Discrete Mathematics 31 (1980) 9–28

ENUMERATIONS OF ORDERED TREES*

Nachum DERSHOWITZ and Shmuel ZAKS**

Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received 12 June 1979
Revised January 1980

We deal with the class T_n of ordered trees with n edges. Several enumeration problems concerning T_n and some of its combinatorial properties are studied.

Closed-form expressions for the following enumerations are given: (1) the number of trees in T_n with k leaves, (2) the number of nodes in T_n with d children, (3) the number of trees in T_n with root degree r, and (4) the number of nodes in T_n on level l with d children.

1. Introduction

Mathematical trees (as well as their natural counterparts) come in a variety of forms. There are rooted trees and unrooted ones; some rooted trees are ordered, others are not; some trees come with labels, others do not. And there are restricted classes of trees, e.g. binary, full binary, k-ary, complete k-ary, etc. In this paper we concentrate on unlabelled ordered trees with no restrictions on the degrees of the nodes and study several combinatorial properties of this class of trees.

For the class T_n of all ordered trees with n edges, the following enumeration functions are defined:

(1) $\mathcal{L}_n(k)$ = number of such trees with exactly k leaves.
(2) $\mathcal{D}_n(d)$ = total number of nodes of degree d in these trees.
(3) $\mathcal{R}_n(r)$ = number of such trees in which the root has degree r.
(4) $\mathcal{N}_n(l, d)$ = total number of nodes of degree d that reside on level l in these trees.

The next section presents the basic definitions and several one-to-one correspondences among ordered trees and other combinatorial objects that are used in subsequent sections. In Section 3 we investigate the functions $\mathcal{L}$, $\mathcal{D}$ and $\mathcal{R}$; we

* Research supported under NSF Grant MCS77-22830.
** Present address: Department of Computer Science, Technion, Haifa, Israel

show that

(1) $\mathscr{L}_n(k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}$,

(2) $\mathscr{D}_n(d) = \binom{2n-1-d}{n-1}$, and

(3) $\mathscr{R}_n(r) = \frac{r}{n}\binom{2n-1-r}{n-1}$.

We also derive other combinatorial and statistical properties; for example: for each tree in this class with k leaves there is a corresponding tree with $n+1-k$ leaves, the expected number of leaves is $\frac{1}{2}(n+1)$, and the expected root degree is $3n/(n+2)$. Sections 2 and 3 include two "reflection" lemmata that illuminate relationships between ordered trees and their corresponding binary-tree representations. For completeness we have included a discussion of the functions $\mathscr{L}$ and $\mathscr{R}$; although the results are not new, our proofs may be of interest.

Section 4 is devoted to the function $\mathscr{N}$; our main result is the closed-form expression

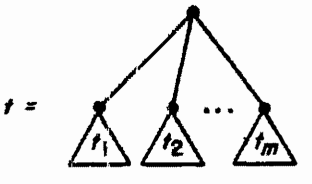
(4) $\mathscr{N}_n(l, d) = \frac{2l+d}{2n-d}\binom{2n-d}{n+l}$.

From it, the expressions for the functions $\mathscr{D}$ and $\mathscr{R}$—independently proved in Section 3—may be derived.

2. Preliminaries

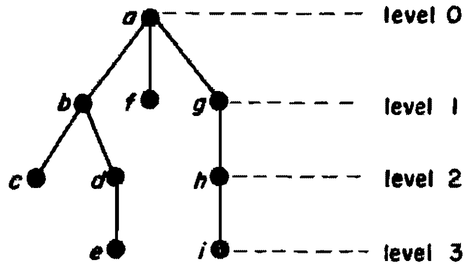
2.1. *Definitions*

The combinatorial structures that we shall be dealing with are (unlabelled) ordered trees. Our terminology is, in the main, borrowed from Knuth [3]. *Ordered trees* may be defined recursively as follows: if $t_1, t_2, \ldots, t_m$ are ordered trees, $m \geq 0$, then



is also an ordered tree. The trees $t_1, t_2, \ldots, t_m$ are *subtrees* of the distinguished node—called the *root* of t—connecting them. The roots of the subtrees are *children* of the root of the tree. The trees are ordered in the sense that the order among subtrees (or children) is significant.

With each node x in a tree t we associate two values: its degree and level. The *degree* of x is the number of children it has, and the *level* of x is its distance (the number of edges separating it) from the root of t. A node of degree 0 is termed a *leaf*, otherwise it is called an *internal* node. The root is the only node on level 0. (See Fig 1.)



leaves (nodes of degree 0): c, e, f, i

internal nodes $\begin{cases} \text{of degree 1}: & d, g, h \\ \text{of degree 2}: & b \\ \text{of degree 3}: & a \end{cases}$

root (level 0): a
level 1 : b, f, g
level 2 : c, d, h
level 3 : e, i

Fig. 1. An ordered tree with eight edges.

2.2. *Correspondences*

Let T_n $(n \geqslant 0)$ denote the set of ordered trees with n edges. As can be seen in Fig. 2, the sizes of these sets form a series $1, 1, 2, 5, 14, 42, \ldots$. These are the well-known Catalan numbers C_n:

$$|T_n| = C_n = \frac{1}{n+1}\binom{2n}{n}$$

(see, for example, Gardner [2]).

There are numerous one-to-one correspondences between elements of these sets of ordered trees and other combinatorial objects (see, for example, Kuchinski [4]). Among them, the correspondences between the following sets will help in our enumerations:

T_n: the set of ordered trees with n edges.

P_n: the set of *legal* sequences of n open and n close parentheses. A parenthetic expression is called "legal" if each open parenthesis has a matching close parenthesis.

I_n: The set of *dominating* sequences of $n+1$ nonnegative integers $(a_j)_0^n$, such that $\sum_{j=0}^{n} a_j = n$ and

$$\sum_{j=0}^{i} a_j \geqslant i \quad \text{for all } i,\ 0 \leqslant i \leqslant n.$$

L_n: the set of *admissible* paths from the point $(0, 0)$ to (n, n) in an $n \times n$ lattice.

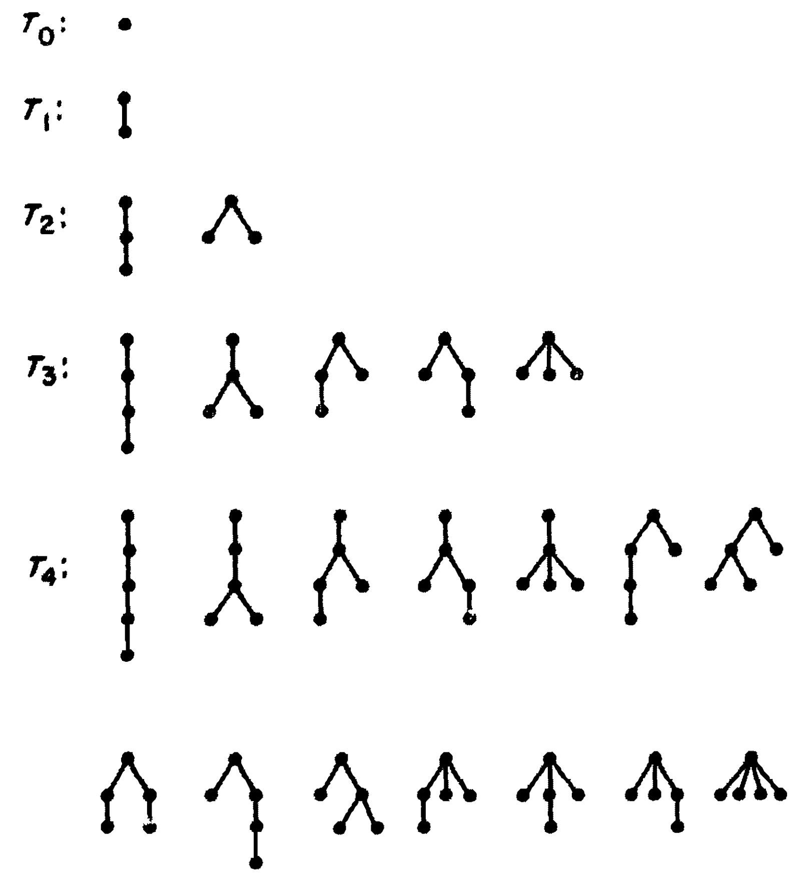


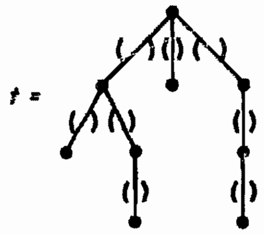
Fig. 2. T_n = {ordered trees with n edges}.

All steps in a lattice path are either up or to the right; a path is "admissible" if it does not pass below the diagonal $y = x$.

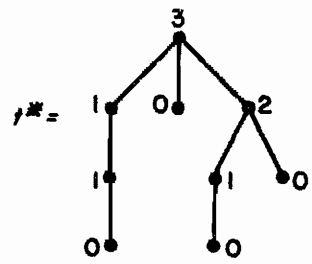
B_n: the set of *full* binary trees with n internal nodes. An ordered tree is "full binary" if all nodes are either of degree 0 (leaves) or 2 (have a *left* child and a *right* child).

The correspondences between these five sets are described below and illustrated in Fig. 3 using a tree $t \in T_8$ (Fig. 3.1) and its *reflection* $t^* \in T_8$ (Fig. 3.2). We shall alternate between the two trees t and t^* for a reason that will become apparent later.

$T_n \leftrightarrow P_n, I_n$: Given a tree $t \in T_n$, traverse it in *preorder* (visit the root, then traverse its subtrees from left to right), writing an open parenthesis for each edge passed on the way down and a close parenthesis for each edge passed on the way up. For the tree t in Fig. 3.1, this yields the legal parenthetic expression $p(t) \in P_8$ shown in Fig. 3.3. Similarly, if the degree of each node in t^* is recorded on the way down, then the dominating sequence $i(t^*) \in I_8$ shown in Fig. 3.4 is obtained. In general, if $t_1, t_2, \ldots, t_m$ are the subtrees of a tree t, then $p(t) = (p(t_1))(p(t_2)) \cdots (p(t_m))$ and $i(t) = m\ i(t_1)\ i(t_2) \cdots i(t_m)$.



3.1. An ordered tree t.



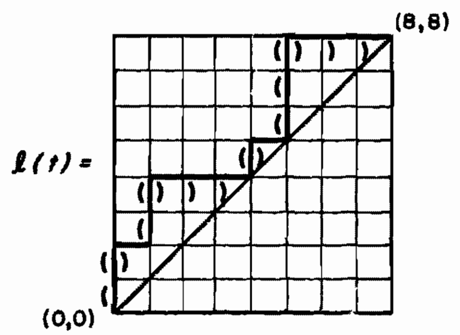
3.2. Its reflection t^*.

$$p(t) = (()(()))()((()))$$

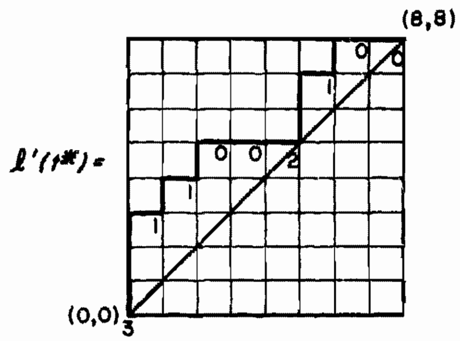
3.3. A legal parenthetic expression.

$$i(t^*) = 311002100$$

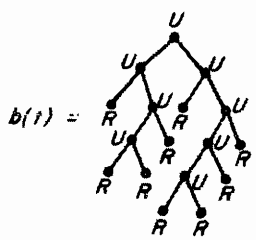
3.4. A dominating sequence of nonnegative integers.



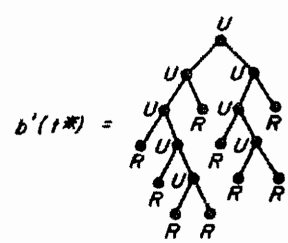
3.5. An admissible lattice path corresponding to $p(t)$.



3.6. An admissible lattice path corresponding to $i(t^*)$.



3.7. A full binary tree corresponding to $l(t)$.



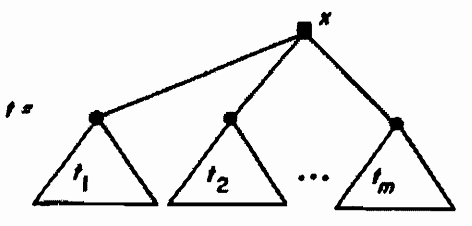
3.8. A full binary tree corresponding to $l'(t^*)$.

Fig. 3. Correspondences between T_n, P_n, I_n, L_n, and B_n.

$P_n, I_n \leftrightarrow L_n$: Lattice paths may be easily obtained from sequences of parentheses or integers. Given a legal parenthetic expression $p \in P_n$, we start at $(0, 0)$ and go up one coordinate for each open parenthesis and go right one coordinate for each close parenthesis. This yields a path $l(t)$ that remains in the upper left half of the lattice and ends at (n, n). Alternatively, given a dominating sequence $i \in I_n$, the function l' defines an admissible lattice path by letting each integer in i determine the number of coordinates to move up before moving right one coordinate. The paths $l(t)$ in Fig. 3.5 and $l'(t^*)$ in Fig. 3.6 may be obtained in such a manner from Figs. 3.3 and 3.4, respectively.

$L_n \leftrightarrow B_n$: Given an admissible lattice path $l \in L_n$ corresponding to an ordered tree $t \in T_n$, a unique full binary tree $b(t) \in B_n$ is constructed in the following manner: Build the binary tree in preorder, each step up on the path corresponding to an internal node and each step to the right corresponding to a leaf. A final leaf must be added. In the same manner, the function b' defines a full binary tree based on the path l'. See Figs. 3.7 and 3.8.

By this construction, if



then $b(t)$ is as shown in Diagram 1 (note that this is the same as the full binary tree obtainable from the binary tree representation given in Knuth [3, Section 2.3.2]), while $b'(t)$ is as shown in Diagram 2.

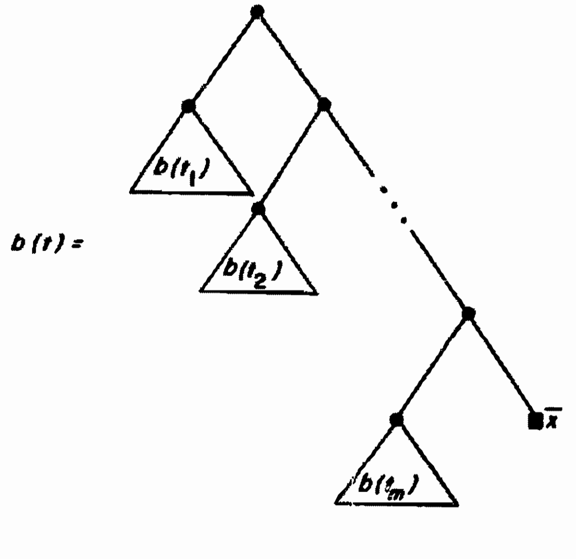


Diagram 1

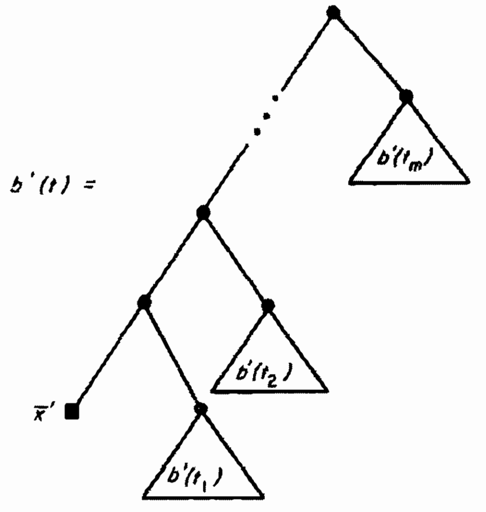


Diagram 2

It follows by induction that

First Reflection Lemma. *The ordered trees $t, t^* \in T_n$ are reflections of each other, if and only if the full binary trees $b(t), b'(t^*) \in B_n$ are reflections of each other.*

For example, the ordered tree t^* in Fig. 3.2 is the reflection of the tree t in Fig. 3.1, and the full binary tree $b'(t^*)$ in Fig. 3.8 is the reflection of the full binary tree $b(t)$ in Fig. 3.7.

Each of the above correspondences is one-to-one and can be used in either direction. Four properties follow directly from these correspondences, and are summarized in the following

Characterization Lemma

(1) The number of leaves in a (not edgeless) tree t

= the number of **()** patterns in $p(t)$

= the number of *corners* (i.e. path segments of the form) in $l(t)$

= the number of left leaves in $b(t)$

= the number of right leaves in $b'(t)$.

(2) The number of internal nodes in a (not edgeless) tree t

= one more than the number of **((** (or equivalently**))**) patterns in $p(t)$

= one more than the number of (or equivalently) path segments in $l(t)$

= the number of right leaves in $b(t)$

= the number of left leaves in $b'(t)$.

(3) The number of nodes of degree d in a tree t
= the number of occurrences of d in $i(t)$
= the number of vertical path segments of length exactly d in $l'(t)$.

(4) The number of nodes of degree d in all the trees in T_n
= the number of occurrences of d in all the sequences in I_n
= the number of occurrences of runs of exactly d('s (or equivalently d)'s) in all the expressions in P_n.

2.3. *The Cycle Lemma*

A sequence p of open and close parentheses is called a *legal prefix* if it is a prefix of a legal parenthetic expression, but neither p nor any initial segment of p is itself a legal parenthetic expression. In other words, a prefix consisting of m('s and n)'s, $m > n$, is legal if at any point within the prefix the number of ('s to the left is greater than the number of)'s. For example, **((()()** is a legal prefix; **)((()(** and **()()((** are not.

The following lemma has been rediscovered a number of times; it is a powerful tool in enumeration arguments.

Cycle Lemma (Dvoretzky and Motzkin [1]). *For any sequence $p_1p_2\cdots p_{m+n}$ of m open parentheses and n close parentheses, where $m > n$, there exist exactly $m-n$ cyclic permutations*

$$p_ip_{i+1}\cdots p_{m+n}p_1\cdots p_{i-1}$$

that are legal prefixes.

For example, of the six cyclic permutations of the sequence **)((()(**, only two are legal prefixes: **((()()** and **(()()(**.

We can use this lemma to determine the number $f(i,j)$ of admissible lattice paths from $(0,0)$ to (i,j). The obvious recurrence for the function f is:

$$f(i,j)=\begin{cases}1 & \text{if } i=0,\\ 0 & \text{if } j<i,\\ f(i,j-1)+f(i-1,j) & \text{otherwise.}\end{cases}$$

Each admissible path corresponds to a legal prefix of $j+1$ open parentheses and i close parentheses: just construct the corresponding parenthetic expression and prefix it with an extra open parenthesis. The total number of ways to arrange the $i+j+1$ parentheses on a line is $\binom{i+j+1}{i}$; by the Cycle Lemma, exactly $(j-i+1)/(j+i+1)$ of these arrangements are legal. Hence, the total number of legal prefixes is

$$f(i,j)=\frac{j-i+1}{j+i+1}\binom{j+i+1}{i}=\frac{j-i+1}{j+1}\binom{i+j}{i}=\binom{i+j}{i}-\binom{i+j}{i-1}.$$

If $i=j$, then we get the Catalan numbers $\binom{2i}{i}/(i+1)$ (See Fig. 4).

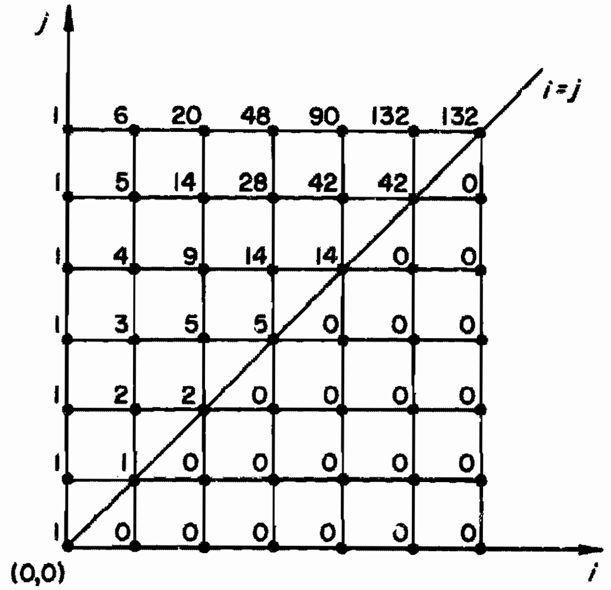


Fig. 4. The lattice function $f(i, j)$.

3. $\mathcal{L}_n(k)$, $\mathcal{D}_n(d)$, and $\mathcal{R}_n(r)$

3.1. *Introduction*

In this section we study the number $\mathcal{L}_n(k)$ of trees with n edges and exactly k leaves, the number $\mathcal{D}_n(d)$ of nodes of degree d among the trees with n edges, and the number $\mathcal{R}_n(r)$ of trees with n edges and root-degree r. Our main results are closed-form expressions for these numbers. Some consequences and additional statistical properties of T_n are discussed.

3.2. $\mathcal{L}_n(k)$

The following closed-form expression for $\mathcal{L}_n(k)$ was given by Narayana [5] in connection with partial orders on partitions:

Theorem 1. *The number $\mathcal{L}_n(k)$ of ordered trees with n edges and k leaves is*

$$\mathcal{L}_n(k)=\frac{1}{n}\binom{n}{k}\binom{n}{k-1}=\frac{1}{k}\binom{n-1}{k-1}\binom{n}{k-1}=\frac{1}{n+1}\binom{n-1}{k-1}\binom{n+1}{k}.$$

Examples. Of the 14 ordered trees with 4 edges, $\mathcal{L}_4(2)=6$ have 2 leaves and $\mathcal{L}_4(3)=6$ have 3 leaves. (See Fig. 2.)

We present here a new combinatorial proof of this result.

Proof. The number of trees in T_n with k leaves is equal to the number of legal parenthetic expressions in P_n with k occurrences of () (Characterization Lemma), which in turn equals the number of legal prefixes with $n+1$ ('s, n)'s, *and* k ()'s. By the Cycle Lemma, of the k possible cyclic arrangements of a parenthetic expression beginning with (, ending with), and having $n+1$ ('s, n)'s, and k ()'s, exactly one is a legal prefix. The total number of such expressions (legal or not) is $\binom{n}{k-1}\binom{n-1}{k-1}$ (the number of ways to partition both the ('s and the)'s into k nonempty runs); thus

$$\mathscr{L}_n(k)=\frac{1}{k}\binom{n}{k-1}\binom{n-1}{k-1}. \qquad \square$$

An alternative lattice-path proof of this theorem is given in the Appendix.

3.3. *Second Reflection Lemma*

From Theorem 1 it follows that

$$\mathscr{L}_n(k)=\frac{1}{n}\binom{n}{k}\binom{n}{k-1}=\mathscr{L}_n(n+1-k),$$

which proves

Theorem 2. *The number of trees in T_n with k leaves is equal to the number of trees in T_n with $n+1-k$ leaves.*

Corollary 2.1. *The expected number of leaves in a tree in T_n is $\frac{1}{2}(n+1)$; the expected number of internal nodes is also $\frac{1}{2}(n+1)$.*

Since the sum of the degrees of the internal nodes in T_n is equal to the number of edges $(n/(n+1))\binom{2n}{n}$ and the number of internal nodes is $\frac{1}{2}\binom{2n}{n}$, we have

Corollary 2.2. *The expected degree of an internal node in a tree in T_n is $2n/(n+1)$.*

We present two additional proofs of Theorem 2; they are both direct and constructive. The two proofs are then related to each other by another reflection lemma.

Second proof. We define a one-to-one correspondence between trees with k leaves and trees with $n+1-k$ leaves in three steps: Given a tree $t\in T_n$ with k leaves and $n+1-k$ internal nodes, the corresponding full binary tree $b(t)\in B_n$ has k left leaves and $n+1-k$ right leaves (by the Characterization Lemma). Thus, the reflection of $b(t)$ has $n+1-k$ left leaves, corresponding to a tree $t'\in T_n$ with exactly $n+1-k$ leaves. $\square$

Third proof. Consider the recursively defined one-to-one correspondence within

T_n, as shown in Diagram 3, between trees with k internal nodes (i.e. $n+1-k$ leaves) and trees with k leaves where $t'_1, t'_2, \ldots, t'_m$ are the trees corresponding to the m subtrees of t. Every internal node x in the tree t corresponds to a leaf x' in the tree t' and vice-versa. □

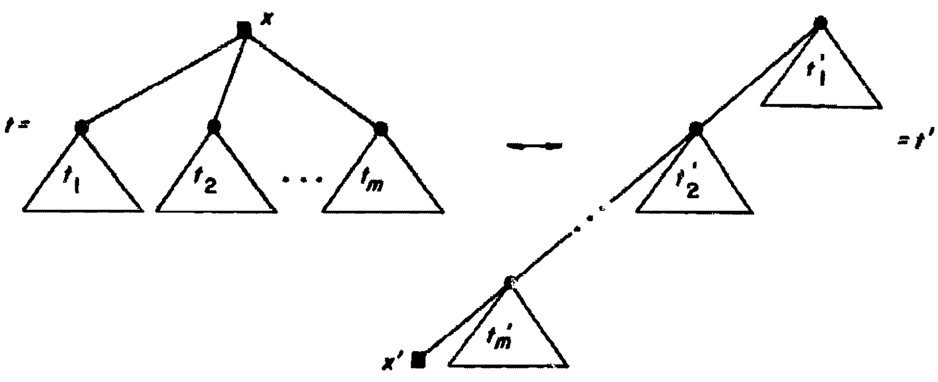


Diagram 3

We extend the First Reflection Lemma of Section 2.2 with a

Second Reflection Lemma. *Two ordered trees $t, t' \in T_n$ correspond to each other in the manner of the third proof (above), if and only if the full binary trees $b(t)$, $b(t') \in B_n$ are reflections of each other.*

Proof. Behold Diagram 4. □

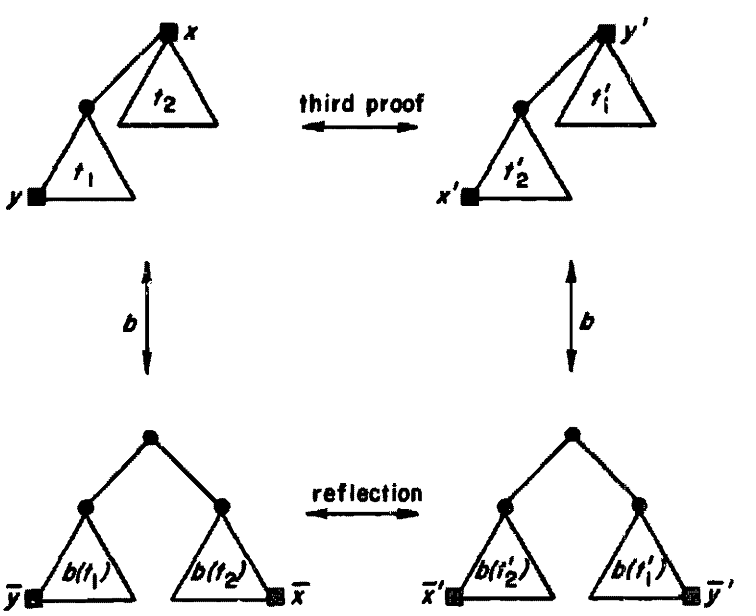


Diagram 4

Note that the second proof uses the three correspondences $t \leftrightarrow b(t) \leftrightarrow b(t') \leftrightarrow t'$, whereas the third proof uses the short-cut $t \leftrightarrow t'$.

Combining the two reflection lemmata, we have the

Corollary. *Let t^* be the reflection of t and $b^*(t)$ be the reflection of $b(t)$. Then $b'(t^*) = b^*(t) = b(t')$.*

3.4. $\mathcal{D}_n(d)$ and $\mathcal{R}_n(r)$

In this subsection we present closed-form expressions for $\mathcal{D}_n(d)$—the number of nodes in T_n of degree d, and $\mathcal{R}_n(r)$—the number of trees in T_n with root of degree r.

Theorem 3. *The total number $\mathcal{D}_n(d)$ of nodes in T_n of degree d is*

$$\mathcal{D}_n(d) = \binom{2n-1-d}{n-1}.$$

Example. There are $\mathcal{D}_4(1) = 20$ unary nodes in the trees in T_4. (See Fig. 2.)

Proof.

$\mathcal{D}_n(d)$ = the total number of runs of exactly d)'s in the parenthetic expressions in P_n (Characterization Lemma)
= the total number of runs of exactly d)'s on cycles of $n+1$ ('s and n)'s (Cycle Lemma)
= the number of ways to arrange $n-1$ ('s and $n-d$)'s on a line (that is what remains after placing $()^d($ on the cycle)
$= \binom{2n-1-d}{n-1}$.

Note that if an expression contains t occurrences of $)^d$, it will be counted exactly t times, as desired. □

Corollary 3.1. *The expected number of nodes of degree d in a tree in T_n is less than $(n+1)/2^d$.*

Proof.

$$\frac{\mathcal{D}_n(d)}{C_n} = \frac{\binom{2n-1-d}{n-1}}{\frac{1}{n+1}\binom{2n}{n}} = \frac{n(n-1)\cdots(n-d+1)}{2n(2n-1)\cdots(2n-d+1)}\frac{n}{(2n-d)}(n+1) \le \left(\frac{1}{2}\right)^d (n+1). \quad \square$$

It follows that the expected number of nodes of degree greater than d is also less than $(n+1)/2^d$. This demonstrates the extreme paucity of nodes of high degree.

Theorem 4. *The number $\mathcal{R}_n(r)$ of trees with n edges and root of degree r is*

$$\mathcal{R}_n(r) = \frac{r}{n}\binom{2n-1-r}{n-1}.$$

Example. $\mathcal{R}_4(2) = 5$ of the 14 trees in T_4 have a root of degree 2. (See Fig. 2.)

Proof. Suppose a tree $t \in T_n$ has a root of degree r. Then the lattice path $l'(t)$ begins with $(0,0)\to(0,r)\to(1,r)$ (see Section 2.2). The number of admissible paths $(1,r)\to(n,n)$ is by symmetry equal to the number of admissible paths $(0,0)\to(n-r,n-1)$: $(r/n)\binom{2n-1-r}{n-1}$ (see Section 2.3). □

Note that $\mathcal{R}_n(r) = (r/n)\mathcal{D}_n(r)$, i.e. exactly r/n of the $\binom{2n-1-r}{n-1}$ nodes of degree r in T_n are roots.

Corollary 4.1. *The expected root degree of trees in T_n is $3n/(n+2)$.*

Proof. We must evaluate

$$\frac{\sum_{r=0}^{n} r\mathcal{R}_n(r)}{|T_n|} = \frac{\frac{1}{n}\sum_{r=0}^{n} r^2\binom{2n-1-r}{n-1}}{C_n}.$$

But

$$\sum_{r=0}^{n} r^2\binom{2n-1-r}{n-1} = \frac{3n}{n+2}\binom{2n}{n-1}$$

(this may be derived from the formula $\sum_{k=0}\binom{s-k}{l}\binom{t+k}{m} = \binom{s+t+1}{l+m+1}$, where $m \geq t \geq 0$ and $l, s \geq 0$; see Knuth [3, Section 1.2.6, Eq. (25)]) and the result follows. □

A recurrence relation and the closed-form expression for $\mathcal{R}_n(r)$ are given in Ruskey and Hu [6], where $\mathcal{R}_n(r)$ counts the number of trees in B_n whose rightmost leaf is on level r. The correspondence between these two interpretations of $\mathcal{R}_n(r)$ follows immediately from the alternative interpretations of the lattice path l'; see Figs. 3.6, 3.2, and 3.7.

Note that

(a) the expected degree of a node in any tree in T_n is (obviously) almost 1 $(n/(n+1))$;

(b) the expected degree of an internal node in T_n is (Corollary 2.2) almost 2 $(2n/(n+1))$;

(c) the expected degree of a root in T_n is (Corollary 4.1) almost 3 $(3n/(n+2))$.

4. $\mathcal{N}_n(l, d)$

In this section, we investigate the function $\mathcal{N}_n(l, d)$. Our main result is a closed-form expression for $\mathcal{N}$:

Theorem 5. *The total number $\mathcal{N}_n(l,d)$ of nodes in T_n of degree d on level l is*

$$\mathcal{N}_n(l,d)=\frac{2l+d}{2n-d}\binom{2n-d}{n+l}=\binom{2n-1-d}{n+l-1}-\binom{2n-1-d}{n+l}.$$

Examples. There are $\mathcal{N}_4(2,1)=5$ unary nodes on level 2 of the 14 trees in T_4 (see Fig. 2); there are $\mathcal{N}_{10}(4,2)=1700$ binary nodes on level 4 of the 16,796 trees in T_{10}.

Proof. We first prove the following four properties for n, l, $d\geqslant 0$:

(1) $\mathcal{N}_{n+1}(l+1,d)=\mathcal{N}_{n+1}(l+1,d+1)+\mathcal{N}_{n+1}(l,d+1)$,

(2) $\mathcal{N}_{n+1}(0,d+1)=\mathcal{N}_{n+1}(0,d+2)+\mathcal{N}_n(0,d)$,

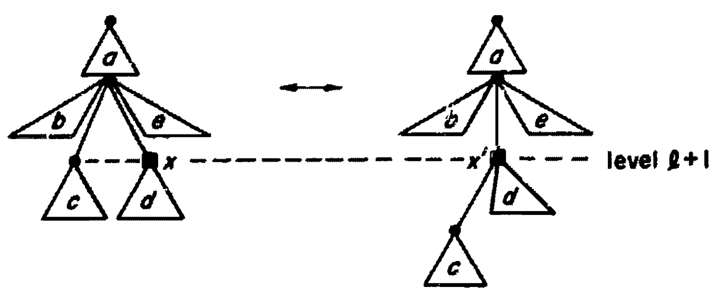
(3) $\mathcal{N}_n(l,d)=\begin{cases}1 & \text{if } d=n-l,\\ 0 & \text{if } d>n-l,\end{cases}$

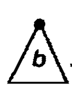
(4) $\mathcal{N}_{n+1}(0,0)=0$.

Together, they form a recurrence relation (1) in the three variables n, l, and d and boundary conditions (2, 3, 4), thus defining $\mathcal{N}_n(l,d)$ for all n, l, $d\geqslant 0$. Note that n and l are increasing while d is decreasing. (The reader should convince himself that the boundary conditions suffice.)

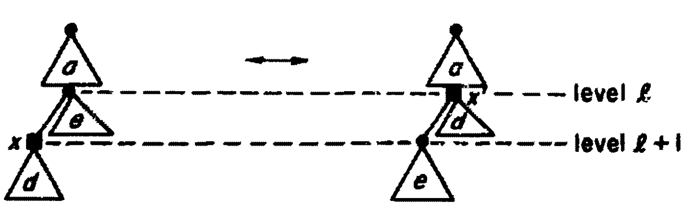
(1) Let x be a node with d children on level $l+1$ of some tree in T_{n+1}.

(a) If x has an older sibling (to its left), then the following correspondence holds between x and a node x' with one more child (and therefore at least one child) on the same level in another tree in T_{n+1}:



All the subtrees , , etc. in this proof are arbitrary.

(b) If x has no older sibling, then it corresponds to a node x' with an additional child one level up in some (perhaps the same) tree in T_{n+1}:



Since x either (a) has an older sibling or (b) does not, and the above correspondences are one-to-one, it follows that

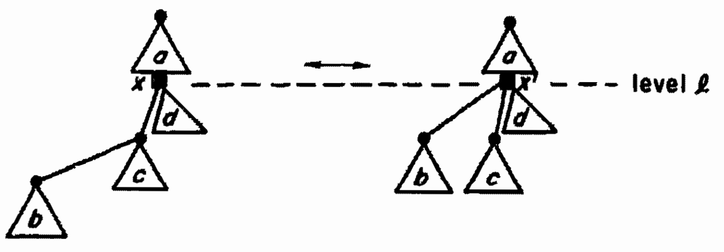
$$\mathcal{N}_{n+1}(l+1, d) = \mathcal{N}_{n+1}(l+1, d+1) + \mathcal{N}_{n+1}(l, d+1).$$

(2) We prove, more generally, that

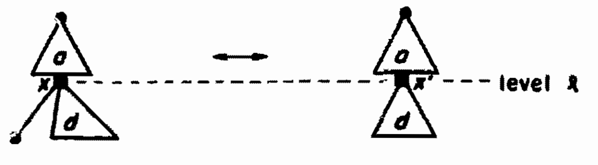
$$\mathcal{N}_{n+1}(l, d+1) = \mathcal{N}_{n+1}(l, d+2) + \mathcal{N}_n(l, d).$$

Let x be a node with $d+1$ children on level l of some tree in T_{n+1}.

(a) If x has at least one grandchild from his eldest child, then it corresponds to a node x' with an additional child (and therefore at least two children) on the same level in another tree in T_{n+1}:



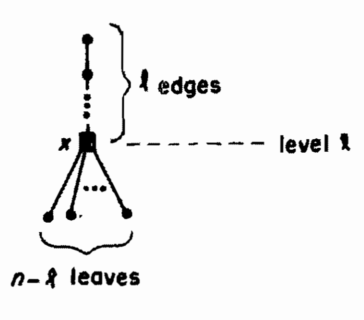
(b) If x's eldest child is childless (a leaf), then it corresponds to a node x' with one less child in a tree with one less edge:



Since x's eldest child either (a) has children or (b) does not, and the above correspondences are one-to-one, it follows that

$$\mathcal{N}_{n+1}(l, d+1) = \mathcal{N}_{n+1}(l, d+2) + \mathcal{N}_n(l, d).$$

(3) Clearly, it takes at least l edges to reach level l, leaving no more than $n-l$ edges for the rest of the tree. So, if the degree of a node x on level l is $n-l$, we have the single possibility:



(4) Obviously the root must be of degree greater than zero if there is at least one edge in the tree.

Since the function $\binom{2n-d}{n+l}(2l+d)/(2n-d)$ also satisfies the recurrence relation (1) together with the boundary conditions (2, 3, 4), it follows that

$$\mathcal{N}_n(l,d)=\frac{2l+d}{2n-d}\binom{2n-d}{n+l}. \qquad \square$$

Recall that the function $f(i,j)=\binom{i+j+1}{i}(j-i+1)/(j+i+1)$, used at the end of Section 2 to yield the number of admissible lattice paths from $(0,0)$ to (i,j), was based on the recurrence $f(i,j)=f(i,j-1)+f(i-1,j)$ with the boundary conditions $f(0,j)=1$ and $f(i,j)=0$ if $j<i$. It turns out that

$$\mathcal{N}_n(l,d)=\frac{2l+d}{2n-d}\binom{2n-d}{n+l}=f(n-d-l,n+l-1).$$

With this in mind, it is easy to construct a Pascal-like half-triangle (Fig. 5) giving the values of $\mathcal{N}$.

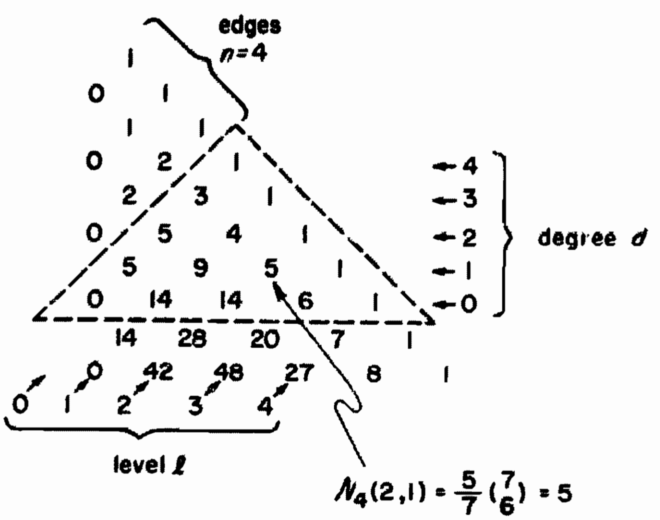


Fig. 5 $\mathcal{N}_n(l,d)$.

Each entry in the triangle is the sum of the two entries above it. Also,

$$\mathcal{N}_n(l,d)=\mathcal{N}_{n+1}(l-1,d+2)=\mathcal{N}_{n-1}(l+1,d-2)=\cdots.$$

Together, these two facts yield various relations among the values of $\mathcal{N}$, e.g.

$$\mathcal{N}_n(l,d)+\mathcal{N}_n(l+1,d)=\mathcal{N}_{n+1}(l,d+1).$$

Some consequences of this theorem are:

Corollary 5.1. *The number of leaves in T_n residing on level l is $(l/n)\binom{2n}{n-l}$.*

Proof. $\mathcal{N}_n(l,0)=(l/n)\binom{2n}{n-l}$. $\square$

Corollary 5.2. *The number of internal nodes in T_n residing on level l is $((l+1)/n)\binom{2n}{n-l-1}$.*

Proof. We have seen (Property (1)) that

$$\mathcal{N}_n(l+1, d) = \mathcal{N}_n(l+1, d+1) + \mathcal{N}_n(l, d+1)$$

for $n > 0$; thus

$$\mathcal{N}_n(l, d+1) = \mathcal{N}_n(l+1, d) - \mathcal{N}_n(l+1, d+1).$$

Hence,

$$\begin{aligned}\sum_{d=1}^{n} \mathcal{N}_n(l, d) &= \mathcal{N}_n(l+1, 0) - \mathcal{N}_n(l+1, n) = \mathcal{N}_n(l+1, 0)\\ &= \frac{2l+2}{2n}\binom{2n}{n+l+1} = \frac{l+1}{n}\binom{2n}{n-l-1}.\end{aligned} \qquad \square$$

Corollary 5.3. *The total number of nodes in T_n residing on level l is*

$$\frac{2l+1}{2n+1}\binom{2n+1}{n-l}$$

Proof.

$$\begin{aligned}\sum_{d=0}^{n} \mathcal{N}_n(l, d) &= \mathcal{N}_n(l, 0) + \sum_{d=1}^{n} \mathcal{N}_n(l, d)\\ &= \mathcal{N}_n(l, 0) + \mathcal{N}_n(l+1, 0) = \mathcal{N}_{n+1}(l, 1)\\ &= \frac{2l+1}{2n+1}\binom{2n+1}{n-l}.\end{aligned} \qquad \square$$

Alternative lattice-path arguments for these results are given in the Appendix.

Corollary 5.4. *The total number $\mathcal{D}_n(d)$ of nodes of degree d in T_n is $\binom{2n-1-d}{n-1}$.*

Proof.

$$\mathcal{D}_n(d) = \sum_{l \geq 0} \mathcal{N}_n(l, d) = \sum_{l \geq 0}\left[\binom{2n-1-d}{n+l-1} - \binom{2n-1-d}{n+l}\right] = \binom{2n-1-d}{n-1}. \qquad \square$$

Corollary 5.5. *The number $\mathcal{R}_n(r)$ of trees in T_n with root of degree r is $(r/n)\binom{2n-1-r}{n-1}$.*

Proof. Since the number of such trees equals the number of roots of degree r, we have

$$\mathcal{R}_n(r) = \mathcal{N}_n(0, r) = \frac{r}{2n-r}\binom{2n-r}{n} = \frac{r}{n}\binom{2n-1-r}{n-1}. \qquad \square$$

The last two results have already been independently and directly proved in the previous sectio ı (Theorems 3 and 4).

5. Discussion

We have investigated various enumerations of the class T_n of ordered trees with n edges. Our main result was the closed-form expression $\binom{2n-d}{n+l}(2l+d)/(2n-d)$ for the number $\mathcal{N}_n(l,d)$ of nodes of degree d on level l in trees in T_n. Other results were derived from this formula, for some of which we also gave independent direct proofs.

In searching for closed-form expressions, we made considerable use of computer-generated data and of the handbook of sequences by Sloane [7].

Appendix

In this appendix we give alternative lattice-path proofs (A) for the number $\mathcal{L}_n(k)$ of trees in T_n with k leaves (Theorem 1) and (B) for the total number of leaves and of internal nodes on level l in T_n (Corollaries 5.1 and 5.2).

(A) We first prove that the number of inadmissible paths from (i,j) to (n,n), with k corners, where $0\le i\le j\le n$, is equal to the number of paths from $(j+1, i-1)$ to (n,n) with k corners. Note that the point $(j+1, i-1)$ is symmetric to the point (i,j) with respect to the line $y=x-1$. The proof is by induction on the number of corners, k.

Consider the inadmissible paths from (i,j) to (n,n) with $k\ge 0$ corners that begin with $(i,j)\to(i',j)\to(i',j+1)$, where $i'>j$. They are in one-to-one correspondence with the paths beginning with $(j+1,i-1)\to(i',i-1)\to(i',j+1)$ and continuing with the same k corners to (n,n) (see Diagram 5). (This serves as the base case of the induction.)

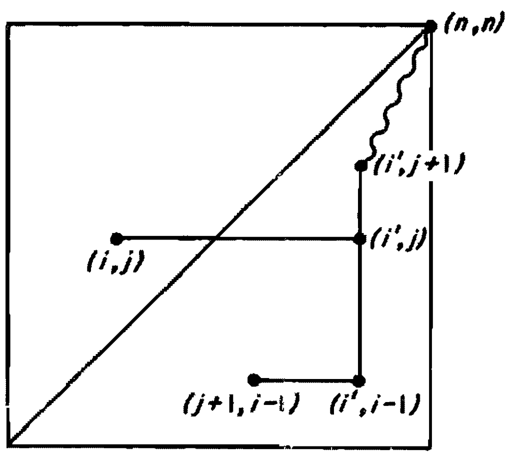


Diagram 5

Now consider those inadmissible paths from (i,j) to (n,n) with $k>0$ corners that begin with $(i,j)\to(i'-1,j)\to(i'-1,j')\to(i',j')$, where $i'-1\le j<j'$. They are in one-to-one correspondence with the paths from $(j+1,\ i-1)$ to (n,n) with k corners beginning with $(j+1,i-1)\to(j',i-1)\to(j',i'-1)\to(j'+1,i'-1)$, since,

by the inductive hypothesis, the number of inadmissible paths from (i', j') to (n, n) with the remaining $k-1$ corners is equal to the number of paths from $(j'+1, i'-1)$ to (n, n) with $k-1$ corners (see Diagram 6).

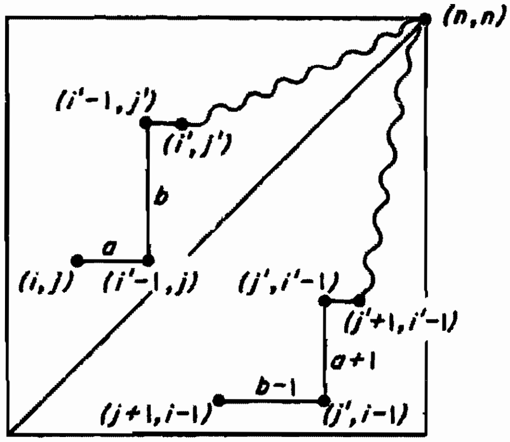


Diagram 6

The number of lattice paths (admissible or not) from $(0, 0)$ to (m, n) with exactly k corners is $\binom{m}{k}\binom{n}{k}$. To see this, choose k x_i's such that $0 \leq x_1 < x_2 < \cdots < x_k < m$ and k y_i's such that $0 < y_1 < y_2 < \cdots < y_k \leq n$; the k corners are at the points (x_i, y_i). Thus, the total number of paths from $(0, 0)$ to (n, n) with k corners is $\binom{n}{k}\binom{n}{k}$, while the total number of paths from $(1, -1)$ to (n, n) with k corners is $\binom{n-1}{k}\binom{n+1}{k}$. It follows that

$$\mathscr{L}_n(k) = \binom{n}{k}\binom{n}{k} - \binom{n-1}{k}\binom{n+1}{k} = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}. \qquad \square$$

(B) Recall that a leaf is characterized by a corner in the lattice path (Characterization Lemma). Accordingly, a leaf on level l corresponds to a corner on the lth diagonal above $y = x$, i.e. a path segment $(i, i+l-1) \to (i, i+l) \to (i+1, i+l)$ (see diagram 7).

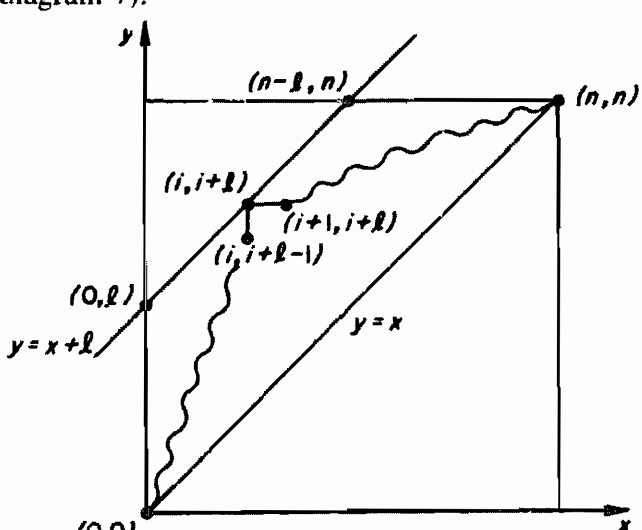


Diagram 7

Recall further (Section 2.3) that the number of admissible paths from $(0, 0)$ to (i, j)—denote that by $[(0, 0) \to (i, j)]$—is $\binom{i+j+1}{i}(j-i+1)/(j+i+1)$. Thus,

number of leaves on level l

$$= \sum_{i=0}^{n-l} [(0,0)\to(i, i+l-1)][(i+1, i+l)\to(n, n)]$$

$$= \sum [(0,0)\to(i, i+l-1)][(0,0)\to(n-i-l, n-i-1)] \quad \text{(by symmetry)}$$

$$= \sum_{i=0}^{n-l} \frac{l}{l+2i}\binom{l+2i}{i}\frac{l}{2n-2i-l}\binom{2n-2i-l}{n-i}$$

$$= \frac{l}{n}\binom{2n}{n-l},$$

using the identity (see, e.g. Knuth [3, Section 1.2.6, Eq. (31)])

$$\sum_{k=0}^{m} \frac{r}{r-kt}\binom{r-kt}{k}\frac{s}{s-mt+kt}\binom{s-mt+kt}{m-k} = \frac{r+s}{r+s-mt}\binom{r+s-mt}{m},$$

for integer m.

Similarly, since an internal node on level l is characterized by a vertical crossing of $y = x + l$, we have

number of internal nodes on level l

$$= \sum_{i=0}^{n-l} [(0,0)\to(i, i+l-1)][(i, i+l+1)\to(n, n)]$$

$$= \sum_{i=0}^{n-l} \frac{l+1}{l+2i}\binom{l+2i}{i}\frac{l+2}{2n-2i-l}\binom{2n-2i-l}{n-i+1}$$

$$= \frac{l+1}{n}\binom{2n}{n-l-1}. \qquad \square$$

Theorem 5 itself, i.e. $\mathcal{N}_n(l, d) = [(0, 0) \to (n-d-l, n+l+1]$, can be proved by a more complicated lattice-path argument.

References

[1] A. Dvoretzky and Th. Motzkin, A problem of arrangements, Duke Math. J. 14 (1947) 305–313.

[2] M. Gardner, Mathematical games: Catalan numbers, Scientific American 234 (6) (1976) 120–125.

[3] D.E. Knuth, The Art of Computer Programming, Vol. 1: Fundamental Algorithms (Addison-Wesley, Reading, MA, 1968).

[4] M.J. Kuchinski, Catalan structures and correspondences, M.S. Thesis, Dept. of Mathematics, West Virginia Univ., Morgantown, WV (May 1977).

[5] T.V. Narayana, A partial order and its application to probability, Sankhya 21 (1959) 91–98.

[6] F. Ruskey and T.C. Hu, Generating binary trees lexicographically, SIAM J. Comput. 6 (4) (1977) 745–758.

[7] N.J.A. Sloane, A Handbook of Integer Sequences (Academic Press, New York, 1973).