Trees and Paths

Nachum Dershowitz and Shmuel Zaks

Feb. 1998

Let L_n^i be the set of lattice paths from $(0,0)$ to (n,n) that have exactly i steps up ($\uparrow$) and i steps to the right ($\rightarrow$) below the diagonal. We use L_n as short for L_n^0, those paths wholly above the diagonal. Let T_n be the set of ordered (plane planted) trees with n edges and B_n the set of binary trees with n internal nodes. These sets are all in 1-1 correspondence (the L_n^i by the Chung-Feller Theorem) and are counted by the Catalan numbers. In what follows, we assume that they are uniformly distributed.

Define $\bar{h}(L)$ to be the expected height (meaning, distance from the diagonal) of a step in the lattice paths in some set L, $\hat{h}(L)$ be the expected height of a path in L (that is, the maximum height of its steps), $\hat{h}(T)$, the expected height (length of maximum path from root) of a tree in T, and $\bar{h}(T)$, the expected level of a node (length of path from root) in T. Define the *girth at a node* of a tree to be the number of nodes on the same level to its right, plus nodes on the next level that are to its left. Define the *girth of a tree* to be the maximum girth at its internal nodes. Let $g(b)$ denote the girth of tree b, $\hat{g}(B)$ the expected girth of a tree in B, and $\bar{g}(B)$ the expected girth at an internal node in B.

Consider the following correspondence between lattice paths in L_n and binary trees: Traverse the path and build a binary tree in *level-order*. Each step up $\uparrow$ corresponds to an internal node $\circ$ of the binary tree, with left and right slots; each step right $\rightarrow$ corresponds to a leaf $\Box$. Level-order means that the nodes are placed in the next available slot going from left-to-right, moving to a lower level after the current one is filled (see [9]). When we're done, we have one slot that must be filled with a leaf to complete the tree.

More generally, one can build a k-tree forest of binary trees in level-order from a sequence $s_1, \ldots, s_m$ of subtrees with slots. Start with k slots in a row and list the slots from left-to-right in a first-in-first-out queue. Repeatedly place the next tree s_i in the sequence in the slot at the head of the queue, and add the slots in the frontier of s_i (from left-to-right) at the rear of the queue. As long as each of the subsequences $s_1, \ldots, s_i$ has at least $i-k$ slots, we end up with a forest of k trees with $k+s-m$ remaining slots, where s is the total number of slots in the given sequence.

Thus, every path in L_n (contributing $2n$ nodes with a total of $2n$ slots) corresponds to a k-tree forest with k slots (that can be filled with leaves at the end). Let $b_k(l)$ denote the k-tree forest obtained from path l in this manner. The construction can be described recursively: Suppose l has d subpaths $l_1, \ldots, l_d$, each one starting at a point $(i, i+1)$ one step after the path touched the diagonal and ending one step before the path is next at the diagonal, then $b_k(l)$ is the forest obtained by glueing the sequence $\circ$, $b_{k+1}(l_1)$, $\Box$, $\circ$, $b_{k+1}(l_2)$, $\Box$, $\ldots$, $\circ$, $b_{k+1}(l_d)$, and $\Box$ together in level-order.

The girth of the resulting forest $b_k(l)$ is the maximum of the girths of the forests corresponding to the subpaths l_i. By induction on the length of the path, it can be shown that the height of l (which is one more than the height of its tallest subpath) is equal to $g(b_k(l)) - k + 1$, for all k. Thus, the height of l is equal to the girth of the corresponding binary tree $b_1(l)$, and, hence, the distribution of height of lattice paths—and of ordered trees—is the same as the distribution of girth for binary trees:

Theorem 1

$$
\begin{aligned}
\bar{g}(B_n) &= \bar{h}(L_n) = \bar{h}(T_n) \\
\hat{g}(B_n) &= \hat{h}(L_n) = \hat{h}(T_n)
\end{aligned}
$$

It is the case (cf. the argument in [6]) that the average heights $\hat{h}(L_n^i)$ for $i = 0, \ldots, n$ can differ by at most 1. Using the following facts:

$$
\begin{aligned}
\bar{l}(T_n) &= \tfrac{1}{2}\left[4^n/\tbinom{2n}{n} - 1\right] && [11,\ 10,\ 1,\ 5] \\
&\approx \tfrac{1}{2}\left[\sqrt{\pi n} - 1\right] \\
\hat{h}(B_n) &\approx 2\sqrt{\pi n} && [7] \\
\hat{h}(T_n) &\approx \sqrt{\pi n} && [2,\ 7,\ 6]
\end{aligned}
$$

we get

Corollary 1

$$\begin{aligned}
\bar{g}(B_n) &\approx \tfrac{1}{2}\sqrt{\pi n} \\
\hat{g}(B_n) &\approx \sqrt{\pi n} \\
\hat{h}(L_n^i) &\approx \sqrt{\pi n}
\end{aligned}$$

The above result implies that the expected (average and worst case) space requirements for the queue needed to implement a level-order traversal of a binary tree is half the expected requirements for the stack used in a naive preorder traversal, since $\hat{h}(B_n) \approx 2\sqrt{\pi n}$, and the same as an intelligent preorder traversal, in which the parent is removed before the second child is explored.

The *Strahler (register) number* of a binary tree b is the height of the largest complete binary tree homeomorphically embeddable in b. It gives the maximum height of a traversal stack, when the branch with smaller Strahler number is always chosen to be traversed first. See [8]. Using the above correspondence, we can show that the number of binary trees with Strahler number r is equal to the number of lattice paths (or ordered trees) with height greater than $2^r - 2$ and less than $2^{r+1} - 1$.

Since the Strahler number of a binary tree b is equal to

$$\lceil \lg g(b) \rceil$$

it follows that the number of lattice paths with height h is equal to the number of binary trees with Strahler number $\lceil \lg h \rceil$ (and that the number of binary trees with Strahler number r is equal to the number of lattice paths with height 2^r).

Note. Given the opportunity, we would also like to mention some related published results from [3, 4, 5, 6] on lattice paths, tree statistics, and the Narayana numbers

$$\frac{1}{n}\binom{n}{k}\binom{n}{k-1}$$

(which count the number of trees in T_n with exactly k leaves).

References

[1] B. Dasarathy and C. Yang. A transformation on ordered trees. *Comp. J.*, 23(2):161–164, 1980.

[2] N. deBruijn, D. E. Knuth, and O. Rice. The average height of planted plane trees. In R. C. Read, editor, *Graph Theory and Computing*, pages 15–22. Academic Press, New York, 1972.

[3] N. Dershowitz and S. Zaks. Enumerations of ordered trees. *Discrete Mathematics*, 31(1):9–28, July 1980.

[4] N. Dershowitz and S. Zaks. Ordered trees and non-crossing partitions (Note). *Discrete Mathematics*, 62(2):215–218, November 1986.

[5] N. Dershowitz and S. Zaks. Patterns in trees. *Discrete Applied Mathematics*, 25(3):241–255, 1989.

[6] N. Dershowitz and S. Zaks. The Cycle Lemma and some applications. *European J. of Combinatorics*, 11(1):35–40, 1990.

[7] P. Flajolet and A. Odlyzko. The average height of binary trees and other simple trees. *J. of Computer and System Sciences*, 25:171–213, 1982.

[8] P. Flajolet, J. C. Raoult, and J. Vuillemin. The number of registers required for evaluating arithmetic expressions. *Theoretical Computer Science*, 9:99–125, 1979.

[9] D. E. Knuth. *Fundamental algorithms*, volume 1. Addison-Wesley, Reading, MA, second edition, 1973.

[10] A. Meir and J. Moon. On the altitude of nodes in random trees. *Can. J. Math.*, 30(5):997–1015, 1978.

[11] J. M. Volosin. Enumeration of the terms of object domains according to the depth of embedding. *Sov. Math. Dokl.*, 15:1777–1782, 1974.