Counting Free Binary Trees Admitting a Given Height

Frank Harary
Computer Science Department
New Mexico State University
Las Cruces, NM 88003, USA

Edgar M. Palmer
Mathematics Department
Michigan State University
East Lansing, MI 48823, USA

Robert W. Robinson
Computer Science Department
University of Georgia
Athens, GA 30602, USA

Dedicated to the memory of R. C. Bose, the combinatorial and statistical pioneer.

Suggested running head: Counting Free Binary Trees

Author to whom proofs should be addressed:

Robert W. Robinson
Computer Science Department
415 GSRC
University of Georgia
Athens, GA 30602

Abstract

Recursive equations are derived for the exact number t_h of nonisomorphic free trees which have some rooting as a binary tree of height *h*. Numerical results are calculated using these formulae.

1. Introduction

A **binary tree** *T* can be defined as a rooted tree in which each node has degree at most 3, except that the root has degree at most 2. The **height** of *T* is the maximum distance from the root node to an endnode. Binary trees are much used in theoretical computer science, with height often being a key parameter directly related to the efficiency of associated algorithms. A **free binary tree** *F* is an unrooted tree which has a node *u* (not necessarily unique) such that *F* is a binary tree when rooted at *u*. Our purpose is to derive formulae for the number of unlabeled free binary trees which have a rooting that produces a binary tree of height *h*; we say that such a tree **admits height** *h*. In general our terminology follows [3]. Unlabeled counting does not distinguish between versions of a tree which differ only in the assignment of labels to the nodes.

A **3-tree** has maximum degree at most 3. It is convenient for our purpose of counting free binary trees by admissible height to consider 3-trees first. Obviously every free binary tree is a 3-tree, and conversely since any node of degree 1 or 2 could serve as the root. Figure 1 shows a free binary tree *F* which has four distinct binary rootings. Rooting *F* at node 5 or 6 gives one binary tree of height 5; at 7 gives height 4; at 3 gives height 3; finally, rooting *F* at 8 or 9 gives a second binary tree of height 5. Thus *F* admits height 3, 4, and 5. In the total of free binary trees of order a admitting height 5, for instance, *F* will be counted just once.

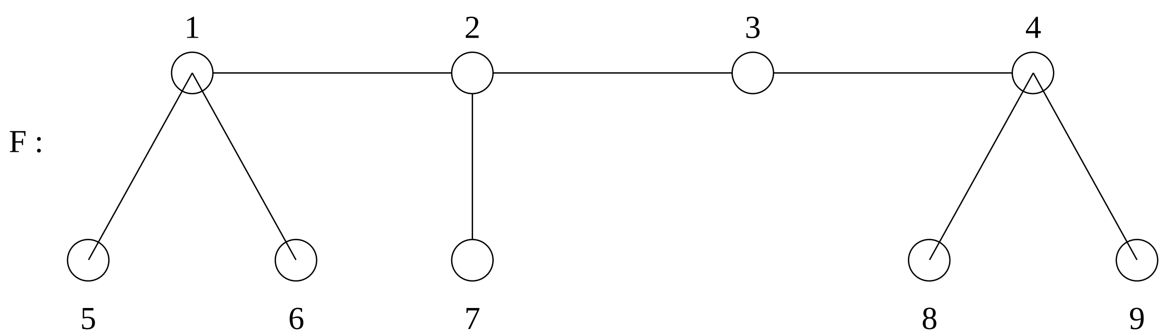

FIGURE 1. A free binary tree which has four binary rootings

Both rooted and unrooted 3-trees have been counted by Cayley and Otter; see [4] for a modern exposition.

2. Planted 3-trees of given height

In a **planted tree**, the root is an endnode. Let p_h be the number of planted 3-trees of height *h*, and let q_h be the number of height less than *h*, including for convenience the empty one with no nodes and no edges.

Then $p_1 = q_1 = 1$, while for all $h \geq 1$,

$$q_{h+1} = q_h + p_h \tag{1}$$

$$p_{h+1} = \binom{1+p_h}{2} + p_h q_h \tag{2}$$

Note that the numbers p_h were known to Etherington [2]; they are sequence number 718 in Sloane's book, [6].

To justify (2), we observe that a planted tree of height $h+1$ has two major subtrees, one of height *h* and the other of height *h* or less. For both to have height *h*, there are $\binom{1+p_h}{2}$ possibilities since we need to select two trees (which may be isomorphic) from among the p_h of height *h*, and their order is immaterial. For the case when one major subtree has height *h* and the other less, the possibilities are enumerated by $p_h q_h$ since the two branches cannot be confused with one another. The empty case admitted by $q_1 = 1$ corresponds to the possibility that the node adjacent to the root has degree 2, so that there is really only one major subtree.

In order to allow for the analysis of free 3-trees, it will be necessary to determine the number $d_{h,i}$ of planted 3-trees of height *h* which have no nodes of degree 1 or 2 at level *i* (distance *i* from the root). Of course all 3-trees of height *h* have one or more nodes of degree 1 at level h and no nodes at any level greater than h, so $d_{h,h} = 0$ and $d_{h,i} = p_h$ for all $i > h$. In fact, our interest will be in the number $(p_h - d_{h,i})$ of 3-trees of height h which do have a node of degree 1 or 2 at level i, for $1 \le i < h$. However the defining equations are more direct when written in terms of $d_{h,i}$. It will also be convenient to identify the quantity

$$e_{h,i} = 1 + \sum_{1 \le j < h} d_{j,i} \quad , \tag{3}$$

which bears the same relation to $d_{h,i}$ that q_h bears to p_h. One can then write the recursively defining equations as

$$d_{h+1,i+1} = \begin{pmatrix} 1+d_{h,i} \\ 2 \end{pmatrix} + d_{h,i}\, e_{h,i} \tag{4}$$

$$e_{h+1,i} = e_{h,i} + d_{h,i} \tag{5}$$

for $h > i \ge 1$. These parallel precisely equations (1) and (2). For boundary conditions we have

$$\begin{aligned} d_{h+1,1} &= p_{h+1} - p_h \quad , \\ e_{h+1,1} &= p_h \end{aligned} \tag{6}$$

for all $h \ge 1$. This is because if a planted tree of height $h + 1$ has a node of degree 1 or 2 adjacent to the root, that node must have degree 2 since $h \ge 1$. By suppressing this node, one obtains a tree of height h in a 1-1 fashion, so that

$$p_{n+1} - d_{h+1,1} = p_h \quad .$$

Now

$$e_{h+1,1} = 1 + \sum_{1 \le k \le h} d_{k,1} = 1 + d_{1,1} + \sum_{2 \le k \le h} (p_k - p_{k-1})$$

$$= 1 + d_{1,1} + p_h - p_1$$

$$= p_h$$

since $p_1 = 1$ and $d_{1,1} = 0$.

3. Free 3-trees by admissible height

It does not appear possible to apply the principle of Otter's dissimilarity characteristic [4, p.56] to obtain the number t_h of free 3-trees which have some rooting as a binary tree of height *h*. Instead, we will make use of the fact that every tree has a unique center consisting of a single node or two adjacent nodes. The possibilities for binary rootings of various heights are enumerated separately for these two cases. This approach was used by Cayley [1] when he first counted trees.

Case 1 The center is a single node.

Assuming a nontrivial tree *T*, the diameter is $2h$ for some $h \ge 1$. Then some two branches at the center must have height *h* and the third branch (if there is one) must have height at most *h*. The number of ways to choose these branches is

$$a_h = \binom{2+p_h}{3} + \binom{1+p_h}{2} q_h \ . \qquad (7)$$

The first term counts the number of ways to choose all three branches to have height *h*, and the second gives the number with two branches of height *h* and either no third branch or else a third branch having some height *k*, $1 \le k < h$.

Suppose now that one of the branches at the center of T has a node of degree 1 or 2 at level i, $i \geq 1$. Then T would have height $h+i$ if rooted at such a node, since any path of maximum length must pass through the center. The number of ways that T could fail to contain such a node is exactly

$$\binom{2+d_{h,i}}{3} + \binom{1+d_{h,i}}{2} e_{h,i} \quad . \tag{8}$$

This is just as for (7) except that every branch must fail to have a node of degree 1 or 2 at level i. Subtracting (8) from (7) will then give the number of 3-trees of diameter $2h$ which have a binary rooting of height $h+i$, $1 \leq i \leq h$.

There remains the possibility of rooting at the central node. The center has degree at most 2 exactly when there are just two branches. In that case the tree has height h when rooted at the center, so we have exactly

$$\binom{1+p_h}{2} \tag{9}$$

3-trees of diameter $2h$ which have a binary rooting of height h.

Case 2 The center consists of two adjacent nodes.

The diameter is $2h-1$ for some $h \geq 1$, and we can obtain any such tree in a unique fashion by joining two trees of height h at the root, then smoothing out the root node. We refer to these two trees as the branches at the bicenter. Of course their order is unimportant, and they may be isomorphic. Hence there are exactly

$$b_h = \binom{1+p_h}{2} \tag{10}$$

3-trees of diameter $2h-1$.

In this case a node of level *i* on one of the branches at the bicenter gives a rooting of height $h+i-1$. The number of 3-trees of diameter $2h-1$ having no node of level *i* of degree 1 or 2 on either branch at the center is just

$$\binom{1+d_{h,i}}{2} \ . \tag{11}$$

Subtracting (11) from (10) then gives the number of 3-trees of diameter $2h-1$ which have a binary rooting of height $h+i-1$.

The total number t_h of free 3-trees with a binary rooting is just the sum of the numbers obtained in Cases 1 and 2, for the appropriate values of *h* and *i*. More explicitly, for $h \geq 1$ we have

$$t_h = \binom{1+p_h}{2} + \sum_{i=1}^{\lfloor h/2 \rfloor} \left\{ a_{h-i} - \binom{2+d_{h-i,i}}{3} - \binom{1+d_{h-i,i}}{2} e_{h-i,i} \right\}$$
$$+ \sum_{i=1}^{\lfloor (h+1)/2 \rfloor} \left\{ b_{h-i+1} - \binom{1+d_{h-i+1,i}}{2} \right\} . \tag{12}$$

4. Numerical results.

Table I lists p_h for $h \leq 11$. Equations (1) and (2) enable us to calculate the sequence $p_1, p_2, \ldots, p_n$ in *O*(n) time.

Table II gives the values of t_h for $h \leq 10$. Note that $p_{h+1} \geq t_h$. This is because any tree with a binary rooting of height *h* corresponds to a planted 3-tree of height $h+1$. This correspondence is obtained by adding a new root of degree one adjacent to the original root node. In general there are trees with more than one binary rooting of height h, so that equality does not hold. (An example is provided by the tree F of Figure 1, which has two different binary rootings of height 5.) However, it is apparent that $p_{h+1} - t_h$ is small compared to t_h as h increases, so that multiple rootings of the same height are relatively rare.

TABLE I *The number of planted 3-trees by height*

h	p_h
1	1
2	2
3	7
4	56
5	2212
6	2595782
7	3374959180831
8	56951835044892390674843871
9	1621755757492238630142053127707136510316873428428 2
10	1315045868479612356871818745780631171143294098975989709465167937766 2208052979598672586922495727505 81
11	8646728181026489602610406537158318670928372786737024641130379069 39 4221138489756289944296330853108308241821596669137971686949329478 33 6661530334430058051973336177293923772027610801794840747988177012

In general, the method employed enables one to compute the values $t_1, t_2, \ldots, t_n$ with $O(n^2)$ integer arithmetic operations and storage of $O(n)$ integers. This analysis of complexity takes no account of the rapid increase in the size of the numbers involved. It is clear that $\log t_n = O(n^2)$, so this has a significant effect.

First, (1) and (2) are applied to compute p_h and q_h for $h \le n$. Simultaneously (7) and (10) are applied to determine a_h and b_h for $h \le n$, and these values are stored. At the same time, (5) and (6) are used to find $d_{h,1}$ and $e_{h,1}$ for $h \le n$, and these too are stored. The calculation proceeds by induction on i, $i = 1, \cdots, \lfloor (n+1)/2 \rfloor$. As the numbers $d_{h,i}$ and $e_{h,i}$ are computed and stored, their contributions to $t_1, \ldots, t_n$ as given in (12) are accummulated. First $d_{h,i+1}$ for $h \le n$ is given by (3), and then $e_{h,i+1}$ for

$h \le n$ is determined from (4).

By computing the values of $d_{h,i}$ in descending order of h, one can overwrite the $d_{h,i}$ array by the $d_{h,i+1}$. Using (4) one calculates the $e_{h,i+1}$ in ascending order, but the $e_{h,i}$ are not needed and so can be overwritten too. In order to avoid separately storing the values $e_{i+1,i}$ needed to start with (4), note that for $i \ge 2$ we have

$$e_{i+1,i} = e_{i,i-1} + p_{i-1},$$

and

$$p_{i-1} = d_{i,i-1}.$$

Now $d_{i,i-1}$ should still be available due to the fact that $d_{h,i}$ only needed computing for $h > i$. This is because $d_{i,i} = 0$ (so can be handled separately) and $d_{h,i}$ for $h < i$ is not called for in (12). For the same reasons $e_{i,i-1}$ should also still be available. Finally, the trees counted by $d_{i,i-1}$ can be obtained in a 1-1 fashion from those of height $i - 1$ by joining two new endnodes to each old endnode. Each new tree then has height i but has only nodes of degree 3 at level $i - 1$. Hence $p_{i-1} = d_{i,i-1}$ as claimed above.

TABLE II *The number of free binary trees by height*

h	t_h
1	2
2	7
3	52
4	2133
5	2590407
6	3374951541062
7	5695183504479116640376509
8	16217557574922386301420514191523784895639577710480
9	131504586847961235687181874578063117114329409897550318273792033024 340388219235081096658023517076950
10	864672818102648960261040653715831867092837278673702464113037906939 422113848975628994429633085310791372806105278543091014135638261111 3325681250718311629163466222152852597067554256522520919973090955

References

1. A. CAYLEY, On the analytical forms called trees, with applications to the theory of chemical combinations, *Rep. Brit. Assoc. Advance. Sci.* **45** (1875), 257-305 = *Math. Papers,* Vol. 9, 427-460.
2. I. M. H. ETHERINGTON, On non-associative combinations, *Proc. Roy. Soc. Edinburgh* **59** (1938/39), 153-162.
3. F. HARARY, "Graph Theory," Addison-Wesley, Reading, 1969.
4. F. HARARY and E. M. PALMER, "Graphical Enumeration," Academic, New York, 1973.
5. R. OTTER, The number of trees, *Ann. of Math.* **49** (1948), 583-599.
6. N. J. A. SLOANE, "A Handbook of Integer Sequences" Academic, New York, 1973.