ELSEVIER

Discrete Mathematics 241 (2001) 471–477

DISCRETE MATHEMATICS

www.elsevier.com/locate/disc

A pairing of the vertices of ordered trees

Seunghyun Seo

Department of Mathematics, Korea Advanced Institute of Science and Technology, Taejon 305-701, South Korea

Abstract

A combinatorial bijection of the terminal vertices (other than roots) of ordered (rooted plane) trees with the internal vertices (including roots) is described (for ordered trees with at least one edge). © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Catalan numbers; Dyck paths; Ordered trees; Leaves; Bijective proof

1. Shapiro's problem

Let $\mathscr{P}_n$ be the set of all rooted plane or *ordered* trees with $n+1$ vertices. Fig. 1 shows all the elements of $\mathscr{P}_3$ with roots at the bottom.

It is well known that $|\mathscr{P}_n| = C_n$, where C_n is the nth Catalan number given by

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

In $\mathscr{P}_3$, we can easily check that there are 20 vertices, of which 10 are terminal vertices other than roots or *leaves*. That half the vertices are leaves is no accident.

Theorem 1. *Among the vertices of ordered trees in $\mathscr{P}_n$, exactly half of them are leaves for $n > 0$.*

Shapiro presented a proof of this result using generating functions in [27], but finding it so attractive, and believing that there must be other, neater, more insightful proofs, offered it also as a problem in *The American Mathematical Monthly* [26]. Our objective is to give a demonstration that is more combinatorial in spirit, although we do so in the setting of *Dyck paths*, rather than that of ordered trees. We find that the factor of one half stems from the reflection symmetry of unrestricted paths, and show how this symmetry implicitly induces a bijection between terminal and internal

E-mail address: shseo@math.kaist.ac.kr (S. Seo).

PII: S0012-365X(01)00130-3

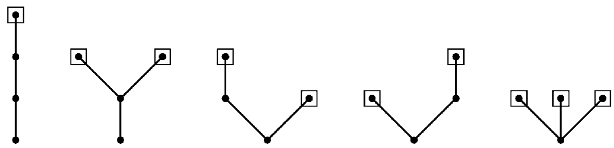

Fig. 1. Ordered trees with 4 vertices.

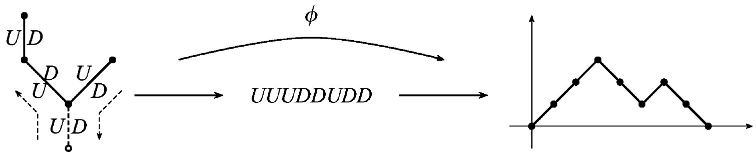

Fig. 2. $\phi : \mathscr{P}_n \to \mathscr{D}_n$.

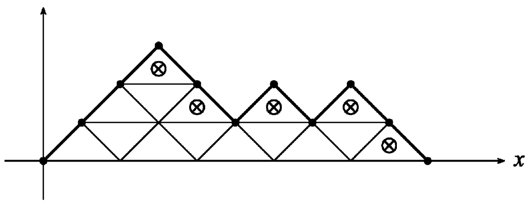

Fig. 3. An elevated Dyck path and exposed triangles.

vertices. Moreover, as Shapiro suspected, the result is not entirely new, so we also provide some discussion of its earlier appearances.

2. Bijective proof of Theorem 1 via Dyck paths

The notion of a *Dyck path* is a familiar one; see, in extenso [4]. By an *elevated* Dyck path is meant a Dyck path in which, with the exception of the first and last steps, all steps are on or above level 1; and by an *unrestricted* path is meant a lattice path in which the steps need not be on or above level 0.

Let $\mathscr{D}_n$ be the set of all the elevated Dyck paths from $(0,0)$ to $(2n+2,0)$; and let $\phi : \mathscr{P}_n \to \mathscr{D}_n$ be the familiar *one-step-at-a-time* bijection illustrated in Fig. 2 (for a reference, consult [32]). We think of U and D as representing the upward and downward passage of *vertices*. This leads to the identification of vertices in trees in $\mathscr{P}_n$ with certain triangles under the paths in $\mathscr{D}_n$. To be more precise, the area between a Dyck path and the x-axis can be divided into unit triangles, as shown in an example in Fig. 3. A triangle in the division of a Dyck path is called *exposed*, if its northeast boundary is a part of the path.

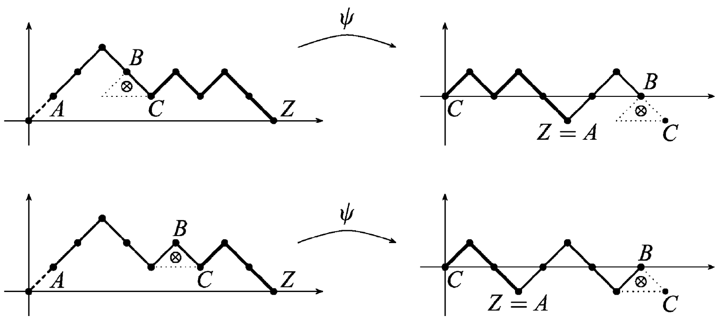

Fig. 4. The map ψ.

Let $\mathscr{V}(\mathscr{P}_n)$ be the set of vertices of all trees in $\mathscr{P}_n$. An element of $\mathscr{V}(\mathscr{P}_n)$ may be thought of as a pair (v, π), where v is a vertex of a tree π in $\mathscr{P}_n$. Similarly, let $\mathscr{E}(\mathscr{D}_n)$ be the set of all exposed triangles in the division of the area under a Dyck path in $\mathscr{D}_n$ just illustrated (as marked by $\otimes$ in Fig. 3). An element of $\mathscr{E}(\mathscr{D}_n)$ may be thought of as a pair (t, μ), where t is an exposed triangle in a path μ in $\mathscr{D}_n$. Triangles in $\mathscr{E}(\mathscr{D}_n)$ where the path is both the northeast and northwest boundary will be referred to *peak* triangles. The following observation is then straightforward.

Lemma 2. *The bijection ϕ extends in a natural way to a bijection between $\mathscr{V}(\mathscr{P}_n)$ and $\mathscr{E}(\mathscr{D}_n)$ in which leaves correspond to peak triangles.*

We now draw on a bijection described in Lemma 3.2 of [21] to show that, for each triangle in $\mathscr{E}(\mathscr{D}_n)$ there is an unrestricted path from $(0,0)$ to $(2n,0)$. Again, it is easier to illustrate this bijection, which we denote ψ. So, consider an exposed triangle in $\mathscr{E}(\mathscr{D}_n)$ situated in the path μ from the origin $O=(0,0)$ to $Z=(2n+2,0)$. Let A be the first point on the path after O; and let B and C be consecutive points on the downward part of the path—refer to Fig. 4, noting how B and C define a triangle in $\mathscr{E}(\mathscr{D}_n)$. To obtain an unrestricted path from $(0,0)$ to $(2n,0)$, we take C as the origin, identify A with Z, and finish at B, which is then the point $(2n,0)$, because of the deletion of the upward step from O to A and the downward step from B to C. In effect, we are just interchanging sections of the path in $\mathscr{D}_n$ with which we started, and it is easy to undo this process.

The examples in Fig. 4 have been chosen to show the different fates of peak triangles and other triangles in $\mathscr{E}(\mathscr{D}_n)$ under the bijection ψ. If B and C define a peak triangle, then the unrestricted path obtained under ψ ends on an upward step; and otherwise this path ends on a downward step.

Lemma 3. *There is a bijection ψ from $\mathscr{E}(\mathscr{D}_n)$ to the unrestricted paths from $(0,0)$ to $(2n,0)$ in which peak triangles correspond to paths whose final step is upward.*

But clearly the set of unrestricted paths from $(0,0)$ to $(2n,0)$ has a mirror symmetry, ρ, say: to any such path p, the mirror image, $\rho(p)$, in the x-axis is another unrestricted

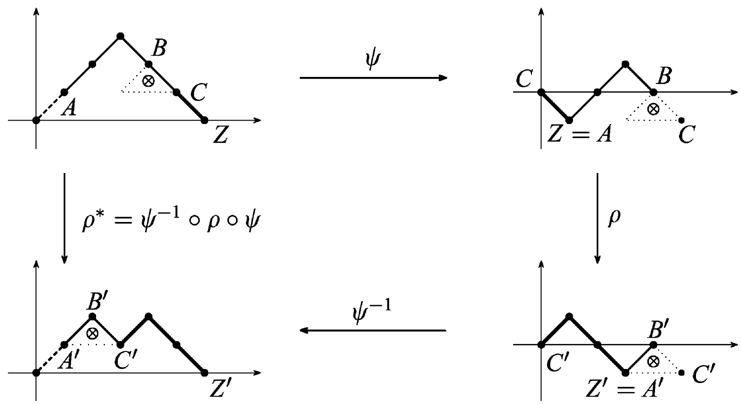

Fig. 5. The map ρ^*.

path between the same endpoints. Moreover, if p ends on an upward step, $\rho(p)$ ends on a downward step. Thus, exactly half these unrestricted paths end on an upward step, and exactly half end on a downward step. Tracing back through the bijections delineated in our lemmas, we see, in turn, that exactly half of the triangles in $\mathscr{E}(\mathscr{D}_n)$ are peak triangles, and then that half the vertices in $\mathscr{V}(\mathscr{P}_n)$ are leaves. This completes our proof of Theorem 1 (Fig. 5).

3. Commentary

First of all, bearing in mind the inspiration given by [21] for our bijective proof of Theorem 1 via Dyck paths, we note that a similar approach yields a combinatorial interpretation for the identity

$$\sum_{k\geqslant 0}\binom{2n}{k}=4^n.$$

Indeed, the cardinality of the collection of all triangles in the Dyck paths in $\mathscr{D}_n$ is 4^n, a fact closely related to the problem given publicity in [34]. Bijections related to the area under Dyck paths have been explored further very recently in [20].

Secondly, as we hinted at the end of our opening section, Theorem 1 is not without antecedents, and we take this opportunity to mention some earlier references of cognate interest, without making any claim to have traced the history of this result in its entirety. Riordan, in a study [23] in generating functions published in 1975, investigated, among other distributions, the number $t(n,k)$ of ordered trees with n edges and k leaves. Riordan shows, in effect, that

$$t(n,k)=\frac{1}{n}\binom{n}{k}\binom{n}{k-1},\quad 1\leqslant k\leqslant n$$

and concludes that this distribution is therefore the same as a more familiar one for ballot paths by length of horizontal segments. However, it is clear, from this expression for $t(n,k)$, that we have the symmetry

$$t(n,k) = t(n,n+1-k), \quad 1 \leqslant k \leqslant n$$

from which Theorem 1 follows as an easy consequence, noting that a tree on n edges with $n+1-k$ leaves has k internal vertices, for $n > 0$. Both this expression and the associated symmetry appeared again, in 1977, in a study [24] of so-called *connective relations*: explicit correspondences are given between these relations and ordered trees, thereby explaining why the distributions are the same for these two types of structure; and *implicit* in these correspondences is a bijection on the vertices of ordered trees with n edges, for $n > 0$.

The result enjoyed a good year in 1980, appearing in one form or another in at least three places. Dershowitz and Zaks, in [7], gave several combinatorial proofs of the above symmetry for $t(n,k)$, inferring as a corollary that expected number of leaves in a tree on n edges is $(n+1)/2$, as is the expected number of internal nodes. They returned, in [8], published the following year, to deduce the symmetry for $t(n,k)$ from a more general result. Interestingly enough, Shapiro, in collaboration with Rogers, came close to Theorem 1 in a discussion of *picture counting*, in [25]. They use this approach to investigate the number of ordered trees with n edges of which k are what they call *eldest branches*, that is, a leftmost, upward edge at a vertex. The distribution is the same as that for $t(n,k)$, which is no surprise, since, as they go on to note, eldest branches correspond to internal vertices. They then use another instance of picture counting to obtain the distribution of ordered trees by edges and leaves more directly.

Dasarathy and Yang [6], also in 1980, start from a ordered tree, T, put it in correspondence with a binary ordered tree, reflect this to obtain another binary ordered tree, and reverse the correspondence to get back to another ordered tree, $R(T)$, say. They then show that the total number leaves in T and $R(T)$ is $n+1$, when T has n edges, and they observe that the average number of leaves in such trees is $(n+1)/2$. Our approach here, through the induced map ρ^* and reflection of unrestricted paths, seems similar in spirit. Clearly, the factor of *one half* might most plausibly be explained in terms of a reflection. The problem is to find an involution on the ordered trees that will do the trick that reflection does for binary ordered trees or for unrestricted paths. However, we revisit Theorem 1 in a sequel [28] in terms of the structure of ordered trees. Moreover, a short survey of combinatorial occurrences of $\frac{1}{2}\binom{2n}{n}$ is given in [12].

For completeness, we also note that Deutsch mentions Theorem 1 in terms of Dyck paths, in [4], where attention is also drawn to [33]. Of related interest here, is the more recent note [5] of Deutsch giving a neat pictorial proof that among ordered trees with n edges the distribution by vertices of degree q is the same as that by vertices at odd level of degree $q-1$. Rogers has pointed out that, in a rather similar way, internal vertices (including roots) at one level can be put into one-to-one correspondence with leaves at the next higher level, again leading to a proof of Theorem 1 (personal communication).

Of course, no one can now write on the Catalan numbers without citing what must be considered the current *locus classicus*, the 66 instances gathered by Stanley in exercise 6.19 in [30]. However, articles featuring this sequence occur frequently in the more *popular* journals, not always cross-referenced. So, in default of an updated bibliography, we list a selection of expository articles from the last 3 years, by year of publication: [34,1,13,31,2,3,22]. A further interesting development in this period has been the emergence of studies of the history of this sequence and its relatives (again listed by year of publication): [29,9,19,10,11,14–18].

Acknowledgements

This paper has its genesis in a seminar given, at the Korea Advanced Institute of Science and Technology, on 26 November, 1999 by D.G. Rogers, where he discussed Shapiro's problem [26]. I am also very much indebted to my supervisor, Professor Dongsu Kim, for drawing [21] to my attention, as well as for his helpful discussion of earlier drafts of this note. It was after the completion of this note that I learnt of the parallel researches of the authors of [12], and I am most grateful to those authors for sharing their paper with me as a preprint.

References

[1] D. Beckwith, Legendre polynomials and polygon dissections, Amer. Math. Monthly 105 (1998) 256–257.
[2] D. Callan, Two uniformly distributed parameters defining Catalan numbers, Amer. Math. Monthly 106 (1999) 948–949.
[3] D. Callan, A combinatorial interpretation of a Catalan numbers identity, Math. Mag. 72 (1999) 295–298.
[4] E. Deutsch, Dyck path enumeration, Discrete Math. 204 (1999) 167–202 (esp. p. 175).
[5] E. Deutsch, A bijection on ordered trees and its consequences, J. Combin. Theory, Ser. A 90 (2000) 210–215.
[6] B. Dasarathy, C. Yang, A transformation on ordered trees, Comput. J. 23 (1980) 161–164 (esp. Theorem 3).
[7] N. Dershowitz, S. Zaks, Enumeration of ordered trees, Discrete Math. 31 (1980) 9–28 (esp. Theorem 2 and Corollary 2.1).
[8] N. Dershowitz, S. Zaks, Applied tree enumeration, in: CAAP81: Trees in Algebra and Programming—6th Colloquium, Genoa, 5–7 March, 1981. Lecture Notes in Computer Science, Vol. 112, Springer, Berlin, 1981, pp. 180–193 (esp. p. 184).
[9] L. Habsieger, M. Kazarian, S. Lando, On the second number of Plutarch, Amer. Math. Monthly 105 (1998) 446.
[10] E.J. Fennessey, P.J. Larcombe, Using old combinatorial problems to teach analysis: Euler's borderline convergence, Congr. Numer. 138 (1999) 211–220.
[11] E.J. Fennessey, P.J. Larcombe, On a finite polynomial generating function for Catalan subsequences: an 18th century observation proved, Congr. Numer. 141 (1999) 49–60.
[12] A. Frosini, R. Pinzani, S. Rinaldi, About half the middle binomial coefficient, preprint.
[13] N. Hungerbühler, Proofs without words: a bijection between certain lattice paths, Math. Mag. 71 (1998) 219.
[14] P.J. Larcombe, On pre-Catalan Catalan numbers: Kotelnikow—addenda to [19], Math. Today 35 (1999) 25.

[15] P.J. Larcombe, On the history of the Catalan numbers: a first record in China, Math. Today 35 (1999) 89.
[16] P.J. Larcombe, The eighteenth century Chinese discovery of the Catalan numbers, Math. Spectrum 32 (1999/2000) 5–7.
[17] P.J. Larcombe, On Catalan numbers and expanding the sine function, Bull. Inst. Combin. Appl. 28 (2000) 39–47.
[18] P.J. Larcombe, A forgotten convolution type identity of Catalan, Util. Math. 57 (2000) 65–72.
[19] P.J. Larcombe, P.D.C. Wilson, On the trail of the Catalan sequence, Math. Today 34 (1998) 114–117.
[20] E. Pergola, Two bijections for the area under Dyck paths, Discrete Math., to appear.
[21] E. Pergola, R. Pinzani, A combinatorial interpretation of the area of Schröder paths, Electron. J. Combin. 6 (1999) #R47.
[22] S.G. Prentice, Stacks, bracketings, and GC-arrangements, Math. Mag. 72 (1999) 321–324.
[23] J. Riordan, Enumeration of plane trees by branches and endpoints, J. Combin. Theory Ser. A 19 (1975) 214–222.
[24] D.G. Rogers, The enumeration of a family of ladder graphs I: connective relations, Quart. J. Math. (Oxford) (2) 28 (1977) 421–431 (Section 5, esp. p. 428).
[25] D.G. Rogers, L.W. Shapiro, Deques, trees, and lattice paths, in: Combinatorial Mathematics VIII: Proceedings of the Eighth Australian Conference on Combinatorial Math., Geelong, 25–29, August, 1980, Lecture Notes in Mathematics, Vol. 884, Springer, Berlin, 1981, pp. 293–303 (esp. pp. 296–297).
[26] L.W. Shapiro, Problem 10753, Amer. Math. Monthly 106 (1999) 777.
[27] L.W. Shapiro, The higher you go, the odder it gets, Congr. Numer. 138 (1999) 93–96.
[28] S. Seo, H. Shin, A perfect matching on vertices of ordered trees, preprint.
[29] R.P. Stanley, Hipparchus, Plutarch, Schröder, and Hough, Amer. Math. Monthly 104 (1997) 344–350.
[30] R.P. Stanley, Enumerative Combinatorics: II, Cambridge University Press, Cambridge, 1999 (esp. pp. 219–229; 256–265).
[31] H.S. Snevily, D.B. West, The bricklayer problem and the strong cycle lemma, Amer. Math. Monthly 105 (1998) 131–143.
[32] J.H. van Lint, Combinatorial Theory Seminar, Eindhoven University of Technology, Lecture Notes in Mathematics, Vol. 382, Springer, Berlin, 1974 (esp. pp. 21–27).
[33] H.S. Wilf, generatingfunctionology, 2nd Edition, Academic Press, New York, NY, 1994 (esp. p. 53, (2.5.11)).
[34] W.-J. Woan, L.W. Shapiro, D.G. Rogers, The Catalan numbers, the Lebesgue integral, and 4^{n-2}, Amer. Math. Monthly 104 (1997) 926–931.