Hagen's Identity from Catalanian Forests

Len Smiley
University of Alaska Anchorage

April 18, 2005

1 Introduction

"Catalania" is a recent coinage [8, p. 256] related to the sequence of Catalan numbers $C_n = \frac{1}{2n+1}\binom{2n+1}{n}$. A visitor to Catalania may also encounter a three-parameter family of positive integers $F(n,t,k) = \frac{k}{tn+k}\binom{tn+k}{n}$ (so that $C_n = F(n,2,1)$). The fun in this discovery is in describing F as a *combinatorial* ,as well as arithmetical, generalization of C.

Given integers $t \geq 2$, $k \geq 1$, consider the set $\mathcal{P}(n,t,k)$ of lattice paths with steps $(1,0)$ and $(0,1)$ in the (x,y) plane from $(0,0)$ to $((t-1)n+k-1,n)$ that never pass above the line $L_t : (t-1)y = x$. It is well known that C_n enumerates these for $t=2$, $k=1$.

Similarly, we have the set $\mathcal{A}(n,t,k)$ of plane (unlabelled) forests of exactly k components, each of which is a t-ary tree (all vertices have either 0 or t descendents), such that the total number of internal (non-leaf) vertices is n.

Again C_n famously counts the case $t=2$, $k=1$.

In this note, we produce a bijection between paths $\mathcal{P}(n,t,k)$ and forests $\mathcal{A}(n,t,k)$, using an n-sequence of integers as intermediary, then give a quick verification that both families are counted by $F(n,t,k)$.

2 Hagen- is it easy?

A plane t-ary forest consists of an ordered set of rooted (ordered) trees whose vertices are either "internal" (having exactly t descendents), or are leaves (having 0 descendents). The vertices are unlabelled, so a forest may be specified by the string $(d_1,\ldots,d_N)$, of 0's and t's, one character per vertex, which results when the forest is traversed in preorder (root first, then descendent-trees; components left-to-right), and the number of descendents of each visited vertex is appended to the string. For example, '03000' has two components: an isolated vertex followed by a single root with its three children. The string '303000000' has three: a tree, the second child of whose root is internal, followed by two isolated vertices. Because a rooted tree has one more vertex than children (edges), it's clear that $N = tn + k$, where n is the number of internal vertices, and k is

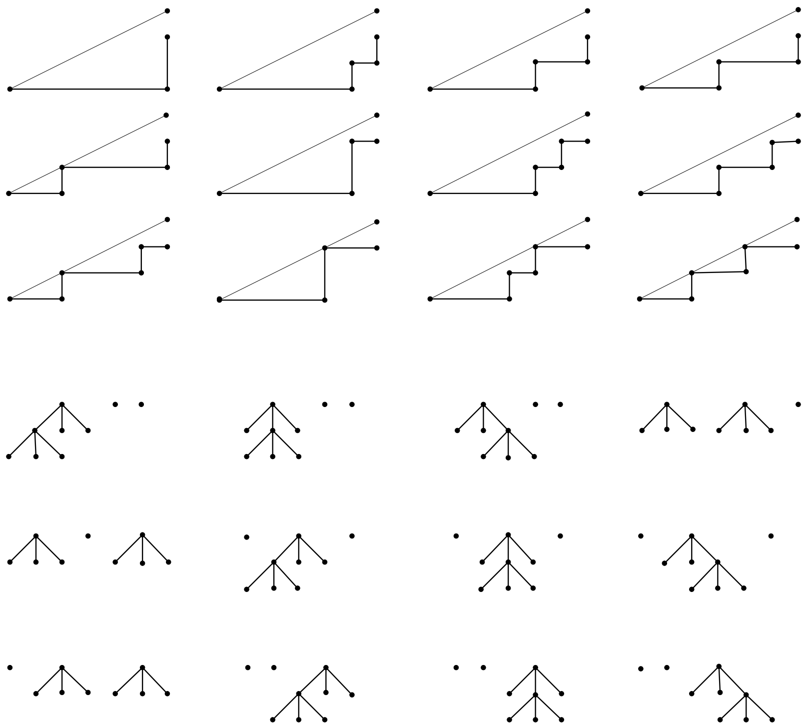

Figure 1: $\mathcal{P}(n,3,3)$ and $\mathcal{A}(n,3,3)$

the number of components, so that the leaf-count $(t-1)n+k$. Not every string comes from a forest, however.

The classical counts of rooted plane forests [8, Th. 5.3.10] and/or lattice paths specialize to give $F(n,t,k) = \frac{k}{nt+k}\binom{nt+k}{n}$.

3 Remarks

It is amusing to note that an ancient identity [4, (3.142)], attributed to Rothe and Hagen [6, 3]

$$\sum_{m=0}^{n} \frac{k}{mt+k}\binom{mt+k}{m}\frac{\ell}{(n-m)t+\ell}\binom{(n-m)t+\ell}{n-m} = \frac{k+\ell}{nt+k+\ell}\binom{nt+k+\ell}{n}$$

becomes almost obvious when interpreted using Catalanian forests. For other proofs, see [5, (1.2.6) Example 4],[1],[10].

In fact Hagen's contribution is the variant

$$\sum_{m=0}^{n} m\cdot\frac{k}{mt+k}\binom{mt+k}{m}\frac{\ell}{(n-m)t+\ell}\binom{(n-m)t+\ell}{n-m} = \frac{kn}{k+\ell}\frac{k+\ell}{nt+k+\ell}\binom{nt+k+\ell}{n}$$

which is also easy using the forest model. To wit, the LHS counts t-ary forests with n internal nodes and $k+\ell$ components with a distinguished internal node within the first k component trees (or none distinguished if the first k trees are isolated vertices). The RHS gives the fraction $\frac{k}{k+\ell}$ of the t-ary forests with n internal nodes and $k+\ell$ components with a distinguished internal node anywhere in the forest. But each of the latter objects may have its components cyclically shifted to produce $k+\ell$ distinct ones (since the distinguished node moves!), exactly k of which satisfy the LHS condition.

From this one obtains the Rothe-Hagen form

$$\sum_{m=0}^{n}(p+qm)\cdot F(m,k,t)\cdot F(n-m,\ell,t) = \frac{p(k+\ell)+kqn}{k+\ell}F(n,k+\ell,t)$$

by adding the $p-$ and $q-$ multiples of the two identities, which combined form, according to a comment of Zeng, implies Abel's binomial identity

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x(x-kz)^{k-1}(y+kz)^{n-k}.$$

Acknowledgement. I owe my introduction to this topic to my editing of Monthly Problem 11071 [2] - essentially the case $F(n,3,2)$ - and thank all the contributors to the solution file.

References

[1] Erik S. Andersen, Mogens E. Larsen, *Rothe-Abel-Jensen-identiter*, Normat, vol. 42, no. 3, 1994, pp. 116-128.

[2] Emeric Deutsch, *Problem 11071*, American Mathematical Monthly, March, 2004.

[3] John G. Hagen, *Synopsis der Hoheren Mathematik, Band I*, Berlin, 1891.

[4] Henry W. Gould, *Combinatorial Identities*, Morgantown, 1972.

[5] Donald E. Knuth, *The Art of Computer Programming, vol. 1, 3rd Ed.*, Addison-Wesley, 1997, pp.126-7.

[6] Heinrich A. Rothe, *Formulae de serierum ...*, Leipzig, 1793.

[7] Richard P. Stanley, http://www-math.mit.edu/~rstan/ec/catalan.pdf, http://www-math.mit.edu/~rstan/ec/catadd.pdf

[8] Richard P. Stanley,*Enumerative Combinatorics, vol. 2*, Cambridge University Press, 1999.

[9] Shmuel Zaks, *Lexicographic generation of ordered trees*, vol, 10, no. 1, 1980, pp. 63-82.

[10] Jiang Zeng, *Multinomial Convolution Polynomials*, Discrete Mathematics, 160, no. 1-3, 1996, pp. 219-228.