

MORE PATTERNS IN TREES: UP AND DOWN, YOUNG AND OLD, ODD AND EVEN*

NACHUM DERSHOWITZ[†] AND SHMUEL ZAKS[‡]

Abstract. We apply the tree-pattern enumeration formulæ of earlier work of ours [N. Dershowitz and S. Zaks, *Discrete Appl. Math.*, 25 (1989), pp. 241–255], and a new extension thereof, to some recent enumerations of distributions of leaves in ordered trees [W. Y. C. Chen, E. Deutsch, and S. Elizalde, *European J. Combin.*, 27 (2006), pp. 414–427] and in bicolored ordered trees [L. H. Clark, J. E. McCanna, and L. A. Székely, *Bull. Inst. Combin. Appl.*, 21 (1997), pp. 33–45], and of distributions of up-down-up subpaths in Dyck lattice paths [Y. Sun, *Discrete Math.*, 287 (2004), pp. 177–186]. Bijections are used to facilitate the derivation of statistics for bicolored trees.

Key words. tree enumerations, tree patterns, node distribution, ordered trees, plane-planted trees, bicolored trees, binary trees, Dyck paths, lattice paths, bridges

AMS subject classification. 05A15

DOI. 10.1137/070687475

the bridge guard’s bucket
upside-down to dry...
fresh leaves
—Issa (1818)

1. Introduction. Over and above their intrinsic combinatorial interest, enumerations of classes of trees have manifold applications to average-case analysis of algorithms. For instance, the performance of various manipulations of tree structures may depend on the distribution of node degrees or of tree heights. For one example, see [12].

Several recent lattice-path and tree enumerations turn out to be amenable to the generic pattern enumeration formula we gave in [7], which was based on Dvoretzky and Motzkin’s cycle lemma [10] (called “penetrating analysis” in [16]; see [8]). One formulation of this lemma is the following.

CYCLE LEMMA (see [10]). *For any sequence of m natural numbers j_0, \dots, j_{m-1} , whose sum is n , with $m > n$, there are exactly $m - n$ offsets π , $0 \leq \pi < m$, such that*

$$j_{\pi \bmod m} + \dots + j_{(\pi+k-1) \bmod m} > k + n - m$$

for all k , $1 \leq k < m$.

For example, the sequence 102001030 has 2 ($= 9 - 7$) such cyclic permutations:

$$\underline{10} \underline{3010} \underline{200} \quad \underline{3010} \underline{200} \underline{10}.$$

Notice that both of these sequences can be read as well-formed Polish-prefix expressions (as indicated by the underscoring), where each number j is followed by j well-formed expressions.

*Received by the editors April 4, 2007; accepted for publication (in revised form) July 28, 2008; published electronically January 16, 2009.

<http://www.siam.org/journals/sidma/23-1/68747.html>

[†]Department of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel (nachum.dershowitz@cs.tau.ac.il). This author’s research was supported in part by the Israel Science Foundation (grant 250/05).

[‡]Department of Computer Science, Technion, Haifa 32000, Israel (zaks@cs.technion.ac.il).

To understand why the lemma holds, note that such a sequence must have at least one occurrence of 0, since $m > n$. Replacing a (cyclically adjacent) pair $j0$, where j is positive, with just $j - 1$, decreases both the sum n and count m by 1. This does not affect the quantity—or starting positions—of valid cyclic permutations, which cannot begin with 0. Repeatedly making such replacements terminates when only zeros remain, at which point the lemma clearly holds true.

Specifically, we show here how Sun’s enumeration [23] of Dyck lattice paths with a given number of subsequences $\nearrow \searrow \nearrow$ (**udu**) can be solved in this manner, and how a convenient extension of our pattern formula applies to the enumeration by Chen, Deutsch, and Elizalde [4] of ordered trees with given numbers of eldest and noneldest childless children (leaves). We also show how to apply the extended formula to count trees with given distributions of leaves on odd and even levels, as done by Clark, McCanna, and Székely [5].

We review the pattern enumeration of [7] in the next two sections, and then extend it in section 4. Our main result, Theorem 4.1 below, counts occurrences of a multiset of patterns in ordered trees, or in other Catalan structures (see Figure 1). Tree patterns are specified by subtrees having empty slots that can be filled with leaves and/or with larger subtrees, and gaps that can be filled by a series of subtrees (see Figure 2).

These formulæ are applied to the problems of Sun [23], of Chen, Deutsch, and Elizalde [4], and of Clark, McCanna, and Székely [5] in sections 5, 6, and 7, respectively. Section 7 includes a new bijection between ordered trees, mapping odd-level nodes to internal nodes, and even-level internal nodes to leftmost leaves, plus some results on their distributions.

In the concluding section, the correspondence between the lattice-path enumeration of [23], the ordered-tree enumeration of [4], and the bicolored tree enumeration of [5] is explicated.

2. Pattern formulæ. *Ordered trees* (that is, plane-planted trees with a root whose children are also ordered in sequence) may be defined as follows:

$$T ::= \langle T^* \rangle,$$

meaning a bracketed sequence of zero or more ordered trees.¹ In other words, T is the set of nonempty *balanced* bracketed expressions, where the first open bracket matches the *last* close bracket. See Figure 1(a). A *subtree* of $t \in T$ is any subsequence of t that is itself a tree in T . Ordered trees are counted by the Catalan/Segner [2, 20] numbers, and are in one-to-one correspondence with many other combinatorial structures, including those in Figure 1(b),(c). See [13].

Patterns come in four basic shapes: \triangle , \diamond , \blacktriangle , and ∞ . The plain triangle *pattern* \triangle matches any subtree, a lozenge \diamond corresponds to any tree leaf $\langle \rangle$, a dark triangle \blacktriangle matches any nonleaf subtree (that is, a subtree rooted at an internal node), and an ellipsis ∞ can match any sequence of (zero or more) subtrees. So, the pattern \triangle matches any subtree matched by either \diamond or \blacktriangle .

Base patterns can be composed to form more complicated shapes. An ellipsis is intended to match a forest (sequence) of trees, rather than a single tree, so it makes sense only within a composite pattern. Thus, tree patterns have the following

¹Some authors (e.g., [24, 5]) follow a convention by which ordered trees have an extra unary (monovalent) node connected by an extra edge to what is the root of our trees. This necessitates minor changes in the parameters of some enumerations.

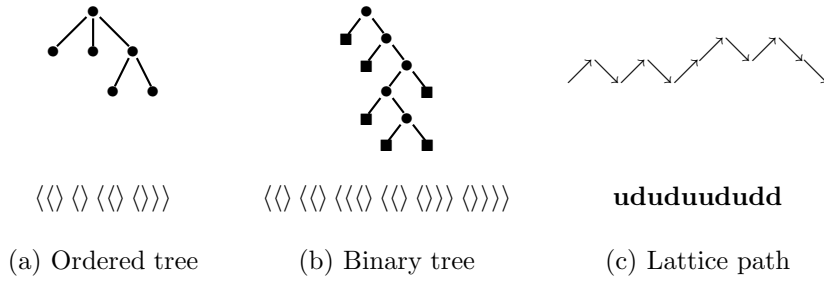


FIG. 1. *Corresponding objects.*

grammar:

$$P ::= \diamond \mid \triangle \mid \blacktriangle \mid \langle Q^+ \rangle,$$

$$Q ::= P \mid \infty,$$

where Q^+ means one or more patterns Q , in sequence. See Figure 2.

A *composite* pattern $\langle p_1 \cdots p_q \rangle$ *matches* a tree $\langle t_1 \cdots t_n \rangle$ if the latter’s immediate subtrees $t_1 \cdots t_n$ can be divided into q (possibly empty) subsequences $t_1 \cdots t_{k_1} \cdots t_{k_2} \cdots t_{k_q} = t_n$ of trees, such that each p_i ($i = 1, \dots, q$) matches the subsequence $t_{k_{i-1}+1} \cdots t_{k_i}$ ($k_0 = 0$). Of course, only the ellipsis pattern ∞ can match a subsequence of more than one subtree; an ellipsis even matches the empty sequence of zero subtrees. In other words, a match is an injection (embedding) of pattern nodes to tree nodes and of pattern edges to tree edges that preserves edge incidence and order, and also edge neighbors—unless the pattern has an ellipsis between the edges in question.

For example, the pattern $\langle \diamond \infty \blacktriangle \rangle$, depicted in Figure 2(a), matches any tree whose root has two or more children, the youngest (rightmost) having children, but the eldest (leftmost) still childless (a leaf). Thus, it matches the tree t , depicted in Figure 1(a). Clearly, the same tree can match many different (base and composite) patterns. In fact, t is also matched by the patterns in Figure 2(b),(c) but not by that in Figure 2(d).

There may be many ways to divide n subtrees into q subsequences when a pattern has more than one ellipsis. For instance, the pattern $\langle \diamond \infty \triangle \infty \rangle$ in Figure 2(b) matches t in two ways, for each of the two younger children.

A pattern *occurs* in a tree if it matches any *subtree*; for example, $p = \langle \diamond \infty \triangle \infty \rangle$ (Figure 2(b)) also occurs at the youngest child of the root of t (Figure 1(a)). A multiset (bag) of patterns occurs in a tree if the patterns match *disjoint* subtrees. The singleton set $\{\langle \diamond \diamond \rangle\}$ of patterns appears exactly once in t , while $\{\diamond, \diamond\}$ appears six times, once for each choice of two of the leaves. The disjointness requirement means that no tree node or edge may be part of more than one pattern match—though triangles (plain \triangle or dark \blacktriangle) of one pattern can be at the root of an occurrence of another pattern. Thus, the pattern multiset $\{p, p\}$ occurs only twice in t : one p at the youngest child and the other at the root in either of two ways.

The number of edges in a pattern is equal to the total number of left (or right) brackets—excluding the first one, plus the total number of triangles, \triangle or \blacktriangle , plus the total number of leaf patterns \diamond . In other words, the number of edges $e(p)$ in a (nonellipsis) pattern p is obtained, recursively, as follows:

$$e(\triangle) = e(\blacktriangle) = e(\diamond) = 0,$$

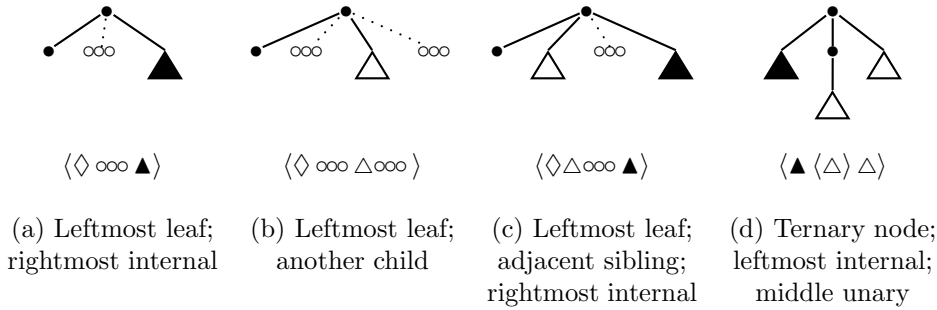


FIG. 2. Some patterns.

$$e(\infty) = -1,$$

$$e(\langle p_1 \cdots p_q \rangle) = e(p_1) + \cdots + e(p_q) + q.$$

The number $v(p)$ of nodes (vertices) in p can be calculated similarly:

$$v(\triangle) = v(\blacktriangle) = v(\infty) = 0,$$

$$v(\diamond) = 1,$$

$$v(\langle p_1 \cdots p_q \rangle) = v(p_1) + \cdots + v(p_q) + 1.$$

For example, the four patterns in Figure 2 comprise a total of $e = 2 + 2 + 3 + 4 = 11$ edges and $v = 2 + 2 + 2 + 2 = 8$ nodes.

Let p_1, \dots, p_q , be some patterns and let their desired multiplicities of occurrence in a tree be n_1, \dots, n_q , respectively. In [7], we provided the pattern enumeration formula

$$(2.1) \quad \frac{1}{u} \binom{u}{n_1, \dots, n_q, u-m} \binom{2n+d+s-2e-m}{n-e}$$

for the number of occurrences of such a multiset of $m = \sum n_i$ patterns among all n -edge ordered trees, where e is the total number of edges in the m patterns, d is the number of plain triangles \triangle therein, s is the number of ellipses ∞ , and $u = n + d - e + 1$. Nonleaf patterns \blacktriangle were not treated in [7].²

This formula can be used to calculate various tree statistics. Let the *size* of a tree be measured by the number of its edges.

Example 2.1. A simple one-pattern case of this formula establishes the expected number of nodes in an ordered tree that are “eldest” leaves (leftmost and childless). The pattern $p_1 = \langle \diamond \infty \rangle$ matches each such leaf. Letting $m = n_1 = e = s = 1$, $d = 0$, $u = n$ in (2.1), yields

$$\frac{1}{n} \binom{n}{1} \binom{2n-2}{n-1} = \binom{2n-2}{n-1}$$

²We use multinomial coefficients $\binom{j}{j_1, \dots, j_k} = \frac{j!}{j_1! \cdots j_k!}$ throughout. Virtually all the formulæ in this paper containing occurrences of the multinomial would benefit from dropping the common (but not universal) requirement that $j = j_1 + \cdots + j_k$, in which case the first two factors of (2.1) would become simply $\binom{u-1}{n_1, \dots, n_q, u-m}$.

for the total number of leftmost leaves among all trees of size n . The number of trees of size n is the Catalan number [14],

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

(see [21, A000108]); so there are, on the average,

$$\frac{\binom{2n-2}{n-1}}{\frac{1}{n+1} \binom{2n}{n}} = \frac{n^2 + n}{4n - 2} \approx \frac{n}{4}$$

leftmost leaves in a tree with $n + 1$ nodes.

Similarly, the average number of “younger” (nonleftmost) leaves is counted by taking $p_1 = \langle \triangle \infty \diamond \infty \rangle$, $m = n_1 = d = 1$, $e = s = 2$, and $u = n$:

$$\frac{\binom{2n-2}{n-2}}{\frac{1}{n+1} \binom{2n}{n}} = \frac{n^2 - 1}{4n - 2} \approx \frac{n}{4}.$$

The leaf subpattern \diamond contributes to the count for *each* nonleftmost leaf child of the node matching $\langle \triangle \infty \diamond \infty \rangle$. \square

Here and throughout, we speak freely of nodes as being “leftmost,” “nonleftmost,” “rightmost,” or “nonrightmost” interchangeably with age-based terminology, “oldest,” “younger,” “youngest,” or “older,” respectively. The root of a tree and the only child of a unary node are leftmost and rightmost at one and the same time.

The following is an easy observation.

PROPOSITION 2.2. *In any ordered tree, the number of leftmost (resp., rightmost) leaves is one more than the number of nonleftmost (resp., nonrightmost) internal nodes.*

The difference of one in the numbers is on account of the root, which is perforce leftmost, as well as rightmost.

This correspondence holds even for the one-node tree, which has one leftmost/rightmost leaf, the root, and no internal nodes at all. It also holds for the two-node tree, which has one leftmost/rightmost leaf and no nonleftmost internal nodes.

Proof. We give two proofs, one “geometric” and one “algebraic.”

1. From each leftmost leaf, travel up leftmost edges as far as possible, reaching the root or else reaching some nonleftmost internal node. In the reverse, from the root or any nonleftmost internal node, one can travel down leftmost edges until encountering the corresponding leftmost leaf.

2. Suppose there are i internal nodes, ℓ leftmost leaves, and k leftmost nonroot internal nodes. Since every internal node has exactly one leftmost child, $i = \ell + k$. Hence $\ell = i - k$, which is one more (for the root) than the number of nonleftmost internal nodes.

(The correspondence of rightmost leaves with nonrightmost internal nodes and the root follows by symmetry.) \square

We will see (Theorem 7.3 below) that leftmost leaves, rightmost leaves, leftmost/rightmost internal nodes, nonleftmost/nonrightmost leaves/internal nodes all occur with almost equal frequency among ordered trees with a given number of nodes (or of edges). All eight cases occur in about one-fourth of the nodes. See the previous example.

3. Tree enumerations. This paper is mainly concerned with counting trees, rather than pattern occurrences.

When the patterns are such that there can be no more than one occurrence of the multiset of patterns in any one tree, then formula (2.1) above counts trees. This is the case, in particular, when the patterns include all the nodes in the tree being counted ($m + e - d = n + 1$), and different patterns do not overlap each other. (Two patterns “overlap” if there is a tree in which both occur and whose occurrences share at least one node.) The latter condition precludes, for instance, sibling ellipses, such as $\langle \infty \infty \triangle \infty \infty \rangle$, which can occur multiple times at the same subtree.

Example 3.1. The number of ordered trees with n edges, ℓ leaves, and no unary or binary nodes is counted by the number of occurrences of $n_1 = \ell$ leaf patterns \diamond and $n_2 = n + 1 - \ell$ patterns $\langle \triangle \triangle \triangle \infty \infty \rangle$ for nodes of (out-) degree at least 3. We have $e = d = 3n - 3\ell + 3$, $s = n + 1 - \ell$, and $m = u = n + 1$, giving

$$\frac{1}{n+1} \binom{n+1}{\ell} \binom{2\ell - n - 3}{n - \ell}.$$

Summing over ℓ , for $2n/3 < \ell \leq n$, counts all n -edge trees sans unary and binary nodes. For $n = 3, 4, \dots$, this is: 1, 1, 1, 4, 8, 13, 31, 71, 144, 318, 729, 1611, 3604, 8249, \dots . Note that this also counts the number of sequences of n natural numbers, excluding 1 and 2, such that the sum of every prefix is no more than its length.³

If we exclude only unary nodes, we get, instead, the n th Riordan number [21, A005043]. See [15, p. 587] and [7, Ex. 3.1.3]; see also [1]. \square

Example 3.2. The number of ordered trees with n edges, r leaves, and i unary nodes is obtained by considering r leaf patterns \diamond , i instances of $\langle \triangle \rangle$, and $n + 1 - r - i$ of $\langle \triangle \infty \infty \rangle$ for the remaining nodes ($e = d = 2n - 2r - i + 2$, $s = n - r - i + 1$, and $m = u = n + 1$):

$$\frac{1}{n+1} \binom{n+1}{r, i, n-r-i+1} \binom{r-2}{n-r-i}. \quad \square$$

When patterns include all the nodes and all the edges in the tree ($e = n$, $m = u = d + 1$, and $s = 0$), enumeration (2.1) simplifies to just

$$(3.1) \quad \frac{1}{m} \binom{m}{n_1, \dots, n_q}.$$

This generalizes [11] and [24], which consider patterns representing the degrees of the nodes only.

Example 3.3. The number of “0-1-2” (“unary-binary”) trees (with maximum outdegree 2) with n edges and ℓ leaves, and, hence, $\ell - 1$ binary nodes and $n - 2\ell + 2$ unary nodes, is $\frac{1}{n+1} \binom{n+1}{\ell, \ell-1, n-2\ell+2}$ ([9]; cf. [19]). Summing over ℓ , we get

$$M_n = \frac{1}{n+1} \sum_{\ell} \binom{n+1}{\ell, \ell-1, n-2\ell+2}$$

for the total number of 0-1-2 trees of size n , which is the n th Motzkin number [21, A001006]. (A different derivation for this enumeration is given in [7, Ex. 3.1.1].) \square

³This sequence was recently added as #A114997 to Neil Sloane’s *The on-line encyclopedia of integer sequences* [21].

If one is interested only in a (lone) pattern occurring at the root of a tree (with no sibling ellipses, so the pattern does not overlap itself), then we have the following tree enumeration (see [7, sect. 3.3]):

$$(3.2) \quad \frac{d+s}{n-e} \binom{2n+d+s-2e-1}{n-e-1}.$$

When $n = e$, this is taken to be 1.

4. Patterns with nonleaf slots. Now we extend (2.1) by incorporating nonleaf patterns \blacktriangle . Like a plain triangle \triangle , a dark triangle \blacktriangle may also overlap an occurrence of another pattern (at its root), but with the added proviso that the latter is not \diamond . This is why we need dark triangles, and cannot simply use $\langle \triangle \infty \rangle$ for nonleaf nodes, instead (something that can be done in the absence of leaf patterns).

For example, the pattern $\langle \diamond \triangle \infty \blacktriangle \rangle$ (see Figure 2(c)) occurs in a tree, such as that in Figure 1(a), at each node of degree at least three, with eldest child a leaf and youngest not a leaf. The pair of patterns $\{ \langle \diamond \triangle \infty \blacktriangle \rangle, \diamond \}$, which includes another leaf pattern, occurs three times in the tree $\langle \langle \rangle \langle \rangle \langle \langle \rangle \langle \rangle \rangle$ of Figure 1(a), once for each of the leaves, excluding the eldest child of the root. It occurs a total of 12 times in the 42 five-edged trees, but only in the following five of them:

$$\langle \langle \bar{\diamond} \bar{\diamond} \bar{\langle \bar{\diamond} \rangle} \rangle \rangle \quad \langle \langle \bar{\diamond} \bar{\langle \bar{\diamond} \rangle} \rangle \rangle \quad \langle \langle \bar{\diamond} \bar{\langle \bar{\diamond} \bar{\diamond} \rangle} \rangle \rangle \quad \langle \langle \bar{\langle \bar{\diamond} \rangle} \bar{\langle \bar{\diamond} \rangle} \rangle \rangle \quad \langle \langle \langle \bar{\diamond} \bar{\langle \bar{\diamond} \rangle} \rangle \rangle \rangle.$$

The composite pattern matches at the high-degree node and the leaf pattern \diamond matches *any* one of the overlined leaves $\bar{\diamond}$.

THEOREM 4.1. *Let $p_1, \dots, p_q, q \geq 0$, be various nonleaf nonellipsis patterns. The number of occurrences among all n -edge ordered trees of n_i of each of the patterns p_i and of $\ell \geq 0$ leaf patterns \diamond is*

$$(4.1) \quad \binom{m}{n_1, \dots, n_q} \binom{u-t}{\ell} \sum_{\ell \leq k \leq u} \frac{1}{u-k+1} \binom{u-k+1}{m} \binom{u-t-\ell}{k-\ell} \binom{n+s-e-1}{u+s-m-k},$$

where $e = \sum e(p_i)$ is the total number of edges in the patterns, d is the number of plain triangles \triangle appearing in them, t is the number of dark triangles \blacktriangle , s is the number of ellipses ∞ , $u = n + d + t - e$, and $m = \sum n_i$. When $m > 0$, this enumeration can be rewritten as

$$(4.1a) \quad \frac{1}{m} \binom{m}{n_1, \dots, n_q} \binom{u-t}{\ell} \sum_{\ell \leq k \leq u} \binom{u-k}{m-1} \binom{u-t-\ell}{k-\ell} \binom{n+s-e-1}{u+s-m-k}.$$

Proof. We count separately for each possible number of “loose” (unattached to composite patterns) tree leaves, $k = \ell, \ell + 1, \dots, u$. Note that the total number of leaves missing from the m patterns cannot exceed $n - e + d \leq u$.

Let $v = \sum v(p_i) + \ell = m + e - d - t + k$ be the number of tree nodes accounted for by the patterns and leaves. (The $m + k$ patterns contain e edges, so the number of nodes and triangles is $v + d + t = e + m + k$.) The proof proceeds in several steps:

1. Arrange the given m nonleaf patterns in a row, in any of $\binom{m}{n_1, \dots, n_q}$ ways.
2. Intersperse $n + 1 - v = u + 1 - m - k$ extra patterns of the form $\langle \triangle \infty \rangle$ among the m patterns, to cover all the missing internal (nonleaf) nodes, in $\binom{u-k+1}{m}$ ways, for a total of $u - k + 1$ patterns.

3. Distribute the $(n - e) - (n + 1 - v) = m - d - t + k - 1$ missing edges (of the $n - e$ missing from the given patterns, $n + 1 - v$ were just added in the previous step), as sequences $\triangle \cdots \triangle$ of triangles, in place of the $s + n + 1 - v$ ellipses (s original and $n + 1 - v$ new), in $\binom{n+s-e-1}{m-d-t+k-1} = \binom{n+s-e-1}{u+s-m-k}$ ways. Note that when there are no missing edges ($n = e$ and $u + 1 - k = m$), this factor is $\binom{s-1}{0} = \binom{s-1}{s-1} = 1$.
4. Place the ℓ distinguished leaves in some of the $d + (n + 1 - v) + (v - e - 1) = u - t$ unrestricted, plain triangles (d in the original patterns, plus $n + 1 - v$ from step 2 and $v - e - 1$ from step 3), in $\binom{u-t}{\ell}$ ways.
5. Place the remaining $k - \ell$ leaves in some of the remaining $u - t - \ell$ unrestricted triangles in $\binom{u-t-\ell}{k-\ell}$ ways.⁴
6. The cyclic arrangement of the resultant $m + (u + 1 - m - k) = u - k + 1$ patterns corresponds to exactly one occurrence of the patterns in a tree. To see this, graft the patterns into one tree by repeatedly picking any pattern in the sequence and inserting it into the closest (rightmost) available triangle slot among the patterns preceding it, wrapping back around from the end when necessary. The $u - k + 1$ patterns contain a total of $d + t + (u + 1 - m - k) + (m - d - t + k - 1) - \ell - (k - \ell) = u - k$ slots. So, in fact, a single tree results from the grafting, with each pattern occurring at the point it ends up in the reconstructed tree. This situation may be viewed as an application of the cycle lemma, given in the introduction. Reading each pattern as the number of its slots, the lemma asserts that each of the $u - k + 1$ cyclic permutations of the patterns gives one and the same grafted outcome. Thus, the enumeration has an additional factor of $\frac{1}{u-k+1}$.

Summing for k , and replacing $\frac{1}{u-k+1} \binom{u-k+1}{m}$ by $\frac{1}{m} \binom{u-k}{m-1}$ when $m > 0$, gives the result. \square

By way of illustration, let us count occurrences in 9-node trees of the quartet of patterns



written linearly as

$$\{\langle \text{ooo} \blacktriangle \rangle, \langle \text{ooo} \blacktriangle \rangle, \langle \triangle \blacklozenge \text{ooo} \rangle, \langle \blacklozenge \rangle\}.$$

The patterns account for $e = 4$ out of $n = 8$ edges and for $v = 5$ of the nodes, including 2 leaves, $\ell = 1$ of which is a pattern on its own. There are $d = 1$ plain triangles and $t = 2$ dark triangles in the patterns, so $u = 8 + 1 + 2 - 4 = 7$. The patterns also have $s = 3$ ellipses.

The construction proceeds as follows:

0. Consider $k = 3$, meaning that we want 2 more leaves in the tree, besides the lone \blacklozenge appearing in the pattern set—and besides the one embedded in the third pattern, for a total of 3 internal nodes and 4 leaves.

⁴Steps 4 and 5 can be supplanted by first, (4') choosing all k leaves in $\binom{n+d-e}{k}$ ways, and only then, (5') selecting $\binom{k}{\ell}$ of them, for the same total contribution of $\binom{n+d-e}{k} \binom{k}{\ell} = \binom{n+d-e}{\ell} \binom{n+d-e-\ell}{k-\ell}$.

1. The $m = 3$ nonleaf patterns can be arranged in $\binom{3}{2,1} = 3$ ways, one of which is

$$\langle \circ \circ \blacktriangle \rangle, \langle \triangle \diamond \circ \circ \rangle, \langle \circ \circ \blacktriangle \rangle.$$

2. We need to add 2 internal-node patterns $\langle \triangle \circ \circ \rangle$ for the remaining 2 as yet unaccounted-for nodes. These extra patterns can be intermingled with the 3 original nonleaf patterns in $\binom{3+2}{3} = 10$ different ways, including, for example, the following sequence:

$$\langle \triangle \circ \circ \rangle, \langle \circ \circ \blacktriangle \rangle, \langle \triangle \circ \circ \rangle, \langle \triangle \diamond \circ \circ \rangle, \langle \circ \circ \blacktriangle \rangle.$$

3. There are only 6 edges in these patterns, so we need to graft in the 2 remaining tree edges, and eliminate all $3 + 2 = 5$ ellipses in the process. This can be done in $\binom{5+2-1}{2} = 15$ ways, such as

$$\langle \triangle \rangle, \langle \triangle \blacktriangle \rangle, \langle \triangle \rangle, \langle \triangle \diamond \triangle \rangle, \langle \blacktriangle \rangle.$$

4. The $\ell = 1$ leaf in the original pattern set can go into any of the 5 ordinary subtree (\triangle) slots—in the middle one, say,

$$\langle \triangle \rangle, \langle \triangle \blacktriangle \rangle, \langle \diamond \rangle, \langle \triangle \diamond \triangle \rangle, \langle \blacktriangle \rangle.$$

5. Similarly, the extra $k - \ell = 2$ leaves can go into any of the remaining $n + d - e - \ell = 4$ plain slots, in $\binom{4}{2} = 6$ combinations, giving something such as this final set of 5 tree patterns, arranged in sequence:

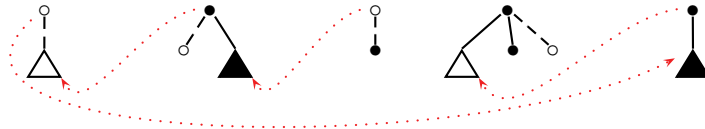
$$\langle \triangle \rangle, \langle \diamond \blacktriangle \rangle, \langle \diamond \rangle, \langle \triangle \diamond \diamond \rangle, \langle \blacktriangle \rangle.$$

We have 4 slots now and 5 patterns. All the leaves have been allocated, so the type of slot no longer matters.

6. Last, these 5 tree pieces are grafted together, step by step, as follows:

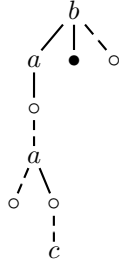
$$\begin{aligned} &\langle \triangle \rangle, \langle \diamond \blacktriangle \rangle, \langle \diamond \rangle, \langle \triangle \diamond \diamond \rangle, \langle \blacktriangle \rangle \\ &\langle \langle \diamond \blacktriangle \rangle \rangle, \langle \diamond \rangle, \langle \triangle \diamond \diamond \rangle, \langle \blacktriangle \rangle \\ &\langle \langle \diamond \langle \diamond \rangle \rangle \rangle, \langle \triangle \diamond \diamond \rangle, \langle \blacktriangle \rangle \\ &\langle \langle \langle \diamond \langle \diamond \rangle \rangle \rangle \rangle, \langle \langle \blacktriangle \rangle \diamond \diamond \rangle \\ &\langle \langle \langle \langle \diamond \langle \diamond \rangle \rangle \rangle \rangle \diamond \diamond \rangle. \end{aligned}$$

Pictorially, this is what happens:



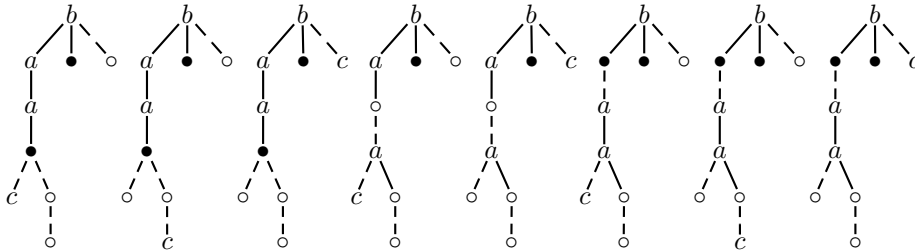
A dashed line in a pattern means that the edge is not from the original pattern set and similarly for an unfilled node. One can see, then, that the original patterns account for 4 out of 8 edges, 2 out of 4 leaves, and 3 out of 5 internal nodes.

The point of the cycle lemma is that the outcome is the same regardless of the order in which (cyclic) neighbors are grafted. The same end result, namely,



would be obtained from any of the 5 cyclic permutations of the starting sequence. The third of the original four patterns, $\langle \triangle \diamond \infty \infty \rangle$, occurs at the root, marked b ; the first two patterns, $\langle \infty \blacktriangle \rangle$, occur at the nodes marked a ; the distinguished leaf is c .

The above theorem says that there are a total of $3 \times 10 \times 15 \times 5 \times 6 \times \frac{1}{5} = 2700$ occurrences of the given patterns among all 490 trees with 8 edges and 4 leaves. For example, the same tree as obtained above happens to contain 8 additional occurrences of the same multiset of patterns, as follows:



Example 4.2. As a trivial example, with no patterns, this formula counts ordered trees of size n , by summing the number of trees with $k = 1, \dots, n + 1$ leaves, since a tree has only one occurrence of an empty pattern set. Letting $q = m = e = d = s = t = \ell = 0$ and $u = n$ in (4.1), we obtain

$$\sum_k \frac{1}{n+1-k} \binom{n}{k} \binom{n-1}{k-1} = \frac{1}{n} \sum_k \binom{n}{k} \binom{n}{k-1} = \frac{1}{2n+1} \binom{2n+1}{n},$$

the n th Catalan number, C_n . Each term $\frac{1}{n} \binom{n}{k-1} \binom{n}{k}$ in the sum is a Narayana number [17], and represents the number of n -edge k -leaf trees, as shown in [6, 18]. \square

Remark 4.3. Suppose there are no dark triangle (\blacktriangle) subpatterns ($t = 0, u = n + d - e$). Then (4.1) reduces to (2.1), but with an additional $n_{q+1} = \ell$ occurrences of the leaf pattern:

$$\begin{aligned} & \binom{m}{n_1, \dots, n_q} \binom{u-1}{\ell} \sum_k \frac{1}{u-k+1} \binom{u-k+1}{m} \binom{u-\ell}{k-\ell} \binom{n+s-e-1}{u+s-m-k} \\ &= \sum_k \frac{1}{u-m-k+1} \binom{u}{n_1, \dots, n_q, \ell, k-\ell, u-m-k} \binom{n+s-e-1}{u+s-m-k} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{u+1} \binom{u+1}{n_1, \dots, n_q, \ell, u-m-\ell+1} \sum_k \binom{u-m-\ell+1}{k-\ell} \binom{n+s-e-1}{u+s-m-k} \\
 &= \frac{1}{u+1} \binom{u+1}{n_1, \dots, n_q, \ell, u-m-\ell+1} \binom{2n+d+s-2e-m-\ell}{n-e}.
 \end{aligned}$$

When the patterns account for all the leaves (as, for instance, when they account for all $n + 1$ nodes), the sum contributes only the $k = \ell$ term, and (4.1a) simplifies to

$$(4.2) \quad \frac{1}{m} \binom{u-\ell+1}{n_1, \dots, n_q, u-\ell-m+1} \binom{n+d-e}{\ell} \binom{n+s-e-1}{u+s-m-\ell}.$$

If they also account for all edges ($n = e$), then $u = d + t = m + k - 1 = m + \ell - 1$, and we get the following mild extension of (3.1):

$$(4.3) \quad \frac{1}{m} \binom{m}{n_1, \dots, n_q} \binom{d}{\ell},$$

with an added factor for the $\binom{d}{\ell}$ ways of filling ℓ of the d unrestricted slots with leaves.

Example 4.4. The number of binary trees with i internal nodes and j left leaves is also counted by the Narayana numbers, as is clear from the standard correspondence [15, sect. 2.3.2] between ordered trees with i edges and binary trees with i internal nodes. Let there be $n_1 = j$ patterns $\langle \diamond \Delta \rangle$, $n_2 = t = i - j$ patterns $\langle \blacktriangle \Delta \rangle$, and $\ell = i + 1 - j$ leaf patterns \diamond . Then put $d = i$, $n = e = 2i$, and $s = 0$ (hence, $m = i$ and $u = 2i - j$) into the formula. Only $k = \ell$ contributes to the sum, and the last factor in (4.2) is $\binom{-1}{0} = 1$, so we get

$$\frac{1}{i} \binom{i}{j} \binom{i}{i+1-j} = \frac{1}{i} \binom{i}{j} \binom{i}{j-1}. \quad \square$$

Example 4.5. The number of ordered trees with n edges, k leaves, and j leftmost (that is, “eldest”) nonroot internal nodes (out of a total of $n - k$ nonroot internal nodes) is counted by $n_1 = j$ patterns p_1 of the form $\langle \blacktriangle \infty \rangle$, $n_2 = n + 1 - j - k$ of the form $p_2 = \langle \diamond \infty \rangle$ (for the internal nodes not covered by p_1), and $\ell = k - n_2 = 2k + j - n - 1$ leaf patterns (for the leaves not in p_2). Putting $m = e = s = n - k + 1$, $t = j$, $d = 0$, and $u = j + k - 1$ into (4.2), we have

$$\begin{aligned}
 &\frac{1}{n-k+1} \binom{n-k+1}{j} \binom{k-1}{2k+j-n-1} \binom{n-1}{n-k} \\
 &= \frac{1}{n} \binom{n}{j, n-j-k+1, n-j-k, 2k+j-n-1}.
 \end{aligned}$$

Letting $i = n - j - k + 1$ yields

$$\frac{1}{n} \binom{n}{i, i-1, j, n-2i-j+1}$$

for the number of trees with j leftmost internal nodes, not counting the root, and $i - 1$ nonleftmost ones. The tree in Figure 1(a), for instance, has one nonleftmost internal node, but no leftmost ones (other than the root). \square

Example 4.6. Looking back at Example 3.2, suppose one wishes to count leftmost leaves separately. Let there be n edges, q internal nodes, i of which are unary, and

$r = n + 1 - q$ leaves, j of which are leftmost. We must, therefore, distinguish between nodes with leftmost leaves, and those without. If x is the number of unary nodes with a lone-leaf child, then we want x occurrences of $\langle \diamond \rangle$, $i - x$ of $\langle \blacktriangle \rangle$, $j - x$ of $\langle \diamond \triangle \infty \rangle$, $\ell = r - j$ loose leaves \diamond , and $n + 1 - i + x - 2j - \ell = q + x - i - j$ of $\langle \blacktriangle \triangle \infty \rangle$, making $m = n - r + 1$, $e = 2n - 2r - i + 2$, $d = s = n - r - i + 1$, $t = n - j - r + 1$, and $u = n - j$. Summing over x , we get

$$\begin{aligned} & \frac{1}{q} \binom{r-1}{r-j} \binom{r-2}{r+i-q-1} \sum_x \binom{q}{x, i-x, j-x, q+x-i-j} \\ &= \frac{1}{q} \binom{r-1}{j-1} \binom{r-2}{q-i-1} \binom{q}{i} \sum_x \binom{i}{x} \binom{q-i}{j-x} \\ &= \frac{1}{q} \binom{q}{i} \binom{q}{j} \binom{r-1}{j-1} \binom{r-2}{q-i-1}. \quad \square \end{aligned}$$

5. Up and down steps. By the standard correspondence between binary trees and Dyck (nonnegative lattice) paths (or *bridges*) [18], the number of paths from $(0, 0)$ to (m, m) not crossing below the baseline (grid diagonal), with exactly j occurrences of the lattice pattern **udu** (a step up, **u**, followed by a step down, **d**, and another step up), as counted in [23], is the same as the number of occurrences of binary trees that have j left leaves with a nonleaf sibling amongst binary trees with m internal nodes. For example, the lattice path in Figure 1(c) has 3 occurrences of **udu**, drawn sideways as $\nearrow \searrow \nearrow$, corresponding to the first (reading left to right) three left leaves in Figure 1(b).

We can count these tree patterns by using (3.1), taking $n_1 = j$ patterns $\langle \diamond \triangle \rangle$ for internal nodes with a left leaf, $n_2 = i$ of $\langle \diamond \diamond \rangle$ for “double” leaves (internal nodes with two leaves), $n_3 = m + 1 - j - 2i$ of $\langle \triangle \diamond \rangle$ for internal nodes with a right leaf, and $n_4 = i - 1$ for the rest $\langle \triangle \triangle \rangle$, for a total of m patterns ($e = 2m$, $d = m - 1$) and

$$(5.1) \quad \frac{1}{m} \binom{m}{j, i, m-2i-j+1, i-1}$$

occurrences. Since the patterns account for all the internal nodes of the tree, no triangle subpattern \triangle can be filled by anything but another binary node.

Summing over all i , we get

$$\begin{aligned} & \frac{1}{m} \binom{m}{j} \sum_i \binom{m-j}{i, i-1, m-2i-j+1} \\ &= \frac{1}{m} \binom{m}{j} \sum_i \binom{m-j}{i} \binom{m-i-j}{i-1} \end{aligned}$$

for the number of binary trees with m binary nodes, j of which have a left leaf child and right nonleaf. For $m = 5$ and $j = 3$, there are $\frac{1}{5} \binom{5}{3} \binom{2}{1} = 4$ such binary trees, including the one portrayed in Figure 1(b).

This enumeration is equivalent to the formula

$$\binom{m-1}{j} M_{m-j-1} = \frac{1}{m-j} \binom{m-1}{j} \sum_i \binom{m-j}{i, i-1, m-2i-j+1}$$

in [23, Thm. 2.1] for Dyck paths with m **u**-steps and m **d**-steps, and including j segments **udu**, where M_{m-j-1} is a Motzkin number (as in Example 3.3).

Alternatively, we can apply (4.1a) with $n_1 = j$ patterns $\langle \diamond \blacktriangle \rangle$ for internal nodes with a leftmost leaf and rightmost nonleaf, $n_2 = i$ double-leaf patterns $\langle \diamond \diamond \rangle$, and $n_3 = m - j - i$ of $\langle \blacktriangle \triangle \rangle$ for the remaining internal nodes, for a total of m patterns ($n = e = 2m$, $d = m - j - i$, $t = m - i$, $s = \ell = 0$, $u = 2m - 2i - j$). Recalling the convention that $\binom{-1}{0} = 1$, we have

$$\begin{aligned} & \frac{1}{m} \binom{m}{j, i, m-j-i} \sum_k \binom{m-j-i}{k} \binom{-1}{m-u+k+1} \\ &= \frac{1}{m} \binom{m}{j, i, m-j-i} \binom{m-j-i}{u-m+1} \\ &= \frac{1}{m} \binom{m}{i, i-1, j, m-2i-j+1}, \end{aligned}$$

as before.

A *low-level* occurrence of a lattice pattern, like **udu**, means that one end of each **u** and **d** step touches the baseline. For example, the lattice path in Figure 1(c) has two low occurrences of **udu** (the first two such occurrences). To count these, we use, this time, the standard correspondence between paths and ordered trees [22], in which the two low occurrences in Figure 1(c) correspond to the two first leaves in Figure 1(a).

For paths from $(0, 0)$ to (n, n) with j low occurrences, we need to count n -edge trees with exactly j nonyoungest (i.e., nonrightmost) leaves sprouting from the root. For example, the pattern $\langle \langle \triangle \infty \infty \rangle \langle \triangle \infty \infty \rangle \diamond \triangle \rangle$ matches a root of degree 4, with the third child having no children, but its two older siblings having children. For a root of degree r , there are, in general, $\binom{r-1}{j}$ such patterns for the different placements of the j leaves (in every place but the last).

Substituting $s = r - j - 1$, $d = s + 1$, and $e = r + s$ into our root formula (3.2), and summing over root degree r , we get

$$\sum_r \frac{2r - 2j - 1}{n + j - 2r + 1} \binom{r-1}{j} \binom{2n - 2j}{n - r - 1}$$

(ignoring the fraction whenever the denominator is 0) for the number of paths with j low-level **udu**'s, as counted in [23, Thm. 3.1]. For example, for $n = 5$ and $j = 2$, we get $\frac{1}{2} \binom{6}{1} \binom{2}{2} + \binom{6}{0} \binom{3}{2} = 6$, including the path in Figure 1(c).

6. Young and old leaves. The number of ordered trees with n edges, i oldest (leftmost) leaves, and j younger (nonleftmost) leaves is given by (4.2), with $n_1 = i$ nodes $\langle \diamond \infty \infty \rangle$ with eldest leaf, $n_2 = t = n + 1 - 2i - j$ nodes $\langle \blacktriangle \infty \infty \rangle$ with nonleaf eldest, $m = e = s = n - i - j + 1$, $d = 0$, and $u = n - i$ (all leaves are accounted for):

$$\begin{aligned} & \frac{1}{n-i-j+1} \binom{n-i-j+1}{i} \binom{i+j-1}{j} \binom{n-1}{n-i-j} \\ (6.1) \quad &= \frac{1}{n} \binom{n}{i, i-1, j, n-2i-j+1} \\ &= \frac{1}{n} \binom{n}{j} \binom{n-j}{i} \binom{n-i-j}{i-1}. \end{aligned}$$

The above formula counts trees, since all nodes are covered by the m patterns and j leaves. It is equivalent to the enumeration in [4, Prop. 2.1], namely,

$$\frac{1}{n} \binom{n}{i} \binom{n-i}{j} \binom{n-i-j}{i-1}.$$

There, a tree-grafting proof, based on [3], and similar to the idea in [7] that has been reused here, is provided. The tree in Figure 1(a) is one of the $\frac{1}{5} \binom{5}{2,1,2} = 6$ five-edge trees with 2 oldest leaves and 2 younger ones.

Summing the formula for all i gives the total number of n -edge trees containing j noneldest leaves:

$$\frac{1}{n} \binom{n}{j} \sum_i \binom{n-j}{i} \binom{n-i-j}{i-1}.$$

By Proposition 2.2, a tree with i leftmost leaves and j nonleftmost ones has $i-1$ nonleftmost interior nodes and $n-2i-j+2$ leftmost ones. Letting $r = n-2i-j+2$ and $s = i-1$, we have that there are

$$\frac{1}{n} \binom{n}{r-1, s, s+1, n-2s-r}$$

trees with n edges, r leftmost interior nodes, and s nonleftmost ones, and that there are a total of

$$\frac{1}{n} \binom{n}{r-1} \sum_s \binom{n-r+1}{s, s+1, n-2s-r}$$

n -edge trees with r leftmost internal nodes and any number of nonleftmost ones.

7. Odd and even levels. The Narayana numbers,

$$\frac{1}{n} \binom{n}{q} \binom{n}{q-1},$$

also happen to count the number of (“bicolored”) n -edge ordered trees with q nodes (leaves or internal) on their odd levels (the root’s children, great-grandchildren, etc.) [24]. (See [5, Thm. 4.3B].) For example, $q = 3$ in Figure 1(a). This enumeration is refined in [5, Thm. 4.4B], where it is shown that there are

$$\frac{1}{r} \binom{r}{\ell} \binom{q}{i} \binom{r-2}{q-i-1} \binom{q-1}{r-\ell-1}$$

trees with q odd nodes, including i leaves, and $r = n+1-q$ even nodes, ℓ of which are leaves. For example, $i = \ell = 2$ for the tree in Figure 1(a).

In [6], we gave the following bijection between n -edge ordered trees with ℓ leaves and those with ℓ internal nodes:

$$\begin{aligned} \langle \rangle^\# &= \langle \rangle, \\ \langle s, t_1, \dots, t_k \rangle^\# &= \langle \langle t_1, \dots, t_k \rangle^\#, s_1, \dots, s_m \rangle, \end{aligned}$$

where

$$s^\# = \langle s_1, \dots, s_m \rangle.$$

Applying that idea to even levels, but leaving odd levels intact, we get a bijection defined as follows:

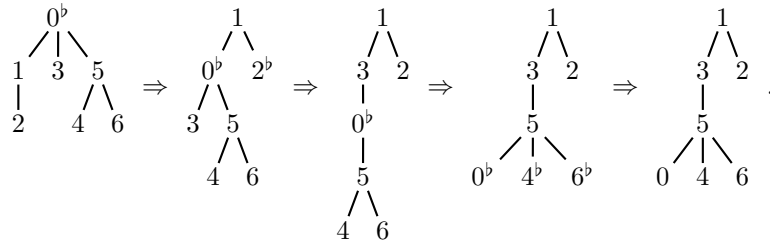
$$(7.1) \quad \begin{aligned} \langle \rangle^b &= \langle \rangle, \\ \langle \langle s_1, \dots, s_m \rangle, t_1, \dots, t_k \rangle^b &= \langle \langle t_1, \dots, t_k \rangle^b, s_1^b, \dots, s_m^b \rangle. \end{aligned}$$

This maps n -edge trees ($n > 0$) with j even internal nodes, ℓ even leaves, and i odd leaves to n -edge trees with j leftmost leaves, ℓ nonleftmost leaves, and i unary nodes. More generally, odd nodes are mapped via this bijection to internal nodes, each with one additional child.

To see the correspondences, consider the following points:

- The above mapping b is applied only to internal nodes from even levels; the recursion continues down the leftmost branch until only a leaf remains. So each even internal node corresponds to a leftmost leaf.
- If one of the s_i is a leaf, then it maps to $s_i^b = \langle \rangle^b = \langle \rangle$, yielding a nonleftmost leaf.
- Each odd node $\langle s_1, \dots, s_m \rangle$ of degree m corresponds to the degree $m + 1$ internal node that results from $\langle \langle t_1, \dots, t_k \rangle^b, s_1^b, \dots, s_m^b \rangle$.

The following example serves to illustrate the process:



Applying the above bijection, we arrive immediately at the conclusion that the formula of Example 4.6, viz.,

$$\begin{aligned} & \frac{1}{q} \binom{q}{i} \binom{q}{r-\ell} \binom{r-1}{\ell} \binom{r-2}{q-i-1} \\ &= \frac{1}{r} \binom{r}{\ell} \binom{q}{i} \binom{r-2}{q-i-1} \binom{q-1}{r-\ell-1} \\ &= \frac{1}{r} \binom{r}{j} \binom{q}{k} \binom{r-2}{k-1} \binom{q-1}{j-1}, \end{aligned}$$

also counts trees with q odd nodes, r even nodes, i odd leaves, j even internal nodes, $k = q - i$ odd internal nodes, and $\ell = r - j$ even leaves, rederiving the enumeration in [5].

There is also a simple bijection between trees that flips odd and even levels, for trees with at least two levels, and preserves the degree of all but two nodes:

$$\begin{aligned} \langle \rangle^\circ &= \langle \rangle, \\ \langle \langle s_1, \dots, s_m \rangle, t_1, \dots, t_k \rangle^\circ &= \langle \langle t_1, \dots, t_k \rangle, s_1, \dots, s_m \rangle. \end{aligned}$$

We have the following theorem.

THEOREM 7.1.

- (1) *The expected degree of even-level nodes among all ordered trees of a given size is exactly 1.*
- (2) *The expected degree of odd-level nodes among all size n ordered trees is $\frac{n-1}{n+1}$.*

Proof. The expected degree of an even node is exactly 1, since

$$\text{average even degree} = \frac{\text{total number of odd nodes}}{\text{total number of even nodes}} = 1.$$

The first equality is by definition; the second is on account of the odd to even bijection. Since there are the same quantities of odd and even nodes, we must have

$$\text{average odd degree} = 2 \times \text{average degree} - \text{average even degree} = 2 \frac{n}{n+1} - 1.$$

The expected degree of an arbitrary node is clearly $\frac{n}{n+1}$, there being $n+1$ nodes in each tree of size n . \square

To summarize, taking Proposition 2.2 into account, we have the following theorem.

THEOREM 7.2.

- (1) *The following distributions in ordered trees of a given size (greater than 0) are identical:*
 - (a) *even-level nodes (per tree);*
 - (b) *odd-level nodes;*
 - (c) *leaves;*
 - (d) *internal nodes.*
- (2) *The following distributions in ordered trees of a given size (greater than 0) are identical:*
 - (a) *even-level leaves;*
 - (b) *younger leaves;*
 - (c) *eldest internal nodes, minus 1.*
- (3) *The following distributions in ordered trees of a given size (greater than 0) are identical:*
 - (a) *even-level internal nodes;*
 - (b) *eldest leaves;*
 - (c) *younger internal nodes, plus 1.*
- (4) *The following distributions in ordered trees of a given size are identical:*
 - (a) *odd-level nodes of degree d ;*
 - (b) *nodes of degree $d+1$.*

Thus, statistics for the distribution of node degrees can be applied to the degrees of odd nodes—with an offset of 1. For example, since the average degree of an internal node is $2n/(n+1) \approx 2$ [6, Cor. 2.2], the average degree of all odd nodes (leaf or internal) is $2n/(n+1) - 1 = (n-1)/(n+1) \approx 1$.

The above correspondences explain why, in fact, the Narayana numbers count trees with a given number of odd (even) nodes, just as they do trees with a given number of internal nodes (leaves). See Example 4.2.

Since the bijection between odd-level nodes and even-level nodes changes the degree of at most two nodes and can turn at most one leaf into an internal node, we also have the following theorem.

THEOREM 7.3.

- (1) *The following distributions in ordered trees of a given size (greater than 0) are identical, give or take 1:*

- (a) *even-level leaves, or younger leaves (per tree);*
 - (b) *even-level internal nodes, or eldest leaves;*
 - (c) *odd-level leaves;*
 - (d) *odd-level internal nodes;*
 - (e) *eldest internal nodes;*
 - (f) *younger internal nodes.*
- (2) *The following distributions in ordered trees of a given size (greater than 0) are identical, give or take 1:*
- (a) *odd-level nodes of degree d ;*
 - (b) *even-level nodes of degree d .*

We know that leaves and internal nodes have the same distributions and that odd and even nodes also do. Furthermore, the cases in which the odd-even bijection changes an odd-level leaf into an even-level internal node are precisely those when the leftmost level-one node is a leaf. There are exactly C_{n-1} such cases for trees of size n . Accordingly, we have the following theorem.

THEOREM 7.4.

- (1) *The expected number of even-level leaves in a size n ordered tree is*

$$\frac{n+1}{4} - \frac{1}{2} \frac{C_{n-1}}{C_n} = \frac{n+1}{4} - \frac{n+1}{4(2n-1)} = \frac{n+1}{4} \left(1 - \frac{1}{2n-1} \right).$$

- (2) *The expected number of odd-level leaves in a size n ordered tree is*

$$\frac{n+1}{4} + \frac{1}{2} \frac{C_{n-1}}{C_n} = \frac{n+1}{4} \left(1 + \frac{1}{2n-1} \right).$$

- (3) *The expected number of even-level internal nodes in a size n ordered tree is*

$$\frac{n+1}{4} \left(1 + \frac{1}{2n-1} \right).$$

- (4) *The expected number of odd-level internal nodes in a size n ordered tree is*

$$\frac{n+1}{4} \left(1 - \frac{1}{2n-1} \right).$$

8. Up-down-ups, younger leaves, and even leaves. It should come as no surprise at this point that the same number, namely,

$$(\star) \quad \frac{1}{n} \binom{n}{i, i-1, j, n-2i-j+1},$$

counts

- a. ordered trees with n edges, i oldest leaves, and j younger ones (see (6.1) and [4]);
- b. binary trees with n internal nodes, i internal nodes with two leaves, and j internal nodes with a leaf only on the left (see (5.1));
- c. Dyck (nonnegative lattice) paths with n **u**'s, $n+1$ **d**'s, i **udd**'s, and j **udu**'s, where the extra **d** is always placed at the end of the path and cannot affect the **udu** count (cf. section 5 and [23]);
- d. ordered trees with n edges, i even internal nodes, and j even leaves (cf. section 7 and [5]); as well as

e. ordered trees with n edges, $j + 1$ eldest internal nodes, and $i - 1$ younger internal nodes (Example 4.5).

For the correspondences between enumerations (a), (b), (c), just consider the standard correspondences [15, 6, 22], under which a leftmost leaf (or, symmetrically, a rightmost leaf) in an ordered tree, a double leaf in a binary tree, and a lattice-path sequence **udd** all correspond to each other, as do a nonleftmost (or nonrightmost) leaf in an ordered tree, a lone leftmost leaf in a binary tree, and a path sequence **udu**. For the correspondence between (a) and (d), use bijection \flat (see (7.1)), matching even internal nodes with oldest leaves and even leaves with younger leaves. For the correspondence between (a) and (e), recall (from Proposition 2.2) that the number of eldest leaves in a tree equals the number of younger internal nodes plus the root, and (from Theorem 7.2(2)) that younger leaves match up with eldest internal nodes minus the root.

Rewriting formula (\star) in terms of a total of $\ell = i + j$ leaves in an n -edge ordered tree, i of which are leftmost, we get

$$\frac{1}{n} \binom{n}{i} \binom{n-\ell}{i-1} \binom{n-i}{n-\ell}.$$

Summing over all i gives the Narayana numbers,

$$\frac{1}{n} \binom{n}{\ell} \binom{n}{\ell-1},$$

for trees with n edges and ℓ leaves (leftmost or otherwise), as well as for trees with ℓ even nodes [24]. This, in turn (see Example 4.2), adds up to the Catalan number,

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

for size n ordered trees with any number of leaves.

Acknowledgment. We thank the referees for their suggested improvements.

REFERENCES

- [1] F. R. BERNHART, *Catalan, Motzkin, and Riordan numbers*, Discrete Math., 204 (1999), pp. 73–112.
- [2] E. C. CATALAN, *Note sur un problème de combinaisons*, J. Math. Pures Appl., 3 (1838), pp. 111–112.
- [3] W. Y. C. CHEN, *A general bijection algorithm for trees*, Proc. Natl. Acad. Sci. USA, 87 (1990), pp. 9635–9639.
- [4] W. Y. C. CHEN, E. DEUTSCH, AND S. ELIZALDE, *Old and young leaves on plane trees*, European J. Combin., 27 (2006), pp. 414–427; available online from <http://www.math.dartmouth.edu/~sergi/papers/oldleaves.pdf>.
- [5] L. H. CLARK, J. E. MCCANNA, AND L. A. SZÉKELY, *A survey of counting bicoloured trees*, Bull. Inst. Combin. Appl., 21 (1997), pp. 33–45.
- [6] N. DERSHOWITZ AND S. ZAKS, *Enumerations of ordered trees*, Discrete Math., 31 (1980), pp. 9–28.
- [7] N. DERSHOWITZ AND S. ZAKS, *Patterns in trees*, Discrete Appl. Math., 25 (1989), pp. 241–255.
- [8] N. DERSHOWITZ AND S. ZAKS, *The cycle lemma and some applications*, European J. Combin., 11 (1990), pp. 35–40.
- [9] R. DONAGHEY AND L. W. SHAPIRO, *Motzkin numbers*, J. Combin. Theory Ser. A, 23 (1977), pp. 291–301.
- [10] A. DVORETSKY AND T. MOTZKIN, *A problem of arrangements*, Duke Math. J., 14 (1947), pp. 305–313.

- [11] A. ERDÉLYI AND I. M. H. ETHERINGTON, *Some problems of non-associative combinations II*, Edinburgh Math. Notes, 32 (1941), pp. 7–12.
- [12] P. FLAJOLET AND J.-M. STEYAERT, *On the analysis of tree-matching algorithms*, in Proceedings of the 7th Colloquium on Automata, Languages and Programming, J. W. Bakker and J. van Leeuwen, eds., Lecture Notes in Comput. Sci. 85, Springer-Verlag, Berlin, 1980, pp. 208–219.
- [13] H. W. GOULD, *Catalan and Bell Numbers: Research Bibliography of Two Special Number Sequences*, 6th ed., Mathematica Monongaliae 12, Combinatorial Research Institute, Morgantown, WV, 1985.
- [14] F. HARARY, G. PRINS, AND W. T. TUTTE, *The number of plane trees*, Indag. Math., 26 (1964), pp. 319–329.
- [15] D. E. KNUTH, *The Art of Computer Programming, Vol. 1: Fundamental Algorithms*, Addison-Wesley, Reading, MA, 1968.
- [16] S. GOPAL MOHANTY, *Lattice Path Counting and Applications*, Academic Press, New York, 1979.
- [17] T. VENKATA NARAYANA, *Sur les treillis formés par les partitions d’un entier et leurs applications à la théorie des probabilités*, C. R. Acad. Sci. Paris, 240 (1955), pp. 1188–1189.
- [18] J. RIORDAN, *Enumeration of plane trees by branches and endpoints*, J. Combin. Theory Ser. A, 19 (1975), pp. 215–222.
- [19] G. ROTE, *Binary trees having a given number of nodes with 0, 1, and 2 children*, Sémin. Lothar. Combin., 38 (1996); available online from http://www.mat.univie.ac.at/~slc/wpapers/s38pr_rote.pdf, 6 pp.
- [20] J. A. VON SEGNER, *Enumeratio modorum, quibus figurae planae rectilineae per diagonales dividuntur in triangula*, Novi Comm. Acad. Scient. Imper. Petropolitanae, 7 (1759), pp. 203–209.
- [21] N. J. A. SLOANE, *The on-line encyclopedia of integer sequences*, <http://www.research.att.com/~njas/sequences> (2006).
- [22] R. P. STANLEY, *Enumerative Combinatorics, Vol. I*, Wadsworth & Brooks/Cole, Monterey, CA, 1986.
- [23] Y. SUN, *The statistic “number of udu’s” in Dyck paths*, Discrete Math., 287 (2004), pp. 177–186.
- [24] W. T. TUTTE AND F. HARARY, *The number of plane trees with a given partition*, Mathematika, 11 (1964), pp. 99–101.