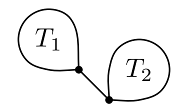
Mike Hendy on Catalan Numbers and Evolutionary Trees

S_0	S_1	S_2	S_3	S_4	S_5	$\cdots$
1	1	2	3	5	8	$\cdots$
2	4	6	10	16	26	$\cdots$
4	7	11	18	29	47	$\cdots$

Mike Hendy of Massey University, Palmerston North, recently reminded me that we each solved a problem brought to New Zealand by a visiting mathematician, Kenneth Stolarsky. Consider the table on the right, where each element in S_1 is the lowest positive integer which has not appeared in a previous row of $S_1 \cup S_2 \cup S_3 \cup S_4 \cup \cdots$, and each element in S_2 is formed by multiplying the S_1 element on the same row by the "Golden Ratio" $\frac{1}{2}(1+\sqrt{5})$ and finding the closest integer. The elements in S_3, S_4 etc are formed using the Fibonacci difference equation $u_n = u_{n-1} + u_{n-2}$. The elements in S_0 are the differences of those in S_1 and S_2. Prove that every positive integer occurs once and only once in $S_1 \cup S_2 \cup S_3 \cup S_4 \cup \cdots$, and that $S_0 = S_1 \cup S_2$.

Mike has sent me the following dissertation on Catalan Numbers and Evolutionary Trees. I am using this unchanged except that I have replaced his figures written in LaTeX picture environment by my own versions written using the package PSTricks because I think this package should be better known. This is where you take over, Mike.

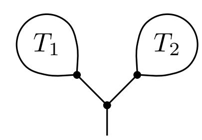
In his most recent visit to New Zealand, Douglas Rogers spent five days in Palmerston North. At my invitation he delivered three guest lectures on Generating Functions to my Honours Course in Discrete Mathematics. The first talk concerned pictorial derivations of generating functions, illustrated by an application of Catalan numbers. Amongst other things, the Catalan number c_n is the number of rooted planar trees with n edges. (The reader may be familiar with the number of legal bracket pairs such as occur when discussing associativity, where c_n is the number of different pairings that can be made in a product of n symbols.)



Douglas' pictorial derivation of the recurrence relation was based on the diagram at the right. For $n = 0$ edges, there is just the trivial tree of one vertex, so $c_0 = 1$. For $n > 0$ let T_1 and T_2 be the subtrees incident to the leftmost edge from the root. If T_1 has k edges, then T_2 has $n - k - 1$ edges. Thus $c_n = \sum_k c_k c_{n-k-1} \quad (n \geq 1; c_0 = 1)$.

(If we use the device of setting $c_n = 0$, for all negative n, then we do not need to set limits to this sum.) Hence the ordinary generating function $C(x) = \sum_n c_n x^n$ satisfies the equation $C(x) = 1 + xC^2(x)$.

This presentation reminded me of my construction of a generating function for the sequence b_n of rooted binary evolutionary trees labelled by $[n] = \{1, 2, \cdots, n\}$. (These are rooted trees, with the leaves labelled by $[n]$, and the other (internal) vertices are each of degree 3. I will refer to these here as $n-$trees.) It is easily seen that $b_1 = 1$ counts the tree of one edge, which joins the root to the leaf 1. We find each $(n+1)-$tree can be obtained from an $n-$tree T, by joining a leaf labelled $n+1$ to an edge of T. This construction increases the number of edges by 2 and internal vertices by 1. Hence each $n-$tree has $2n-1$ edges and $n-1$ internal vertices, giving the recursion $b_{n+1} = (2n-1)b_n, \quad (n \geq 1; b_1 = 1.)$ Thus we have the well known result that $b_n = (2n-3)!!$, the product of the first $n-1$ odd positive integers.

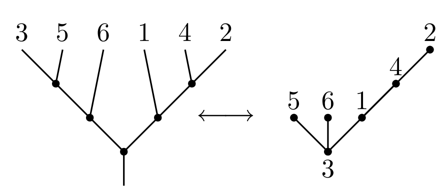


However I wanted to illustrate generating functions to my class. For $n > 1$, an $n-$tree T can be decomposed into two subtrees, a $k-$tree T_1 (labelled by $S \subset [n]$ of order k), and an $(n-k)-$tree T_2 (labelled by $[n] \setminus S$ of order $n-k$) by deleting the root. When we sum over all such pairs of complementary trees and over all $\binom{n}{k}$ labelling subsets S, we count each pairing twice (the trees are not embedded in the plane, so we do not distinguish left from right). Hence we obtain the recursion $b_n = \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k} b_k b_{n-k}$ for $n > 1$ with $b_1 = 1$. This sum corresponds to half the coefficient of x^n in the square of the exponential generating function $B(x) = \sum_n b_n \frac{x^n}{n!}$, and implies $B(x) = x + \frac{1}{2}B^2(x)$.

In comparing these equations Douglas noted that $B(x) = xC(x/2)$, so comparing coefficients this implies $2^{n-1}b_n = n!c_{n-1}$ for $n \geq 1$. A hurried numerical check of the first few values: $b_1 = 1$, $b_2 = 1$, $b_3 = 3$, $b_4 = 15$, $b_5 = 105$ and $c_0 = 1$, $c_1 = 1$, $c_2 = 2$, $c_3 = 5$, $c_4 = 14$, confirmed this observation. Of course numerical agreement demands a constructive bijection.

We then noted that as a planar tree with $n-1$ edges has n distinguished vertices, so $2^n c_{n-1}$ counts the number of planar trees with $n-1$ edges, with the vertices labelled by $[n]$. The $n-$trees (with leaves labelled by $[n]$) are not planar, indeed a Biologist wishing to present such a tree on paper has a left/right choice at each of the $n-1$ internal vertex, hence there are 2^{n-1} distinct embeddings of $n-$trees in the plane.

Thus we seek a bijection between the set C of planar rooted trees of n vertices, with the vertices labelled by $[n]$, and the set B of planar rooted binary trees of n leaves, with the leaves labelled by $[n]$. For a planar $n-$tree in B, this is achieved by shrinking the left edge above each internal vertex of T to a vertex, and deleting the root edge, to produce a tree in C. Conversely, given any tree in C, at each labelled non-leaf vertex v, add a new edge (v, v') up and left, and move all but the rightmost subtree above v (including the labels) to v'. Finally add a root edge below the root of T. This produces a tree in B.



John Butcher butcher@math.auckland.ac.nz