J. Korean Math. Soc. **56** (2019), No. 5, pp. 1247–1263
https://doi.org/10.4134/JKMS.j180503
pISSN: 0304-9914 / eISSN: 2234-3008

PICK TWO POINTS IN A TREE

Hana Kim and Louis W. Shapiro

Dedicated to the loving memory of our friend and colleague, Hana Kim (1987-2018)

Abstract. In ordered trees, two randomly chosen vertices are said to be dependent if one lies under the other. If not, we say that they are independent. We consider several classes of ordered trees with uniform updegree requirements and find the generating functions for the trees with two marked dependent/independent vertices. As a result, we compute the probability for two vertices being dependent/independent. We also count such trees by the distance between two independent vertices.

1. Introduction

In this paper we pick two points in an ordered tree and ask various questions. We start with ordered trees and can ask, if the two points are picked at random, what is the chance one will lie under the other. We will call such a pair of points *dependent* and will use the same term for the trees in which they lie. Otherwise we use the term *independent*[1]. We draw our ordered trees going up and use the terms points and vertices interchangeably.

One application of this might go as follows. You are examining a large family with, say 1000 edges. There are a small number of vertices where some phenomenon is observed. Say 15 such vertices. Then check to see which of the $\binom{15}{2} = 105$ pairs are dependent. If the proportion is higher than expected, then there might be an indication that the phenomenon is hereditary.

The questions we will ask in this paper are as follows.

- What is the proportion of ordered trees with two marked dependent/independent vertices?

Received July 25, 2018; Accepted October 11, 2018.

2010 *Mathematics Subject Classification.* 05A15, 05A19, 60C05.

Key words and phrases. ordered trees, probability, dependent vertices, independent vertices, Riordan matrices.

The first author was supported by the National Research Foundation of Korea funded by the Korea government (NRF-2016R1C1B1014356 and MSIP-2016R1A5A1008055).

[1]Note that our definition of dependence/independence applies only to ordered trees, not to general graphs or trees.

- When do the two independent vertices meet at a specific height? The root is at height 0 and by meeting point we mean the highest point beneath the two vertices that were selected.

We use only a few simple tools such as the $V = TL$ equation and these tools lead quickly to many results where V, L, and T denote the generating functions for trees with a marked vertex, for trees with a marked leaf, and for the number of trees respectively, see [2].

We generalize the prime example of ordered trees to trees that have the same updegree possibilities at each vertex. This is a known concept but with various names. We use *UUR* as in uniform updegree requirement. In a more limited context [3,4,10] the terms simply generated, simple variety, and S-tree are used.

One well known example is Motzkin trees where the updegree of each vertex is $0, 1$, or 2 with 0 indicating a leaf.

The *A-sequence* of a family of trees is $(a_0, a_1, a_2, a_3, \ldots)$ where a_k is the number of possibilities for updegree k. Thus $a_k = 0$ means not allowed, $a_k = 1$ means we can have updegree k, $a_k = 2$ means that all the k edges up from that vertex are all red or all green. Arbitrary weights are allowed. Here is a short list for the A-sequences of trees with uniform conditions on updegrees.

Type of UUR tree	A-sequence
Ordered tree	$1, 1, 1, 1, 1 \ldots$
Motzkin tree	$1, 1, 1, 0, 0, \ldots$
incomplete binary tree	$1, 2, 1, 0, 0, \ldots$
complete binary tree	$1, 0, 1, 0, 0, \ldots$
even tree	$1, 0, 1, 0, 1, 0, \ldots$
incomplete ternary tree	$1, 3, 3, 1, 0, 0, \ldots$
complete ternary tree	$1, 0, 0, 1, 0, 0, 0, \ldots$
Schröder tree	$1, 2, 2, 2, 2, \ldots$
spoiled child tree	$1, 2, 1, 1, 1, \ldots$
Hex tree	$1, 3, 1, 0, 0, \ldots$
Gamma tree	$1, 0, 1, 1, 1, \ldots$
simple tree (path)	$1, 1, 0, 0, 0, \ldots$

We will call the generating function for the A-sequence the *updegree function* and we denote it as $A(z)$ or simply A.

We observe that a tree with n edges has $(n+1)$ vertices so $V = \sum_{n\geq 0}(n+1)t_n z^n$ where t_n is the number of trees with n edges. Thus $V = (zT)' = T + zT'$.

In [1], it is shown that two vertex-related statistics of UUR trees can be expressed as so-called Riordan matrices. A *Riordan matrix* is an infinite lower triangular matrix whose kth column is defined by the generating function $g(z)f(z)^k$ for $g(z), f(z) \in \mathbb{C}[[z]]$ such that $f(0) = 0$. The Riordan matrix is denoted by $(g(z), f(z))$ or (g, f). If $B(z)$ is the generating function for the column vector obtained by multiplying a Riordan matrix (g, f) by a column vector $[h_0, h_1, \ldots]^T$ with $h(z) = \sum_{n\geq 0} h_n z^n$, then $B(z) = g(z) \cdot (h \circ f)(z)$.

This is called the *fundamental theorem for Riordan matrices* (*FTRM*) and we will use the notation $B = (g, f) * h$ for $[b_0, b_1, \ldots]^T = (g, f)[h_0, h_1, \ldots]^T$. By the FTRM, the usual matrix multiplication of two Riordan matrices can be computed in terms of generating functions as:

$$(g, f) * (h, \ell) = (g \cdot (h \circ f), \ell \circ f).$$

Indeed, it is proved in [1] that if $v_{n,k}$ ($\ell_{n,k}$, resp.) is the number of vertices (leaves, resp.) at height k over UUR trees with n edges, then the corresponding matrices can be expressed as Riordan matrices as follows:

$$\mathbb{V} := [v_{n,k}]_{n,k\geq 0} = (T, L_1) \quad \text{and} \quad \mathbb{L} := [\ell_{n,k}]_{n,k\geq 0} = (1, L_1)\,, \tag{1}$$

where $L_1 = zT'/V$. It is also shown in [1] that a matrix version of $V = TL$ is given by

$$\mathbb{V} = \mathbb{T}\mathbb{L},$$

where $\mathbb{T} = (T, z)$. It turns out that many statistics related to UUR trees with a dependent/independent pair of vertices can be expressed in terms of Riordan matrices.

2. Dependence and independence of a pair of vertices

In order to count the UUR trees with two dependent/independent vertices by the height of certain distinguishing points, we first refine the leaf function by height.

Let L_k be the generating function for UUR trees with a marked leaf at height k. Then we have $L = L_0 + L_1 + L_2 + \cdots$ since every leaf must be at some height. We also observe that $L_k = (L_1)^k$ from (1). Hence

$$L = \mathbb{L} * \frac{1}{1-z} = (1, L_1) * \frac{1}{1-z} = \frac{1}{1-L_1}$$

and inverting gives

$$L_1 = 1 - \frac{1}{L} = 1 - \frac{1}{V/T} = \frac{V-T}{V} = \frac{(zT)' - T}{V} = \frac{zT'}{V}.$$

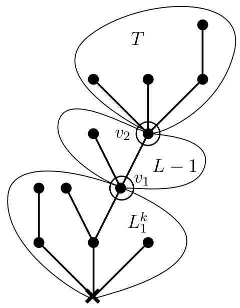
We also find that $L_1 = \sum_{j\geq 1} a_j \binom{j}{1} z(zT)^{j-1}$ where we choose one of the j edges to be the leaf and can attach trees to the other $j-1$ edges at the root. Thus $L_1 = z(A' \circ (zT))$.

We can now "uplift" the result at the root to height 1 (or k) by multiplying by L_1 (or $(L_1)^k$). This property will be called the *uplift principle*.

Theorem 2.1. *Let $\mathcal{T}$ be the set of UUR trees and let $d_{n,k}$ be the number of trees with n edges in $\mathcal{T}$ that have two marked dependent vertices, the lower one of which is at level k. Then*

$$[d_{n,k}]_{n,k\geq 0} = (T(L-1), L_1) = (V - T, L_1)\,.$$

Proof. If one vertex lies above the other at level k, our guiding picture is:



Suppose that we mark one vertex v_1 at level k first and then mark another vertex v_2 above v_1. Since two marked vertices should be distinct, we obtain the generating function $L_1^k(L-1)$. Also, since v_2 need not to be a leaf, we multiply $L_1^k(L-1)$ by T for an arbitrary tree rooted at v_2. Then $[d_{n,k}]_{n,k\geq 0} = (T(L-1), L_1)$. $\square$

By the FTRM, the generating function for the number of trees with n edges in $\mathcal{T}$ that have two marked dependent vertices is $(T(L-1), L_1) * \frac{1}{1-z} = T(L-1) \cdot \frac{1}{1-L_1} = T(L-1)L$.

What happens when we choose two independent vertices?, i.e., not one above the other. Each such pair will have a meeting point, the vertex of maximum height that is beneath both points.

Theorem 2.2. *Let $\mathcal{T}$ be the set of UUR trees with the updegree function A and let $m_{n,k}$ be the number of trees with n edges in $\mathcal{T}$ that have two marked independent vertices meeting at height $k \geq 0$. Then*

$$[m_{n,k}]_{n,k\geq 0} = \left(\frac{(zV)^2(A'' \circ (zT))}{2}, L_1\right).$$

Proof. We first consider the case where $k = 0$, i.e., two marked independent vertices meet at the root. If the updegree of the root is n, then there are $\binom{n}{2}$ possibilities to choose two principal subtrees each of which includes a marked vertex. Since the remaining $n-2$ principal subtrees can be arbitrary, we have the generating function $\binom{n}{2}a_n z^n V^2 T^{n-2}$. Summing over all $n \geq 2$, we obtain

$$\begin{aligned}\sum_{n\geq 2}\binom{n}{2}a_n z^n V^2 T^{n-2} &= \frac{L^2}{2}\sum_{n\geq 2} n(n-1)a_n (zT)^n \\ &= \frac{z^2L^2T^2}{2}\cdot \sum_{n\geq 2} n(n-1)a_n(zT)^{n-2} \\ &= \frac{(zV)^2}{2}\cdot (A'' \circ (zT)).\end{aligned}$$

The generating function for the case $k \geq 1$ follows from the uplift principle. $\square$

By the FTRM, the generating function for the total number of UUR trees with two marked independent vertices is

$$\begin{aligned}
(2) \qquad & \left(\frac{(zV)^2(A''\circ(zT))}{2}, L_1\right) * \frac{1}{1-z} \\
&= \frac{(zV)^2(A''\circ(zT))}{2}\cdot\frac{1}{1-L_1} \\
&= \frac{(zV)^2(A''\circ(zT))L}{2}.
\end{aligned}$$

We now have basic equations for UUR trees with one or two marked vertices. For one marked vertex the basic equation is the One Point Equation

$$V = TL \quad\text{or}\quad (zT)' = TL.$$

For two marked vertices we have the Two Point Equation

$$\frac{z(zT)''}{2} = L(L-1)T + \frac{(zV)^2}{2}\cdot(A''\circ(zT))L.$$

This holds since the generating function for the number of UUR trees with two distinct marked vertices is $\frac{1}{2}\sum_{n\geq 2}(n+1)nt_nz^n = \frac{z}{2}(zT)''$. The two terms on the right hand side come from the dependent and independent cases.

The rest of the present paper is devoted to several statistics related to UUR trees with two marked dependent/independent vertices using the results obtained in this section.

Example A. Ordered trees

We first examine the most basic case, ordered trees. Here $T = C = 1 + zC^2$, the generating function for the Catalan numbers. Since $C = \sum_{n\geq 0}\frac{1}{n+1}\binom{2n}{n}z^n$ we have $V = (zC)' = \sum_{n\geq 0}\binom{2n}{n}z^n$. These are the central binomial coefficients with the generating function $B = \frac{1}{\sqrt{1-4z}}$. Since $V = B$ and $T' = C' = BC^2$ we have $L_1 = zT'/V = zC^2$. We use the following facts [5]:

$$\begin{aligned}
[z^n]C^k &= \frac{k}{2n+k}\binom{2n+k}{n}, \\
[z^n]BC^k &= \binom{2n+k}{n}, \\
[z^n]B^{-1} &= [z^n]\frac{1-4z}{\sqrt{1-4z}} = \binom{2n}{n} - 4\binom{2n-2}{n-1}.
\end{aligned}$$

By Theorem 2.1, the number of ordered trees with two marked dependent vertices by the height of the lower vertex starts with

$$\begin{aligned}\left(B-C, zC^2\right) &= \left[\frac{n-k}{n+k+1}\binom{2n}{n-k}\right]\\ &= \begin{bmatrix} 0 &&&&&&& \\ 1 & 0 &&&&&& \\ 4 & 1 & 0 &&&&& \\ 15 & 6 & 1 & 0 &&&& \cdots \\ 56 & 28 & 8 & 1 & 0 &&& \\ 210 & 120 & 45 & 10 & 1 & 0 && \\ 792 & 495 & 220 & 66 & 12 & 1 & 0 & \\ &&& \cdots & \cdots &&& \ddots \end{bmatrix}. \end{aligned} \tag{3}$$

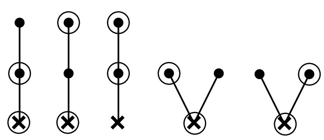
Recall that $L = B/C$ and $B = 1 + 2zCB$ so $L - 1 = \frac{B-1}{2} = zCB$ and $L = \frac{B+1}{2}$. Thus the generating function counting ordered trees with two marked dependent vertices is

$$\begin{aligned} L(L-1)T &= \frac{B^2-1}{4}\cdot C = \frac{B^2-1}{4}\cdot\frac{B-1}{2zB} = \frac{1}{8z}\left(B^2 - 1 - B + \frac{1}{B}\right)\\ &= z + 5z^2 + 22z^3 + 93z^4 + 386z^5 + 1586z^6 + \cdots \qquad (A000346)^2 \end{aligned}$$

and

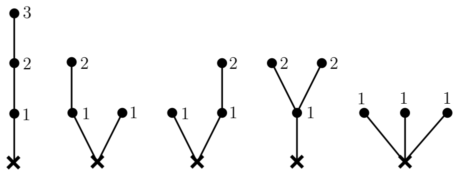
$$\begin{aligned} [z^n]L(L-1)T &= \frac{1}{8}[z^{n+1}]\left(B^2 - 1 - B + \frac{1}{B}\right)\\ &= \frac{1}{8}\left(4^{n+1} - \binom{2n+2}{n+1} + \left(\binom{2n+2}{n+1} - 4\binom{2n}{n}\right)\right)\\ &= \frac{1}{8}\left(4^{n+1} - 4\binom{2n}{n}\right). \end{aligned}$$

To check when $n = 2$ we have 5 possibilities correspond to the row sum of the third row of (3).



This is also the total height of all the vertices. For instance, if $n = 3$ we have

[2]The A number (A******) denotes the corresponding sequence in OEIS [8]. If there is no A number, the sequence is new and doesn't appear in OEIS so far.

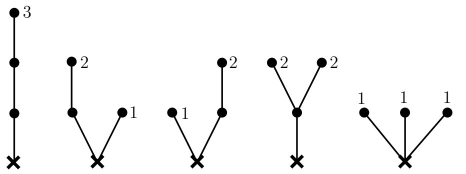


$$6 \quad + \quad 4 \quad + \quad 4 \quad + \quad 5 \qquad + \quad 3 \quad = \quad 22.$$

If we pick a leaf and a vertex beneath the leaf, that is, one of the two marked dependent vertices is a leaf, then we get the more elegant generating function

$$L(L-1) = \frac{B+1}{2} \cdot \frac{B-1}{2} = \frac{B^2-1}{4} = \sum_{n \geq 1} 4^{n-1} z^n.$$

This also gives the total height of all the leaves. Again if $n = 3$ we have



$$3 \quad + \quad 3 \quad + \quad 3 \quad + \quad 4 \qquad + \quad 3 \quad = \quad 4^2.$$

We are now interested in the probability of two randomly chosen vertices being dependent. The generating function for ordered trees with two marked distinct vertices is

$$\begin{aligned} \frac{z}{2}(zC)'' = \frac{z}{2}B' = \frac{z}{2}2B^3 = zB^3 &= z + 6z^2 + 30z^3 + 140z^4 + 630z^5 + \cdots \\ &= \sum_{n \geq 1} \frac{n}{2}\binom{2n}{n} z^n. \qquad (A002457) \end{aligned}$$

The ratio of ordered trees with two marked dependent vertices to all ordered trees with two distinct vertices is

$$\frac{\frac{1}{8}\left(4^{n+1} + 4\binom{2n}{n}\right)}{\frac{n}{2}\binom{2n}{n}} \sim \frac{1}{4n} \cdot \frac{4^{n+1}}{\frac{4^n}{\sqrt{\pi n}}} + \frac{8}{n} = \sqrt{\frac{\pi}{n}} + \frac{8}{n} \sim \sqrt{\frac{\pi}{n}}.$$

For example, when $n = 100$ we have $\sqrt{\frac{\pi}{100}} \doteqdot 0.17725$ while the actual ratio is numerically 0.18747.

On the other hand, since $A'' = 2(1-z)^{-3}$, by Theorem 2.2 the number of ordered trees with $n+2$ edges which have two marked independent vertices meeting at height k starts with

$$[m_{n+2,k}]_{n,k\geq 0} = \left(B^2C^3, zC^2\right) = \begin{bmatrix} 1 & & & & & & \\ 7 & 1 & & & & & \\ 37 & 9 & 1 & & & & \cdots \\ 176 & 56 & 11 & 1 & & & \\ 794 & 299 & 79 & 13 & 1 & & \\ 3473 & 1471 & 470 & 106 & 15 & 1 & \\ & & & \cdots & & & \ddots \end{bmatrix}.$$

We note that the matrix is shifted two rows up by removing z^2 from the first generating function $z^2B^2C^3$.

A routine, but lengthy, computation yields

$$m_{n,k} = \sum_{j=0}^{n-k} \frac{2k+3}{2j+2k+3}\binom{2j+2k+3}{j}4^{n-k-j}.$$

For instance, when two vertices meet at height 1 we have for $n \geq 3$

$$m_{n,1} = \frac{1}{2}\left(4^n - \frac{5n^2+5n+2}{(n+2)(n+1)}\binom{2n}{n}\right).$$

By (2), the generating function for ordered trees with two marked independent vertices is

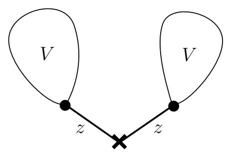
$$\begin{aligned} &\frac{(zB)^2}{2} \cdot \frac{2}{(1-zC)^3} \cdot \frac{B}{C} \\ &= z^2B^3C^2 \\ &= z^2 + 8z^3 + 47z^4 + 244z^5 + 1186z^6 + 5536z^7 + \cdots \quad (A029760) \end{aligned}$$

whose nth coefficient is $\frac{1}{2}\left((n+1)\binom{2n}{n} - 4^n\right)$.

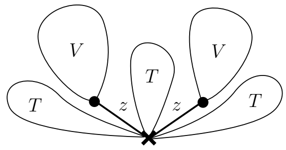
So the probability that two independent vertices meet at height 1 is

$$\frac{m_{n,1}}{\frac{1}{2}\left((n+1)\binom{2n}{n} - 4^n\right)} = \frac{4^n - \frac{5n^2+5n+2}{(n+2)(n+1)}\binom{2n}{n}}{(n+1)\binom{2n}{n} - 4^n} \sim \sqrt{\frac{\pi}{n}}.$$

For ordered trees, we have a simple graphical proof of Theorem 2.2 since the root (also any vertex) can have any degree. First we have the picture that guarantees that the meeting point is the root.



Filling in the rest gives us



and the generating function is $z^2T^3V^2$, i.e.,

$$\begin{aligned} z^2C^3B^2 &= C(zBC)^2 = C\left(\frac{B-1}{2}\right)^2 = \frac{1}{2z}\left(1-\frac{1}{B}\right)\left(\frac{B-1}{2}\right)^2 \\ &= \frac{1}{8z}\left(B^2-3B+3-\frac{1}{B}\right) \\ &= z^2+7z^3+37z^4+176z^5+794z^6+3473z^7+\cdots \quad (A006419). \end{aligned}$$

Thus for $n \geq 1$,

$$\begin{aligned} [z^n]z^2C^3B^2 &= \frac{1}{8}[z^{n+1}]\left(B^2-3B+3-(B-4zB)\right) \\ &= \frac{1}{8}\left(4^{n+1}-4\binom{2n+2}{n+1}+\binom{2n}{n}\right) = \frac{1}{2}\left(4^n-\frac{3n+1}{n+1}\binom{2n}{n}\right). \end{aligned}$$

We end the first example with an additional observation. What is the chance that one of two independent vertices is on the left most branch? By the left most branch we mean the left most principal (rooted at the root of the whole tree) subtree.

Since the generating function counting the total number of independent pairs of points over all ordered trees is $z^2B^3C^2$, the generating function for such independent pairs to both be in the left most branch is

$$\begin{aligned} z(z^2B^3C^2)C &= \left(\frac{B-1}{2}\right)^3 \\ &= z^3+9z^4+57z^5+312z^6+1578z^7+7599z^8+\cdots \quad (A045720). \end{aligned}$$

On the other hand, the generating function for the case where one vertex is in the left most branch while the other vertex is not in the left most branch is

$$\begin{aligned} zB(B-C) &= \left(\frac{B-1}{2}\right)^2 \\ &= z^2+6z^3+29z^4+130z^5+562z^6+2380z^7+\cdots \quad (A008549). \end{aligned}$$

We already showed that

$$[z^n]z^2B^3C^2 = \frac{1}{2}\left((n+1)\binom{2n}{n}-4^n\right) \sim \frac{n}{2}\binom{2n}{n}.$$

Similar computation yields

$$[z^n]\left(\frac{B-1}{2}\right)^3 = \frac{1}{8}\left((2n+4)\binom{2n}{n}-3\cdot 4^n\right) \sim \frac{n}{4}\binom{2n}{n}$$

and

$$[z^n]\left(\frac{B-1}{2}\right)^2=\frac{1}{4}\left(4^n-2\binom{2n}{n}\right)\sim\frac{\sqrt{\pi n}}{4}\binom{2n}{n}.$$

Thus of all pairs of independent vertices, in the limit, one half will have both in the left most branch.

Example B. Oldest Child (or Schröder) trees

The Schröder numbers count the number of lattice paths from $(0,0)$ to $(2n,0)$ using steps $U=(1,1)$, $D=(1,-1)$, and $H=(2,0)$ with the provision that the paths never go below the x-axis. If we denote the generating function as $r(z)$ or r we have

$$\begin{aligned} r=1+zr+zr^2&=\frac{1}{1-z-zr}=\frac{1-z-\sqrt{1-6z+z^2}}{2z}\\ &=1+2z+6z^2+22z^3+90z^4+\cdots. \qquad (A006318)\end{aligned}$$

These are called the large Schröder numbers.

If no horizontal steps are allowed along the x-axis, then we have the small Schröder numbers with generating function

$$s(z)=s=1+zrs=\frac{1}{1-zr}=\frac{r+1}{2}.$$

Note also that $r=\frac{s}{1-zs}$.

If we do allow the paths to go below the x-axis, we have the Delannoy numbers with generating function

$$\begin{aligned} D(z)=D=1+zD+2zrD&=\frac{1}{\sqrt{1-6z+z^2}}\\ &=s+szD+szrD \qquad (4)\\ &=\frac{s}{1-zs+zsr}=\frac{s}{1-zs(r+1)}=\frac{s}{1-2zs^2}.\end{aligned}$$

To see (4), we consider paths which stay above the x-axis with no horizontal steps on the x-axis (s) or such paths followed by a horizontal step (szD) or such paths followed by down step ($szrD$). Similar reasoning leads to $D=r+zr^2D=\frac{r}{1-zr^2}$. All this material is well known and discussed in detail in [10, Ch.6] and with some history in [9].

More novel is the connection with oldest child (Schröder) trees. These are ordered trees where the left most edge from a vertex can be either red or green. In terms of family sociology, an oldest child might be spoiled or not, hence the name. The A-sequence for such trees is $1,2,2,2,\ldots$ so $A(z)=\frac{1+z}{1-z}$.

The one point equation $V=TL$ has $T=r$ and $V=(zr)'=r+zr'$. But $r'=r^2\frac{1+r}{1-zr^2}$ so that

$$V=r+zr^2\frac{1+r}{1-zr^2}=r+zr(1+r)D=r+2zrsD$$

and this equation gives us that

$$L = \frac{V}{r} = 1 + 2zsD = 1 + 2zs\frac{s}{1-2zs^2} = \frac{1}{1-2zs^2}.$$

Then $L - 1 = \frac{2zs^2}{1-2zs^2}$ so

$$\begin{aligned} L(L-1) &= 2z\left(\frac{s}{1-2zs^2}\right)^2 = 2zD^2 \\ &= \frac{2z}{1-6z+z^2} = 2z + 12z^2 + 70z^3 + 408z^4 + \cdots \end{aligned}$$

and this counts all leaf heights.

Incidentally the sequence $1, 6, 35, 204, \dots$ corresponding to $\frac{1}{1-6z+z^2}$ counts many things, see (A001109) in [8]. Among them is the incircle radius for pythagorean triangles where the short sides differ by 1. The series starts with the $3:4:5$ triangle having an incircle of radius 1 while the $20:21:29$ triangle has an incircle of radius 6.

The generating function for trees with two dependent points is

$$\begin{aligned} L(L-1)T &= \frac{2zr}{1-6z+z^2} \\ &= 2z + 16z^2 + 106z^3 + 664z^4 + 4058z^5 + \cdots. \quad (A110099) \end{aligned}$$

For independent points we start with

$$A''(zr) = \frac{4}{(1-zr)^3} = 4s^3.$$

By Theorem 2.2, a lengthy computation of generating functions gives the Riordan matrix defined by simple generating functions

$$\begin{aligned} [m_{n+2,k}]_{n,k\ge 0} &= \left(2sr^2D^2, 2zs^2\right) \\ &= \begin{bmatrix} 2 & & & & & & \\ 22 & 4 & & & & & \\ 176 & 52 & 8 & & & & \cdots \\ 1248 & 468 & 120 & 16 & & & \\ 8330 & 3620 & 1220 & 272 & 32 & & \\ 53678 & 25832 & 10064 & 2992 & 608 & 64 & \\ & & & \cdots & & & \ddots \end{bmatrix}. \end{aligned}$$

For the total, by applying the FTRM to the above matrix but shifted two rows down, we obtain

$$\begin{aligned} \left(2z^2sr^2D^2, 2zs^2\right) * \frac{1}{1-z} &= 2z^2r^2D^2 \cdot \frac{s}{1-2zs^2} = 2z^2r^2D^3 \\ &= 2z^2 + 26z^3 + 236z^4 + 1852z^5 + \cdots. \end{aligned}$$

For instance consider oldest child trees with 7 edges. Then the probable number of pairs of independent vertices with a meeting point at height k are tabulated as follows, accurate to 4 places:

k	0	1	2	3	4	5
$\frac{[z^n]2z^2sr^2D^2(2zs^2)^k}{[z^n]2z^2r^2D^3}$	0.5757	0.2771	0.1079	0.0321	0.0065	0.0007

3. Distance between two vertices

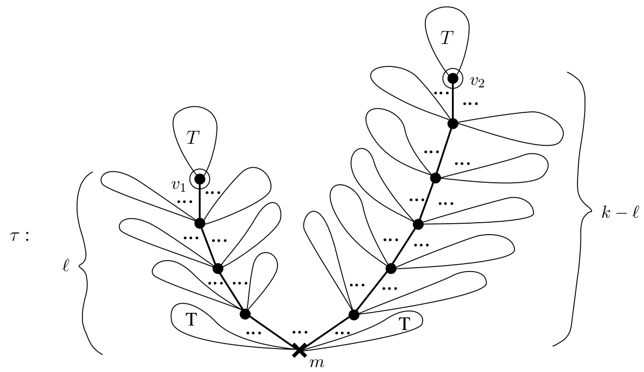
Another question we may ask concerns the distance between two vertices. For complete binary trees, the distance between two vertices measures the similarity between two species in a phylogenetic tree [6, 7].

Theorem 3.1. *Let $\mathcal{T}$ be the set of UUR trees with the updegree function A and let $h_{n,k}$ be the number of trees with $n+2$ edges in $\mathcal{T}$ that have two marked independent vertices at distance $k+2$. Then*

$$\mathbb{V}_{ind} := [h_{n,k}]_{n,k\geq 0} = \left(\frac{LT^2(A''\circ(zT))}{2}, L_1\right)D$$

where $D = \mathrm{diag}(1, 2, 3, \ldots)$.

Proof. Suppose that the two marked independent vertices meet at height 0, i.e., the root is the meeting point m. An UUR tree τ that has such two vertices is of the form



where $1 \leq \ell \leq k-1$ and the other subtrees contribute a T. Let P_1 and P_2 be the paths from the root m of τ to v_1 and v_2, respectively. Consider the bottom subtree of τ that has m as its root and one edge towards each of v_1 and v_2. Then such a subtree contributes $\sum_{j\geq 2}\binom{j}{2}a_j z^2(zT)^{j-2} = \frac{z^2}{2}(A''\circ(zT))$.

On the other hand, consider a tree whose root $v(\neq v_1,\ v_2,\ m)$ is a vertex lying on P_1 or P_2, and has another vertex w on the path as a child of v which is a leaf at the same time. There are $(\ell-1)+(k-\ell-1) = k-2$ such trees and the generating function is $\sum_{j\geq 1}\binom{j}{1}a_j z(zT)^{j-1} = z(A'\circ(zT)) = L_1$.

For each $\ell = 1, 2, \ldots, k-1$ the generating function for τ is

$$\frac{(zT)^2}{2}(A'' \circ (zT))(z(A' \circ (zT)))^{k-2}$$

including the two Ts for trees with roots v_1 and v_2, so total we have

$$(k-1)\frac{(zT)^2}{2}(A'' \circ (zT))(z(A' \circ (zT)))^{k-2}.$$

By multiplying L to uplift the meeting point to any height, and removing z^2 to start from $n = 2$, we obtain the desired result. $\square$

If we remove the two top trees rooted at v_1 and v_2 from the tree τ in the proof of Theorem 3.1, then we obtain the results on two leaves at a given distance. Note that any two distinct leaves are independent.

Corollary 3.2. *Let $\mathcal{T}$ be the set of UUR trees with the updegree function A and let $b_{n,k}$ be the number of trees with $n+2$ edges in $\mathcal{T}$ that have two marked leaves at distance $k+2$. Then*

$$\mathbb{L}_{ind} := [b_{n,k}]_{n,k\geq 0} = \left(\frac{L(A'' \circ (zT))}{2}, L_1\right) D$$

where $D = \mathrm{diag}(1, 2, 3, \ldots)$.

By Riordan multiplication we see that the matrices counting UUR trees with two marked independent vertices and two marked leaves by distance are related as

$$\mathbb{T}^2 \mathbb{V}_{ind} = \mathbb{L}_{ind}$$

which is in contrast to $\mathbb{V} = \mathbb{T}\mathbb{L}$.

Example C. Complete binary trees

For complete binary trees, $A = 1 + z^2$ and $L = B(z^2)$ so the generating function for the complete binary trees with two marked leaves is

$$z^2 L^3 = \frac{z^2}{(1-4z^2)^{3/2}} = z^2 + 6z^4 + 30z^6 + 140z^8 + 630z^{10} + \cdots \quad (A002457)$$

and $[z^n]z^2L^3 = (n-1)!/\left((\frac{n}{2}-1)!\right)^2$ for even $n \geq 2$.

It follows from Corollary 3.2 that the number of complete binary trees on n edges with two marked leaves at distance k is

$$\begin{bmatrix} 1 & & & & \\ 0 & & & & \\ 2 & 2 & & & \\ 0 & 0 & & & \cdots \\ 6 & 6 & 4 & & \\ 0 & 0 & 0 & & \\ 20 & 20 & 16 & 8 & \\ & & & \cdots & \ddots \end{bmatrix}$$

whose (n,k)-entry is

$$[z^n](k-1)2^{k-2}z^{2k-2}\cdot\frac{1}{\sqrt{1-4z^2}}\cdot\left(\frac{1-\sqrt{1-4z^2}}{2z^2}\right)^{k-2}$$

$$= \begin{cases} 0 & \text{if } n \text{ is odd,} \\ (k-1)2^{k-2}\binom{n-k}{\frac{n-2}{2}} & \text{if } n \text{ is even.} \end{cases} \tag{5}$$

For instance, consider complete binary trees with 10 edges. If we pick two leaves randomly then they have about 29% chance of being at distance 4, see Figure 1.

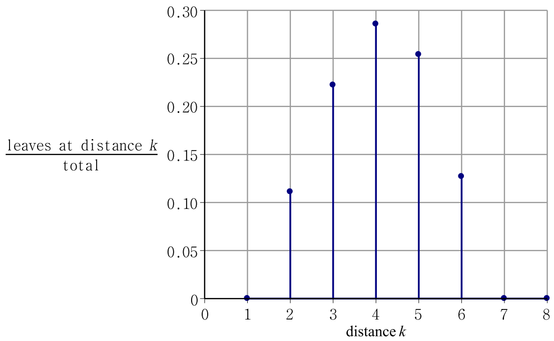


FIGURE 1. Distribution of pairs of leaves in complete binary trees with 10 edges by distance.

Figure 2 represents the ratio of complete binary trees on n edges with two marked leaves at a given distance. For example, the line for the case $k = 4$ passes through $(6, 0.4)$. This means that 40% of pairs of leaves in complete binary trees with 6 edges are at distance 4.

As one can see in Figure 2, the maximum ratio of complete binary trees with two marked leaves at distance k is attained at different values of k depending on the number of edges. Specifically, if we consider complete binary trees with n edges and the maximum ratio occurs at $k_n^{\max}$, then $k_2^{\max} = 2$ and $k_4^{\max} = 3$. For a fixed $n \geq 6$, the ratio function is given by

$$\rho_n(k) = \frac{(k-2)2^{k-3}\binom{n-k+1}{\frac{n-2}{2}}}{(k-1)2^{k-2}\binom{n-k}{\frac{n-2}{2}}} = \frac{1}{2} - \frac{n(k-3)}{2(k-1)(2k-n-4)}.$$

Note that the minimum (maximum, resp.) possible distance between two marked leaves in a complete binary tree with n edges is 2 ($n/2+1$, resp.). The ratio function $\rho_n(k)$ is monotone increasing in the interval $[2, n/2+1]$. To find $k_n^{\max}$, it suffices to find k such that $\rho_n(k) > 1$ which is equivalent to $k^2 - 3k - n + 2 < 0$, or $k < (3+\sqrt{1+4n})/2$. Now we compare the numbers in (5) at two integers $k = (3+\sqrt{1+4n})/2$ and $(1+\sqrt{1+4n})/2$.

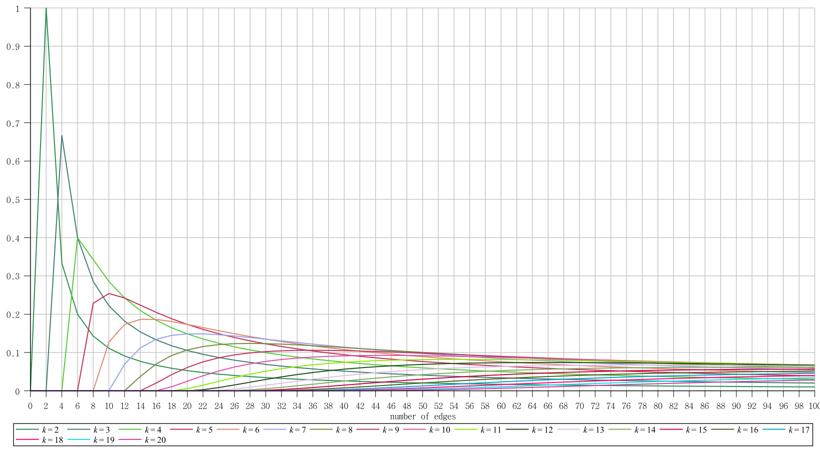


FIGURE 2. Distribution of complete binary trees with two marked leaves at a given distance by the number of edges. Different colors are used to distinguish the distances between two leaves.

If $n = m(m+1)$, then $k = m+2$, and the number in (5) is the same at both $k = m+2$ and $k = m+1$. This implies $k_n^{\max} = m+1, m+2$. On the other hand, $n = (m+1)(m+2)$ gives $k = m+3$ which can be covered by the case we already computed. Further, the number in (5) attains a bigger value at $k = m+2$ than at $k = m+1$. So consequently we obtain

$$k_n^{\max} = \begin{cases} m+1,\ m+2 & \text{if } n = m(m+1) \text{ for } m \geq 2, \\ m+2 & \text{if } m(m+1) < n < (m+1)(m+2) \text{ for } m \geq 2. \end{cases}$$

For example, $k_{10}^{\max} = 4$ means that for complete binary trees with 10 edges the likeliest distance between two leaves is 4.

Theorem 3.3. *Let $\mathcal{T}$ be the set of UUR trees with the updegree function A and let $c_{n,k}$ be the number of trees with $n+1$ edges in $\mathcal{T}$ that have two marked dependent vertices at distance $k+1$. Then*

$$[c_{n,k}]_{n,k\geq 0} = (V, L_1)\,.$$

Proof. Suppose that one of the two marked dependent vertices is the root. We have the same figure for an UUR tree τ that has such two vertices as in the proof of Theorem 3.1 but v_1 and v_2 in τ lie on a single path and there is no meeting point. Then the generating function for the trees with n edges that have two marked dependent vertices at distance k is $TLL_1^k = TL\left(z(A' \circ (zT))\right)^k = V\left(z(A' \circ (zT))\right)^k$ for $k \geq 1$. $\square$

Example D. Motzkin trees

Motzkin trees have the updegree function $A = 1 + z + z^2$ so

$$T = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z^2} \quad \text{and} \quad L = \frac{1}{\sqrt{1 - 2z - 3z^2}}.$$

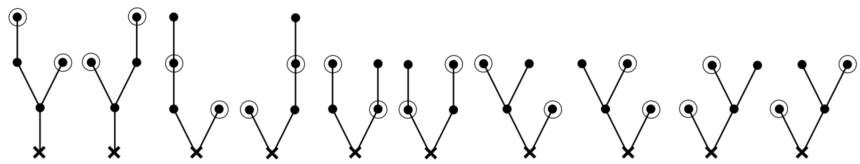
What are the chances of two vertices in a Motzkin tree with n edges having distance 3? By Theorems 3.1 and 3.3, the generating functions for such pairs of independent and dependent vertices are respectively

$$\begin{aligned} \mathcal{I}_3(z) &:= 2(zT)^2(z(1+2zT)) = \frac{1}{z^2}\big(z^2 + 3z - 2 + (3z^3 - 2z^2 - 5z + 2)L\big) \\ &= 2z^3 + \mathbf{10}z^4 + 36z^5 + 120z^6 + 380z^7 + \cdots \end{aligned}$$

and

$$\begin{aligned} \mathcal{D}_3(z) &:= V(z(1+2zT))^3 \\ &= -\frac{1}{2z^2}\big(3z^3 - 10z^2 - 12z + 8 + (-9z^4 - 21z^3 + 14z^2 + 20z - 8)L\big) \\ &= z^3 + 8z^4 + 36z^5 + 132z^6 + 445z^7 + \cdots . \end{aligned}$$

For example, there are 10 Motzkin trees with 4 edges which have two marked independent vertices of distance 3:



Singularity analysis leads to

$$\begin{aligned} \mathcal{I}_3(z) &= \sqrt{3}(1-3z)^{-\frac{1}{2}} + O((1-3z)^{\frac{1}{2}}), \\ \mathcal{D}_3(z) &= \frac{3\sqrt{3}}{2}(1-3z)^{-\frac{1}{2}} + O((1-3z)^{\frac{1}{2}}) \end{aligned}$$

which give $h_{n,3} \sim \sqrt{3}\sqrt{\frac{1}{\pi n}} \cdot 3^n$ and $c_{n,3} \sim \frac{3\sqrt{3}}{2}\sqrt{\frac{1}{\pi n}} \cdot 3^n$. It follows from Theorems 2.1 and 2.2 that the generating function for Motzkin trees with two marked vertices is

$$(zV)^2 L = \frac{3\sqrt{3}}{8}(1-3z)^{-\frac{3}{2}} + O((1-3z)^{-\frac{1}{2}})$$

for the independent case and is

$$V(L-1) = \frac{9}{4}(1-3z)^{-1} + O((1-3z)^{-\frac{1}{2}})$$

for the dependent case. Thus the ratio of two independent vertices at distance 3 to all pairs of independent vertices in Motzkin trees is

$$\frac{h_{n,3}}{\text{total}} \sim \frac{\frac{\sqrt{3}}{\sqrt{\pi n}} \cdot 3^n}{\frac{3\sqrt{3}}{4}\sqrt{\frac{n}{\pi}} \cdot 3^n} = \frac{4}{3n}.$$

For the dependent case, we obtain

$$\frac{c_{n,3}}{\text{total}} \sim \frac{\frac{3\sqrt{3}}{2}\sqrt{\frac{1}{\pi n}} \cdot 3^n}{\frac{9}{4} \cdot 3^n} = \frac{2\sqrt{3}}{3}\sqrt{\frac{1}{\pi n}}.$$

For example, we have $\frac{h_{500,3}}{\text{total}} \sim 0.00266$ and $\frac{c_{500,3}}{\text{total}} \sim 0.02913$ while the actual ratios are to 6 decimal places 0.00303 and 0.03101, respectively.

References

[1] G.-S. Cheon, H. Kim, and L. W. Shapiro, *Profiles of ordered trees with mutation and associated Riordan matrices*, Linear Algebra Appl. **511** (2016), 296–317. `https://doi.org/10.1016/j.laa.2016.09.006`

[2] G.-S. Cheon and L. W. Shapiro, *Protected points in ordered trees*, Appl. Math. Lett. **21** (2008), no. 5, 516–520. `https://doi.org/10.1016/j.aml.2007.07.001`

[3] M. Drmota, *Random Trees*, Springer Verlag, Wien, 2009. `https://doi.org/10.1007/978-3-211-75357-6`

[4] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009. `https://doi.org/10.1017/CBO9780511801655`

[5] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, second edition, Addison-Wesley Publishing Company, Reading, MA, 1994.

[6] N. C. Jones and P. A. Pevzner, *An Introduction to Bioinformatics Algorithms*, MIT Press, 2004.

[7] C. Semple and M. Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and its Applications, **24**, Oxford University Press, Oxford, 2003.

[8] N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, http://oeis.org.

[9] R. P. Stanley, *Hipparchus, Plutarch, Schröder, and Hough*, Amer. Math. Monthly **104** (1997), no. 4, 344–350. `https://doi.org/10.2307/2974582`

[10] ———, *Enumerative Combinatorics. Vol. 2*, Cambridge Studies in Advanced Mathematics, **62**, Cambridge University Press, Cambridge, 1999. `https://doi.org/10.1017/CBO9780511609589`

Hana Kim
Applied Algebra and Optimization Research Center
Sungkyunkwan University
Suwon 16419, Korea

Louis W. Shapiro
Department of Mathematics
Howard University
Washington, DC 20059, USA
Email address: `lshapiro@howard.edu`