Applied Mathematics Letters 25 (2012) 1010–1015

Contents lists available at SciVerse ScienceDirect

Applied Mathematics Letters

journal homepage: www.elsevier.com/locate/aml

The uplift principle for ordered trees☆

Gi-Sang Cheon [a], Louis Shapiro [b,*]

[a] Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Republic of Korea

[b] Department of Mathematics, Howard University, Washington, DC 20059, USA

A R T I C L E I N F O

Article history:
Received 14 November 2011
Accepted 14 November 2011

Keywords:
Ordered tree
Uplift principle
Mutator
Riordan matrix

A B S T R A C T

In this paper, we describe the uplift principle for ordered trees which lets us solve a variety of combinatorial problems in two simple steps. The first step is to find the appropriate generating function at the root of the tree, the second is to lift the result to an arbitrary vertex by multiplying by the leaf generating function. This paper, though self contained, is a companion piece to Cheon and Shapiro (2008) [2] though with many more possible applications. It also may be viewed as an invitation, via the symbolic method, to the authoritative 800 page book of Flajolet and Sedgewick (2009) [8]. Our examples, with one exception, are different from those in this excellent reference.

1. Introduction

By a tree, we will mean an ordered tree and we will use an $\times$ to mark the root. They will be drawn going up. The number of trees with n edges is $C_n = \frac{1}{n+1}\binom{2n}{n}$, the nth Catalan number. The generating function for these is:

$$C = C(z) = \sum_{n\geq 0} C_n z^n = 1 + z + 2z^2 + 5z^3 + \cdots = \frac{1-\sqrt{1-4z}}{2z}.$$

For much more information about the various number sequences that occur here refer to [1]. We list the A numbers of the various sequences that occur to facilitate using that reference. For instance, the Catalan numbers are sequence A000108 in [1].

If we consider trees with one of its $n+1$ vertices marked, we have

$$(n+1)\frac{1}{n+1}\binom{2n}{n} = \binom{2n}{n}$$

possible trees. The generating function for such trees with a marked vertex, which we call a *mutator*, is

$$B = B(z) = \sum_{n\geq 0}\binom{2n}{n} z^n = 1 + 2z + 6z^2 + 20z^3 + \cdots = \frac{1}{\sqrt{1-4z}}. \quad [A000984]$$

For classes of trees with uniform conditions on updegrees e.g., ordered trees, even trees, $0 \cdot 1 \cdot 2$ or Motzkin trees, complete or incomplete binary trees, complete or incomplete ternary trees, hex trees, we get the basic equation

$$V = LT \tag{1}$$

where V, L and T are generating functions for trees with the mutator m, for trees with a marked leaf vertex and for the number of trees respectively, see [2]. The hex trees arise from putting benzene molecules together with no three meeting

☆ This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-0003187).

* Corresponding author.
E-mail addresses: gscheon@skku.edu (G.-S. Cheon), lshapiro@howard.edu (L. Shapiro).

0893-9659/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2011.11.018

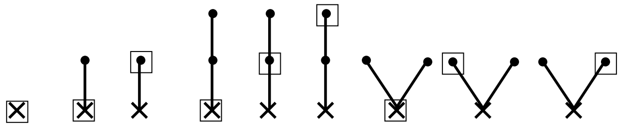

Fig. 1. Ordered trees with the mutator marked with a square.

at point. As trees these are trees with every vertex having outdegree 0,1, or 2 but the 1 case can be left, vertical, or right but if two they must be left and right edges, see [3].

Here is the idea of the proof for (1). Consider a tree with a mutator marked m. Snip the tree in two parts at the mutator. Then we can easily get $V = LT$ since the mutator just became a leaf and the top tree has the same updegree conditions.

For the class of all ordered trees, $V = B$ and $T = C$ so that

$$L = \frac{B}{C} = 1 + z + 3z^2 + 10z^3 + 35z^4 \cdots . \quad [A088218]$$

We consider trees with a marked vertex and we either have the trivial tree consisting of just the root or there is a left most edge, e, at the root. The subtree above e will be denoted U_e and the tree to the right of e will be denoted R_e. Since the marked vertex can be in either U_e or in R_e, we see that $B = 1 + 2CB$. If we classify the marked vertex by height, we get

$$B = C + zC^3 + z^2C^5 + \cdots = \sum_{k=0}^{\infty} z^k C^{2k+1} = \frac{C}{1 - zC^2}.$$

Here the term $z^k C^{2k+1}$ counts vertices at height k with z^k accounting for the k edges on the path from the root to the marked vertex. Then there are $k + 1$ vertices along this path and subtrees can grow on either the left or right side of the path giving C^{2k} possibilities. There can also be one more subtree on top of the marked vertex contributing the last factor C.

The other identity used in this paper is easily shown:

$$\frac{B}{C} = \frac{B+1}{2}.$$

Also recall that

$$[z^n]C^s = \frac{s}{2n+s}\binom{2n+s}{n} \quad \text{and} \quad [z^n]BC^s = \binom{2n+s}{n},$$

where $[z^n]$ is the coefficient extraction operator defined by $[z^n]A = a_n$ for $A = \sum_{n\geq 0} a_n z^n$.

Let $\tilde{B}(z) = B(z^2)$ and let $\tilde{C}(z) = C(z^2)$. If we have paths using n down steps and $n + s$ up steps with $U = (1, 1)$ and $D = (1, -1)$ then the number of paths is $\binom{2n+s}{n}$ and the appropriate generating function is $\tilde{B}(z\tilde{C})^s$ with $\tilde{B}$ counting paths up to the last visit to the x-axis, $\tilde{B}(z\tilde{C})$ counting paths until the last visit to the line $y = 1$, $\tilde{B}(z\tilde{C})^2$ counting paths until the last visit to the line $y = 2$, and so on. This gives the result for BC^s and the other result about C^s is the ballot number problem somewhat disguised and can be proved via Andre's reflection principle. See [4–7].

2. The uplift principle

We will look at ordered trees again and with $L = B/C$ in hand, we can examine some question at the root and then uplift to an arbitrary vertex. The uplift idea has two steps.

The uplift principle First, find the generating function for whatever is being counted at the root. Then uplift the result at the root to an arbitrary vertex by multiplying by the leaf generating function L.

We will examine some examples to illustrate the uplift principle. It leads to several integer sequences and seems to give new combinatorial interpretations for these numbers.

Example 2.1 (*Mutator*)**.** Let us consider trees with a mutator. For instance, see Fig. 1. Suppose all the points above the mutator including the mutator are infected. How many infected points are there? We could say these infected points are new type or superior type or polluted or unpolluted points depending on the context.

Step 1. There are $B = \sum_{n\geq 0}\binom{2n}{n} z^n = \frac{1}{\sqrt{1-4z}}$ vertices counting the root and all the vertices above the root.

Step 2. Pick a point anywhere in the tree to be the mutator and multiply by $L = B/C$. By the uplift principle the generating function is

$$\frac{B}{C} \cdot B = \frac{B^2}{C} = 1 + 3z + 11z^2 + 42z^3 + \cdots . \quad [A032443] \qquad (2)$$

Since $\frac{B}{C} = \frac{B+1}{2}$, we can write this as

$$B \cdot \frac{B+1}{2} = \frac{B^2+B}{2} = \frac{1}{2}\sum_{n\geq 0}\left\{4^n + \binom{2n}{n}\right\} z^n.$$

Hence the expected number of infected points is

$$\frac{[z^n](B^2+B)/2}{[z^n]B} = \frac{1/2\left\{4^n+\binom{2n}{n}\right\}}{\binom{2n}{n}}.$$

Stirling's formula gives us $4^n \sim \sqrt{\pi n}\binom{2n}{n}$ so that as n increases, the expected number of points approaches $\frac{1}{2}(\sqrt{\pi n}+1) \sim \frac{\sqrt{\pi n}}{2}$.

What if we require that the mutator has to be on the rightmost path from the root?

The leaf generating function when the mutator is on the rightmost path is just C since any ordered tree has a unique leaf on the rightmost path. The number of trees with this right mutator condition is C^2 and the expected number of infected points is

$$\frac{[z^n]BC}{[z^n]C^2} = \frac{\binom{2n+1}{n}}{\frac{2}{2n+2}\binom{2n+2}{n}} = \frac{n+2}{2}.$$

This seems unintuitive as $\frac{n+2}{2}$ becomes much larger than $\frac{\sqrt{\pi n}}{2}$. If every mutator is on a path, why is the rightmost path special? The answer is that the rightmost path is short on average and mutators near the root generate more infected points.

What if we reverse the conditions so that any vertex can be the mutator but only those points on the rightmost path from the mutator are infected?

Assume that the mutator is the root and the rightmost path has 2 edges; then there are 3 infected points. The number of trees with a rightmost path of length 3 has the generating function z^2C^2. Summing over all possible lengths of the rightmost path gives us

$$1 + 2zC + 3(zC)^2 + \cdots = \frac{1}{(1-zC)^2} = C^2.$$

If the rightmost path starts at the origin and we are counting vertices, the average length is the same as the number of infected vertices and is

$$\frac{[z^n]C^2}{[z^n]C} = \frac{\frac{2}{2n+2}\binom{2n+2}{n}}{\frac{1}{2n+1}\binom{2n+1}{n}} = \frac{(4n+2)}{(n+2)} \longrightarrow 4.$$

Next we can "uplift" from the origin to an arbitrary vertex by multiplying by B/C to get the generating function

$$BC = \sum_{n\geq 0}\binom{2n+1}{n} z^n = 1 + 3z + 10z^2 + 35z^3 \cdots. \quad [A088218]$$

Thus the expected number of right path infected points is

$$\frac{[z^n]BC}{[z^n]B} = \frac{\binom{2n+1}{n}}{\binom{2n}{n}} = \frac{2n+1}{n+1} \longrightarrow 2.$$

The ratio of the limits is of course $2:1$. We mentioned earlier that the uplift idea applied to any class of ordered trees with the uniform updegree requirement. Here is a table comprising these right path results for a variety of such classes. The proofs are omitted but are of a similar nature.

Class	Root mutator	Arbitrary mutator	Ratio
Ordered trees	4	2	$2:1$
$0 \cdot 1 \cdot 2$ Motzkin trees	6	3	$2:1$
Incomplete binary trees	3	2	$3:2$
Even trees	3	3/2	$2:1$
Oldest child syndrome trees	$2+\sqrt{2}$	$\frac{2}{3}\left(2+\sqrt{2}\right)$	$3:2$
Path	n	$n/2$	$2:1$

Even trees have all vertices with even updegree, and that oldest child trees have the oldest child of each vertex as either spoiled or not, (or either red or green).

Example 2.2 (*Vertices by Updegree*). Let us consider all ordered trees with n edges. What is the generating function counting all vertices of updegree k ?

If $k = 0$, clearly it is the leaf function $L = B/C$. Let $k \geq 1$. Since the generating function is $(zC)^k$ when k is updegree at the root, by the uplift principle, the generating function for all vertices with updegree k is

$$\frac{B}{C} \cdot (zC)^k = z^k BC^{k-1}.$$

Then we have

$$[z^n]z^k BC^{k-1} = [z^{n-k}]BC^{k-1} = \binom{2(n-k)+k-1}{n-k} = \binom{2n-k-1}{n-1}.$$

Example 2.3 (*Twigs*). We define a *twig* to be a vertex with 2 children and no grandchildren. What is the generating function counting all twigs in all ordered trees with n edges?

For step 1, we figure out the generating function at the root. The only possibility is a twig with generating function z^2. We uplift to the general case by multiplying by the leaf function $L = \frac{B}{C}$. Thus by the uplift principle, the generating function is

$$z^2 \frac{B}{C} = z^2 \frac{B+1}{2} = z^2 + z^3 + 3z^4 + 10z^5 + 35z^6 + \cdots. \quad [A088218]$$

If t_n is the number of twigs from all trees with n edges, then $t_n = \frac{1}{2}\binom{2(n-2)}{n-2}$ for $n \geq 2$. We note that the average number of twigs is

$$\frac{t_n}{C_n} = \frac{\frac{1}{2}\binom{2n-4}{n-2}}{\frac{1}{n+1}\binom{2n}{n}} = \frac{n+1}{2} \cdot \frac{n(n-1)n(n-1)}{(2n)(2n-1)(2n-2)(2n-3)}$$
$$= \frac{(n+1)n(n-1)}{8(2n-1)(2n-3)} \longrightarrow \frac{n}{32}.$$

Informally, when n is large, 32 new edges get you about one more twig.

Example 2.4 (*Children vs. Grandchildren*). Does the average vertex in the average ordered tree have more children or more grandchildren?

A child at the root produces the generating function zC^3 with the z for the edge connecting the root to the child, and the three possible subtrees are to the left of this edge, to the right of this edge, and on top of this edge. Then the uplift principle gives

$$\frac{B}{C} \cdot zC^3 = zBC^2 = \sum_{n\geq 0} \frac{n}{n+1}\binom{2n}{n} = z + 4z^2 + 15z^3 + 56z^4 \cdots. \quad [A001791]$$

Since $[z^n]zC^3 = \frac{3n}{n+2}C_n$, the average number of children of the root is $\frac{3n}{n+2} \to 3$ as $n \to \infty$. The result which also gives the average number of returns to the x-axis for Dyck paths is well known and one recent reference is [8, pages 174 and 632]. But the average number of children is $\frac{n}{n+1}\binom{2n}{n} / \binom{2n}{n} = \frac{n}{n+1} \to 1$ as $n \to \infty$.

An alternate approach is to note that a tree with n edges has $n+1$ vertices and all but the root are someone's child. Thus there are $\frac{n}{n+1}$ children per vertex on average.

Similarly, counting grandchildren at the root leads to the generating function z^2C^5. Since

$$[z^n]z^2C^5 = \frac{5n(n-1)}{(n+2)(n+3)}\binom{2n}{n},$$

the average number of grandchildren of the root is $\frac{5n(n-1)}{(n+2)(n+3)} \to 5$ as $n \to \infty$.

Thus the root definitely has more grandchildren than children on average. The situation changes after the uplift to an arbitrary vertex. The uplift gives us the generating function

$$\frac{B}{C} \cdot z^2C^5 = z^2BC^4 = \sum_{n\geq 2}\binom{2n}{n-2} z^n = z^2 + 6z^3 + 28z^4 + 120z^5 + \cdots. \quad [A002694]$$

Since

$$[z^n]z^2BC^4 = \binom{2n}{n-2} = \frac{n(n-1)}{(n+1)(n+2)}\binom{2n}{n},$$

the average number of grandchildren is $\frac{n}{n+1} \cdot \frac{n-1}{n+2} \to 1$ as $n \to \infty$.

There are three possibly surprising conclusions here. One is that there are fewer grandchildren than children. Also that both limits tend to one. Third is that all these limits are integers.

3. Riordan group elements

In several combinatorial counting problems, we have seen the appearance of an element in the Riordan group, for example see [2,9].

This, briefly, is a group of infinite lower triangular matrices defined by two generating functions:

$$\begin{aligned} &g(z) = g = 1 + g_1 z + g_2 z^2 + \cdots \quad \text{and} \\ &f(z) = f = f_1 z + f_2 z^2 + \cdots \quad \text{with } f_1 \neq 0. \end{aligned}$$

Then $\ell_{n,k} = [z^n](gf^k)$ and $L = [\ell_{n,k}]_{n,k\geq 0}$ is a Riordan matrix which is denoted by (g, f). Multiplying (g, f) by a column vector $(h_0, h_1, \ldots)^T$ with the generating function $h(z)$, we obtain the generating function $g(z)h(f(z))$ for the resulting column vector. Simply we write $(g, f)h = gh(f)$ and we call this the *fundamental property for the Riordan matrix*. This property leads to the group operation for the Riordan group which is just a matrix multiplication and is expressed as

$$(g(z), f(z))(G(z), F(z)) = (g(z)G(f(z)), F(f(z))).$$

As observed in Example 2.1, the generating function for the total number of infected points above the mutator including the mutator is $\frac{B^2}{C}$, see (2). This same sequence occurs as the main diagonal in the 2^n enhanced version of Pascal's rectangle ([A032443]) :

$$\begin{bmatrix} \underline{1} & 1 & 1 & 1 & 1 & \\ 2 & \underline{3} & 4 & 5 & 6 & \\ 4 & 7 & \underline{11} & 16 & 22 & \cdots \\ 8 & 15 & 26 & \underline{42} & 64 & \\ 16 & 31 & 57 & 99 & \underline{163} & \\ & & \cdots & & & \end{bmatrix}.$$

After some rearranging, we have the Riordan matrix given by

$$\left(\frac{B^2}{C}, zC\right) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 11 & 4 & 1 & 0 & 0 \\ 42 & 16 & 5 & 1 & 0 \\ 163 & 64 & 22 & 6 & 1 \\ & & \cdots & & \end{bmatrix}.$$

If we want to carry more information, we can classify vertices by height above the root. To have a mutator at height k, we replace L by $(zC^2)^k$. This gives us another Riordan matrix

$$\mathbb{V} = (v_{n,k})_{n,k\geq 0} = (C, zC^2) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & 1 & 0 & 0 \\ 5 & 9 & 5 & 1 & 0 \\ 14 & 28 & 20 & 7 & 1 \\ & & \cdots & & \end{bmatrix}$$

where $v_{n,k}$ is the number of vertices at height k among all trees with n edges. Similarly, if we classify leafs by height we obtain the Riordan matrix

$$\mathbb{L} = (l_{n,k})_{n,k\geq 0} = (1, zC^2) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 5 & 4 & 1 & 0 \\ 0 & 14 & 14 & 6 & 1 \\ & & \cdots & & \end{bmatrix}$$

where $l_{n,k}$ is the number of leaves at height k among all trees with n edges. The matrix version of $V = TL$ is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & 1 & 0 & 0 \\ 5 & 9 & 5 & 1 & 0 \\ 14 & 28 & 20 & 7 & 1 \\ & & \cdots & & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 & 0 \\ 5 & 2 & 1 & 1 & 0 \\ 14 & 5 & 2 & 1 & 1 \\ & & \cdots & & \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 5 & 4 & 1 & 0 \\ 0 & 14 & 14 & 6 & 1 \\ & & \cdots & & \end{bmatrix}.$$

Example 2.2 also provides an interesting Riordan matrix with combinatorial implications. We have

$$\left(\frac{B}{C}, zC\right) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 \\ 10 & 6 & 3 & 1 & 0 \\ 35 & 20 & 10 & 4 & 1 \\ & & \cdots & & \end{bmatrix}$$

where (n, k)-entry is the number of vertices of ordered trees with n edges and updegree k (also see [A100100]).

We can use this matrix to compute many other statistics concerning ordered trees. If, for instance, we wanted to know how many vertices have at least 2 children, we could start with

$$\left(\frac{B}{C}, zC\right)(0, 0, 1, 1, \ldots)^T.$$

Applying the fundamental property yields

$$\begin{aligned}\left(\frac{B}{C}, zC\right)\frac{z^2}{1-z} &= \frac{B}{C}\frac{(zC)^2}{1-zC} = z^2 B\frac{C}{1-zC} = z^2BC^2 \\ &= z^2 + 4z^3 + 15z^4 + \cdots = \sum_{n\geq 2}\binom{2n-2}{n-2}z^n.\end{aligned}$$

The probability that a randomly chosen vertex from a randomly chosen ordered tree with n edges has updegree at least 2 is

$$\frac{\binom{2n-2}{n-2}}{\binom{2n}{n}} = \frac{n-1}{2(2n-1)} \to \frac{1}{4}$$

as $n \to \infty$.

Acknowledgment

We thank the referee for some quite helpful comments.

References

[1] N.J.A. Sloane, The on-line encyclopedia of integer sequences, http://www.research.att.com/~njas/sequences.
[2] G.-S. Cheon, L.W. Shapiro, Protected points in ordered trees, Appl. Math. Lett. 21 (2008) 516–520.
[3] F. Harary, R.C. Read, The enumeration of tree-like polyhexes, Proc. Edinb. Math. Soc. (2) 17 (1970) 1–13.
[4] R.L. Graham, D.E. Knuth, O. Patashnik, Concrete Mathematics, second ed., Addison-Wesley Publishing Company, 1994, p. 201.
[5] J. Riordan, Combinatorial Identities, Wiley, 1968, pp. 153–154 (problem 2).
[6] H.S. Wilf, Generatingfunctionology, Academic Press, Harcourt Brace Jovanovich, 1990.
[7] R.P. Stanley, Enumerative Combinatorics, vol. 2, Cambridge Univ. Press, 1999.
[8] P. Flajolet, R. Sedgewick, Analytic Combinatorics, Cambridge University Press, 2009.
[9] L.W. Shapiro, S. Getu, W.-J. Woan, L. Woodson, The Riordan group, Discrete Appl. Math. 34 (1991) 229–239.