

[Graduate Theses, Dissertations, and Problem Reports](https://researchrepository.wvu.edu/etd) 

2005

# Enumeration of the generalized Catalan numbers

Steven L. Richardson Jr. West Virginia University

Follow this and additional works at: [https://researchrepository.wvu.edu/etd](https://researchrepository.wvu.edu/etd?utm_source=researchrepository.wvu.edu%2Fetd%2F2192&utm_medium=PDF&utm_campaign=PDFCoverPages) 

#### Recommended Citation

Richardson, Steven L. Jr., "Enumeration of the generalized Catalan numbers" (2005). Graduate Theses, Dissertations, and Problem Reports. 2192. [https://researchrepository.wvu.edu/etd/2192](https://researchrepository.wvu.edu/etd/2192?utm_source=researchrepository.wvu.edu%2Fetd%2F2192&utm_medium=PDF&utm_campaign=PDFCoverPages)

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

#### Enumeration of the generalized Catalan numbers

Steven L. Richardson, Jr

A thesis submitted to the Eberly College of Arts and Sciences at West Virginia University in partial fulllment of the requirements for the degree of Master of Science in Mathematics

> Henry W. Gould, Chair Harvey Diamond, Ph.D. John Goldwasser, Ph.D. Michael Mays, Ph.D. Department of Mathematics

Morgantown, West Virginia 2005 Keywords: Catalan, sequences, Dyck Paths, enumeration

#### ABSTRACT

#### Enumeration of the generalized Catalan numbers

Steven L. Richardson, Jr

The sequence of numbers given by the equation

$$
C(n) = \frac{1}{n+1} \binom{2n}{n}
$$

are widely known as the Catalan numbers, because they were first studied by Catalan. The ways to enumerate these numbers are similarly widely known and studied, as can be seen in Gould's catalog of Catalan numbers. Our objective is to find ways to enumerate this larger and generalized set of sequences, using multiple methods that will build on known enumerations of the Catalan numbers.

#### TABLE OF CONTENTS

Page i - Title Page

Page ii - Abstract

Page iii - Table of Contents

Page 1 - Introduction

Page 2 - Enumeration by Solutions to Inequalities

Page 7 - Enumeration by Dyck Paths and Binary Trees

Page 8 - Enumeration by Diagonalization of Polygons

Page 8 - Appendix 1: Justification for Exhaustive Tables

Page 9 - Appendix 2: Exhaustive Tables

Page 33 - Acknowledgments

Page 33 - References

## 1 Introduction

The sequence of numbers given by the equation

$$
C(n) = \frac{1}{n+1} \binom{2n}{n} \tag{1}
$$

are widely known as the Catalan numbers, because they were first studied by Catalan [1,2] The ways to enumerate these numbers are similarly widely known and studied, as can be seen in Gould's [1] catalog of Catalan numbers.

Gould, in 1956 [3], developed the following equation

$$
A_n(a,b) = \frac{a}{a + bn} \binom{a + bn}{n}
$$
 (2)

and it is easily seen to be a generalization of  $(1)$ :  $(1)$  is  $(2)$  with a=1 and b=2. Gould also showed the following convolution

$$
\sum_{k=0}^{n} A_k(a, b) A_{n-k}(c, b) = A_n(a+c, b)
$$
 (3)

for  $(3)$ . This also can be used to produce a convolution for  $(1)$ , again with  $a=1$  and  $b=2$ , adding  $c=1$ .

Thus, it can be seen that the Catalan numbers are simply a special case of this larger set of sequences, which we will call Rothe numbers from this point forward. Also, equation (3) can be known as the Rothe convolution.

The generalized Catalan numbers, seen in the Appendix, will be submitted as a set to the Online Encyclopedia of Integer Sequences (OEIS) along with some of these enumerations.

These sequences have been studied and have occurred in various sources, but have never been studied in conjunction with the Catalan sequences or together in this way. For example, in studying the cell growth problem [7], Harary discovered some of these sequences (specically those sequences with a=b), but didn't link them to equation (2) in the way that we will.

Naiomi Cameron, as we discuss later, developed [9] representations for binary and ternary trees. These, as we will see, can also be generalized to be enumerated by these Rothe numbers.

Our objective is to find ways to enumerate this larger and generalized set of sequences, using multiple methods that will build on known enumerations of the Catalan numbers.

# 2 Enumeration by Solutions to Inequalities

The enumeration of the Catalan numbers that we will be build on was shown by Carlitz [2]. He proved that the number of solutions to the inequality

$$
1 \le a_1 \le a_2 \le \ldots \le a_n \le n+1, \ a_i \le i \tag{4}
$$

was  $(1)$ .

Our goal is now to enumerate all sequences produced by (2) by the solutions to inequalities such as (4). First, let's examine the sequences which satisfy (4) and represent the Catalan numbers.

```
1 2 2 4
1 1 3 4
1 2 3 4
```
All subsequent sequence illustrations can be seen in the Appendix.

It can be seen, as we start to generalize this inequality, that the number of solutions to the following inequality will give the numbers corresponding to  $(2)$ , with a=1 and b=3:

$$
1 \le a_1 \le a_2 \le \ldots \le a_{2n} \le n+1, \ a_i \le \frac{i+1}{2} \tag{5}
$$

In fact, for a=1 and a general b, it can be seen that the number of solutions to this inequality will give the numbers corresponding to (2) with those a and b values.

$$
1 \le a_1 \le a_2 \le \ldots \le a_{(b-1)n} \le n+1, \ a_i \le \frac{i+(b-2)}{b-1} \tag{6}
$$

Finally, for a general a and b, the number of solutions to this inequality will give the numbers corresponding to  $(2)$ .

$$
1 \le a_1 \le a_2 \le \ldots \le a_{(b-1)n+(a-1)} \le n+1, \ a_i \le \frac{i+(b-2)}{b-1} \qquad (7)
$$

This inequality and restriction works for any positive integers a, b, and n except for b=1, where there is no restriction on the elements in the inequality.

To prove that this is true, we have to examine the maximum value that an element can have in the array that follows:

$$
(a_1, a_2, \ldots, a_{(b-1)n+(a-1)}) \tag{8}
$$

In general, the maximum values are:

$$
\underbrace{1 \dots 1}_{n} \underbrace{2 \dots 2}_{n} \dots \underbrace{n \dots n}_{n+1} \underbrace{(n+1) \dots (n+1)}_{n}
$$
 (9)

where the first n groups contain b-1 elements and the last group contains a-1 elements.

First, we'll align the maximum elements in

$$
A_n(a-1,b) \tag{10}
$$

under the above maximum elements.

$$
\underbrace{1 \dots 1}_{1} \underbrace{2 \dots 2}_{1} \dots \underbrace{n \dots n}_{n} \overbrace{(n+1) \dots (n+1)}_{n+1}
$$
\n
$$
(11)
$$

where, in the bottom equation, there are instead a-2 elements in the last group.

It is easily seen that every sequence that fulfills the bottom restrictions also fulfills the top restrictions. Thus, we can start to enumerate the included sequences by including every sequence that fulfills the bottom restrictions. As previously stated, this is

$$
A_n(a-1,b)
$$

Now, we need to enumerate all the sequences that this doesn't include.

To accomplish this, we just need to look at the places that the 2 arrays differ. There are n places, counting down from n to 1 from the right hand side. To enumerate the sequences to include when we move n to  $n+1$  (the first difference), we see that we need to include the sequences with the following maximum elements:

$$
\underbrace{1 \dots 1} \underbrace{2 \dots 2} \dots \underbrace{n \dots n} \tag{12}
$$

where the groups again have b-1 elements. This is easily seen to be

$$
A_n(1,b)
$$

Next, we move to the difference to the left of this one. This includes the sequences with the following maximum elements:

$$
\underbrace{1 \dots 1}_{\infty} \underbrace{2 \dots 2}_{\infty} \dots \underbrace{(n-1) \dots (n-1)}_{\infty}
$$
 (13)

(again with each group containing b-1 elements) on the left and the sequences with the following maximum elements:

$$
\underbrace{1\ldots1}_{2\ldots2} \underbrace{2\ldots2}_{}
$$
 (14)

(with the first group containing b-1 elements and the last group containing a-2 elements) on the right. Thus, this enumeration is

$$
(A_{n-1}(1,b)) * (A_1(a-1,b))
$$

This process continues until we arrive at the last difference, where we see the following enumeration: The sequences with the following maximum elements:

$$
\underbrace{1\ldots1}_{}
$$
 (15)

(containing b-1 elements) on the left and the sequences with the following maximum elements:

$$
\underbrace{1\ldots1}_{2\ldots2}\underbrace{2\ldots\widehat{n\ldots n}}\tag{16}
$$

(with the all groups containing b-1 elements except the last group, which contains a-2 elements) on the right.

Thus, we can represent the enumeration with this sum:

$$
\sum_{k=0}^{n} A_k (a-1, b) A_{n-k} (1, b) \tag{17}
$$

which is easily seen to be a specific case of  $(3)$ . Thus, by  $(3)$ , this sum is

$$
A_n(a,b) \tag{18}
$$

This proves that the solutions to the above inequality do enumerate the numbers

$$
A_n(a,b) = \frac{a}{a + bn} \binom{a + bn}{n}
$$
 (19)

The following triangle was first discovered in 1961 by Forder [8] and is known as Catalan's triangle:



Each term can be seen to follow the following equation:

$$
T(n,k) = \sum_{j=0}^{k} T(n-1,j)
$$
 (20)

or, in other words, the sum of all the entries above and to the left of the term in question.

It is easily seen that the values on the diagonal are the Catalan numbers. What is interesting, though, is that each column has a meaning in the list of sequences defined above. Referring to the above list of sequences that enumerate the Catalan numbers, we can see that the columns represents the largest value in each sequence. For instance, for  $n=3$ , there is 1 sequences with largest value 1, 2 sequences with largest value 2, and 2 sequences with largest value 3. In the triangle above, we can see that 5 is derived by adding  $1+2+2.$ 

Even more interesting, though, is that we can see all sequences

$$
A_n(a,2) = \frac{a}{a+2n} \binom{a+2n}{n} \tag{21}
$$

in this triangle. Of course, the Catalan sequences (the above with  $a=2$ ) can be seen on the first 2 diagonals. However, the case with  $a=3$  is on the third diagonal down (3, 9, 28, ...), and it can be seen that all sequences correspond to the ath diagonal.

For b=3, the following array can be seen to enumerate the sequences

$$
A_n(a,3) = \frac{a}{a+3n} \binom{a+3n}{n} \tag{22}
$$



Again, the case with  $a=2$  can be seen on the first "diagonal"  $(1, 3, 12, ...)$ 55...) and the ath case can be seen in a similar way. Note, though, that in order to track the correct sequence, we have to move in a "knights move,"

that is, 1 over and 2 down. Each column, again, represents the largest element in the sequence that will enumerate the proper element.

In fact, an array such as this can be constructed for each value of b. Below is the array for b=4.

In addition, these triangles give rise to convolution codes, which are different than the standard convolution code that is seen above and uses the relation of the triangles, (20).

Looking at the ternary triangle above, you can see that the last element of the even numbered rows is a number that satisfies (2) with  $a=1$  and  $b=3$ . I'll label these elements as  $D(i)$ . Using (20), you can see that  $D(2) = 3D(1)$ and  $D(3) = 3D(2) + 3D(1)$ . From these, we can start to accumulate a convolution. In this case, the first four terms are  $3,3,10,42$ .

A conjecture for this convolution code, in general, is:

$$
A_n(1,b) = \sum_{j=1}^{n-2} \left(\frac{1}{(b-1)j+1} \binom{bj}{j} \frac{b-1}{n-j} \binom{b(n-j)-b}{n-j-1}\right) + \frac{b}{(b-1)(n-1)+1} \binom{b(n-1)}{n-1} \tag{23}
$$

## 3 Enumeration by Dyck Paths and Binary Trees

Eric Weisstein [10] defines a Dyck path as "a staircase walk from  $(0,0)$  to  $(n,n)$  which never crosses (but may touch) the diagonal y=x. The number of Dyck paths of order n is given by the Catalan number:"

$$
C_{n-1} = \frac{1}{n+1} \binom{2n}{n} \tag{24}
$$

Cameron [9] defines, in a similar vein, ternary paths as "the set of all positive paths from  $(0,0)$  to some  $(3n,0)$  with  $S = (1,1),(1,-2),$ " and a generalized t-Dyck path as "a positive path from  $(0,0)$  to  $((t+1)n,0)$  with S =  $(1,1),(1,-t)$ ." Here are some ternary paths:

Cameron shows that the number of ternary paths is given by this representation:

$$
t_n = \frac{1}{2n+1} \binom{3n}{n} \tag{25}
$$

This can be seen to be our  $A(n)$  with a=1 and b=3. In the same way, the generalized t-Dyck path can be seen to be represented by  $A(n)$  with a=1 and b=t.

## 4 Enumeration by Diagonalization of Polygons

It has also been documented that the Catalan numbers (our numbers with  $a=1$  and  $b=2$ ) can be enumerated by the number of ways to diagonalize a polygon into triangles. Again, this can be generalized to fit our discussion. It can easily be seen that our numbers with  $a=1$  can be enumerated by the number of ways to diagonalize a polygon into  $(b-1)$ gons.

#### 5 Appendix 1: Justication for Exhaustive Tables

On Thursday, August 24, 2004, Donald Knuth wrote the following, as part of an email to Henry Gould:

"We know that exhaustive listings of set partitions for small n have been made for a long time, notably in Japan.

But what about the Catalan numbers? There are oodles of combinatorial problems with C(n) objects of order n; yet I can't recall seeing any tables of the individual objects, rather than tables of the cardinalities, going back before say 1950.

This is in sharp contrast to extremely voluminous books with tables of permutations, combinations, and partitions. Euler didn't exhibit the triangulations of a small n-gon; Catalan and Rodrigues and their colleagues didn't tabulate the possibilities for parentheses. Not even in the Hindenburg school, where people like Rothe wrote highly relevant papers, do I recall seeing tables of Catalan-counted objects  $\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{1-\frac{1}{\sqrt{$ day, combinatorialists usually made combinatorial lists!"

Also, Derrick Norman Lehmer, in his book "List of Prime Numbers from 1 to 10,006,721," Carnegie Institute, Washington, Pub. 165 (1914), has the following to say in the preface:

"In spite of the contention of certain eminent scientists that mathematics is a science that has nothing to do with observation and experiment, the history of the Theory of Numbers has been chiefly made by those who followed methods closely allied to those of the student of natural science. Gauss himself, the most successful investigator in this field, was an indefatigable computer, as may be seen by consulting the long list of tables in his collected works. Jacobi was also a tireless maker of tables. It is hardly likely, indeed, that any theorem of importance in the Theory of Numbers was ever discovered which was not found in the first place by observation of listed results. This fact is itself sufficient warrant for the publication of factor tables and lists of primes, even at the present time, when the theory of the sequence of prime numbers has been worked out to remarkable perfection. Every investigator is familiar with the fact that a table constructed for one purpose almost always suggests other directions in which research may be protably conducted. Also, no matter how extraordinary the powers of the pure analyst may be, he ought not to neglect the check on his results which an accurate table may afford. Much depends, therefore, on the absolute accuracy of a table such as this, in which the values of a discontinuous function are listed. Clearly, in such a table an entry can not be computed from the entries adjacent to it; no interpolation scheme is practicable, and each entry must stand on its own merits."

## 6 Appendix 2: Exhaustive Tables

The Generalized Catalan Numbers for  $n=1$  to 10,  $a=1$  to 5,  $b=1$  to 5:



 $A(1,5)$ 







A(3,5)





 $A(5,5)$ 

1 1

 $n=2$  (

 $n=3$  (

1 1 1 1 1 1  $1 \t1 \t1$ 

Sequences for small values of n:





 $\frac{A_n(1)}{A_n}$ 

 $, 4)$  (27)



 $A_n(2,2)$  (28)





 $\frac{A_n(2)}{2}$ 

 $\frac{1}{2}$  (29)



 $\frac{A_n(2)}{2}$  $, 4)$  (30)

```
1 1 1 1 1 2 2
1 1 1 1 2 2 2
1 1 1 2 2 2 2
1 1 1 1 1 1 3
1 1 1 1 1 2 3
1 1 1 1 2 2 3
1 1 1 2 2 2 3
n=3 (52):
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 2
1 1 1 1 1 1 1 1 2 2
1 1 1 1 1 1 1 2 2 2
1 1 1 1 1 1 2 2 2 2
1 1 1 1 1 2 2 2 2 2
1 1 1 1 2 2 2 2 2 2
1 1 1 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 3
1 1 1 1 1 1 1 1 2 3
1 1 1 1 1 1 1 2 2 3
1 1 1 1 1 1 2 2 2 3
1 1 1 1 1 2 2 2 2 3
1 1 1 1 2 2 2 2 2 3
1 1 1 2 2 2 2 2 2 3
1 1 1 1 1 1 1 1 3 3
1 1 1 1 1 1 1 2 3 3
1 1 1 1 1 1 2 2 3 3
1 1 1 1 1 2 2 2 3 3
1 1 1 1 2 2 2 2 3 3
1 1 1 2 2 2 2 2 3 3
1 1 1 1 1 1 1 3 3 3
1 1 1 1 1 1 2 3 3 3
1 1 1 1 1 2 2 3 3 3
1 1 1 1 2 2 2 3 3 3
1 1 1 2 2 2 2 3 3 3
1 1 1 1 1 1 3 3 3 3
1 1 1 1 1 2 3 3 3 3
1 1 1 1 2 2 3 3 3 3
1 1 1 2 2 2 3 3 3 3
```
# $A_n(3,2)$  (31)

1 2 2 2 1 1 1 3



 $\frac{A_n(3)}{2}$  $\frac{1}{2}$  (32)

n=1 (3):

$$
\underline{A_n(3,4)}\tag{33}
$$



 $\frac{A_n(4)}{2}$ 

 $, 2)$  (34)









 $\frac{1}{2}$  (35)

> $\frac{A_n(4)}{2}$  $, 4)$  (36)



#### ACKNOWLEDGMENTS

The author wishes to thank Henry Gould, for the wisdom, knowledge and guidance and my wife, without whom none of this would be possible.

#### REFERENCES

1. E. Catalan, "Note sur un probleme de combinaisons," J. Math. Pures Appl., (1)3(1838), pp. 111-112.

2. E. Catalan, "Melanges Mathematiques," Mem. Soc. Sci. Liege (2)12(1885); orig. publ. 1868.

3. H. W. Gould, "Combinatorial Identities." Revised Edition, Morgantown, WV, 1972.

4. H. W. Gould, "Binomial Coefficients, The Bracket Function, and Compositions with Relatively Prime Summands," The Fibonacci Quarterly, (2)4(1964), pp. 241-259

5. H. W. Gould, "Some Generalizations of Vandermonde's Convolution," Amer. Math. Monthly, (68)2(1956), pp. 84-91.

6. L. Carlitz, "Enumeration of Two-Line Arrays," Fibonacci Quarterly, (2)11(1973), pp. 113-130.

7. F. Harary, et al, "On the Cell-Growth Problem for Arbitrary Polygons," Discrete Mathematics 11(1975), pp. 371-389.

8. H. G. Forder, Some problems in combinatorics, Math. Gazette, vol. 45, 1961, 199-201.

9. N. Cameron, "Random Walks, Trees and Extensions of Riordan Group Techniques," Annual Joint Mathematics Meetings, Baltimore, MD, Jan. 2003.

10. Eric W. Weisstein "Dyck Path." From Mathworld-A Wolfram Web Resource. http://mathworld.wolfram.com/DyckPath.html.