Catalan Numbers

Tom Davis
tomrdavis@earthlink.net
http://www.geometer.org/mathcircles
October 24, 2001

We begin with a set of problems that will be shown to be completely equivalent. The solution to each problem is the same sequence of numbers called the Catalan numbers. Later in the document we will derive relationships and explicit formulas for the Catalan numbers in many different ways.

1 Problems

1.1 Balanced Parentheses

Suppose you have n pairs of parentheses and you would like to form valid groupings of them, where "valid" means that each open parenthesis has a matching closed parenthesis. For example, $(()())$ is valid, but $())()($ is not. How many groupings are there for each value of n?

Table 1 shows the possible groupings for $0 \leq n \leq 5$.

$n = 0$:	*	1 way
$n = 1$:	()	1 way
$n = 2$:	()(), (())	2 ways
$n = 3$:	()()(), ()(()), (())(), (()()), ((()))	5 ways
$n = 4$:	()()()(), ()()(()), ()(())(), ()(()()), ()((())), (())()(), (())(()), (()())(), ((()))(), (()()()), (()(())), ((())()), ((()())), (((())))	14 ways
$n = 5$:	()()()()(), ()()()(()), ()()(())(), ()()(()()), ()()((())), ()(())()(), ()(())(()), ()(()())(), ()((()))(), ()(()()()), ()(()(())), ()((())()), ()((()())), ()(((()))), (())()()(), (())()(()), (())(())(), (())(()()), (())((())), (()())()(), (()())(()), ((()))()(), ((()))(()), (()()())(), (()(()))(), ((())())(), ((()()))(), (((())))(), (()()()()), (()()(())), (()(())()), (()(()())), (()((()))), ((())()()), ((())(())), ((()())()), (((()))()), ((()()())), ((()(()))), (((())())), (((()()))), ((((()))))	42 ways

Table 1: Balanced Parentheses

* It is useful and reasonable to define the count for $n = 0$ to be 1, since there is exactly one way of arranging zero parentheses: don't write anything.

1.2 Mountain Ranges

How many “mountain ranges” can you form with n upstrokes and n downstrokes that all stay above the original line? If, as in the case above, we consider there to be a single mountain range with zero strokes, Table 2 gives a list of the possibilities for $0 \leq n \leq 3$:

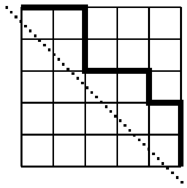
$n = 0$:	*	1 way
$n = 1$:	/\	1 way
$n = 2$:	/\ /\/\, / \	2 ways
$n = 3$:	/\ /\ /\ /\/\ / \ /\/\/\, /\/ \, / \/\, / \, / \	5 ways

Table 2: Mountain Ranges

Note that these must match the parenthesis-groupings above. The “(” corresponds to “/” and the “)” to “\”. The mountain ranges for $n = 4$ and $n = 5$ have been omitted to save space, but there are 14 and 42 of them, respectively. It is a good exercise to draw the 14 versions with $n = 4$.

1.3 Diagonal-Avoiding Paths

In a grid of $n \times n$ squares, how many paths are there of length $2n$ that lead from the upper left corner to the lower right corner that do not touch the diagonal dotted line from upper left to lower right? In other words, how many paths stay on or above the main diagonal?



```
  /\  /\/\
/  \/    \
```

Figure 1: Corresponding Path and Range

This is obviously the same question as in the example above, with the mountain ranges running diagonally. In Figure 1 we can see how one such path corresponds to a mountain range.

Another equivalent statement for this problem is the following. Suppose two candidates for election, A and B, each receive n votes. The votes are drawn out of the voting urn one after the other. In how many ways can the votes be drawn such that candidate A is never behind candidate B?

1.4 Multiplication Orderings

Suppose you have a set of $n + 1$ numbers to multiply together, meaning that there are n multiplications to perform. Without changing the order of the numbers themselves, you can multiply the numbers together in many orders. Here are the possible multiplication orderings for $0 \le n \le 4$ multiplications. The groupings are indicated with parentheses and dot for multiplication in Table 3.

$n = 0$	a	1 way
$n = 1$	$(a\cdot b)$	1 way
$n = 2$	$((a\cdot b)\cdot c),\ (a\cdot (b\cdot c))$	2 ways
$n = 3$	$(((a\cdot b)\cdot c)\cdot d),\ ((a\cdot b)\cdot (c\cdot d)),\ ((a\cdot (b\cdot c))\cdot d),$ $(a\cdot ((b\cdot c)\cdot d)),\ (a\cdot (b\cdot (c\cdot d)))$	5 ways
$n = 4$	$((((a\cdot b)\cdot c)\cdot d)\cdot e),\ (((a\cdot b)\cdot c)\cdot (d\cdot e)),\ (((a\cdot b)\cdot (c\cdot d))\cdot e),$ $((a\cdot b)\cdot ((c\cdot d)\cdot e)),\ ((a\cdot b)\cdot (c\cdot (d\cdot e))),\ (((a\cdot (b\cdot c))\cdot d)\cdot e),$ $((a\cdot (b\cdot c))\cdot (d\cdot e)),\ ((a\cdot ((b\cdot c)\cdot d))\cdot e),\ ((a\cdot (b\cdot (c\cdot d)))\cdot e),$ $(a\cdot (((b\cdot c)\cdot d)\cdot e)),\ (a\cdot ((b\cdot c)\cdot (d\cdot e))),\ (a\cdot ((b\cdot (c\cdot d))\cdot e)),$ $(a\cdot (b\cdot ((c\cdot d)\cdot e))),\ (a\cdot (b\cdot (c\cdot (d\cdot e))))$	14 ways

Table 3: Multiplication Arrangements

To convert the examples above to the parenthesis notation, erase everything but the dots and the closed parentheses, and then replace the dots with open parentheses. For example, if we wish to convert $(a\cdot (((b\cdot c)\cdot d)\cdot e))$, first erase everything but the dots and closed parentheses: $\cdot\cdot)\cdot)\cdot))$. Then replace the dots with open parentheses to obtain: $(()()())$.

The examples in Table 3 are arranged in exactly the same order as the entries in Table 1 with the correspondence described in the previous paragraph. Try to convert a few yourself in both directions to make certain you understand the relationships.

1.5 Polygon Triangulation

If you count the number of ways to triangulate a regular polygon with $n + 2$ sides, you also obtain the Catalan numbers. Figure 2 illustrates the triangulations for polygons having 3, 4, 5 and 6 sides.

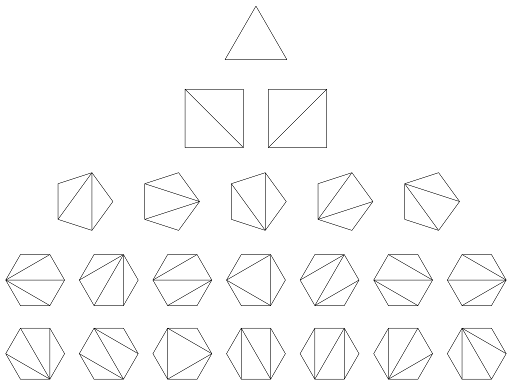


Figure 2: Polygon Triangulations

As you can see, there are 1, 2, 5, and 14 ways to do this. The “2-sided polygon” can also be triangulated in exactly 1 way, so the case where $n = 0$ also matches.

1.6 Hands Across a Table

If $2n$ people are seated around a circular table, in how many ways can all of them be simultaneously shaking hands with another person at the table in such a way that none of the arms cross each other? Figure 3 illustrates the arrangements for 2, 4, 6 and 8 people. Again, there are 1, 2, 5 and 14 ways to do this.

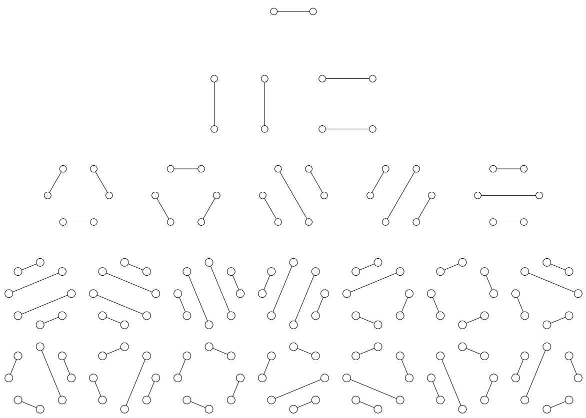


Figure 3: Hands Across the Table

1.7 Binary Trees

The Catalan numbers also count the number of rooted binary trees with n internal nodes. Illustrated in Figure 4 are the trees corresponding to $0 \leq n \leq 3$. There are $1, 1, 2$, and 5 of them. Try to draw the 14 trees with $n = 4$ internal nodes.

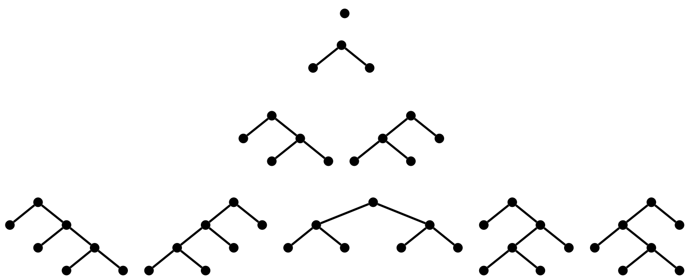


Figure 4: Binary Trees

1.8 Plane Rooted Trees

Figure 5 shows a list of the plane rooted trees with n edges, for $0 \leq n \leq 3$. Try to draw the 14 trees with $n = 4$ edges.

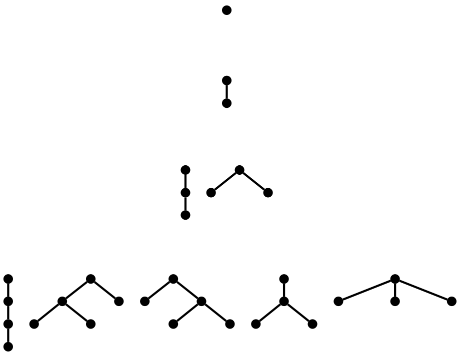


Figure 5: Plane Rooted Trees

1.9 Skew Polyominos

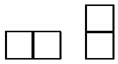
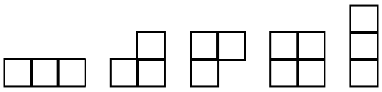
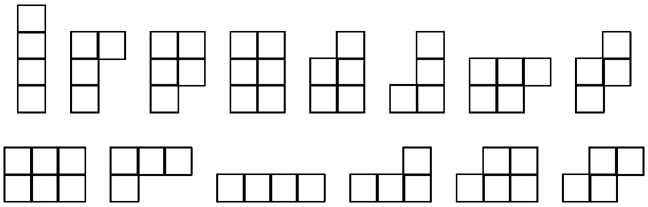
$n = 1$	
$n = 2$	
$n = 3$	
$n = 4$	

Table 4: Skew Polyominos with Perimeter $2n + 2$

A polyomino is a set of squares connected by their edges. A skew polyomino is a polyomino such that every vertical and horizontal line hits a connected set of squares and such that the successive columns of squares from left to right increase in height—the bottom of the column to the left is always lower or equal to the bottom of the column to the right. Similarly, the top of the column to the left is always lower than or equal to the top of the column to the right. Table 4 shows a set of such skew polyominos.

Another amazing result is that if you count the number of skew polyominos that have a perimeter of $2n + 2$, you will obtain C_n. Note that it is the perimeter that is fixed—not the number of squares in the polyomino.

2 A Recursive Definition

If you have convinced yourself that all the problems in the previous section are equivalent, it is only necessary to count one of them to have a count for all. If you have no idea how to begin, one good way is to write down a formula that relates the count for a given n to previously-obtained counts. It is usually easy to count the configurations for $n = 0$, $n = 1$, and $n = 2$ directly, and from there, you can count more complex versions.

In this section, we'll use the example with balanced parentheses discussed and illustrated in Section 1.1. Let us assume that we already have the counts for $0, 1, 2, 3, \cdots, n-1$ pairs and we would like to obtain the count for n pairs. Let C_i be the number of configurations of i matching pairs of parentheses, so $C_0 = 1$, $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, and $C_4 = 14$.

We know that in any balanced set, the first character has to be "(". We also know that somewhere in the set is the matching ")" for that opening one. In between that pair of parentheses is a balanced set of parentheses, and to the right of it is another balanced set:

$$(A)B,$$

where A is a balanced set of parentheses and so is B. Both A and B can contain up to $n-1$ pairs of parentheses, but if A contains k pairs, then B contains $n-k-1$ pairs.

Thus we can count all the configurations where A has 0 pairs and B has $n-1$ pairs, where A has 1 pair and B has $n-2$ pairs, and so on. Add them up, and we get the total number of configurations with n balanced pairs.

Here are the formulas. It is a good idea to try plugging in the numbers you know to make certain that you haven't made a silly error. In this case, the formula for C_3 indicates that it should be equal to $C_3 = 2 \cdot 1 + 1 \cdot 1 + 1 \cdot 2 = 5$.

$$\begin{aligned}
C_1 &= C_0C_0 && (1)\\
C_2 &= C_1C_0 + C_0C_1 && (2)\\
C_3 &= C_2C_0 + C_1C_1 + C_0C_2 && (3)\\
C_4 &= C_3C_0 + C_2C_1 + C_1C_2 + C_0C_3 && (4)\\
\cdots & \qquad\quad \cdots && \\
C_n &= C_{n-1}C_0 + C_{n-2}C_1 + \cdots + C_1C_{n-2} + C_0C_{n-1} && (5)
\end{aligned}$$

Consider others (the triangulation of the polygons in Section 1.5, for example) and see how exactly the same formulas relating the C_n arise.

Using the formula, it is easy to obtain the first few Catalan numbers: 1, 1, 2, 5, 14, 42, 132, 429, $1430, 4862, 16796, 58786, 208012, 742900, 2674440, 9694845, 35357670, 129644790, 477638700, 1767263190, 6564120420, 24466267020, 91482563640, 343059613650, 1289904147324, \ldots$

2.1 Counting Diagonal-Avoiding Paths

Perhaps the easiest way to obtain an explicit formula for the Catalan numbers is to analyze the number of diagonal-avoiding paths discussed in Section 1.3. We will do so by counting the total number of paths through the grid and then subtract off the number of paths that hit the diagonal.

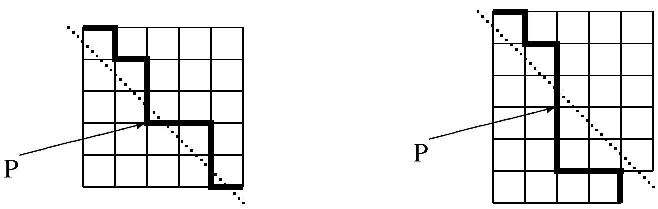


Figure 6: Modifying a Bad Path

Figure 6 illustrates a typical path that we do not want to count since it crosses the dotted diagonal line. Such a path may cross that line multiple times, but there is always a first time; in the figure, point P is the first grid point it touches on the wrong side of the diagonal. There will always be such a point P for every bad path.

For every such path, reflect the path beginning at P—every time the original path goes to the right, go down instead, and when the original path goes down, go to the right. It is clear that by the time the path reaches the point P it will have traveled one more step down than across, so it will have moved k steps to the right and $k+1$ steps down. The total path has n steps across and down, so there remain $n-k$ steps to the right and $n-k-1$ steps down. But since we swap steps to the right and steps down, the modified path with have a total of $(k)+(n-k-1) = n-1$ steps to the right and $(k+1)+(n-k) = n+1$ steps down. Thus every modified path ends at the same point, $n-1$ steps to the right and $n+1$ steps down.

Every bad path can be modified this way, and every path from the original starting point to this point $n-1$ to the right and $n+1$ down corresponds to exactly one bad path. Thus the number of bad paths is the total number of routes in a grid that is $(n-1)$ by $(n+1)$.

There are $\binom{m+k}{m}$ paths through an $k \times m$ grid[1]. Thus the total number of paths through the $n \times n$ grid is $\binom{2n}{n}$ and the total number of bad paths is $\binom{2n}{n+1}$. Thus C_n, the n^{th} Catalan number, or the total number of diagonal-avoiding paths through an $n \times n$ grid, is given by:

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

3 Counting Mountain Ranges—Method 1

A very similar argument can be made as in the previous section if we use the interpretation of the Catalan numbers based on the count of mountain ranges as described in Section 1.2. In that section, we are seeking arrangements of n up-strokes and n down-strokes that form valid mountain ranges.

[1] To see this, remember that there are m steps down that need to be taken along the $k+1$ possible paths going down. Thus the problem reduces to counting the number of ways of putting m objects in $k+1$ boxes which is $\binom{m+k}{m}$.

If we completely ignore whether the path is valid or not, we have n up-strokes that we can choose from a collection of $2n$ available slots. In other words, ignoring path validity, we are simply asking how many ways you can rearrange a collection of n up-strokes and n down-strokes. The answer is clearly $\binom{2n}{n}$.

Now we have to subtract off the bad paths. Every bad path goes below the horizon for the first time at some point, so from that point on, reverse all the strokes—replace up-strokes with down-strokes and vice-versa. It is clear that the new paths will all wind up 2 steps above the horizon, since they consist of $n+1$ up-strokes and $n-1$ down-strokes. Conversely, every path that ends two steps above the horizon must be of this form, so it corresponds to exactly one bad path.

How many such bad paths are there? The same number as there are ways to choose the $n+1$ up-strokes from among the $2n$ total strokes, or $\binom{2n}{n+1}$.

Thus the count of valid mountain ranges, or C_n, is given by exactly the same formula:

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

4 Counting Mountain Ranges—Method 2

Here is a different way to analyze the mountain problem. This time, imagine that we begin with $n+1$ up-strokes and only n down-strokes—we add an extra up-stroke to our collection.

First we solve the problem: How any arrangements can be made of these $2n+1$ symbols, without worrying about whether they form a "valid" mountain range (whatever that means with an unbalanced number of up-strokes and down-strokes). Clearly, if the ordering does not matter, there are $\binom{2n+1}{n}$ ways to do this.

One thing is certain, however. No matter how they are arranged, they mountain range will be one unit higher at the end, since we take $n+1$ steps up and only n steps down.

Let's look at a specific example with $n = 3$ (and $2n+1 = 7$): **up up down up up down down**. In Figure 4, we have arranged this sequence over and over and you can see that every 7 steps, the mountain range is one unit higher.

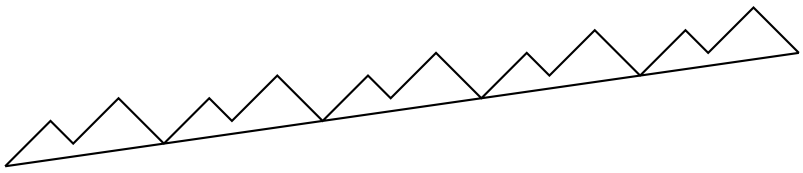


Figure 7: Growing Mountains

Since it is a repeating pattern, it's clear that we can draw a straight line below it that touches the bottom-most points of the growing mountain range.

In our example, this touching line seems to hit only once per complete set of 7 strokes, and we will show that this will always be the case, for any unbalanced number of up-strokes and down-strokes.

We can draw our mountain range on a grid, and it's clear that the slope of the line is $1/(2n+1)$ (it goes up 1 unit in every complete cycle of the pattern of $2n+1$ strokes. But lines with slope $1/(2n+1)$ can only hit lattice points every $2n+1$ units, so there is exactly one touching in each complete cycle.

If you have a series of $2n+1$ strokes, you can cycle that around to $2n+1$ arrangements. For example, the arrangement //\/\ can be cycled to four other arrangements: /\/\/, \/\//, /\//\ and \//\/. That means the complete set of arrangements can be divided into equivalence classes of size $2n+1$, where two arrangements are equivalent if they are cycled versions of each other.

If we consider the version among these $2n+1$ cycles, the only one that yields a valid mountain range is the one that begins at the low point of the $2n+1$ arrangement. Thus, to get a count of valid mountain ranges with n up-strokes and n down-strokes, we need to divide our count of $2n+1$ stroke arrangements by $2n+1$:

$$C_n = \frac{1}{2n+1}\binom{2n+1}{n} = \frac{1}{2n+1}\cdot\frac{(2n+1)!}{n!(n+1)!} = \frac{1}{n+1}\cdot\frac{(2n)!}{n!\,n!} = \frac{1}{n+1}\binom{2n}{n}.$$

5 Generating Function Solution

Using the formulas 1 through 5 in Section 2, we can obtain an explicit formula for the Catalan numbers, C_n using the technique known as generating functions.

We begin by defining a function $f(z)$ that contains all of the Catalan numbers:

$$f(z) = C_0 + C_1 z + C_2 z^2 + C_3 z^3 + \cdots = \sum_{i=0}^{\infty} C_i z^i.$$

If we multiply $f(z)$ by itself to obtain $[f(z)]^2$, the first few terms look like this:

$$[f(z)]^2 = C_0C_0 + (C_1C_0 + C_0C_1)z + (C_2C_0 + C_1C_1 + C_0C_2)z^2 + \cdots.$$

The coefficients for the powers of z are the same as those for the Catalan numbers obtained in equations 1 through 5:

$$[f(z)]^2 = C_1 + C_2 z + C_3 z^2 + C_4 z^3 + \cdots. \qquad (6)$$

We can convert Equation 6 back to $f(z)$ if we multiply it by z and add C_0, so we obtain:

$$f(z) = C_0 + z[f(z)]^2. \qquad (7)$$

Equation 7 is just a quadratic equation in $f(z)$ which we can solve using the quadratic formula. In a more familiar form, we can rewrite it as: $zf^2 - f + C_0 = 0$. This is the same as the quadratic equation: $af^2 + bf + c = 0$, where $a = z$, $b = -1$, and $c = C_0$. Plug into the quadratic formula and we obtain:

$$f(z) = \frac{1 - \sqrt{1-4z}}{2z}. \qquad (8)$$

Notice that we have used the $-$ sign in place of the usual $\pm$ sign in the quadratic formula. We know that $f(0) = C_0 = 1$, so if we replaced the $\pm$ symbol with $+$, as $z \to 0$, $f(z) \to \infty$.

To expand $f(z)$ we will just use the binomial formula on

$$\sqrt{1-4z} = (1-4z)^{1/2}.$$

If you are not familiar with the use of the binomial formula with fractional exponents, don't worry—it is exactly the same, except that it never terminates.

Let's look at the binomial formula for an integer exponent and just do the same calculation for a fraction. If n is an integer, the binomial formula gives:

$$(a+b)^n = a^n + \frac{n}{1}a^{n-1}b^1 + \frac{n(n-1)}{2\cdot 1}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3\cdot 2\cdot 1}a^{n-3}b^3 + \cdots.$$

If n is an integer, eventually the numerator is going to have a term of the form $(n-n)$, so that term and all those beyond it will be zero. If n is not an integer, and it is $1/2$ in our example, the numerators will pass zero and continue. Here are the first few terms of the expansion of $(1-4z)^{1/2}$:

$$\begin{aligned}(1-4z)^{1/2} \quad = \quad & 1 - \frac{\left(\frac{1}{2}\right)}{1}4z + \frac{\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)}{2\cdot 1}(4z)^2 - \frac{\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)}{3\cdot 2\cdot 1}(4z)^3 + \\ & \frac{\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)}{4\cdot 3\cdot 2\cdot 1}(4z)^4 - \frac{\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\left(-\frac{7}{2}\right)}{5\cdot 4\cdot 3\cdot 2\cdot 1}(4z)^5 + \cdots\end{aligned}$$

We can get rid of many powers of 2 and combine things to obtain:

$$(1-4z)^{1/2} = 1 - \frac{1}{1!}2z - \frac{1}{2!}4z^2 - \frac{3\cdot 1}{3!}8z^3 - \frac{5\cdot 3\cdot 1}{4!}16z^4 - \frac{7\cdot 5\cdot 3\cdot 1}{5!}32z^5 - \cdots \tag{9}$$

From Equations 9 and 8:

$$f(z) = 1 + \frac{1}{2!}2z + \frac{3\cdot 1}{3!}4z^2 + \frac{5\cdot 3\cdot 1}{4!}8z^3 + \frac{7\cdot 5\cdot 3\cdot 1}{5!}16z^4 + \cdots \tag{10}$$

The terms that look like $7\cdot 5\cdot 3\cdot 1$ are a bit troublesome. They are like factorials, except they are missing the even numbers. But notice that $2^2\cdot 2! = 4\cdot 2$, that $2^3\cdot 3! = 6\cdot 4\cdot 2$, that $2^4\cdot 4! = 8\cdot 6\cdot 4\cdot 2$, et cetera. Thus $(7\cdot 5\cdot 3\cdot 1)\cdot 2^4 4! = 8!$. If we apply this idea to Equation 10 we can obtain:

$$f(z) = 1 + \frac{1}{2}\left(\frac{2!}{1!1!}\right)z + \frac{1}{3}\left(\frac{4!}{2!2!}\right)z^2 + \frac{1}{4}\left(\frac{6!}{3!3!}\right)z^3 + \frac{1}{5}\left(\frac{8!}{4!4!}\right)z^4 + \cdots = \sum_{i=0}^{\infty}\frac{1}{i+1}\binom{2i}{i}z^i.$$

From this we can conclude that the i^{th} Catalan number is given by the formula

$$C_i = \frac{1}{i+1}\binom{2i}{i}.$$