

On the distribution of depths in increasing trees

Markus Kuba

Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria
kuba@dmg.tuwien.ac.at

Stephan Wagner

Department of Mathematical Sciences
Stellenbosch University
Private Bag X1, Matieland 7602, South Africa
swagner@sun.ac.za

Submitted: Oct 28, 2009; Accepted: Oct 1, 2010; Published: Oct 15, 2010
Mathematics Subject Classifications: 05A19; 05C05 60C05

Abstract

By a theorem of Dobrow and Smythe, the depth of the k th node in very simple families of increasing trees (which includes, among others, binary increasing trees, recursive trees and plane ordered recursive trees) follows the same distribution as the number of edges of the form $j - (j + 1)$ with $j < k$. In this short note, we present a simple bijective proof of this fact, which also shows that the result actually holds within a wider class of increasing trees. We also discuss some related results that follow from the bijection as well as a possible generalization. Finally, we use another similar bijection to determine the distribution of the depth of the lowest common ancestor of two nodes.

1 Introduction

Increasing trees are rooted labeled trees where the nodes of a tree of size n are labeled by distinct integers from the set $\{1, \dots, n\}$ in such a way that the sequence of labels along any branch starting at the root is increasing. There are various important families of increasing trees, such as *binary increasing trees*, *recursive trees* or *plane-oriented recursive trees*. A general framework for these instances is given by what is known as *simple families of increasing trees* [3]; such a family \mathcal{T} is characterized by a sequence of non-negative

numbers $(\varphi_k)_{k \geq 0}$, where $\varphi_0 > 0$. This sequence is called the degree-weight sequence. We assume that there exists a $k \geq 2$ with $\varphi_k > 0$ to avoid trivialities.

Now we assign a *weight* $w(T)$ to any ordered tree T by $w(T) := \prod_v \varphi_{d(v)}$, where v ranges over all nodes of T and $d(v)$ is the out-degree of v . Furthermore, let $L(T)$ be the number of increasing labelings of T with integers $1, 2, \dots, |T|$, as explained above, and define the total weights by $T_n := \sum_{|T|=n} w(T) \cdot L(T)$. It follows that the exponential generating function $T(z) := \sum_{n \geq 1} T_n \frac{z^n}{n!}$ satisfies the autonomous first order differential equation

$$T'(z) = \Phi(T(z)), \quad T(0) = 0, \quad (1)$$

where $\Phi(t) = \sum_{n=0}^{\infty} \varphi_n t^n$. This equation follows easily from the fact that one can describe a tree as a root node with several subtrees from the same family attached to it (see for instance [3] or [4]).

Important special cases include $\Phi(t) = (1+t)^2$, which corresponds to binary increasing trees, $\Phi(t) = e^t$ (recursive trees), and $\Phi(t) = \frac{1}{1-t}$ (plane-oriented recursive trees). In all these cases, the total weight can simply be interpreted as the number of trees of given size within the family. Binary trees are essentially equivalent to binary search trees, which in turn serve as an analytic model for the famous Quicksort algorithm [8]. Plane-oriented recursive trees, on the other hand, are a special instance of the well known Barabási-Albert model [2] for scale-free networks (see also [5]), which is used as a simplified growth model of the world wide web [1].

From a combinatorial point of view, it is interesting to note that full binary increasing trees ($\Phi(t) = 1+t^2$) are enumerated by the tangent numbers (see [10] for various interesting bijections), while there are $(n-1)!$ recursive trees, $n!$ binary increasing trees and $(2n-3)!!$ plane-oriented recursive trees with n nodes.

A specific subclass of increasing trees is known as *very simple families* [11] of increasing trees. The three aforementioned examples all belong to this subclass, which is essentially characterized by the fact that the function $\Phi(t)$ is either of the form $(1+ct)^\alpha$ for constants c, α of the same sign ($\alpha < 0$ or $\alpha \in \{2, 3, \dots\}$) or of the form e^{ct} for some positive constant c . These specific families have the property that they can be described via a tree evolution process, as pointed out by Panholzer and Prodinger in [11].

A remarkable result by Dobrow and Smythe [7] states that the depth of the k th node (i.e., the distance from the root) in a random increasing tree from one of the very simple families follows the exact same distribution as the number of edges between two nodes whose labels are $\leq k$ and differ by exactly 1 (henceforth, we will simply call such edges *1-edges*). See also [11]. The aim of this note is to show that this holds more generally for simple families of increasing trees, and to present a simple bijective proof of this fact. Several further corollaries follow as well, and the bijection can also be generalized, see Section 3.

Finally, we present another similar bijection and use it to determine the distribution of the depth of the lowest common ancestor of two nodes i and j , $i < j$. It turns out (somewhat surprisingly) that the distribution only depends on i , and that it converges to a geometric distribution if $i, j \rightarrow \infty$.

2 The bijection

Let us now describe a bijection B_k on the set of ordered increasing trees as follows:

- If node $j - 1$ lies on the unique path from 1 to k in T and ℓ is its successor on this path, then j takes the position of ℓ in $B_k(T)$ (i.e., if ℓ is the r th child of $j - 1$ in T , then j is the r th child of $j - 1$ in $B_k(T)$).
- If $j \leq k$ but node $j - 1$ does not lie on this path, then the parent of j in $B_k(T)$ is the same as the parent of $j - 1$ in T , and the position as a child is the same as well (as before).
- If $j > k$, then the parents (and positions) of j in T and $B_k(T)$ are the same.

The inverse operation B_k^{-1} is equally simple:

- If $j < k$ and nodes j and $j + 1$ are connected in T , then j lies on the path from 1 to k in $B_k^{-1}(T)$, and the successor ℓ of j on this path takes the position of $j + 1$ (i.e., if $j + 1$ is the r th child of j in T , then ℓ is the r th child of j in $B_k^{-1}(T)$).
- If $j < k$ but nodes j and $j + 1$ are not connected, then the parent of j in $B_k^{-1}(T)$ is the same as the parent of $j + 1$ in T , and the position as a child is the same as well.
- If $j > k$, then the parents (and positions) of j in T and $B_k^{-1}(T)$ are the same.

It is easy to see that both operations are well-defined and inverses of each other. Figure 1 shows an example with $k = 9$.

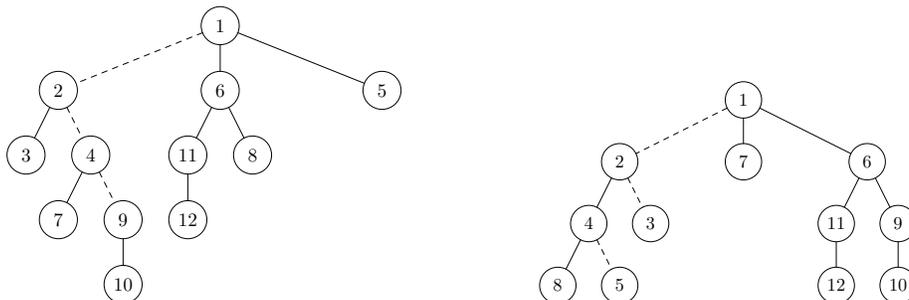


Figure 1: The bijection in an example: T (left) and $B_9(T)$ (right).

The following properties of the bijection are immediate:

- For any increasing tree T with $n \geq k$ nodes, $B_k(T)$ is an increasing tree with n nodes and the same outdegrees.
- Edges on the path between the root and k are mapped to 1-edges in $B_k(T)$ whose ends are labeled with numbers $\leq k$.

Since all outdegrees remain the same, the weights $w(T)$ and $w(B_k(T))$ are also always the same, regardless of the degree-weight sequence φ . The following results are obtained as a consequence. For very simple families of increasing trees, these theorems occur in the aforementioned paper by Dobrow and Smythe [7]. Our bijection provides a simple combinatorial explanation for these results, which were obtained by probabilistic techniques in [7], with the additional benefit that they generalize to a wider range of increasing trees, namely to all simple families.

Theorem 1 (cf. Dobrow/Smythe, Theorem 5)

In a random increasing tree with n nodes from a simple family, the probability that k is attached to j is exactly the probability that the last 1-edge with labels $\leq k$ is between j and $j + 1$.

More generally, the following holds:

Theorem 2 *In a random increasing tree with n nodes from a simple family, the probability that the ancestors of k are j_1, j_2, \dots, j_s in this order ($j_1 > j_2 > \dots > j_s$) is the same as the probability that the only 1-edges with labels between j_s and k are $j_1 - (j_1 + 1), j_2 - (j_2 + 1), \dots, j_s - (j_s + 1)$.*

Theorem 3 (cf. Dobrow/Smythe, Theorem 7)

In a random increasing tree with n nodes from a simple family, the distribution of the depth of node k is the same as the distribution of the number of 1-edges with labels $\leq k$. Furthermore, the probability that node j lies on the unique path between 1 and k is the same as the probability that there is an edge between j and $j + 1$.

In particular, one has the following corollary:

Corollary 4 *The probability that j lies on the path between 1 and k does not depend on k .*

Remark 1 None of the above theorems depends on the size of the increasing tree. In the case of very simple families, which can be generated by a growth process, this is essentially trivial, but it is quite surprising that this remains true within the more general setting of simple families of increasing trees.

3 Generalization

Our bijection can be generalized further to prove the following:

Theorem 5 (cf. Dobrow/Smythe, Theorem 6)

In a random increasing tree with n nodes from a simple family, the distribution of the distance between nodes i and k ($i < k$) is the same as the distribution of the sum of the distance between i and $i + 1$ and the number of 1-edges with labels between $i + 1$ and k .

To this end, consider a bijection $B_{i,k}$ that is defined as follows:

- If $i + 1 < j$, node $j - 1$ lies on the unique path from 1 to k in T and ℓ is its successor on this path, then j takes the position of ℓ in $B_{i,k}(T)$ (i.e, if ℓ is the r th child of $j - 1$ in T , then j is the r th child of $j - 1$ in $B_{i,k}(T)$).
- If $i + 1 < j \leq k$ but node $j - 1$ does not lie on this path, then the parent of j in $B_{i,k}(T)$ is the parent of $j - 1$ in T , and the position as a child is the same as well.
- If $j \leq i$ or $j > k$, then the parents (and positions) of j in T and $B_{i,k}(T)$ are the same.
- Finally, we have to specify the parent of $i + 1$: let ℓ be the node in T that lies on the path between i and k and has the smallest label $> i$. Suppose further that ℓ is the r th child of node h in T . Then $i + 1$ is the r th child of node h in $B_{i,k}(T)$.

See Figure 2 for an example with $i = 4$ and $k = 12$. Note that the path between i and k is mapped to the path between i and $i + 1$ and a collection of 1-edges, thereby proving Theorem 5.

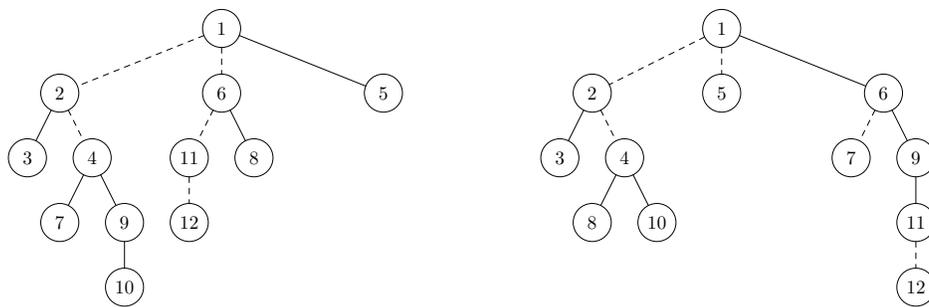


Figure 2: The generalized bijection in an example: T (left) and $B_{4,12}(T)$ (right).

4 Common ancestors

Note that the distance between two nodes i and j ($i < j$) equals the sum of their depths, minus the depth of their lowest common ancestor, which we will henceforth denote by $i \wedge j$. Hence it is natural to study the distribution of the depth of the lowest common ancestor. It turns out that this distribution has a discrete limit if we let $i, j \rightarrow \infty$ (a geometric limit distribution, to be precise), as opposed to the Gaussian limit that follows from the decomposition in Theorem 3 for very simple families. Perhaps more surprisingly, the distribution only depends on the label i , but not on j , which is shown by yet another similar bijection:

Theorem 6 *In a random increasing tree with n nodes from a simple family, the distribution of the depth of the lowest common ancestor of nodes i and j , $i < j$, is independent of j .*

Proof: Clearly it suffices to show that the distribution is the same for $i \wedge j$ and $i \wedge (j + 1)$. To this end, construct the following involution on the set of increasing trees: if j is the parent of $j + 1$, nothing changes. Otherwise, interchange j and $j + 1$. This is clearly possible without violating the condition that the labels along a path from the root increase. Furthermore, the collection of outdegrees and thus the weight of the tree remains the same. If j is the parent of $j + 1$, then they have the same common ancestors with i ; otherwise, the common ancestors of i and j become the common ancestors of i and $j + 1$, and vice versa. This proves the theorem and shows that it is sufficient to study the lowest common ancestor of two nodes with successive labels i and $i + 1$. ■

Let us study the distribution of the depth of $(n - 1) \wedge n$, i.e., the lowest common ancestor of the nodes $n - 1$ and n , in a simple increasing tree of size n ; for very simple trees that can be described by a growth process, this is also the distribution if the size of the tree is greater than n , since the lowest common ancestor cannot change if more nodes are added. We apply an approach via generating functions: let $T(z, u)$ be the bivariate generating function in which z marks the size of the tree, and u marks the depth of $(n - 1) \wedge n$ in a tree of size n . If $n - 1$ and n are in distinct branches, then the depth is 0, otherwise it is 1 plus the depth in the subtree that contains the two. Let $(\varphi_k)_{k \geq 0}$ and $\Phi(t)$ be a sequence and the associated power series, as explained in the introduction, and let $t_n(u)$ be the n th coefficient of $T(z, u)$ (which is—up to normalization—the probability generating function of the depth of $(n - 1) \wedge n$). Then we have, for $n > 2$,

$$\begin{aligned}
 t_n(u) &= \sum_{k \geq 1} \varphi_k \sum_{r_1+r_2+\dots+r_k=n-1} \binom{n-1}{r_1, r_2, \dots, r_k} t_{r_1}(1)t_{r_2}(1) \cdots t_{r_k}(1) \\
 &+ \sum_{k \geq 1} \varphi_k \sum_{r_1+r_2+\dots+r_k=n-3} \sum_{j=1}^k \binom{n-3}{r_1, r_2, \dots, r_k} \\
 &\quad \cdot t_{r_1}(1)t_{r_2}(1) \cdots (ut_{r_{j+2}}(u) - t_{r_{j+2}}(1)) \cdots t_{r_k}(1).
 \end{aligned}$$

The first summand accounts for the case that the depth is 0. In the second summand, we consider all possible trees with the property that nodes n and $n - 1$ are in the same (the j th) branch. It follows easily that $T(z, u) = \sum_{n \geq 1} t_n(u) \frac{z^n}{n!}$ satisfies

$$T'''(z, u) = T'''(z, 1) + (uT''(z, u) - T''(z, 1)) \cdot \Phi'(T(z, 1)),$$

where all derivatives are with respect to z . Since

$$T''(z, 1) = \frac{d}{dz} T'(z, 1) = \frac{d}{dz} \Phi(T(z, 1)) = T'(z, 1) \cdot \Phi'(T(z, 1)),$$

we can rewrite this as

$$T'''(z, u) = T'''(z, 1) + (uT''(z, u) - T''(z, 1)) \cdot \frac{T''(z, 1)}{T'(z, 1)}.$$

Solving the linear differential equation yields

$$T''(z, u) = T''(z, 1) + (u - 1)T'(z, 1)^u \int_0^z \frac{T''(y, 1)^2}{T'(y, 1)^{u+1}} dy.$$

With the additional conditions $T(0, u) = 0$ and $T'(0, u) = 1$ (which are not essential, though), $T(z, u)$ is uniquely determined. In general, there is no explicit expression for the integral, but for very simple families of increasing trees, the formula simplifies:

- If $\Phi(t) = e^{ct}$ for some $c > 0$, then $T(z, 1) = \frac{1}{c} \log \frac{1}{1-cz}$, and after some simplifications

$$T''(z, u) = \frac{c}{(2-u)(1-cz)^2} + \frac{c(1-u)}{2-u}(1-cz)^{-u}.$$

Now we can extract the coefficient of z^n to obtain the probability generating function

$$\frac{t_n(u)}{t_n(1)} = \frac{[z^n]T(z, u)}{[z^n]T(z, 1)} = \frac{1}{2-u} \left(1 - \binom{n-3+u}{n-1} \right).$$

The precise probabilities can now be expressed in terms of Stirling numbers of the second kind, and it is also obvious that this probability generating function converges to $\frac{1}{2-u}$ for all $u < 2$, which shows that the limit distribution for $n \rightarrow \infty$ is $\text{Geom}(\frac{1}{2})$. The average depth is precisely $\frac{n-2}{n-1}$.

- If $\Phi(t) = (1+ct)^\alpha$, then $T(z, 1) = \frac{1}{c} \left((1+c(1-\alpha)z)^{1/(1-\alpha)} - 1 \right)$, and we obtain

$$T''(z, u) = \frac{\alpha(1-\alpha)c}{1-2\alpha+\alpha u} (1+c(1-\alpha)z)^{1/(1-\alpha)-2} + \frac{\alpha^2 c(u-1)}{1-2\alpha+\alpha u} (1+c(1-\alpha)z)^{\alpha u/(1-\alpha)},$$

and the formula

$$\frac{t_n(u)}{t_n(1)} = \frac{1-\alpha}{1-2\alpha+\alpha u} + \frac{\alpha(u-1)}{1-2\alpha+\alpha u} \binom{\alpha u/(1-\alpha)}{n-2} / \binom{1/(1-\alpha)-2}{n-2}$$

for the probability generating function follows immediately. Again, one obtains a geometric limit distribution, since the probability generating function tends to $\frac{1-\alpha}{1-2\alpha+\alpha u}$ as $n \rightarrow \infty$; the average equals $\frac{\alpha(n-2)}{2-\alpha+(\alpha-1)n}$ in this case.

Let us combine these two examples into a theorem:

Theorem 7 *The limit distribution of the depth of $i \wedge j$, as $i, j \rightarrow \infty$, is $\text{Geom}(\frac{1}{2})$ for recursive trees, $\text{Geom}(\frac{1-\alpha}{1-2\alpha})$ for generalized plane oriented trees (i.e., $\Phi(t) = (1-t)^\alpha$, $\alpha < 0$), and $\text{Geom}(\frac{d-1}{2d-1})$ for d -ary increasing trees (i.e., $\Phi(t) = (1+t)^d$, $d = 2, 3, \dots$).*

Let us finish with a few remarks:

Remark 2 The last result can be easily generalized to the lowest common ancestor of several nodes $i_1 < i_2 < \dots < i_r$, and an analogous bijective argument shows that the distribution only depends on i_1 .

Remark 3 In the case of recursive trees, there is a simple relation to another combinatorial problem: the number of recursive trees with $n + 1$ nodes for which the depth of $n \wedge (n + 1)$ is $k - 1$ ($k \geq 1$) is also exactly the number of permutations of $1, 2, \dots, n$ for which n is an element of the k th cycle (where cycles are sorted in the canonical way, i.e., by their smallest elements), cf. [6, p.258]. This can be seen directly as follows: For a given recursive tree, let the nodes on the path from 1 to $n + 1$ be $1 = i_0, i_1, \dots, i_k = n + 1$. Then we can decompose the recursive tree into disjoint subtrees rooted at i_0, i_1, \dots, i_k . Since there is a bijection between recursive trees and permutations, we can map each of these subtrees (except for the last one, which only consists of the single node $n + 1$) to a cycle to obtain a permutation of n elements. This correspondence is clearly bijective. If n is in the k th cycle, then the depth of $n \wedge (n + 1)$ is $k - 1$, and vice versa.

Remark 4 Unfortunately it seems that, even though our bijections apply to the wider class of simple families of increasing trees, it remains difficult to obtain precise distribution results if the variety under consideration is none of the very simple families, cf. [9, 11]. The generating function approach that led to Theorem 7 applies to all simple families, but only if the lowest common ancestor of the two highest-labeled nodes is considered (which is sufficient for families that arise from a growth process).

References

- [1] R. Albert, H. Jeong, and A.-L. Barabási. The diameter of the world wide web. *Nature*, 401:130–131, 1999.
- [2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [3] F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP '92 (Rennes, 1992)*, volume 581 of *Lecture Notes in Comput. Sci.*, pages 24–48. Springer, Berlin, 1992.
- [4] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial species and tree-like structures*. Cambridge University Press, Cambridge, 1998.
- [5] B. Bollobás and O. M. Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34. Wiley-VCH, Weinheim, 2003.
- [6] L. Comtet. *Advanced combinatorics*. D. Reidel Publishing Co., Dordrecht, enlarged edition, 1974.
- [7] R. P. Dobrow and R. T. Smythe. Poisson approximations for functionals of random trees. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 79–92, 1996.
- [8] C. A. R. Hoare. Quicksort. *Comput. J.*, 5:10–15, 1962.
- [9] M. Kuba and A. Panholzer. On the distribution of distances between specified nodes in increasing trees. *Discrete Appl. Math.*, 158(5):489–506, 2010.

- [10] A. G. Kuznetsov, I. M. Pak, and A. E. Postnikov. Increasing trees and alternating permutations. *Uspekhi Mat. Nauk*, 49(6(300)):79–110, 1994.
- [11] A. Panholzer and H. Prodinger. Level of nodes in increasing trees revisited. *Random Structures Algorithms*, 31(2):203–226, 2007.