, and $(w_i)_{i \in \mathbb{N}}$, the weight sequence of \mathcal{T}_n^w , be such that there is a $k \in \mathbb{N}$ such that $w_i = 1$ for all $i > k$. Then we can construct \mathcal{T}_n^w from a Hoppe tree with root weight $\theta = \sum_{i=1}^k w_i$ by a coupling construction, more precisely by splitting the root into k nodes. To avoid confusion we will write i for node i in the Hoppe tree and i^* for node i in the reconstructed tree.

We now describe the coupling construction: First construct a Hoppe tree on $n - k + 1$ nodes and with θ , the weight of the root, equal to $\sum_{i=1}^k w_i$. Then construct a WRT of size k corresponding to $(w_i)_{i \in \{1, \dots, k\}}$. Now we replace the root of the Hoppe tree by this weighted recursive tree of size k in the following way: nodes $1^*, 2^*, \dots, k^*$ are the nodes of the WRT of size k we just constructed. For $i \geq 2$, node i in the Hoppe tree becomes node $(i + k - 1)^*$ in the reconstructed tree, so we shift the names of the rest of the nodes by $k - 1$. Then, for all $i \geq 2$, if i is a child of node 1 in the Hoppe tree, $(i + k - 1)^*$ becomes a child of one of the nodes $1^*, \dots, k^*$ in the reconstructed tree, proportionally to their weights. This means that if i is a child of node 1 in the Hoppe tree, for $1 \leq j^* \leq k$, node $(i + k - 1)^*$ will become a child of a node j^* in the reconstructed tree with probability $\frac{w_j}{\sum_{\ell=1}^k w_\ell}$.

Let us check that this gives the attachment probabilities corresponding to the WRT we aim to construct.

For $1 \leq i^* \leq k < j^*$,

$$\begin{aligned} \mathbb{P}(j^* \text{ is child of } i^*) &= \mathbb{P}(j - k + 1 \text{ is child of } 1, i^* \text{ is chosen as the parent of } j^*) \\ &= \frac{\sum_{\ell=1}^k w_\ell}{j - k + 1 - 2 + \sum_{\ell=1}^k w_\ell} \frac{w_i}{\sum_{\ell=1}^k w_\ell} = \frac{w_i}{j - 1 - k + \sum_{\ell=1}^k w_\ell}. \end{aligned}$$

For $k < i^* < j^*$,

$$\begin{aligned} \mathbb{P}(j^* \text{ is child of } i^*) &= \mathbb{P}(j - k + 1 \text{ is child of } i - k + 1) \\ &= \frac{1}{j - k + 1 - 2 + \sum_{\ell=1}^k w_\ell} = \frac{1}{j - 1 - k + \sum_{\ell=1}^k w_\ell}. \end{aligned}$$

The following two sections will be using the coupling construction just described.

3. Use of the coupling in WRT statistics

In this section we apply the coupling construction of the previous section to study the number of leaves in a WRT and the height of a Hoppe tree. Later, in Section 4, the coupling construction will also be used in order to understand the size of the largest branch in a WRT.

3.1. Number of leaves

A node in a tree with degree one is said to be a leaf. Focusing on the number of leaves, the reconstruction process in our coupling construction implies that all the leaves of the Hoppe tree are still leaves in the reconstructed tree, since we do not change any relation among the nodes $2, \dots, n - k + 1$ of the Hoppe tree or, respectively, nodes $(k + 1)^*, \dots, n^*$ of the reconstructed tree. There can be at most $k - 1$ additional leaves among the first k nodes. Thus, we can bound the number of leaves \mathcal{L}_n^w of \mathcal{T}^w by the number of leaves \mathcal{L}_n^θ of \mathcal{T}^θ :

$$\mathcal{L}_{n-k+1}^\theta \leq \mathcal{L}_n^w \leq \mathcal{L}_{n-k+1}^\theta + k - 1. \tag{1}$$

In [10] the following results about the leaves of Hoppe trees are given. In this theorem and below, \mathcal{G} denotes a standard normal random variable.

Theorem 1. ([10].) *Let \mathcal{L}_n^θ denote the number of leaves of a Hoppe tree with $n \geq 2$ nodes. Then*

$$\begin{aligned} \mathbb{E}[\mathcal{L}_n^\theta] &= \frac{n}{2} + \frac{\theta - 1}{2} + \mathcal{O}\left(\frac{1}{n}\right), \\ \text{Var}(\mathcal{L}_n^\theta) &= \frac{n}{12} + \frac{\theta - 1}{12} + \mathcal{O}\left(\frac{1}{n}\right), \\ \mathbb{P}(|\mathcal{L}_n^\theta - \mathbb{E}[\mathcal{L}_n^\theta]| \geq t) &\leq 2 \exp\left\{-\frac{6t^2}{n + \theta + 1}\right\} \quad \text{for all } t > 0, \text{ and} \\ \frac{\mathcal{L}_n^\theta - \mathbb{E}[\mathcal{L}_n^\theta]}{\sqrt{\text{Var}(\mathcal{L}_n^\theta)}} &\longrightarrow_d \mathcal{G} \text{ as } n \rightarrow \infty. \end{aligned}$$

Using the above theorem we can derive results on the number of leaves in WRTs whose nodes have constant weight after a certain index.

Theorem 2. *Let \mathcal{L}_n^w denote the number of leaves of a WRT of size n with weight sequence $(w_i)_{i \in \mathbb{N}}$ such that there is a $k \in \mathbb{N}$ such that for all $i > k$ we have $w_i = 1$. Then*

$$\begin{aligned} \mathbb{E}[\mathcal{L}_n^w] &= \frac{n}{2} + C + \mathcal{O}\left(\frac{1}{n}\right) \text{ with } |C| \leq \frac{\sum_{i=1}^k w_i + k}{2}, \\ \text{Var}(\mathcal{L}_n^w) &= \frac{n}{12} + \mathcal{O}(\sqrt{n}), \\ \mathbb{P}(|\mathcal{L}_n^w - \mathbb{E}[\mathcal{L}_n^w]| \geq t) &\leq 2 \exp\left\{-\frac{6(t - 2k + 2)^2}{n - k + 2 + \sum_{i=1}^k w_i}\right\} \quad \text{for all } t > 2k - 2, \text{ and} \\ \frac{\mathcal{L}_n^w - \mathbb{E}[\mathcal{L}_n^w]}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} &\longrightarrow_d \mathcal{G} \text{ as } n \rightarrow \infty. \end{aligned}$$

Proof. First, for the expected value we get, from Theorem 1 and (1),

$$\begin{aligned} \mathcal{L}_{n-k+1}^\theta &\leq \mathcal{L}_n^w \leq \mathcal{L}_{n-k+1}^\theta + k - 1 \\ &\Rightarrow \frac{n - k + 1}{2} + \frac{\sum_{i=1}^k w_i - 1}{2} + \mathcal{O}\left(\frac{1}{n}\right) \leq \mathbb{E}[\mathcal{L}_n^w] \\ &\leq \frac{n - k + 1}{2} + \frac{\sum_{i=1}^k w_i - 1}{2} + k - 1 + \mathcal{O}\left(\frac{1}{n}\right) \\ &\Rightarrow \frac{n}{2} + \frac{\sum_{i=1}^k w_i - k}{2} + \mathcal{O}\left(\frac{1}{n}\right) \leq \mathbb{E}[\mathcal{L}_n^w] \leq \frac{n}{2} + \frac{\sum_{i=1}^k w_i + k - 2}{2} + \mathcal{O}\left(\frac{1}{n}\right) \\ &\Rightarrow \mathbb{E}[\mathcal{L}_n^w] = \frac{n}{2} + C + \mathcal{O}\left(\frac{1}{n}\right), \end{aligned}$$

where C depends on k and w_1, \dots, w_k , and we have $|C| \leq \frac{\sum_{i=1}^k w_i + k}{2}$.

Equation (1) also directly gives a concentration inequality. We have that

$$\begin{aligned} &\mathbb{P}(|\mathcal{L}_n^w - \mathbb{E}[\mathcal{L}_n^w]| \geq t) \\ &\leq \mathbb{P}(|\mathcal{L}_n^w - \mathcal{L}_{n-k+1}^\theta| + |\mathcal{L}_{n-k+1}^\theta - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]| + |\mathbb{E}[\mathcal{L}_{n-k+1}^\theta] - \mathbb{E}[\mathcal{L}_n^w]| \geq t) \\ &\leq \mathbb{P}(|\mathcal{L}_{n-k+1}^\theta - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]| \geq t - 2k + 2) \\ &\leq 2 \exp\left\{-\frac{6(t - 2k + 2)^2}{n - k + 2 + \sum_{i=1}^k w_i}\right\}. \end{aligned}$$

The coupling also gives us an approximation for the variance. Let Y denote the number of additional leaves among the first k nodes in the reconstructed tree. Then $\text{Var}(Y) = \mathcal{O}(k^2)$ since $Y \leq k$. Since $\text{Var}(\mathcal{L}_n^w) = \text{Var}(\mathcal{L}_{n-k+1}^\theta + Y) = \text{Var}(\mathcal{L}_{n-k+1}^\theta) + \text{Var}(Y) + 2\text{Cov}(\mathcal{L}_{n-k+1}^\theta, Y)$, we get, by the Cauchy–Schwarz inequality,

$$\text{Var}(\mathcal{L}_n^w) = \frac{n}{12} + \mathcal{O}(\sqrt{n}).$$

In a similar way, one can make conclusions about the asymptotic distribution. For this we will need Slutsky’s theorem: Let X_n and Y_n be sequences of random variables such that $X_n \rightarrow_d X$ and $Y_n \rightarrow c$ in probability for $c \in \mathbb{R}$. Then

$$\lim_{n \rightarrow \infty} X_n + Y_n =_d X + c.$$

In order to derive a central limit theorem for \mathcal{L}_n^w , we write

$$\begin{aligned} \frac{\mathcal{L}_n^w - \mathbb{E}[\mathcal{L}_n^w]}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} &= \frac{\mathcal{L}_n^w - \mathcal{L}_{n-k+1}^\theta}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} + \frac{\mathcal{L}_{n-k+1}^\theta - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} \\ &\quad + \frac{\mathbb{E}[\mathcal{L}_{n-k+1}^\theta] - \mathbb{E}[\mathcal{L}_n^w]}{\sqrt{\text{Var}(\mathcal{L}_n^w)}}. \end{aligned}$$

Now, by Theorem 1 and our previous result on $\text{Var}(\mathcal{L}_n^w)$ we have

$$\frac{\mathcal{L}_{n-k+1}^\theta - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]}{\sqrt{\text{Var}(\mathcal{L}_{n-k+1}^\theta)}} \sqrt{\frac{\text{Var}(\mathcal{L}_{n-k+1}^\theta)}{\text{Var}(\mathcal{L}_n^w)}} \rightarrow_d \mathcal{G} \quad \text{as } n \rightarrow \infty,$$

and since $\text{Var}(\mathcal{L}_n^w) \rightarrow \infty$ as $n \rightarrow \infty$ we have, by (1), that

$$\left| \frac{\mathcal{L}_n^w - \mathcal{L}_{n-k+1}^\theta}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} + \frac{\mathbb{E}[\mathcal{L}_{n-k+1}^\theta] - \mathbb{E}[\mathcal{L}_n^w]}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} \right| \leq \frac{2k}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} \rightarrow 0 \text{ almost surely.}$$

Now we can apply Slutsky's theorem and conclude that

$$\frac{\mathcal{L}_n^w - \mathbb{E}[\mathcal{L}_n^w]}{\sqrt{\text{Var}(\mathcal{L}_n^w)}} \rightarrow_d \mathcal{G} \text{ as } n \rightarrow \infty.$$

□

Remark 1. It might be possible to get results on the number of leaves of a general WRT \mathcal{T}_n^w by writing \mathcal{L}_n^w as the sum of $\mathbf{1}(\ell_i^w)$, where ℓ_i^w denotes the event that node i is a leaf in \mathcal{T}_n^w . It follows from the construction principle that

$$\mathbb{P}(\ell_i^w) = \prod_{j=i+1}^n \left(1 - \frac{w_j}{w_1 + \dots + w_{j-1}} \right).$$

After some manipulation this expression becomes

$$\mathbb{P}(\ell_i^w) = \frac{w_1 + \dots + w_{i-1}}{w_1 + \dots + w_{n-1}} \prod_{j=i+1}^{n-1} \left(1 + \frac{w_j - w_i}{w_1 + \dots + w_{j-1}} \right).$$

An exact expression for the expectation of the number of leaves of a θ^k -RT can be obtained by writing

$$\mathbb{E}[\mathcal{L}_n^{\theta^k}] = \sum_{i=2}^n \mathbb{E}[\mathbf{1}(\ell_i^{\theta^k})]$$

and using the expression above. After some computations we get

$$\mathbb{E}[\mathcal{L}_n^{\theta^k}] = \frac{n}{2} + \frac{k(\theta - 1)}{2} + \frac{k\theta(1 - k\theta)}{2(k(\theta - 1) + n - 1)} + \frac{k - 1}{2} \prod_{i=1}^{n-1-k} \frac{\theta(k - 1) + i}{\theta k + i}.$$

Remark 2. For θ^k -RTs it is also possible to obtain results about the expectation, variance, and concentration rate of the number of leaves by using a martingale argument similar to the one in [10]; for details, see [9].

3.2. Height

As a second example we discuss the height of a WRT, which is defined as the length of the longest path from the root to a leaf. Let \mathcal{H}_n^w denote the height of a WRT with weight sequence $(w_i)_{i \in \mathbb{N}}$ such that $w_i = 1$ for $i > k$. Let, moreover, $\mathcal{D}_{1,i}^w$ and $\mathcal{D}_{1,i}^\theta$ denote the distance between the root and node i in the reconstructed weighted tree and the original Hoppe tree respectively. For $i \leq k$, $\mathcal{D}_{1,i}^w$ is at most $k - 1$. Also, for any $i > k$, the path from the root to i^* corresponds to the path from the root to the corresponding node in the original Hoppe tree, i.e. the distance from the root to $i - k + 1$, except that we might have an additional path among the first k nodes instead of the first edge. Thus, $\mathcal{D}_{1,i}^w$ is at least as big as $\mathcal{D}_{i-k+1}^\theta$.

Also, $\mathcal{D}_{1,i}^w$ is at most $k - 1$ bigger than the distance between the root and the corresponding node in the original tree. Let $j - k + 1$ be the first node on the path from node 1 to $i - k + 1$ in

the original tree. Then, in the reconstructed tree j^* will be attached to some h^* , where $1 \leq h \leq k$. Thus, when we consider the path consisting of the path from the root to h^* in the reconstructed tree, the edge from h^* to j^* and the path from $j - k + 1$ to $i - k + 1$ in the Hoppe tree, we get a path from 1^* to i^* in the reconstructed tree. Thus, for all $k + 1 \leq i \leq n$, there is some $h \leq k$ such that

$$D_{1,i}^w = D_{1,h}^w + 1 + D_{j-k+1,i-k+1}^\theta = D_{1,h}^w + D_{1,i-k+1}^\theta.$$

Also, $D_{1,h}^w \leq k - 1$, so we have

$$D_{1,i-k+1}^\theta \leq D_{1,i}^w \leq D_{1,i-k+1}^\theta + k - 1,$$

which implies that

$$\max_{i=1,\dots,n-k+1} \{D_{1,i}^\theta\} \leq \max_{i=1,\dots,n} \{D_{1,i}^w\} \leq \max_{i=1,\dots,n-k+1} \{D_{1,i}^\theta\} + k - 1.$$

Thus, from the definition of the height as $\mathcal{H}_n = \max_{i=1,\dots,n} \{D_{1,i}\}$, we have

$$\mathcal{H}_{n-k+1}^\theta \leq \mathcal{H}_n^w \leq \mathcal{H}_{n-k+1}^\theta + k - 1. \tag{2}$$

We have the following result about the height of Hoppe trees.

Theorem 3. ([10].) *Let \mathcal{H}_n^θ denote the height of a Hoppe tree with n nodes. Then*

$$\mathbb{E}[\mathcal{H}_n^\theta] = e \ln(n) - \frac{3}{2} \ln \ln n + \mathcal{O}(1),$$

$$\text{Var}(\mathcal{H}_n^\theta) = \mathcal{O}(1).$$

Together with the coupling, this allows us to derive the following theorem.

Theorem 4. *Let \mathcal{H}_n^w denote the height of a WRT of size n with weight sequence $(w_i)_{i \in \mathbb{N}}$ such that there is a $k \in \mathbb{N}$ such that, for all $i > k$, we have $w_i = 1$. Then*

$$\mathbb{E}[\mathcal{H}_n^w] = e \ln(n) - \frac{3}{2} \ln \ln(n) + \mathcal{O}(1),$$

$$\text{Var}(\mathcal{H}_n^w) = \mathcal{O}(1).$$

Proof. According to Theorem 3 we have $\mathbb{E}[\mathcal{H}_{n-k+1}^\theta] = e \ln(n - k + 1) - \frac{3}{2} \ln \ln(n - k + 1) + \mathcal{O}(1)$. Note that

$$\begin{aligned} e \ln(n - k + 1) - \frac{3}{2} \ln \ln(n - k + 1) &= e \left(\ln(n) + \ln \left(1 - \frac{k-1}{n} \right) \right) \\ &\quad - \frac{3}{2} \left(\ln \ln(n) + \ln \left(1 + \frac{\ln \left(1 - \frac{k-1}{n} \right)}{\ln(n)} \right) \right) \\ &= e \ln(n) - \frac{3}{2} \ln \ln(n) + o(1). \end{aligned}$$

Thus, we get, by (2) for $n > k$,

$$e \ln(n) - \frac{3}{2} \ln \ln(n) + \mathcal{O}(1) \leq \mathbb{E}[\mathcal{H}_n^w] \leq e \ln(n) - \frac{3}{2} \ln \ln(n) + \mathcal{O}(1) + k - 1,$$

which implies

$$\mathbb{E} [\mathcal{H}_n^w] = e \ln(n) - \frac{3}{2} \ln \ln(n) + \mathcal{O}(1).$$

As before, we might define $Y := \mathcal{H}_n^w - \mathcal{H}_{n-k+1}^\theta$. Then, by (2), we have $Y < k$, so

$$\begin{aligned} \text{Var}(\mathcal{H}_n^w) &= \text{Var}(\mathcal{H}_{n-k+1}^\theta + Y) \\ &= \text{Var}(\mathcal{H}_{n-k+1}^\theta) + \text{Var}(Y) + 2\text{Cov}(\mathcal{H}_{n-k+1}^\theta, Y) \\ &= \mathcal{O}(1). \end{aligned}$$

□

4. Largest branch

4.1. Permutations view

In this section we focus on standard Hoppe trees and study the size of the largest branch in this tree model. The results will sharpen the corresponding observations of [6] for URTs. Before moving on to largest branches, we need to discuss constructions of Hoppe trees via random permutations, in particular permutations that are generated via the Ewens sampling formula.

For each n , the Ewens distribution gives a family of distributions over the vectors

$$C^{(n)} = (C_1^{(n)}, C_2^{(n)}, \dots, C_n^{(n)})$$

with $\sum_{i=1}^n iC_i^{(n)} = n$. In particular, for given $\theta > 0$, the Ewens distribution $\text{EW}(\theta)$ is defined with the probabilities

$$\mathbb{P}_\theta(C^{(n)} = (c_1, \dots, c_n)) = \mathbf{1} \left(\sum_{j=1}^n jc_j = n \right) \frac{n!}{\theta_{(n)}} \prod_{j=1}^n \binom{\theta}{j}^{c_j} \frac{1}{c_j!}, \quad c_1, \dots, c_n \in \mathbb{N},$$

where $\theta_{(n)} = \theta(\theta + 1) \cdots (\theta + n - 1)$.

Below we call a permutation whose cycle structure results from $\text{EW}(\theta)$ a Hoppe permutation. In this setting, c_i stands for the number of cycles in the permutation of size i . In order to relate the topic to Hoppe trees, we need to discuss a recursive construction of Hoppe permutations. The discussion here is similar to the one in [1]. For convenience, the permutation will be constructed on the label set $\{2, 3, \dots, n\}$. We first begin with the permutation (2) with only one cycle. Then 3 either joins the first cycle to the right of 2 with probability $\frac{1}{\theta+1}$, or starts the second cycle with probability $\frac{\theta}{\theta+1}$. Once we have constructed a permutation on $\{2, 3, \dots, k-1\}$, k either starts a new cycle with probability $\frac{\theta}{\theta+k-2}$, or is inserted to the right of a randomly chosen integer already assigned to a cycle. The resulting permutation then has the cycle structure from the distribution $\text{EW}(\theta)$ [2].

Next, we construct a Hoppe tree based on the permutation construction of the previous paragraph. Begin with node 1 as the root and node 2 attached to it. Then, if node 3 begins a new cycle in the corresponding Hoppe permutation, attach it to node 1, and otherwise attach it to node 2. Then, for node $k \geq 4$, if k starts a new cycle in the Hoppe permutation, attach node k to node 1, and otherwise attach it to node j where k was inserted to the right of j in the corresponding permutation.

It is then clear that this gives a bijection between Hoppe permutations and Hoppe trees. In particular, the cycles in a Hoppe permutation are in a one-to-one relation with the branches in the corresponding tree. This reduces the study of the size of the largest branch in a Hoppe tree to the study of the largest cycle in its permutation correspondence.

4.2. Size of the largest branch

For a given tree \mathcal{T} on n vertices, the number of branches is the number of children of the root, i.e. all nodes that are attached to the root. If a node i is attached to the root, then node i with its descendants is said to form a branch of the tree. Let $\mathcal{B}_{n,i}(\mathcal{T})$ be the number of branches of size i in \mathcal{T} . Also define

$$v_n(\mathcal{T}) := \max\{i \in [n - 1] : \mathcal{B}_{n,i}(\mathcal{T}) \geq 1\}$$

to be the number of nodes in the largest branch of a given tree, \mathcal{T} . In [6], it was shown that

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(v_n(\mathcal{T}_n) \geq \frac{n}{2}\right) = \ln 2, \tag{3}$$

when \mathcal{T}_n is a URT on n vertices. The first purpose of this section and the next theorem is to extend the result of [6] to Hoppe trees, and to provide more details about the asymptotic distribution, via exploiting the relation between Hoppe trees and Hoppe permutations. Further, the result in (3) is now extended to an explicit expression for $\lim_{n \rightarrow \infty} \mathbb{P}(v_n(\mathcal{T}_n) \geq cn)$ for $c \in [1/2, 1]$. Once we have the results for the Hoppe tree case, the coupling construction of Section 2 will also generalize these results to WRTs.

Theorem 5.

- (i) Let \mathcal{T}_n^θ be a Hoppe tree. Then $\frac{v_n(\mathcal{T}_n^\theta)}{n}$ converges weakly to a random variable v with Poisson–Dirichlet distribution, whose cumulative distribution function is given by

$$F_\theta(x) = 1 + \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k!} \int \dots \int_{\mathcal{S}_k(x)} \frac{\left(1 - \sum_{j=1}^k y_j\right)^{\theta-1}}{y_1 \dots y_k} dy_1 \dots dy_k, \quad x > 0,$$

where γ is Euler’s constant [7] and

$$\mathcal{S}_k(x) = \left\{y_1 > x, \dots, y_k > x, \sum_{j=1}^k y_j < 1\right\}.$$

- (ii) When $\theta = 1$, for any $c \in [\frac{1}{2}, 1]$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(v_n(\mathcal{T}_n) \leq cn) = 1 - \ln(c^{-1}).$$

In particular,

$$\mathbb{E}[v] \geq \frac{1}{2}.$$

Proof.

(i) First, we translate the problem into the random permutation setting. We have

$$v_n(\mathcal{T}_n^\theta) =_d \max\{i \in [n - 1] : C_{n-1,i}(\theta) \geq 1\} =: \alpha_n(\theta),$$

where $C_{n-1,i}(\theta)$ is the number of cycles of length i in a θ -biased Hoppe permutation. In this setting, Lemma 5.7 in [2] shows that $\frac{\alpha_n(\theta)}{n}$ converges in distribution to a random variable α with cumulative distribution function

$$F_\theta(x) = 1 + \sum_{k=1}^\infty \frac{(-\theta)^k}{k!} \int \cdots \int_{\mathcal{S}_k(x)} \frac{\left(1 - \sum_{j=1}^k y_j\right)^{\theta-1}}{y_1 \cdots y_k} dy_1 \cdots dy_k, \quad x > 0,$$

where γ is Euler’s constant [7] and

$$\mathcal{S}_k(x) = \left\{ y_1 > x, \dots, y_k > x, \sum_{j=1}^k y_j < 1 \right\}.$$

This proves the first part.

(ii) Setting $\theta = 1$ in the argument of (i), and recalling that the random permutation in this case reduces to a uniformly random permutation, yields that $\frac{v_n(\mathcal{T}_n)}{n}$ converges weakly to a random variable v whose cumulative distribution function is given by

$$F_1(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 + \sum_{k=1}^\infty \frac{(-1)^k}{k!} \int \cdots \int_{\mathcal{S}_k(x)} \frac{dy_1 \cdots dy_k}{y_1 \cdots y_k}, & \text{if } x \in [0, 1], \\ 1, & \text{if } x > 1, \end{cases}$$

where $\mathcal{S}_k(x)$ is as before.

Note that if $x \geq \frac{1}{2}$, $\mathcal{S}_k(x) = \emptyset$ for any $k \geq 2$. Thus the expression simplifies for $\frac{1}{2} \leq x \leq 1$ to

$$F_1(x) = 1 - \int_x^1 \frac{1}{y_1} dy_1.$$

This in particular implies that the derivative of $F_1(x)$ over $[1/2, 1]$ simplifies to

$$f_1(x) = \frac{1}{x}.$$

Hence, for any $c \in [\frac{1}{2}, 1]$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(v_n(\mathcal{T}_n) \geq cn) = \mathbb{P}(v \geq c) = \int_c^1 \frac{1}{x} dx = \ln(1/c).$$

Finally, we have

$$\mathbb{E}[v] \geq \int_{1/2}^1 x \frac{1}{x} dx = \frac{1}{2}. \tag{3}$$

Remark 3. The value $\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{\nu_n(\mathcal{T}_n)}{n} \right]$ when \mathcal{T}_n is a URT is known to be the Golomb–Dickman constant in the literature. Its exact value is known to be 0.62432998854 . . .

Now, let \mathcal{T}_n^w be a WRT of size n with weight sequence $(w_i)_{i \in \mathbb{N}}$ such that there is a $k \in \mathbb{N}$ such that, for all $i > k$, we have $w_i = 1$. Since we can find a coupling of a Hoppe tree to a WRT in which the number of nodes in the largest branch differs at most by k , the following now follows immediately.

Theorem 6. Let \mathcal{T}_n^w be a WRT of size n with weight sequence $(w_i)_{i \in \mathbb{N}}$ such that there is a $k \in \mathbb{N}$ such that, for all $i > k$, we have $w_i = 1$. Then, for any c in $[\frac{1}{2}, 1]$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\nu_n(\mathcal{T}_n^w) \leq cn) = 1 - \ln(c^{-1}).$$

5. Depth of node n

The depth of node n is defined as the length of the path from the root to n . We call all the nodes on the single path from the root to n ancestors of n , so the depth of n is also equal to the number of ancestors of n . Note that in this and the next section we have no restrictions on the weight sequence $(w_i)_{i \in \mathbb{N}}$.

Theorem 7. Let \mathcal{D}_n^w denote the depth of node n in a WRT \mathcal{T}_n^w , and let Z_n^w denote the set of ancestors of n . Define $A_{i,n}^w := \mathbf{1}(i \in Z_n^w)$. Then

$$\mathcal{D}_n^w = 1 + \sum_{i=2}^{n-1} A_{i,n}^w.$$

The $A_{i,n}^w$ are mutually independent Bernoulli random variables with

$$\mathbb{P}(A_{i,n}^w = 1) = \frac{w_i}{\sum_{j=1}^i w_j}.$$

This directly yields the expectation and the variance:

$$\begin{aligned} \mathbb{E}[\mathcal{D}_n^w] &= \sum_{i=1}^{n-1} \frac{w_i}{\sum_{j=1}^i w_j}, \\ \text{Var}(\mathcal{D}_n^w) &= \sum_{i=2}^{n-1} \frac{w_i}{\sum_{j=1}^i w_j} \left(1 - \frac{w_i}{\sum_{j=1}^i w_j} \right). \end{aligned}$$

Proof. Each claim will follow easily once we show that \mathcal{D}_n^w can be written as a sum of independent Bernoulli random variables. For this purpose, we first observe that in a given rooted tree the depth of a node is equal to its number of ancestors, since these determine the path from the root to the node. Using that node 1 is definitely an ancestor of n , in the notation of the theorem we thus get

$$\mathcal{D}_n^w = 1 + \sum_{i=2}^{n-1} A_{i,n}^w.$$

We will first find the distribution law of the $A_{i,n}^w$ and then show mutual independence. For the distribution law we will use the method used in [6]: we first find the values for $n - 1$ and $n - 2$ and then proceed by induction.

Node $n - 1$ can only be an ancestor of n if it is the parent of n , so we get

$$\mathbb{P}(A_{n-1,n}^w = 1) = \frac{w_{n-1}}{\sum_{i=1}^{n-1} w_i}.$$

Similarly, $n - 2$ can only be an ancestor of n if it is the parent of n or it is the grandparent of n , in which case $n - 2$ needs to be the parent of $n - 1$, which needs to be the parent of n . This gives

$$\mathbb{P}(A_{n-2,n}^w = 1) = \underbrace{\frac{w_{n-2}}{\sum_{i=1}^{n-1} w_i}}_{n-2 \text{ is parent of } n} + \underbrace{\frac{w_{n-2}}{\sum_{i=1}^{n-2} w_i} \frac{w_{n-1}}{\sum_{i=1}^{n-1} w_i}}_{n-2 \text{ is grandparent of } n} = \frac{w_{n-2}}{\sum_{i=1}^{n-2} w_i}.$$

We will now show by induction that, for all $j = 2, \dots, n - 1$,

$$\mathbb{P}(A_{j,n}^w = 1) = \frac{w_j}{\sum_{i=1}^j w_i}.$$

Let the above be true for all $j \geq i + 1$, and let $C_{i,j}^w$ denote the event that j is a child of i . Then

$$\mathbb{P}(A_{i,n}^w = 1) = \sum_{j=i+1}^{n-1} \mathbb{P}(A_{j,n}^w = 1, C_{i,j}^w) + \mathbb{P}(C_{i,n}^w).$$

Since $C_{i,j}^w$ only relates to the j th step of the construction process and $A_{j,n}^w$ only depends on the $(j + 1)$ th, \dots , n th steps, these two events are independent. Thus,

$$\begin{aligned} \mathbb{P}(A_{i,n}^w = 1) &= \sum_{j=i+1}^{n-1} \mathbb{P}(A_{j,n}^w = 1) \mathbb{P}(C_{i,j}^w) + \mathbb{P}(C_{i,n}^w) \\ &= \sum_{j=i+1}^{n-1} \left(\frac{w_j}{\sum_{k=1}^j w_k} \frac{w_i}{\sum_{k=1}^{j-1} w_k} \right) + \frac{w_i}{\sum_{j=1}^{n-1} w_j}. \end{aligned}$$

To simplify this expression we first note that we can factor out w_i and, by some elementary operations, get:

$$\frac{w_{i+1}}{\sum_{k=1}^{i+1} w_k} \frac{1}{\sum_{k=1}^i w_k} + \frac{w_{i+2}}{\sum_{k=1}^{i+2} w_k} \frac{1}{\sum_{k=1}^{i+1} w_k} = \frac{w_{i+1} + w_{i+2}}{\sum_{k=1}^{i+2} w_k} \frac{1}{\sum_{k=1}^i w_k}.$$

In general, the following holds for $l \in \mathbb{N}$:

$$\frac{w_{i+1} + w_{i+2} + \dots + w_{i+l}}{\sum_{k=1}^i w_k} \frac{1}{\sum_{k=1}^{i+l} w_k} + \frac{w_{i+l+1}}{\sum_{k=1}^{i+l+1} w_k} \frac{1}{\sum_{k=1}^{i+l} w_k} = \frac{w_{i+1} + \dots + w_{i+l+1}}{\sum_{k=1}^{i+l+1} w_k} \frac{1}{\sum_{k=1}^i w_k}.$$

By using this equality $n - i - 2$ times, we thus get

$$\mathbb{P}(A_{i,n}^w = 1) = w_i \left(\frac{w_{i+1} + \dots + w_{n-1}}{\sum_{k=1}^i w_k} \frac{1}{\sum_{k=1}^{n-1} w_k} + \frac{1}{\sum_{k=1}^{n-1} w_k} \right) = \frac{w_i}{\sum_{k=1}^i w_k}.$$

Now we will show that the random variables $A_{i,n}^w$ are mutually independent for $j = 2, \dots, n - 1$. For this we will use the method used in [10]: for any $2 \leq k \leq n - 2$ and $2 \leq j_k < \dots <$

$j_2 < j_1 \leq n - 1$, consider the event that all the j_i and only the j_i are ancestors of n . We will denote this event by E . Then

$$E := \mathbf{1}(A_{j_i,n}^w = 1, A_{j,n}^w = 0, \text{ for } j \neq j_i, i = 1, \dots, k).$$

By the structure of the recursive tree, to realize this event, n must be a child of j_1 , j_1 a child of j_2, \dots, j_{k-1} a child of j_k , and j_k a child of 1. In general, for $i = 1, \dots, k - 1$, j_i must be a child of j_{i+1} . It does not matter to which nodes $j \neq j_i$ attach. Hence, by the attachment probabilities we get

$$\begin{aligned} \mathbb{P}(E) &= \mathbb{P}(A_{j_i,n}^w = 1, A_{j,n}^w = 0, \text{ for } j \neq j_i, i = 1, \dots, k) \\ &= \underbrace{\frac{w_{j_1}}{\sum_{\ell=1}^{n-1} w_\ell}}_{n \text{ child of } j_1} \prod_{i=1}^{k-1} \underbrace{\frac{w_{j_{i+1}}}{\sum_{\ell=1}^{j_i-1} w_\ell}}_{j_i \text{ child of } j_{i+1}} \underbrace{\frac{w_1}{\sum_{\ell=1}^{j_k-1} w_\ell}}_{j_k \text{ child of } 1} \\ &= w_1 w_{j_1} \cdots w_{j_k} \prod_{i=1}^{n-1} \frac{1}{\sum_{\ell=1}^i w_\ell} \prod_{\substack{1 < j < n \\ j \neq j_i, i=1, \dots, k}} \left(\sum_{\ell=1}^{j-1} w_\ell \right) \\ &= \prod_{i=1}^k \frac{w_{j_i}}{\sum_{\ell=1}^{j_i} w_\ell} \prod_{\substack{1 < j < n \\ j \neq j_i, i=1, \dots, k}} \left(\frac{\sum_{\ell=1}^{j-1} w_\ell}{\sum_{\ell=1}^j w_\ell} \right) \frac{w_1}{\sum_{\ell=1}^1 w_\ell} \\ &= \prod_{i=1}^k \frac{w_{j_i}}{\sum_{\ell=1}^{j_i} w_\ell} \prod_{\substack{1 < j < n \\ j \neq j_i, i=1, \dots, k}} \left(1 - \frac{w_j}{\sum_{\ell=1}^j w_\ell} \right) \\ &= \prod_{i=1}^k \mathbb{P}(A_{j_i,n}^w = 1) \prod_{\substack{1 < j < n \\ j \neq j_i, i=1, \dots, k}} \mathbb{P}(A_{j,n}^w = 0). \end{aligned}$$

This implies that the random variables $A_{i,n}^w$ are mutually independent Bernoulli random variables. Expectation and variance formulas for \mathcal{D}_n^w follow from this observation right away. □

The following central limit theorem now follows.

Theorem 8. *If $\mathbb{E}[\mathcal{D}_n^w] \rightarrow \infty$ and $\limsup_{n \rightarrow \infty} \frac{w_n}{\sum_{i=1}^n w_i} < 1$, then we have*

$$\frac{\mathcal{D}_n^w - \mathbb{E}[\mathcal{D}_n^w]}{\sqrt{\text{Var}(\mathcal{D}_n^w)}} \xrightarrow{d} \mathcal{G} \quad \text{as } n \rightarrow \infty.$$

Proof. By Lyapunov’s central limit theorem for sums of independent Bernoulli random variables, if $\text{Var}(\mathcal{D}_n^w) \rightarrow \infty$ then we have $\frac{\mathcal{D}_n^w - \mathbb{E}[\mathcal{D}_n^w]}{\sqrt{\text{Var}(\mathcal{D}_n^w)}} \xrightarrow{d} \mathcal{G}$ as $n \rightarrow \infty$. Now let $p_n = \frac{w_n}{\sum_{j=1}^n w_j}$. Since $\limsup_{n \rightarrow \infty} p_n < 1$, there is a $0 < \varepsilon < 1$ and an $N \in \mathbb{N}$ such that, for all $n > N$, $1 - p_n > \varepsilon$. Then,

$$\text{Var}(\mathcal{D}_n^w) = \sum_{i=1}^n p_i(1 - p_i) = \sum_{i=1}^N p_i(1 - p_i) + \sum_{i=N+1}^n p_i(1 - p_i) > \varepsilon \sum_{i=N+1}^n p_i.$$

Since we know that $\mathbb{E}[\mathcal{D}_n^w] = \sum_{i=1}^n p_i \rightarrow \infty$ as $n \rightarrow \infty$, this implies that $\text{Var}(\mathcal{D}_n^w) \rightarrow \infty$ as $n \rightarrow \infty$. □

Example 1.

- (i) If the weights are limited from below and above, the expectation and variance of the depth of node n will still be equal to $\mathcal{O}(\ln(n))$ asymptotically.
- (ii) For the Hoppe tree, we write \mathcal{D}_n^θ for the depth of node n and have

$$\mathbb{E}[\mathcal{D}_n^\theta] = 1 + \sum_{i=1}^{n-2} \frac{1}{\theta + i} = \log(n) + \mathcal{O}(1),$$

$$\text{Var}(\mathcal{D}_n^\theta) = \sum_{i=1}^{n-2} \frac{\theta + i - 1}{(\theta + i)^2} = \log(n) + \mathcal{O}(1).$$

Notice that the depth in a URT and a Hoppe tree are asymptotically equivalent. The same conclusion also holds for a wide range of statistics whose dependence on the root is small and vanishes asymptotically. This makes the asymptotic study of Hoppe trees slightly uninteresting. Another such example is the number of leaves in a Hoppe tree, which was studied earlier.

- (iii) There are weight sequences for which the behaviour of the depth of the WRT is totally different from the URT case. Let $(w_i)_{i \in \mathbb{N}} = \left(\frac{1}{i^2}\right)_{i \in \mathbb{N}}$. Then

$$\mathbb{E}[\mathcal{D}_n^w] = \sum_{i=1}^{n-1} \frac{\frac{1}{i^2}}{\sum_{j=1}^i \frac{1}{j^2}},$$

and since for all $i \in \mathbb{N}$ we have $1 < \sum_{j=1}^i \frac{1}{j^2} < \frac{\pi^2}{6}$, we get, for $n \geq 2$,

$$1 \leq \mathbb{E}[\mathcal{D}_n^w] \leq \sum_{i=1}^{n-1} \frac{1}{i^2} \implies 1 \leq \mathbb{E}[\mathcal{D}_n^w] \leq \frac{\pi^2}{6}.$$

Also, since

$$\text{Var}(\mathcal{D}_n^w) = \sum_{i=2}^{n-1} \frac{\frac{1}{i^2} \sum_{j=1}^{i-1} \frac{1}{j^2}}{\left(\sum_{j=1}^i \frac{1}{j^2}\right)^2},$$

and, for all $i \geq 2$,

$$\frac{4^2}{5^2} \leq \frac{\sum_{j=1}^{i-1} \frac{1}{j^2}}{\left(\sum_{j=1}^i \frac{1}{j^2}\right)^2} \leq 1,$$

we get

$$\frac{4^2}{5^2} \sum_{i=2}^{n-1} \frac{1}{i^2} \leq \text{Var}(\mathcal{D}_n^w) \leq \sum_{i=2}^{n-1} \frac{1}{i^2} \implies \frac{4}{5^2} \leq \text{Var}(\mathcal{D}_n^w) \leq \frac{\pi^2}{6} - 1.$$

6. Number of branches

Finally, we study the number of branches in a WRT. Recall that the number of branches of a tree is equal to the number of children of the root. Our results are summarized in the following theorem.

Theorem 9. *Let $(w_i)_{i \in \mathbb{N}}$ be a sequence of positive weights. Denote the number of branches in \mathcal{T}_n^w by \mathcal{B}_n^w .*

(i) *We have*

$$\mathbb{E}[\mathcal{B}_n^w] = \sum_{i=2}^n \frac{w_1}{\sum_{k=1}^{i-1} w_k}, \quad \text{Var}(\mathcal{B}_n^w) = \sum_{i=2}^n \frac{w_1}{\sum_{k=1}^{i-1} w_k} \left(1 - \frac{w_1}{\sum_{k=1}^{i-1} w_k} \right).$$

(ii) *If $\mathbb{E}[\mathcal{B}_n^w]$ diverges then a central limit theorem results:*

$$\frac{\mathcal{B}_n^w - \mathbb{E}[\mathcal{B}_n^w]}{\sqrt{\text{Var}(\mathcal{B}_n^w)}} \rightarrow_d \mathcal{G} \quad \text{as } n \rightarrow \infty.$$

(iii) *Further, one has*

$$d_W \left(\frac{\mathcal{B}_n^w - \mathbb{E}[\mathcal{B}_n^w]}{\sqrt{\text{Var}(\mathcal{B}_n^w)}}, \mathcal{G} \right) \leq \frac{1}{\sqrt{\text{Var}(\mathcal{B}_n^w)}} \frac{\sqrt{28} + \sqrt{\pi}}{\sqrt{\pi}},$$

where, for any two probability measures, the Wasserstein metric $d_W(\mu, \nu)$ is defined as $\sup_{h \in \mathcal{H}} \left| \int h(x) d\mu(x) - \int h(x) d\nu(x) \right|$, where $\mathcal{H} = \{h : \mathbb{R} \rightarrow \mathbb{R} : |h(x) - h(y)| \leq |x - y|\}$.

Proof. Letting $b_i = \mathbf{1}(\text{node } i \text{ attaches to node } 1)$, observe that $\mathcal{B}_n^w = \sum_{i=2}^n b_i$ and the b_i are independent with

$$\mathbb{E}[b_i] = \frac{w_1}{\sum_{j=1}^{i-1} w_j}.$$

The formulas for $\mathbb{E}[\mathcal{B}_n^w] = \sum_{i=2}^n \mathbb{E}[b_i]$ and $\text{Var}(\mathcal{B}_n^w) = \sum_{i=2}^n \text{Var}(b_i)$ are then clear.

The central limit theorem follows from Lyapunov’s central limit theorem. If $\mathbb{E}[\mathcal{B}_n^w] \rightarrow \infty$, we show that this implies that $\text{Var}(\mathcal{B}_n^w) \rightarrow \infty$ too. Notice that, for all $i \geq 2$,

$$1 - \frac{w_1}{\sum_{k=1}^i w_k} \geq \frac{w_2}{w_1 + w_2}.$$

Hence, one obtains

$$\text{Var}(\mathcal{B}_n^w) \geq \frac{w_2}{w_1 + w_2} \sum_{i=2}^n \frac{w_1}{\sum_{k=1}^{i-1} w_k} = \frac{w_2}{w_1 + w_2} (\mathbb{E}[\mathcal{B}_n^w] - 1) \rightarrow \infty.$$

The convergence rates can be obtained by using Theorem 3.6 in [15] for independent random variables: Let Y_1, Y_2, \dots, Y_n be independent random variables for which $\mathbb{E}[Y_i^4] < \infty$ and $\mathbb{E}[Y_i] = 0$ holds. Set $\sigma^2 = \text{Var}(\sum_{i=1}^n Y_i)$ and define $W := \sum_{i=1}^n \frac{Y_i}{\sigma}$. Then

$$d_W(W, \mathcal{G}) \leq \frac{1}{\sigma^3} \sum_{i=1}^n \mathbb{E}[|Y_i|^3] + \frac{\sqrt{28}}{\sqrt{\pi} \sigma^2} \sqrt{\sum_{i=1}^n \mathbb{E}[Y_i^4]}.$$

Defining $Y_i = b_{i+1}^w - \mathbb{E}[b_{i+1}^w]$, simple manipulations give $\mathbb{E}[|Y_i|^3] < \text{Var}(b_{i+1}^w)$ and $\mathbb{E}[Y_i^4] < \text{Var}(b_{i+1}^w)$; for details, see [9]. Plugging these results into the expression of the theorem, we get the bound. \square

Remark 4.

- (i) When $w_1 = \theta > 0$, and $w_i = 1$ for $i = 2, 3, \dots$, we obtain

$$\frac{\mathcal{B}_n - \theta \ln n}{\sqrt{\theta \ln n}} \rightarrow_d \mathcal{G} \quad \text{as } n \rightarrow \infty.$$

In particular, when $\theta = 1$ as well, one recovers the central limit theorem for the URT case.

- (ii) When there exist α_1, α_2 such that $0 < \alpha_1 \leq \sup_i w_i \leq \alpha_2 < \infty$, it can be shown that both the expectation and the variance are still of order $\ln n$ and a central limit theorem holds.
- (iii) When the w_i are not bounded, there can be big differences compared to the case of URTs. One such extreme case is when $w_i = i$, where $\mathbb{E}[\mathcal{B}_n^w] \sim 2$ and $\text{Var}(\mathcal{B}_n^w) \sim 14 - \frac{4\pi^2}{3}$. Another one is when $w_i = 1/i$, in which case

$$\frac{6}{\pi^2}(n-1) \leq \mathbb{E}[\mathcal{B}_n^w] \leq n-1, \quad \frac{6}{5\pi^2}(n-2) \leq \text{Var}(\mathcal{B}_n^w) \leq \frac{1}{4}(n-2).$$

- (iv) It is well known that the number of branches and the depth of node n have the same distribution in a URT. This is not the case when the tree is non-uniform. Indeed, this is intuitively clear since having more branches increases the chance that node n attaches to a node at a lower level.

References

- [1] ALTOK, S. AND IŞLAK, Ü. (2017). On leaf related statistics in recursive tree models. *Statist. Prob. Lett.* **121**, 61–69.
- [2] ARRATIA, R., BARBOUR, A. AND TAVARÉ, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach*. European Mathematical Society, Zürich.
- [3] BERGERON, F., FLAJOLET, P. AND SALVY, B. (1992). Varieties of increasing trees. In *Proc. 17th Colloq. Trees in Algebra and Programming*, pp. 24–48.
- [4] DEVROYE, L., FAWZI, O. AND FRAIMAN, N. (2011). Depth properties of scaled attachment random recursive trees. *Random Structures Algorithms* **41**, 66–98.
- [5] DRMOTA, M. (2009). *Random Trees: An Interplay between Combinatorics and Probability*. Springer, Vienna.
- [6] FENG, Q., SU, C. AND HU, Z. (2005). Branching structure of uniform recursive trees. *Science China Ser. A* **48**, 769–784.
- [7] FINCH, S. R. (2003). *Mathematical Constants*. Cambridge University Press.
- [8] GASTWIRTH, J. L. (1977). A probability model of a pyramid scheme. *Amer. Statistician* **31**, 79–82.
- [9] HIESMAYR, E. (2017). On asymptotics of two non-uniform recursive tree models. Master’s thesis. Available at <https://arxiv.org/abs/1710.01402>.
- [10] LECKEY, K. AND NEININGER, R. (2013). Asymptotic analysis of Hoppe trees. *J. Appl. Prob.* **50**, 228–238.
- [11] SMYTHE, R. T. AND MAHMOUD, H. M. (1995). A survey of recursive trees. *Theory Prob. Math. Statist.* **51**, 1–27.
- [12] MARZOUK, C. (2016). Fires on large recursive trees. *Stoch. Process. Appl.* **126**, 265–289.
- [13] NAJOCK, D. AND HEYDE, C. C. (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Prob.* **19**, 675–680.
- [14] PITTEL, B. (1994). Note on the heights of random recursive trees and random m -ary search trees. *Random Structures Algorithms*. **5**, 337–347.
- [15] ROSS, N. (2011). Fundamentals of Stein’s method. *Prob. Surv.* **8**, 210–293.
- [16] SZYMANSKI, J. (1987). On a nonuniform random recursive tree. In *Annals of Discrete Mathematics* 33, ed. A. Barlotti, M. Biliotti, A. Cossu, G. Korchmaros, and G. Tallini, North-Holland, Amsterdam, pp. 297–306.