

Limiting probabilities for vertices of a given rank in 1-2 trees

Miklós Bóna*

Department of Mathematics
University of Florida
Gainesville, FL, U.S.A.

bona@ufl.edu

István Mező

School of Mathematics and Statistics
Nanjing University of Information Science and Technology
Nanjing, People's Republic of China

istvanmezo@gmail.com

Submitted: Mar 20, 2018; Accepted: Aug 16, 2019; Published: Aug 30, 2019

© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

We consider two varieties of labeled rooted trees, namely non-plane and plane 1-2 trees. In these tree varieties, we study the probability that a vertex chosen from all vertices of all trees of a given size uniformly at random has a given rank. We prove that this probability converges to a limit as the tree size goes to infinity.

Mathematics Subject Classifications: 05A05, 05A15

1 Introduction

Various parameters of many models of random rooted trees are fairly well understood *if they relate to a near-root part of the tree or to global tree structure*. The first group includes, for instance, the numbers of vertices at given distances from the root, the immediate progeny sizes for vertices near the top, and so on. See [10] for a comprehensive treatment of these results. The tree height and width are parameters of global nature. A few important papers in that area are [16], [17], [18], [23], and [24].

Let \mathcal{T} be a class of rooted labeled trees. If v is a vertex of a tree $T \in \mathcal{T}$, then let the *rank* of v be the number of edges in the shortest path from v to a leaf of T that is a descendant of v . So leaves are of rank 0, neighbors of leaves are of rank 1, and so on. The rank of a vertex is fundamentally different from other well-studied statistics such as the height or the depth of a vertex, because it is more “local” in that it only depends on a smaller neighborhood of each vertex than those classic statistics. The rank of a vertex has been the subject of a significant number of research papers in the last decade, such as [3], [4], [6], [8], [9], [12], [14], [20], and [21].

*Supported by Simons Collaboration Grant 421967.

For a fixed n , consider all vertices of all trees in \mathcal{T} that have n vertices, and choose one vertex uniformly at random. Let $a_{n,k}$ be the probability that the chosen vertex is of rank k . It is then natural to ask whether the limiting probability

$$a_k = \lim_{n \rightarrow \infty} a_{n,k}$$

exists.

For one tree variety, *decreasing binary trees*, it has been shown [3], [8] that these limits a_k exist, and the values of a_k were explicitly computed in [4] for $k \leq 6$. Recursive trees are discussed in [14]. However, the methods that were successful for these trees are often unsuccessful for other tree varieties if $k > 1$. This is because many of the relevant differential equations cannot be solved, or even, explicitly stated, caused by the fact that many of the relevant functions lack an elementary antiderivative. We will explain this phenomenon in Section 2.

This raises the intriguing question whether we can prove that a_k exists for some of these tree varieties, *even though we cannot explicitly compute its value*. In this paper we will answer that question in the affirmative for the two labeled tree varieties that we consider. For *non-plane 1-2 trees*, this result is given in Theorem 15. These trees are related to many important objects in combinatorics, such as the *ab*-index and *cd*-index of partially ordered sets, min-max trees, set partitions, and alternating permutations. Richard Stanley devotes Section 1.6 of his book [26] to these questions.

For *plane 1-2 trees*, the proof is very similar, and the only step that is different is carried out in Propositions 20 and 21. The main result for such trees is given in Theorem 17. Note that for $k = 0$ and $k = 1$, we are able to compute the exact values of a_k . These are given in Theorems 4 and 6 for non-plane 1-2 trees and in formulae (20) and (21) for plane 1-2 trees.

Results concerning these two kinds of trees are interesting for two reasons. One is that while their definitions are only somewhat different from *decreasing binary trees* [3, 4, 8], which are in bijection with permutations, the methods needed for their study are fundamentally different from the methods needed to deal with those trees. The two varieties of 1-2 trees that we study in this paper prove to be significantly more difficult from this perspective, even though some of the same results seem to hold. The other reason to study these trees is that they are equinumerous to a large collection of combinatorial objects, so any statistic evaluated for these trees is likely to have a meaning for those objects as well. See sequences A000111 and A080635 in [25] for lists of these objects.

2 Non-plane 1-2 trees

Our first example is the class of *labeled non-plane 1-2 trees*. These are rooted trees in which every non-leaf vertex has at most two children, the vertices are bijectively labeled with the elements of $[n] = \{1, 2, \dots, n\}$ so that the label of each vertex is less than that of its parent, and the set of children of any given vertex is unordered. See Figure 1 for the five such trees on vertex set [4]. In this section, when we write *tree*, we will always mean a labeled decreasing non-plane 1-2 tree on vertex set $[n]$.

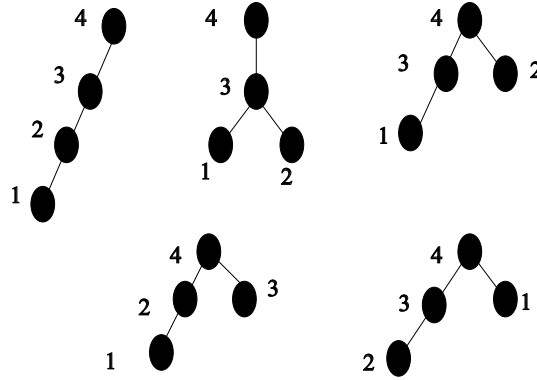


Figure 1: The five decreasing non-plane 1-2 trees on vertex set [4].

It is well-known [10] that the number of labeled non-plane 1-2 trees on vertex set $[n]$ is the n th Euler number E_n , and that the identity

$$E(z) = \sum_{n \geq 0} E_n \frac{z^n}{n!} = \sec z + \tan z \quad (1)$$

holds, where we set $E_0 = 1$.

The first values of E_n are as follows:

n	0	1	2	3	4	5	6	7	8	9	10
E_n	1	1	1	2	5	16	61	272	1385	7936	50521

It follows from (1) that $E(z)$ has two singularities of smallest modulus, at $z = \pi/2$ and at $z = -\pi/2$. Therefore, the exponential order of growth of the Euler numbers is $2/\pi$. In order to find the growth rate of the Euler numbers more precisely, note that at both of these singularities, we can find the residue of $E(z)$ by the following well-known formula (see for instance Theorem 83.2 in [7]).

Proposition 1. *Let $H(z) = f(z)/g(z)$ be a function so that $f(z)$ and $g(z)$ are analytic functions at z_0 , and $f(z_0) \neq 0$, while $g(z) = 0$ and $g'(z) \neq 0$. Then*

$$\text{Res}H(z)|_{z_0} = \frac{f(z_0)}{g'(z_0)}.$$

We can apply Proposition 1 to $E(z)$ if we note that $E(z) = \frac{1+\sin z}{\cos z}$. Then Proposition 1 implies that $\text{Res}E(z)|_{\pi/2} = -2$, and $\text{Res}E(z)|_{-\pi/2} = 0$.

Now observe that

$$\frac{R}{z-a} = \frac{R}{-a} \cdot \frac{1}{1 - \frac{z}{a}} \quad (2)$$

$$= \frac{R}{-a} \sum_{n \geq 0} \frac{z^n}{a^n}. \quad (3)$$

Applying this observation to $E(z)$ with $a = \pi/2$ and $R = -2$, we get that the dominant term of $E(z)$ is of the form $\frac{4}{\pi} \sum_{n \geq 0} z^n (2/\pi)^n$, so

$$\frac{E_n}{n!} \sim \frac{4}{\pi} \cdot \left(\frac{2}{\pi}\right)^n. \quad (4)$$

Here $f_n \sim g_n$ means that $\lim_{n \rightarrow \infty} f_n/g_n = 1$. Formula (4) can be deduced in many ways; the reader may consult pages 268–269 of [10] for one of these.

Now we proceed to determine a_0 and a_1 . For these small values of k , we can explicitly determine a_k , but we will also see why the same approach fails for larger values of k .

2.1 Leaves

Let $A_{0,n}$ denote the total number of *leaves* in all non-plane 1-2 trees on vertex set $[n]$. Then $A_{0,0} = 0$, $A_{0,1} = A_{0,2} = 1$, while $A_{0,3} = 3$, $A_{0,4} = 9$, $A_{0,5} = 35$, and $A_{0,6} = 155$.

Theorem 2. *Let $A_0(z) = \sum_{n \geq 0} A_{0,n} \frac{z^n}{n!}$. Then*

$$A_0(z) = \frac{z - 1 + \cos z}{1 - \sin z}. \quad (5)$$

Proof. Let (v, T) be an ordered pair in which T is a non-plane 1-2 tree on vertex set $[n]$ and v is a leaf of T . Then $A_0(z)$ is the exponential generating function counting such pairs. Let us first assume that $n > 1$, and let us remove the root of T . On the one hand, this leaves a structure that is counted by $A'_0(z)$. On the other hand, this leaves an ordered pair consisting of a non-plane 1-2 tree with a leaf marked, and a non-plane 1-2 tree. By the Product formula of exponential generating functions (see [2, Theorem 8.21]), such ordered pairs are counted by the generating function $A_0(z)E(z)$. Finally, if $n = 1$, then no such ordered pair is formed, while $A'_0(z)$ has constant term 1. This leads to the linear differential equation

$$A'_0(z) = E(z)A_0(z) + 1, \quad (6)$$

with the initial condition $A_0(0) = 0$. Solving this equation we get formula (5) for $A_0(z)$. \square

In order to determine the growth rate of the numbers $A_{0,n}$, we will need the following lemma, which is an enhanced version of Proposition 1. It can easily be proved using basic facts in complex analysis (see Theorem 1 in [7]).

Lemma 3. *Let $H(z) = \frac{f(z)}{g(z)}$ be a function so that f and g are analytic functions at z_0 , and $f(z_0) \neq 0$, while $g(z_0) = g'(z_0) = 0$, and $g''(z) \neq 0$. Then the Laurent series decomposition of $H(z)$ is of the following form.*

$$H(z) = \frac{2f(z_0)}{g''(z_0)} \cdot \frac{1}{(z - z_0)^2} + \frac{h_{-1}}{z - z_0} + h_0 + \cdots,$$

where h_{-1} and h_0 are constants.

Theorem 4. *The equality*

$$a_0 = \lim_{n \rightarrow \infty} \frac{A_{0,n}}{nE_n} = 1 - \frac{2}{\pi} \approx 0.3633802278$$

holds. In other words, for large n , the average number of leaves in a random non-plane 1-2 tree on vertex set $[n]$ is about $n \cdot (1 - \frac{2}{\pi})$.

Proof. We know from (5) that $A_0(z)$ has a unique singularity of smallest modulus, at $z = \pi/2$, hence the exponential growth rate of its coefficients is $2/\pi$. Also note that at that point, the denominator of $A_0(z)$ has a double root. Therefore, Lemma 3 applies and we get that the coefficient of the $(z - \pi/2)^{-2}$ term in the Laurent series of $A_0(z)$ is

$$2 \cdot \frac{\pi/2 - 1 + \cos(\pi/2)}{\sin(\pi/2)} = \pi - 2.$$

Now observe that for any constant D ,

$$\begin{aligned} \frac{D}{(z-a)^2} &= \frac{D}{a^2} \cdot \frac{1}{(1 - \frac{z}{a})^2} \\ &= \frac{D}{a^2} \cdot \sum_{n \geq 0} (n+1) \frac{z^n}{a^n}. \end{aligned}$$

Applying this to the dominant term of $A_0(z)$ with $D = \pi - 2$ and $a = \pi/2$, we get that

$$\frac{A_{0,n}}{n!} \sim (n+1)(\pi-2) \cdot \left(\frac{2}{\pi}\right)^{n+2}. \quad (7)$$

The proof of our claim is now immediate by comparing formulas (7) and (4). □

It is worth pointing out that $\int A_0(z) = 1 - (1-z)(\tan z + \sec z)$, which implies the identity $A_{0,n} = (n+1)E_n - E_{n+1}$. See sequence A034428 in the Online Encyclopedia of Integer Sequences [25], where a different interpretation is given (the number of permutations of descent pattern up-up-down-up-down...). Our result leads to the following natural question.

Question 5. Is there a natural bijection from the set of all *vertices* of non-plane 1-2 trees on vertex set $[n+1]$ in which the root has only one child to the union of the set of all leaves of all non-plane 1-2 trees on vertex set $[n]$ and the set of all non-plane 1-2 trees on vertex set $[n+1]$?

2.2 Vertices of rank 1

Let $A_{1,n}$ be the total number of vertices in all non-plane 1-2 trees on vertex set $[n]$ that are of rank 1. Note that such vertices are neighbors of a leaf. If $n > 1$, then each leaf has exactly one rank-1 vertex as a neighbor, while some rank-1 vertices have not only one, but two leaves as neighbors.

The first few members of the sequence $A_{1,n}$ are $A_{1,0} = 0$, $A_{1,1} = 0$, $A_{1,2} = 1$, $A_{1,3} = 2$, $A_{1,4} = 8$, $A_{1,5} = 30$, and $A_{1,6} = 135$.

Theorem 6. *The equality*

$$a_1 = \lim_{n \rightarrow \infty} \frac{A_{1,n}}{nE_n} = 2 - \frac{\pi^2}{24} - \frac{4}{\pi} \approx 0.3155269391$$

holds.

Proof. Let $A_1(z) = \sum_{n \geq 0} A_{1,n} \frac{z^n}{n!}$. Let (v, T) be an ordered pair in which T is a non-plane 1-2 tree on vertex set $[n]$ and v is a vertex of T that is of rank 1. If v is not the root of T , then removing the root of T decomposes (v, T) into two structures, one of which is again a non-plane 1-2 tree with a vertex of rank 1 marked, and the other one of which is simply a non-plane 1-2 tree. If $n > 1$, and v is the root of T , then removing v , we get two structures, one of which is a leaf, and the other one is a non-plane 1-2 tree. These two structures are distinguishable unless the original tree had three vertices, and its root had two children. That tree contributed $z^3/6$ to the generating function $A_1(z)$, but that contribution was counted twice. This leads to the linear differential equation

$$A_1'(z) = A_1(z) \cdot E(z) + zE(z) - \frac{z^2}{2!}, \quad (8)$$

with the initial condition $A_1(0) = 0$.

Solving this equation we get

$$A_1(z) = \frac{1}{6} \cdot \frac{12z \sin z + 12 \cos z - 12 - 3z^2 \cos z - z^3}{1 - \sin z}. \quad (9)$$

The above formula for $A_1(z)$ shows that $A_1(z)$ has a unique singularity of smallest modulus, at $z = \pi/2$. Therefore, the exponential growth rate of the coefficients of $A_1(z)$ is $2/\pi$. At $z = \pi/2$, the power series $A_1(z)$ has a pole of order two, since the denominator has a double root at that point, while the numerator is non-zero there.

Therefore, we can apply Lemma 3 with $f(z) = 12z \sin z + 12 \cos z - 12 - 3z^2 \cos z - z^3$ and $g(z) = 6(1 - \sin z)$. Then $f(\pi/2) = 6\pi - \frac{\pi^3}{8} - 12$, while $g''(\pi/2) = 6$. Hence Lemma 3 shows that the dominant term of $A_1(z)$ is of the form

$$\frac{2\pi - \frac{\pi^3}{24} - 4}{(z - \frac{\pi}{2})^2} = \frac{2\pi - \frac{\pi^3}{24} - 4}{(\pi/2)^2} \cdot \sum_{n \geq 0} (n+1) \frac{z^n}{(\pi/2)^n}.$$

This implies that

$$\frac{A_{1,n}}{n!} \sim n \cdot \frac{2\pi - \frac{\pi^3}{24} - 4}{(\frac{\pi}{2})^{n+2}}.$$

Comparing this to (4), we obtain the statement of the theorem. \square

Remark 7. Note that it directly follows from the argument we used to prove (8) that if $R_1(z)$ is the exponential generating function for the number of non-plane 1-2 trees on vertex set $[n]$ whose root is of rank 1, then

$$R_1'(z) = zE(z) - \frac{z^2}{2} = z \tan z + z \sec z - \frac{z^2}{2}.$$

Unfortunately, this closed form for $R'_1(z)$ does not lead to an expression for $R_1(z)$ in terms of elementary functions, since the integral $\int z \tan z \, dz$ is known to be non-elementary [5, p. 163].

2.3 Vertices of rank ≥ 2

The methods that we used to enumerate vertices of rank 0 and rank 1 will fail for vertices of higher rank, because we are not able to obtain, let alone, solve the linear differential equations analogous to (8) in a closed form, since the relevant functions have no elementary antiderivatives. Remark 7 shows how early these kinds of problems start occurring; we are not even able to state the equation analogous to (8) in an explicit form.

Therefore, we apply a new method to prove that the limit $a_k = \lim_{n \rightarrow \infty} a_{n,k}$ exists. We will then be able to approximate a_k from above and below. Our first simple notion is the following. Each vertex of a tree is the root of a unique subtree, which we will call *the subtree of the vertex*. In other words, the subtree of a vertex v consists of all descendants of v , including v itself. The subtrees of leaves consist of one vertex only.

For a fixed positive integer r , let $V_{n,r}$ be the probability that a randomly selected vertex in a random non-plane 1-2 tree of size n is the root of a subtree of size exactly r . For instance, if $n = 3$, then $V_{n,1} = 1/2$, $V_{n,2} = 1/6$, and $V_{n,3} = 1/3$. Indeed, there are two non-plane 1-2 tree of size three, (a path and a wedge) with a total of six vertices. Three of those vertices are leaves, hence $V_{3,1} = \frac{3}{6} = \frac{1}{2}$, one of them (the middle vertex of the path) has one proper descendant, so $V_{3,2} = \frac{1}{6}$, while the remaining two vertices are the two roots, yielding $V_{3,3} = \frac{2}{6} = \frac{1}{3}$.

Vertices of a given subtree size are much easier to enumerate than vertices of a given rank, because the number of ways in which a vertex can have a subtree of size r is a fixed number, namely the Euler number E_r , once the set of labels in that subtree is selected.

Proposition 8. *For all positive integers r , the limit*

$$v_r = \lim_{n \rightarrow \infty} V_{n,r}$$

exists.

Proof. Let $V_r(z) = \sum_{n \geq 0} V_{n,r} z^n / n!$. Then by the Product formula, we have

$$V'_r(z) = V_r(z)E(z) + f'_r(z),$$

where $f_r(z)$ is the generating function for the number of trees in which *the root* has a “subtree” of size r . That is, $f_r(z) = E_r z^r / r!$, so $f'_r(z) = E_r z^{r-1} / (r-1)!$, and the last displayed equation becomes

$$V'_r(z) = V_r(z)E(z) + E_r \frac{z^{r-1}}{(r-1)!}. \quad (10)$$

This is a first order linear differential equation with initial condition $V_r(0) = 0$. Its solution is

$$V_r(z) = \frac{E_r}{(r-1)!} \frac{L_r(z)}{1 - \sin z},$$

where $L_r(z)$ is the function that satisfies $L'_r(z) = z^r(1 - \sin z)$ and has a constant term so that $V_r(0) = 0$ holds.

The integral in the numerator can be explicitly computed using the well-known formula

$$\begin{aligned} \int z^r \sin z \, dz &= \cos z \sum_{i=0}^{r/2} (-1)^{i+1} z^{r-2i} \frac{r!}{(r-2i)!} \\ &+ \sin z \sum_{i=0}^{(r-1)/2} (-1)^i z^{r-2i-1} \frac{r!}{(r-2i-1)!}. \end{aligned}$$

This means that $V_r(z)$ has a unique singularity of smallest modulus (a double pole) at $z_0 = \pi/2$. The rest of the argument uses Lemma 3 at $z_0 = \pi/2$ to determine v_r , in the same way as we did in the proofs of Theorems 4 and 6. \square

Remark 9. Note that we are able to explicitly solve the linear differential equation (10) because its correction term, that is, the summand that does not contain $V'_r(z)$ or $V_r(z)$, is a *polynomial*. The same argument used here would work for any polynomial instead of $f'_r(z) = E_r z^{r-1}/(r-1)!$.

Proposition 8 shows that the limit v_r exists for every fixed positive integer r . As the v_r are all positive real numbers, and $\sum_r v_r \leq 1$, the sum $\sum_{r=1}^{\infty} v_r$ is convergent. However, *what* is the value of that sum? The exact formulas we obtain for each v_r from Proposition 8 are too complicated to be useful for the computation of that sum. Note that it is not true in every tree variety that the analogously defined sum is equal to 1. A simple counterexample is the family of rooted trees in which every non-leaf vertex has exactly one child. In that case, there is only one tree of each size, but $v_r = 0$ for all r , since $V_{n,r} = \frac{1}{n}$ for $r \leq n$. Therefore, $\sum_{r=1}^{\infty} v_r = 0$ for that tree variety.

However, for our non-plane 1-2 trees, the sum turns out to be 1, although not in a trivial way. This is the content of the following theorem that has been conjectured by the present authors, and has recently been proved by Svante Janson [15].

Theorem 10. *The equality*

$$\sum_{r=1}^{\infty} v_r = 1$$

holds.

In order to prove Theorem 10, we first need the following fact.

Proposition 11. *For all n , the expected number of leaves in a random non-plane 1-2 tree on vertex set $[n]$ is at least $n/4$.*

Note that in Theorem 4, we proved that much more is true for large n . However, the statement of Proposition 11 is true for *every* n . With a little bit of additional work, it is possible to prove that in fact, the expected number of leaves is at least $n/3$ for every n , but the weaker claim of Proposition 11 suffices for our purposes.

Proof. (of Proposition 11.) Every such tree contains exactly one more leaves than vertices with two children. Therefore, it suffices to show that the expected number of vertices with one child is not more than $n/2$. We prove this by induction on n . Let M_n be the expected number of vertices with one child in a random tree on vertex set $[n]$. Then $M_1 = 0$, $M_2 = 1$, and $M_3 = 1$, so the statement holds if $n \leq 3$. Now let us assume that $n > 3$. Let p_n be the probability of the event P_n that the *root* of a non-plane 1-2 tree on $[n]$ has exactly one child. Furthermore, let $M_{n,A}$ denote the expected number of vertices with one child in a random non-plane 1-2 tree of size n given that event A occurs. Now let A be the event that the root of such a random tree has one child, and let \bar{A} denote the event that that root has 0 or 2 children. Then by Bayes' theorem for expectations,

$$M_n = M_{n,A}p_n + M_{n,\bar{A}}(1 - p_n).$$

So, by our induction hypothesis,

$$M_n \leq \left(\frac{n-1}{2} + 1 \right) \cdot p_n + \frac{n-1}{2} \cdot (1 - p_n) = \frac{n-1}{2} + p_n \leq \frac{n}{2}.$$

The last inequality follows, since $p_n = E_{n-1}/E_n$, the ratio of two consecutive Euler numbers, and the Euler numbers are known to be log-convex [19]. So the sequence of the numbers p_n is decreasing. As $p_3 = 1/2$, it follows that $p_n \leq 1/2$ if $n \geq 3$. \square

Let Z_n be a random variable defined on the set of all vertices of all non-plane 1-2 trees on vertex set $[n]$, so that $Z_n(v)$ is the size of the subtree rooted at v .

Lemma 12. *Let v be a vertex of a nonplane 1-2 tree chosen uniformly at random. Then for all n , the inequality $E(\sqrt{Z_n}) \leq 100$ holds.*

Proof. We will use strong induction to prove the stronger inequality

$$E(\sqrt{Z_n}) \leq 100 - \frac{90}{\sqrt{n}}. \quad (11)$$

This inequality clearly holds for $n = 1$. Now let us assume that it holds for all positive integers $r < n$ and prove it for n .

Let T be a tree of size n , selected uniformly at random. Let v be a vertex of T , selected in the same way.

Then by Proposition 11, there is an at least $1/4$ chance of v being a leaf. There is an $1/n$ chance of v being the root. There is a less than $3/4$ chance of v being a vertex other than a leaf or the root, in which case the induction hypothesis applies to the subtree of v , with some $r < n$ playing the role of n . Therefore,

$$\begin{aligned} E(\sqrt{Z_n}) &\leq \frac{3}{4} \cdot \left(100 - \frac{90}{\sqrt{n}} \right) + \frac{1}{4} + \frac{1}{n} \cdot \sqrt{n} \\ &= 75.25 - \frac{66.5}{\sqrt{n}} \end{aligned}$$

$$\leq 100 - \frac{90}{\sqrt{n}},$$

where the last estimate holds as $n \geq 1$. □

Now we are ready to prove Theorem 10.

Proof. (of Theorem 10) By Lemma 12 and by Markov's inequality, we know that for all positive constants C , and for all n , we have

$$\Pr(\sqrt{Z_n} > 100C) \leq 1/C,$$

so

$$\Pr(Z_n > 10000C^2) \leq 1/C. \quad (12)$$

Let us now assume that $\sum_{k=1}^{\infty} v_k = \alpha < 1$. That means that for all N , the inequality $\sum_{k=1}^N v_k < \alpha$ holds. In other words, if n is large enough, then a random 1-2 tree on $[n]$ has on average at least $(1 - \alpha)n/2$ vertices whose subtree is of size more than N . That is,

$$\Pr(Z_n > N) \geq \frac{1 - \alpha}{2}. \quad (13)$$

Now select C to be a positive integer so that $1/C < (1 - \alpha)/2$, then select $N = 10000C^2$. Then inequality (12) forces

$$\Pr(Z_n > N) \leq 1/C < (1 - \alpha)/2$$

for all n , while inequality (13) forces $\Pr(Z_n > N) \geq (1 - \alpha)/2$ for n sufficiently large, which is clearly a contradiction. □

The following is an obvious corollary of Theorem 10 that we will need later.

Corollary 13. *Let $U_{n,r} = 1 - \sum_{i=1}^r V_{n,i}$ be the probability that a random vertex of a random non-plane 1-2 tree has a subtree of size larger than r . Then clearly,*

$$u_r := \lim_{n \rightarrow \infty} U_{n,r} = 1 - \sum_{i=1}^r v_i.$$

Furthermore,

$$\lim_{r \rightarrow \infty} u_r = 1 - \lim_{r \rightarrow \infty} \sum_{i=1}^r v_i = 0. \quad (14)$$

We now return to our main goal, that is, to proving that the limit $a_k = \lim_{n \rightarrow \infty} a_{n,k}$ exists. For the rest of this section, we fix the rank k of vertices we are studying, and, to alleviate notation, we do *not add* the index k to all parameters related to these vertices.

Our main idea is the following. The set \mathcal{R}_k of all vertices of all trees of size n *contains* the set $\cup_{i=1}^r \mathcal{W}_{n,i}$, where $\mathcal{W}_{n,i}$ is the set of all vertices of all trees of size n that are of rank k and have a subtree of size i . On the other hand, \mathcal{R}_k *is contained* in the set $(\cup_{i=1}^r \mathcal{W}_{n,i}) \cup (\cup_{i>r} \mathcal{V}_{n,i})$, where $\mathcal{V}_{n,i}$ is the set of all vertices in all trees of size n whose subtree is of size i (but are of any rank).

Let $W_{n,i}$ be the probability that a random vertex of a random tree of size n is of rank k and is the root of a subtree of size i .

Definition 14. Let

$$w_i = \lim_{n \rightarrow \infty} W_{n,i}.$$

The limits w_i exist, because the exponential generating functions of the numbers $W_{n,i}$ satisfy a linear differential equation like (10), and, as explained in Remark 9, we can explicitly solve those differential equations, since their “correction term” is a polynomial. Indeed, there are only a finite number of ways that a subtree of a vertex can be of rank k and have a subtree of size i , once the set of labels going into that subtree is selected.

As the w_i are positive real numbers, and for all r , the inequality $\sum_{i=1}^r w_i \leq 1$ holds, the sum

$$w = \sum_{i=1}^{\infty} w_i$$

exists.

Now we are ready to state and prove our main theorem.

Theorem 15. *For all nonnegative integers k , the limit*

$$a_k := \lim_{n \rightarrow \infty} a_{n,k}$$

exists. Furthermore,

$$a_k = w.$$

Proof. First notice that for all n and r , the inequality

$$\sum_{i=1}^r W_{n,i} \leq a_{n,k}$$

holds, since the left-hand side is the probability of a random vertex having a more restrictive property (rank k , subtree size at most r) than the property represented on the right-hand side (rank k). Therefore,

$$\sum_{i=1}^r w_r \leq \liminf_{n \rightarrow \infty} a_{n,k}, \quad (15)$$

and so

$$w \leq \liminf_{n \rightarrow \infty} a_{n,k}. \quad (16)$$

Now notice that for all n and r , the inequality

$$a_{n,k} \leq \sum_{i=1}^r W_{n,i} + \sum_{i>r} V_{n,i} = \left(\sum_{i=1}^r W_{n,i} \right) + U_{n,r}$$

holds. Indeed, the right-hand side is the probability of a random vertex being of rank k and having a subtree of size at most r , or simply having a subtree of size more than r

(and any rank). A particular way of this occurring is when the random vertex is of rank k , which is the event whose probability is represented on the left-hand side.

This implies that for all r , we have

$$\limsup_{n \rightarrow \infty} a_{n,k} \leq \sum_{i=1}^r w_i + \left(1 - \sum_{i=1}^r v_i\right). \quad (17)$$

As r goes to infinity, the first sum on the right-hand side goes to w , while the second sum goes to 0, as we saw in Corollary 13. This proves that

$$\limsup_{n \rightarrow \infty} a_{n,k} \leq w. \quad (18)$$

Comparing inequalities (16) and (18), we see that

$$\limsup_{n \rightarrow \infty} a_{n,k} \leq w \leq \liminf_{n \rightarrow \infty} a_{n,k},$$

completing the proof of the theorem. \square

For numerical approximations, one can use the following corollary, which is an immediate consequence of Theorem 15 that we have just proved, and inequalities (15) and (17) that we have used in the proof of that theorem.

Corollary 16. *For all r , the chain of inequalities*

$$\sum_{i=1}^r w_i \leq a_k \leq \sum_{i=1}^r w_i + \left(1 - \sum_{i=1}^r v_i\right)$$

holds.

3 Plane 1-2 trees

The next tree class we study is the class of decreasing *plane* 1-2 trees on vertex set $[n]$. These are similar to the trees of the previous section, except that now the order of the children of each vertex matters. See Figure 2 for an illustration. We denote the number of such trees on $[n]$ by b_n .

Our goal in this section is to show the following analogue of Theorem 15. Let us choose a vertex of the set of all vertices of all plane labeled trees on vertex set $[n]$ uniformly at random. Let $a_{n,k}$ be the probability that the chosen vertex is of rank k , and let us keep all definitions of the previous section, just take them for *plane* 1-2 trees instead of non-plane ones.

Theorem 17. *For all nonnegative integers k , the limit*

$$a_k := \lim_{n \rightarrow \infty} a_{n,k}$$

exists. Furthermore,

$$a_k = w.$$

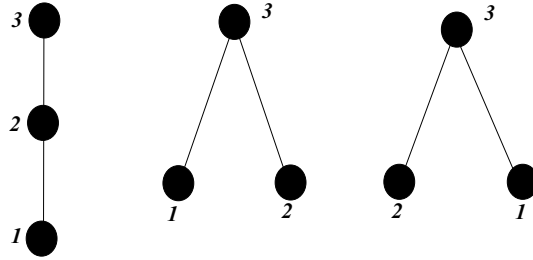


Figure 2: The three decreasing plane 1-2 trees on vertex set [3].

Proof. Most steps are similar to what we saw in Section 2, but there will be one step that requires a separate argument. In this section, if we simply say “tree”, we always mean a decreasing plane 1-2 tree.

The first few values of the sequence b_n , starting with b_1 , are 1, 1, 3, 9, and 39. These trees were studied in [1]. Set $b_0 = 1$, and consider the exponential generating function

$$B(z) = \sum_{n=0}^{\infty} b_n \frac{z^n}{n!}.$$

Removing the root of such a tree results in a structure counted by $B'(z)$. On the other hand, if the tree had more than one vertex, then it results in a pair of such trees, except in the case when the root had only one child. This leads to the separable differential equation

$$B'(z) = 1 - B(z) + B^2(z).$$

Solving this equation yields

$$B(z) = \frac{1}{2} + \frac{\sqrt{3}}{2} \tan \left(\frac{\sqrt{3}}{2} z + \frac{\pi}{6} \right).$$

The power series form of $B(z)$ leads to the asymptotic formula

$$\frac{b_n}{n!} \sim \frac{3^{3/2}}{2\pi} \left(\frac{3^{3/2}}{2\pi} \right)^n. \quad (19)$$

Just as it was the case for non-plane trees, we can determine the values of a_0 and a_1 for plane 1-2 trees as well.

3.1 Leaves

We can find the ratio of leaves among all vertices in a way that is analogous to that for non-plane 1-2 trees.

Theorem 18. *The exponential generating function of the numbers of leaves in decreasing plane 1-2 trees is*

$$B_0(z) = \sum_{n=0}^{\infty} b_{0,n} \frac{z^n}{n!} = \frac{6z + \sqrt{3} \sin(\sqrt{3}z) + 3 \cos(\sqrt{3}z) - 3}{-3\sqrt{3} \sin(\sqrt{3}z) + 3 \cos(\sqrt{3}z) + 6},$$

which satisfies the differential equation

$$B'_0(z) = 2B_0(z)(B(z) - 1) + B_0(z) + 1.$$

Proof. Just like in our proofs for analogous results in Section 2, we count ordered pairs (v, T) , where v is a leaf of the tree T . Let us remove the root of T . Then there are two cases, namely, either the removed root was v , or it was not.

If v is not the root, and we got two trees, one with the marked vertex, then the Product formula yields the generating function $2B_0(z)(B(z) - 1)$, as the order of the components matters.

If v is not the root, and the root has only one child, then removing the root, we get only one tree, with a marked vertex, which contributes the generating function $B_0(z)$.

Finally, if v is the root, the only possible tree is the one-point graph. The removal of that root leads to the empty graph, represented by $1 = \left(\frac{z^1}{1!}\right)'$ in the differential equation. \square

By the generating function we can determine the first values of $b_{0,n}$:

n	0	1	2	3	4	5	6	7	8	9	10
$b_{0,n}$	0	1	1	5	17	93	513	3 477	25 569	212 733	1 929 393

A simple application of Lemma 3 yields that

$$\frac{b_{0,n}}{(n+1)n!} \sim \frac{\frac{4\pi}{9\sqrt{3}} - \frac{1}{3}}{\left(\frac{2\pi}{3\sqrt{3}}\right)^{n+2}},$$

and, by recalling (19),

$$a_0 = \lim_{n \rightarrow \infty} \frac{b_{0,n}}{nb_n} = \frac{2}{3} - \frac{\sqrt{3}}{2\pi} \approx 0.391. \quad (20)$$

3.2 Neighbors of leaves

This case is similar to that of leaves, with some subtle differences. Let $b_{1,n}$ denote the number of all vertices of rank 1 in all trees on vertex set $[n]$, and let $B_1(z)$ be the exponential generating function of these numbers.

Theorem 19. *The generating function $B_1(z)$ satisfies the differential equation*

$$B'_1(z) = 2B_1(z)(B(z) - 1) + B_1(z) + 2z(B(z) - 1) + z - z^2 \quad (B_1(0) = 0).$$

Therefore,

$$B_1(z) = \frac{6z^3 + \sqrt{3}(3z^2 - 15z - 5)\sin(\sqrt{3}z) + 3(3z^2 + 5z - 5)\cos(\sqrt{3}z) + 15}{9(\sqrt{3}\sin(\sqrt{3}z) - \cos(\sqrt{3}z) - 2)}.$$

Proof. Let us count ordered pairs (v, T) , where v is a vertex of rank 1 in a tree T . Let us remove the root of T . The case when v is not the root is the same as in Section 3.1, contributing the term $2B_1(z)(B(z) - 1) + B_1(z)$. When v is the root, then removing it we obtain a leaf and a tree. If this tree is not empty, we must distinguish whether it was on the left or right hand side, so we must add a term $2z(B(z) - 1)$. If, in turn, the subtree is empty, we must add the term representing the path of length one, that is $\left(\frac{z^2}{2!}\right)' = z$. Finally, we must realize that the terms $2B_1(z)(B(z) - 1)$ and $2z(B(z) - 1)$ both contain the two trees on three points where the root has two children. Therefore we must subtract $2\left(\frac{z^3}{3!}\right)' = z^2$. \square

The first values of $b_{1,n}$ are as follows.

n	0	1	2	3	4	5	6	7	8	9	10
$b_{1,n}$	0	0	1	3	15	75	435	2883	21447	177435	1613835

The asymptotic expression for the total number of rank one vertices can be found easily:

$$\frac{b_{1,n}}{(n+1)n!} \sim \frac{540\sqrt{3}\pi - 16\sqrt{3}\pi^3 - 1215}{2187\left(\frac{2\pi}{3\sqrt{3}}\right)^{n+2}},$$

and

$$\lim_{n \rightarrow \infty} \frac{b_{1,n}}{nb_n} = \frac{10}{9} - \frac{5}{2\sqrt{3}\pi} - \frac{8\pi^2}{243} \approx 0.3267. \quad (21)$$

3.3 Vertices of higher rank

If we try to apply the method of Sections 3.1 and 3.2 for vertices of rank k , for $k \geq 2$, we will fail, because yet again, the relevant generating functions will not have elementary antiderivatives. However, the method that we used in Section 2.3 to prove that the limits a_k exist will work again, as we will show.

Let us define the limits v_r as we did in Proposition 8 and w_r exactly as we did in Definition 14, except that now the trees are plane. If we try to follow the argument of the non-plane case, we see that the first step towards proving the existence of v_r and w_r is to show that we can explicitly solve the linear differential equation

$$f'(z) = 2f(z)(B(z) - 1) + f(z) + P(z),$$

where $P(z)$ is a *polynomial* function. Indeed, we get differential equations of the above kind when we attempt to find the probabilities $V(n, r)$ or $W(n, r)$.

Bringing the above differential equation to standard form, we get

$$f'(z) + (1 - 2B(z))f(z) = P(z). \quad (22)$$

In order to solve (22), we multiply both sides by the integrating factor

$$Q(z) = \exp \left(\int (1 - 2B(z)) dz \right) = \frac{1}{2} + \frac{\cos(\sqrt{3}z)}{4} - \frac{\sqrt{3} \sin(\sqrt{3}z)}{4}.$$

Multiplying both sides of (22) by $Q(z)$, we get the equation

$$(f(z)Q(z))' = Q(z)P(z),$$

which we can explicitly solve as long as we can integrate $Q(z)P(z)$. In the present case, we can certainly do that, since $P(z)$ is a polynomial function of z , hence it is a polynomial function of $\sqrt{3}z$ as well, so a substitution $t = \sqrt{3}z$ will result in a function consisting of the sums of summands in the form $t^m \sin t$ and $t^n \cos t$. In the end, we obtain

$$f(z) = \frac{\int Q(z)P(z)}{Q(z)} = \frac{K(z)}{Q(z)}, \quad (23)$$

a meromorphic function. The asymptotic behavior of meromorphic functions is well understood. See Theorem IV.10 of [10] for the most important results. In our case, the numerator $K(z)$ of $f(z)$ in (23) is an entire function, while the denominator has a zero at $z = 2\sqrt{3}\pi/9$ that has multiplicity two. Therefore, the asymptotics of the coefficients of $f(z)$ can be computed using Lemma 3. So the existence of v_r and w_r can also be proved in the same way as it was in Section 2 for non-plane 1-2 trees.

The next step is to prove the plane analogue of Theorem 10, that is, the equality $\sum_{r=1}^{\infty} v_r = 1$ for plane 1-2 trees. There is one step in that proof that needs an argument that is different from its non-plane analogue, which is Proposition 11. Therefore, we announce and prove it separately as follows.

Proposition 20. *For all n , the expected number of leaves in a random plane 1-2 tree on vertex set $[n]$ is at least $n/4$.*

Proof. We prove that the expected number of leaves in plane 1-2 trees on vertex set $[n]$ is at least as large as the expected number of leaves in non-plane 1-2 trees on vertex set $[n]$. As the latter has been proved to be at least $n/4$ in Proposition 11, this will be sufficient.

Clearly, in both tree varieties, the number of vertices with two children is one less than the number of leaves. Therefore, it suffices to prove that average the number of vertices with one child is *at most as large* in plane 1-2 trees on $[n]$ as it is on non-plane 1-2 trees on $[n]$. We use a well-known inequality, known as the *Chebyshev sum inequality* or (a special case of) the *rearrangement inequality*. See [11] or [13] for a proof.

Proposition 21. *(Chebyshev sum inequality.) Let $r_1 \leq r_2 \leq \dots \leq r_u$ and $t_1 \geq t_2 \geq \dots \geq t_u$ be nonnegative real numbers so that $t_1 > 0$. Then the inequality*

$$\frac{r_1 + r_2 + \dots + r_u}{u} \geq \frac{r_1 t_1 + r_2 t_2 + \dots + r_u t_u}{t_1 + t_2 + \dots + t_u}$$

holds.

Let us return to the proof of Proposition 20. Consider all E_n non-plane trees on vertex set $[n]$. Let s_1, s_2, \dots, s_{E_n} denote the number of vertices with exactly one child in each of these trees, and let us order the set of these E_n trees so that the sequence of the s_i is non-decreasing, that is, $s_1 \leq s_2 \leq \dots \leq s_{E_n}$. Then the average number of vertices with one child in all non-plane 1-2 trees on $[n]$ is

$$M_n = \frac{\sum_{i=1}^{E_n} s_i}{E_n}. \quad (24)$$

On the other hand, if such a tree T has s_i vertices with one child, then it has $(n-1-s_i)/2$ vertices with two children, (and, though we will need this only later, T has $(n+1-s_i)/2$ leaves). Therefore, there are exactly $2^{(n-1-s_i)/2}$ *plane* 1-2 trees on vertex set $[n]$ that are identical to T as non-plane trees, and each of those trees has s_i vertices with one child. This proves that the average number of vertices with exactly one child in all *plane* 1-2 trees on vertex set $[n]$ is

$$m_n = \frac{\sum_{i=1}^{E_n} s_i 2^{(n-1-s_i)/2}}{\sum_{i=1}^{E_n} 2^{(n-1-s_i)/2}}. \quad (25)$$

Finally, note that the sequences $s_1 \leq s_2 \leq \dots \leq s_{E_n}$ and $2^{(n-1-s_1)/2} \geq 2^{(n-1-s_2)/2} \geq \dots \geq 2^{(n-1-s_{E_n})/2}$ satisfy the requirements of Proposition 21, so $M_n \geq m_n$ holds. So the expected number of vertices with one child in a random non-plane 1-2 tree is at least as large as the expected number of vertices with one child in a random plane 1-2 tree of the same size. Therefore, the expected number of leaves in a *plane* 1-2 tree is at least as large as the expected number of leaves in a random non-plane 1-2 tree of the same size. The proof of our claim is now immediate, since we saw in Proposition 11 that the expected number of leaves in a random non-plane 1-2 tree on vertex set $[n]$ is at least $n/4$. \square

All remaining steps of Theorem 15 can be carried out without any extra effort, showing that the limits a_k exist for all k , for the variety of plane 1-2 trees as well. \square

4 Approximations

Corollary 16 makes numerical approximations of a_k possible. As the upper bound provided by (17) was obtained by a rather crude estimate, it is reasonable to assume that the lower bound in that corollary is a better estimate for w than the upper bound. It follows from our methods that both the upper and the lower bounds will be of the form $\pi^{-2}F(\pi)$, where F is a polynomial function with rational coefficients. For instance, selecting $k = 2$ and $r = 12$ leads to a lower bound of $0.188285 \leq a_2$. On the other hand, less rigorous but more extensive computations carried out by Jay Pantone [22] suggest the approximate values $a_2 \approx 0.20278137$, $a_3 \approx 0.0893474$, and $a_4 \approx 0.0243854$.

Acknowledgements

The authors are indebted to Svante Janson for the proof of Theorem 10. They are also grateful to Jay Pantone who helped them obtain the numerical results of the last section.

References

- [1] BERGERON, F., FLAJOLET, P., AND SALVY, B., Varieties of Increasing Trees, in Lecture Notes in Computer Science vol. 581, ed. J.-C. Raoult, Springer 1992, pp. 24–48.
- [2] BÓNA, M. (2011) A Walk Through Combinatorics, third edition, World Scientific.
- [3] BÓNA, M. (2014) k -protected vertices in binary search trees, *Adv. in Appl. Math.* **53**, 1–11.
- [4] BÓNA, M. AND PITTEL, B. (2017) On a random search tree: asymptotic enumeration of vertices by distance from leaves, *Adv. in Appl. Probab.* **49** (2017), no. 3, 850–876.
- [5] BRONSTEIN, M. (1997) Symbolic Integration I, Springer.
- [6] COPENHAVER, K. (2017) k -protected vertices in unlabeled rooted plane trees. *Graphs Combin.* **33** (2), 347–355.
- [7] CHURCHILL, R. V. AND BROWN, J. W. (2014) Complex Variables and Applications, ninth edition. McGrawHill, New York, NY.
- [8] DEVROYE, L. AND JANSON, S. (2014). Protected nodes and fringe subtrees in some random trees, *Electron. Commun. Probab.* **19**, no. 6, 10 pages.
- [9] DU, R. R. AND PRODINGER, H. (2012). Notes on protected nodes in digital search trees, *Appl. Math. Lett.* **25** 1025–1028.
- [10] FLAJOLET, P. AND SEDGEWICK, R. (2009) Analytic Combinatorics, *Cambridge University Press*.
- [11] HARDY, G. H.; LITTLEWOOD, J. E.; PÓLYA, G. (1988) Inequalities. Cambridge Mathematical Library. *Cambridge University Press*.
- [12] HEUBERGER, C. AND PRODINGER, H. (2017) Protection number in plane trees. *Appl. Anal. Discrete Math.*, **11** (2), 314–326.
- [13] HOLSTERMANN, J. (2017), A Generalization of the Rearrangement Inequality. *Mathematical Reflections* **5**, 4 pages.
- [14] HOLMGREN, C. AND JANSON, S. (2017). Fringe trees, Crump-Mode-Jagers branching processes and m -ary search trees. *Probability Surveys* **14** 53–154.
- [15] JANSON, S. Personal communication, February, 2018.
- [16] KESTEN H. AND PITTEL, B. (1996). A local theorem for the number of nodes, the height and the number of final leaves in a critical branching process tree, *Random. Struct. Algorithms*. **8**, 243–299.
- [17] KOLCHIN, V. F. (1978). Moment of degeneration of a branching process and height of a random tree, *Math. Notes Acad. Sci. USSR*, **24**, 954–961.
- [18] MAHMOUD, H. AND PITTEL, B. (1984). On the most probable shape of a search tree grown from a random permutation, *SIAM J. Algebraic Discrete Methods*, **5**, 69–81.

- [19] LIU, L. L. AND WANG, Y. On the log-convexity of combinatorial sequences. *Adv. in Appl. Math.* **39** (2007), no. 4, 453–476.
- [20] MAHMOUD, H. AND WARD, M. (2012). Asymptotic distribution of two-protected nodes in random binary search trees, *Appl. Math. Letters*, **25**, no. 12, 2218–2222.
- [21] MAHMOUD, H. AND WARD, M. (2015). Asymptotic properties of protected notes in random recursive trees, *J. Appl. Prob.*, **52**, 290–297.
- [22] PANTONE, J. Personal communication, January, 2018.
- [23] PITTEL, B. (1984). Growing random binary trees, *J. Appl. Math. Anal. Appl.*, **103**, 461–480.
- [24] PITTEL, B. (1994). Note on the heights of random recursive trees and random m -ary search trees, *Random Struct. & Algorithms*. **5**, 337–347.
- [25] ONLINE ENCYCLOPEDIA OF INTEGER SEQUENCES, online database, <http://www.oeis.org>.
- [26] STANLEY, R. (2011) Enumerative Combinatorics, Volume 1, second edition. *Cambridge University Press*, 2011.