Binary trees having a given number of nodes with 0, 1, and 2 children.

Günter Rote*

13. August 1997

Zusammenfassung

We give three combinatorial proofs for the number of binary trees having a given number of nodes with 0, 1, and 2 children. We extend these results to ordered trees with given distribution of nodes according to their numbers of children.

1 Introduction

A *binary tree* is either empty, or it consists of a *root node* and two binary trees, called the left subtree and the right subtree. A nonempty binary tree can also be regarded as a rooted directed tree graph (arborescence) where the outgoing arcs of each node are labeled by distinct labels from the set $\{\mathrm{L}, \mathrm{R}\}$. (We take the arcs as directed away from the root.)

Theorem 1 *The number of binary trees having i nodes with 2 children, j nodes with 1 child, and $k = i + 1$ nodes without children, is*

$$2^j \binom{2i+j}{j} b_i = \binom{n}{i\ j\ k} \frac{2^j}{n},$$

where $b_i = \binom{2i}{i-1}/i$ is the i-th Catalan number and $n = i + j + k$ is the total number of nodes.

Note that k must always equal $i + 1$ in a binary tree.

Prodinger [P] recently computed the probability that a random binary tree with n nodes has i nodes with 2 children (and hence $i + 1$ nodes without children and $n - 2i - 1$ nodes with 1 child). Since the total number of binary trees with n nodes is known—it is b_n—, his formulas can be derived easily from the above theorem and vice versa. Prodinger proved his results by simplifying sums of expressions involving binomial coefficients that were derived by Mahmoud [M].

*Institut für Mathematik, Technische Universität Graz, Steyrergasse 30, A-8010 Graz, Austria. Electronic mail: `rote@opt.math.tu-graz.ac.at`.
World-Wide Web: `http://www.tu-graz.ac.at/rote`.

This paper contains three independent combinatorial proofs of Theorem 1. The first proof is very simple; it it based on the fact that nodes with one child don't really add very much to the combinatorial structure of binary trees and can be easily eliminated, thus reducing the problem to the case where $j = 0$, for which the answer is known to be the Catalan numbers. The second proof relates the trees to be counted to labeled trees with given vertex degrees; the third proof relates them to integer sequences with certain properties. These sequences can be counted very easily by observing that, in each equivalence class under cyclic rotations, there is precisely one sequence that has the required properties (Rotation Principle).

The second and third proofs can be generalized to ordered trees with arbitrary node degrees.

2 The first proof — binary trees

Consider one of the binary trees B which are counted in the theorem. A vertex with 1 child in a binary tree can be eliminated by "short-cutting" it, connecting its child directly with its parent. To avoid problems when trying to eliminate the root node in this way, let us attach to B an artificial new root, whose only child is the original root node, thus getting a *planted* binary tree. Now, by eliminating all j nodes with one child (except the artifical root), we get a unique planted tree B' with the artificial root plus $2i + 1$ nodes, each having either 0 or 2 children. (Viewed as graphs, B and B' are homeomorphic.) Each edge of B' corresponds to a chain of one or more edges in B. The number of possible trees B' is b_i. To reconstruct B from its transformed tree B', we have to distribute j new nodes over the $2i + 1$ edges of B'. There are $\binom{j+2i}{j}$ ways to do this. The factor 2^j accounts for the fact that each new node can have either a left child or a right child.

3 The second proof — d-ary trees

The second proof establishes a correspondences with *labeled trees* with n vertices of given degrees $d_1, \ldots, d_n$, whose number is the multinomial coefficient

$$\binom{n-2}{d_1 - 1 \quad d_2 - 1 \quad \cdots \quad d_n - 1}, \tag{1}$$

see for example Moon [M].

This proof works more generally for d-ary trees, and the resulting theorem generalizes Theorem 1. A nonempty *d-ary tree* is a rooted directed tree where the (at most d) outgoing arcs of each node are labeled by distinct labels from the set $\{1, \ldots, d\}$.

Theorem 2 *The number of d-ary trees with n nodes, having n_i nodes with i children, for $i = 0, 1, \ldots, d$, is*

$$\frac{1}{n} \cdot \binom{n}{n_0 \; n_1 \; \cdots \; n_d} \cdot \binom{d}{0}^{n_0} \binom{d}{1}^{n_1} \binom{d}{2}^{n_2} \cdots \binom{d}{d}^{n_d},$$

if $\sum_{i=0}^{d} n_i i = n - 1$. If this equation does not hold, there are no such trees.

Proof. We augment the d-ary tree to a planted tree as in the previous section by adding an artificial root, and then we *label* the augmented tree B by giving the special label 0 to the artificial root, and the labels $(i, 1), \ldots, (i, n_i)$ to the nodes with i children, in arbitrary order. This procedure gives rise to a *labeled d-ary tree.* The number of labeled d-ary trees obtained in this way is $n_0!n_1!\ldots n_d!$ times the number of trees that we want to determine. On the other hand, a labeled d-ary tree can be constructed from a rooted labeled tree T (viewed as a graph), where $n_0 + 1$ vertices have degree 1 (including the artificial root), n_1 vertices have degree 2, and n_2 vertices have degree 3, etc. They have a total of $n + 1$ nodes, and by (1), their number is

$$\frac{(n-1)!}{0!^{n_0+1}1!^{n_1}2!^{n_2}\ldots d!^{n_d}}.$$

For each such tree, we have to select labels for the arcs going from each node to its children. This can be done in

$$1^{n_0} \cdot d^{n_1} \cdot \Big(d(d-1)\Big)^{n_2} \cdots (d!)^{n_d} = \frac{d!^n}{d!^{n_0}(d-1)!^{n_1}(d-2)!^{n_2}\ldots 0!^{n_d}}$$

ways. Putting everything together, we get the theorem. ∎

Corollary 1 *The number of d-ary trees with n nodes is*

$$\binom{dn}{n-1}\frac{1}{n}.$$

This result was obtained by many authors independently in different forms. For example, Harary, Palmer, and Read [HPR] considered partitions of a convex polygon into $(d+1)$-gons without introducing additional vertices. Beineke and Pippert [BP1, BP2] considered partitions of a d-ball into d-simplices. All these proofs use generating functions.

4 The third proof — ordered trees

An *ordered tree* is a rooted tree where the arcs from each node to its children are numbered consecutively starting with 1. So the first child is distinguished from the second child and so on.

Theorem 3 *The number of ordered trees with n nodes, having n_i nodes with i children, for $i = 0, 1, \ldots,$ is*

$$\frac{1}{n}\binom{n}{n_0\ n_1\ n_2\ \cdots},$$

if $\sum_{i\geq 0} n_i(i-1) = -1$. If this equation does not hold, there are no such trees.

This theorem is equivalent to Theorem 2, since in a d-ary tree the arcs from a node to its i children can be labeled by any subset of i numbers from the set $\{1, \ldots, d\}$, not just by the labels $\{1, \ldots, i\}$. Therefore, to get the number of d-ary trees, the number of ordered trees must simply be multiplied by if $\prod_{i=0}^{d}\binom{d}{i}^{n_i}$ to account for the additional choices of labels.

Tutte [T] obtained this result by using generating functions, see also Goulden and Jackson [GJ, 2.7.7, pp. 112–113].

However, we give another completely independent and elementary proof.

Proof of Theorem 3. We construct a bijection between ordered trees and certain sequences of integers. A *preorder traversal* of an ordered tree starts at the root, and then successively traverses the first subtree, the second subtree, etc., in preorder. If we note the number of children of each node in preorder, we get a sequence $(c_1, \dots, c_d)$ containing n_i elements equal to i, for $i = 0, 1, \dots$. From this sequence, the ordered tree can be uniquely reconstructed: We start at the root and generate its c_1 children. Then we go to the first child and add c_2 children to it, and so on. When we come to a node with $c_k = 0$ children, we backtrack to the next node for which the number of children has not been determined, and generate c_{k+1} children for it.

However, not every sequence corresponds to a tree. Consider a sequence starting with $2, 0, 1, 0, 0, 4, 0, \dots$. After the fourth node has been finished with $c_4 = 0$ children, there is no fifth node whose number of children is to be determined as $c_5 = 0$. One can characterize this situation by observing the number t of *"unfinished" nodes* as the construction proceeds. Initially we have $t_0 = 1$ unfinished node, the root. In the k-th step we finish one node by determining its number of children, but we generate c_k new unfinished nodes. Thus, $t_k = t_{k-1} - 1 + c_k$. The construction can go on as long as there is at least one unfinished node, i.e., $t_k \geq 1$, except at the last step, where we have $t_n = 0$, by the condition on the numbers n_i. This is summarized in the following lemma, where we have for later convenience substituted the sequence c_k by $a_k := c_k - 1$ and replaced t_k by $s_k := t_k - 1$.

Lemma 1 *The number of trees in Theorem 3 equals the number of sequences $a_1, \dots, a_n$ containing n_i elements equal to $i - 1$, for $i = 0, 1, \dots$, such that the partial sums $s_k = a_1 + \cdots + a_k$ are nonnegative, for $k = 1, \dots, n - 1$.* ■

The total number of sequences of n elements with n_i elements equal to $i - 1$, for $i = 0, 1, \dots$, without regarding the nonnegativity condition, is $\binom{n}{n_0\ n_1\ n_2\ \cdots}$. This is by a factor n more than the number that we want. Thus, it appears that one out of n sequences should satisfy the nonnegativity condition of the partial sums. This is proved by the next lemma, which I call the Rotation Principle, in analogy to André's Reflection Principle [A], see [GJ, exercise 5.3.7, pp. 533–534], which is another principle that is useful for counting sequences subject to certain geometric restictions on the paths they generate. In contrast to this principle, which is applicable to ± 1-sequences (possibly with the inclusion of zeros), the following lemma does not have this restriction.

Lemma 2 (Rotation Principle) *Consider a sequence $a_1, \dots, a_n$ of integers with $a_1 + \cdots + a_n = -1$.*

1. *All n cyclic rotations $a_{j+1}, a_{j+2}, \dots, a_n, a_1, \dots, a_j$ are distinct.*
2. *Among the n cyclic rotations there is precisely one whose first $n - 1$ partial sums are nonnegative.*

Proof. Let us draw a polygonal path on a plane grid with a sequence of steps in the directions $(1, a_i)$, starting at the point $(0, 0)$. Figure 1 shows two consecutive cycles

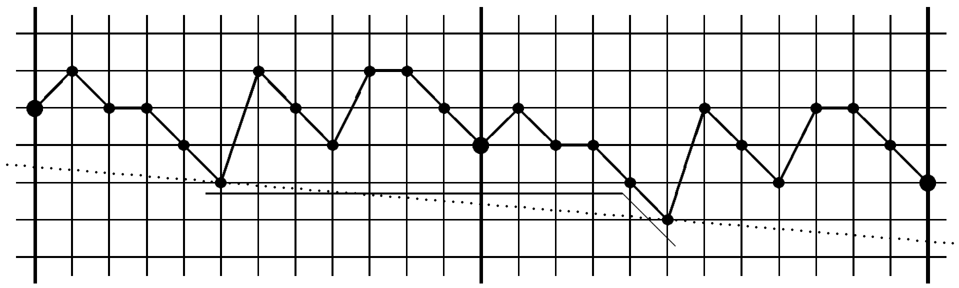

Abbildung 1: The path corresponding to two cycles of a sequence.

corresponding to the sequence $(1, -1, 0, -1, -1, 3, -1, -1, 2, 0, -1, -1)$. The sequence goes through the points (k, s_k) and terminates at the point $(n, -1)$. Each successive cycle is shifted by the vector $(n, -1)$. Now, the condition $s_k \geq 0$ for $1 \leq k < n$ is equivalent to $s_k \geq -k/n$ for all k, i.e., all points (k, s_k) must lie above the line of slope $-1/n$ through the origin. This condition holds not only for the n points obtained from the original sequence, but for the infinite set of points obtained by continuing the given sequence indefinitely in a cyclic fashion. For a cyclically shifted sequence starting at a_{j+1}, the condition is fulfilled if and only if all points lie above the line of slope $-1/n$ through the point (j, s_j). Thus we find the unique starting point by approaching the point set from below with a line of slope $-1/n$ until it touches the first point. This line contains only a single point (among the first n points), and therefore the rotated sequence with nonnegative partial sums is unique.

It we take this sequence as the starting sequence, then all its n cyclic rotations are distinct, since the amount by which they were rotated from the starting sequence can be recovered uniquely from each rotated sequence. ■

If the total sum of -1 is replaced by some other number, the first statement of the lemma remains true as long as that number is relatively prime to n. One may even take arbitrary integral vectors instead of just the vectors of the form $(1, a_i)$.

The idea to select a good starting point for going through a cyclic sequence is not new; it appears for example as the solution to an algorithmic problem of recreational mathematics, see Dewdney [D].

With the help of the two previous lemmas, the proof of the theorem can now be completed easily. If we regard sequences that can be obtained from each other by cyclic rotations as equivalent, then, by Lemma 2.1, each equivalence class contains n sequences, and by Lemma 2.2, each equivalence class contains precisely one sequence which corresponds to a tree. Therefore the total number of sequences must be divided by n, and we get the theorem. ■

One can also follow the argument of Theorem 2 in the reverse direction, and one gets then from Theorem 3 an independent proof for the number (1) of labeled trees with given degrees.

Corollary 2 [?] *The number of ordered trees with n nodes is the $(n-1)$-st Catalan number $b_{n-1} = \binom{2n-2}{n-1}/n$.*

Literatur

[A] D. André. Solution directe du problème résolu par M. Bertrand. *C. R. Acad. Sci. Paris* **105** (1887), 436–437.

[BP1] L. W. Beineke and R. R. Pippert. Enumerating labeled k-dimensional trees and ball dissections. in: *Proc. 2nd Chapel Hill Conf. on Combinatorial Mathematics and its Applications*, Univ. North Carolina, Chapel Hill, 1970, pp. 12–26.

[BP2] L. W. Beineke and R. R. Pippert. The number of labeled dissections of a k-ball. *Math. Ann.* **191** (1971), 87–98.

[D] A. K. Dewdney. Computer Recreations. *Scientific American* **256**, 6 (June 1987), 106–109. Solution in **257**, 5 (November 1987), p. 122. German translation: Computer-Kurzweil. *Spektrum der Wissenschaft*, September 1987, 6–10; February 1988, p. 13.

[GJ] I. P. Goulden and D. M. Jackson. *Combinatorial Enumeration.* Wiley-Interscience, New York 1983.

[HPR] F. Harary, E. M. Palmer, and R. C. Read. On the cell-growth problem for arbitrary polygons. *Discr. Math.* **11** (1975), 371–389.

[M] H. M. Mahmoud The joint distribution of the three types of nodes in uniform binary trees. *Algorithmica* **13** (1995), 313–323.

[M] J. W. Moon. *Counting Labelled Trees.* Canadian Mathematical Monographs, no. 1. Canadian Mathematical Congress, London and Beccles, 1970.

[P] H. Prodinger. A note on the distribution of the three types of nodes in uniform binary trees. *Séminaire Lotharingien de Combinatoire* **38b**, (1996), 5 pages. `http://cartan.u-strasbg.fr/~slc/wpapers/s38proding.html`

[T] W. T. Tutte. The number of planted plane trees with a given partition. *Amer. Math. Monthly* **71** (1964), 272–277.