Catalan Numbers

Tom Davis
tomrdavis@earthlink.net
http://www.geometer.org/mathcircles
February 19, 2016

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?

Perhaps a more precise definition of the problem would be this: A string of parentheses is valid if there are an equal number of open and closed parentheses and if you begin at the left as you move to the right, add 1 each time you pass an open and subtract 1 each time you pass a closed parenthesis, then the sum is always non-negative.

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

$n = 0$:	*	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. It will become clear later that this is exactly the right interpretation.

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

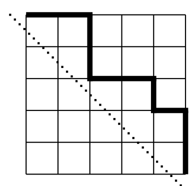
Table 2: Mountain Ranges

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 $n = 4$.

In our formal definition of a valid set of parentheses, we stated that if you add one for open parentheses and subtract one for closed parentheses that the sum would always remain non-negative. The mountain range interpretation is that the mountains will never go below the horizon.

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 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.

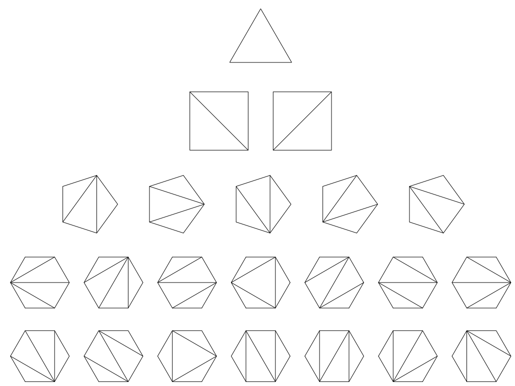


Figure 2: Polygon Triangulations

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.

1.5 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.

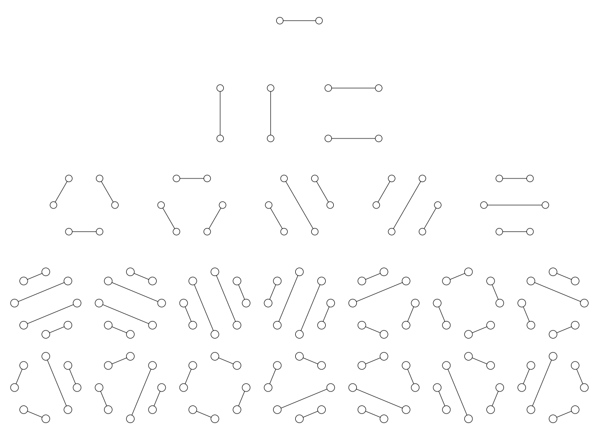


Figure 3: Hands Across the Table

1.6 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 $0 \le n \le 3$. There are $1, 1, 2$, and 5 of them. Try to draw the 14 trees with $n = 4$ internal nodes.

A rooted binary tree is an arrangement of points (nodes) and lines connecting them where there is a special node (the root) and as you descend from the root, there are either two lines going down or zero. Internal nodes are the ones that connect to two nodes below.

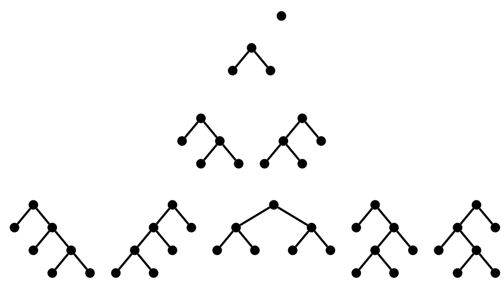


Figure 4: Binary Trees

1.7 Plane Rooted Trees

A plane rooted tree is just like the binary tree above, except that a node can have any number of sub-nodes; not just two.

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

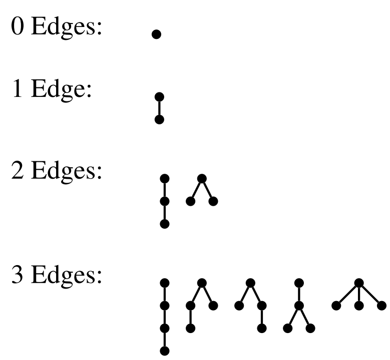


Figure 5: Plane Rooted Trees

1.8 Skew Polyominos

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

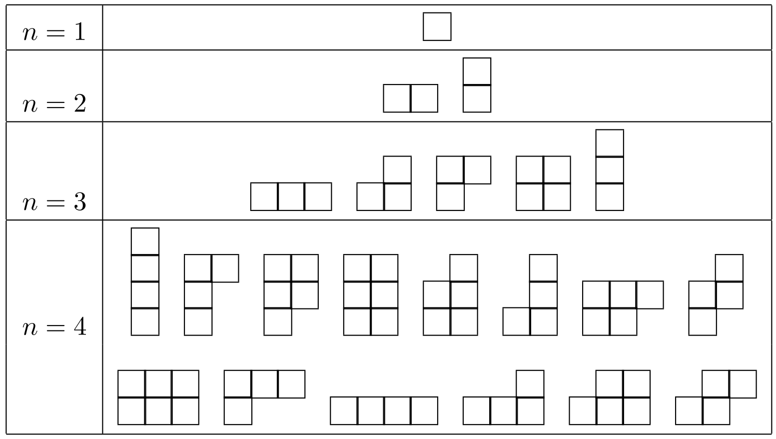


Table 3: Skew Polyominos with Perimeter $2n + 2$

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 3 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.

1.9 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 4.

$n = 0$	(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 4: 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 4 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.

2 A Recursive Definition

The examples above all seem to generate the same sequence of numbers. In fact it is obvious that some are equivalent: parentheses, mountain ranges and diagonal-avoiding paths, for example. Later on, we will prove that the other seqences are also the same. Once we're convinced that they are the same, we only need to have a formula that counts any one of them and the same formula will count them all.

If you have no idea how to begin with a counting problem like this, one good approach 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 $C_4 = 14$, which can be obtained by direct counts.

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. Notice that we will allow either A or B to contain zero pairs, and that there is exactly one way to do so: don't write down any parentheses.

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 $C_3 = 2\cdot 1 + 1\cdot 1 + 1\cdot 2 = 5$.

$$\begin{aligned}
C_1 &= C_0C_0 && (1)\\
C_2 &= C_1C_0 + C_0C_1 && (2)\\
C_3 &= C_2C_0 + C_1C_1 + C_0C_2 && (3)\\
C_4 &= C_3C_0 + C_2C_1 + C_1C_2 + C_0C_3 && (4)\\
\cdots & \qquad\qquad \cdots
\end{aligned}$$

$$C_n \quad = \quad C_{n-1}C_0 + C_{n-2}C_1 + \cdots + C_1C_{n-2} + C_0C_{n-1} \tag{5}$$

Beginning in the next section, we will be able to use these recursive formulas to show that the counts of other configurations (triangulations of polygons, rooted binary trees, rooted tress, et cetera) satisfy the same formulas and thus must generate the same sequence of numbers.

But simply by using the formulas above and a bit of arithmetic, 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, ...

2.1 Counting Polygon Triangulations

It is not hard to see that the polygon triangulations discussed in section 1.4 can be counted in much the same way as the balanced parentheses. See Figure 6.

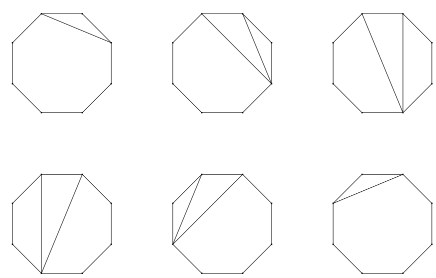


Figure 6: Octagon Triangulations

In the figure we consider the octagon, but it should be clear that the same argument applies to any convex polygon. Consider the horizontal line at the top of the polygon. After triangulation, it will be part of exactly one triangle, and in this case, there are exactly six possible triangles of which it can be a part. In each case, once that triangle is selected, there is a polygon (possibly empty) on the right and the left of the original triangle that must itself be triangulated.

What we would like to show is that a convex polygon with $n > 3$ sides can be triangulated in C_{n-2} ways. Thus the octagon should have $C_{8-2} = C_6$ triangulations.

For the example in the upper left of Figure 6, the triangle leaves a 7-sided figure on the left and an empty figure (essentially a two-sided polygon) on the right. This triangulation can be completed by triangulating both sides; the one on the left can be done in C_5 ways and the empty one on the right, C_0 ways, for a total of $C_5 \cdot C_0$. The middle example on the top leaves a pentagon and a triangle that, in total, can be trianguated in $C_4 \cdot C_1$ ways. Similar arguments can be made for all six positions of the triangle containing the top line, so we conclude that:

$$C_6 = C_5 \cdot C_0 + C_4 \cdot C_1 + C_3 \cdot C_2 + C_2 \cdot C_3 + C_1 \cdot C_4 + C_0 \cdot C_5,$$

which is exactly how the Catalan numbers are defined for the nested parentheses.

Convince yourself that a similar argument can be made for any size original convex polygon.

2.2 Counting Non-Crossing Handshakes

To count the number of hand-shakes discussed in Section 1.5 we can use an analysis similar to that used in section 2.1.

If there are $2n$ people at the table pick any particular person, and that person will shake hands with somebody. To admit a legal pattern, that person will have to leave an even number of people on each side of the person with whom he shakes hands. Of the remaining $n-1$ pairs of people, he can leave zero on the right and $n-1$ pairs on the left, 1 on the right and $n-2$ on the left, and so on. The pairs left on the right and left can independently choose any of the possible non-crossing handshake patterns, so again, the count C_n for n pairs of people is given by:

$$C_n = C_{n-1}C_0 + C_{n-2}C_1 + \cdots + C_1C_{n-2} + C_0C_{n-1},$$

which, together with the fact that $C_0 = C_1 = 1$, is just the definition of the Catalan numbers.

2.3 Counting Trees

Counting the binary trees discussed in Section 1.6 is similar to what we've done previously. Obviously there is one way to make a rooted binary tree with zero or one internal node. To work out the number of trees with n internal node, note that one of those n nodes is the root node, and then the $n-1$ additional internal nodes must be distributed on the left or the right below the root node. These can be distributed as 0 on the left and $n-1$ on the right, 1 on the left and $n-2$ on the right, and so on, yielding exactly the same formula that we had in every previous example.

To count the rooted plane trees discussed in Section 1.7 we use the same strategy. There is one example each for trees with zero and one edge, so the counts here are the same: $C_0 = C_1 = 1$.

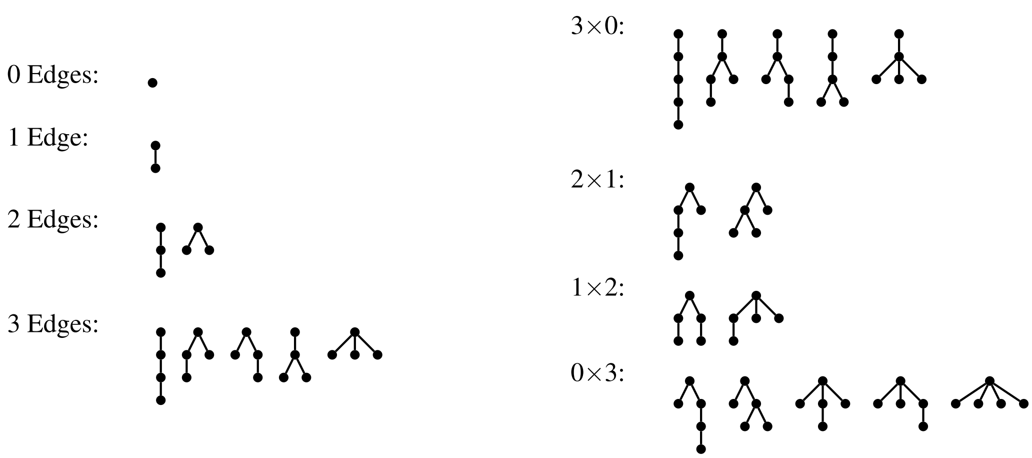


Figure 7: Plane Rooted Trees With 4 Edges

Now, to count the number of plane rooted trees with $n > 1$ edges we again begin from the root. There is at least one edge going down (leaving us with $n-1$ edges to draw). The remaining $n-1$ edges can be placed below that initial edge or hooked directly to the root node to the right of that edge. The $n-1$ edges, as before, can be distributed to these two locations as 0 and $n-1$, as 1 and

$n-2$, et cetera. It should be clear that the same formula defining the Catalan numbers will apply to the count of rooted plane trees.

In Figure 7 the table on the left dupliates the structure of trees with 3 or fewer edges and the table on the right shows how the trees with 4 edges are generated from them.

2.4 Counting Diagonal-Avoiding Paths

Up to now we do not have an explicit formula for the Catalan numbers. We know that a large collection of problems all have the same answers, and we have a recursive formula for those numbers, but it would be nice to have an explicit form.

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.

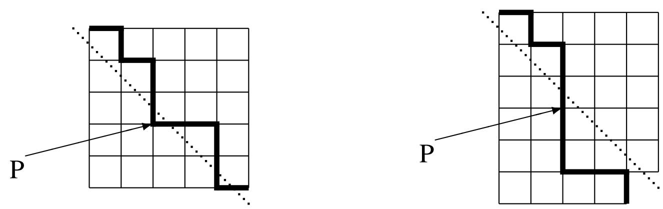


Figure 8: Modifying a Bad Path

Figure 8 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:

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

[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 $\binom{m+k}{m}$.

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.

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 below the horizon, since they consist of $n+1$ down-strokes and $n-1$ up-strokes. Conversely, every path that ends two steps below 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.

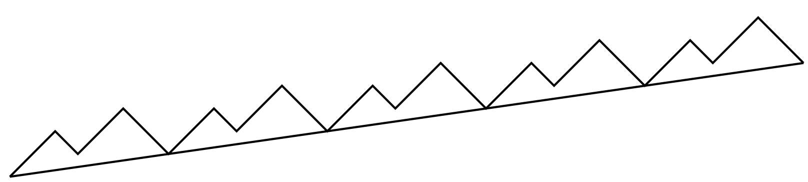


Figure 9: Growing Mountains

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 $2n+1$:

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

Finally, note that when the line is drawn that touches the bottom edge of the range of mountains with one more "up" than "down", the first steps after the touching points are two "ups", since an "up-down" would immediately dip below the line. It should be clear that if one of the two initial "up" moves is removed, the resulting series will stay above a horizontal line.

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:

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

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

$$f(z) = C_0 + z[f(z)]^2. \tag{7}$$

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:

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

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:

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

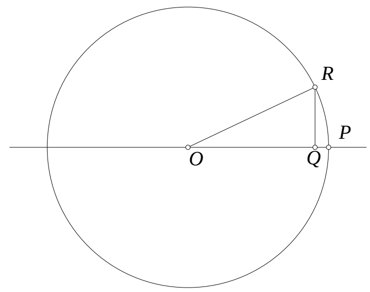
The terms that look like $7\cdot 5\cdot 3\cdot 1$ are a bit troublesome. They are like factorials, except they are missing the even numbers. But notice that $2^2 \cdot 2! = 4\cdot 2$, that $2^3\cdot 3! = 6\cdot 4\cdot 2$, that $2^4\cdot 4! = 8\cdot 6\cdot 4\cdot 2$, et cetera. Thus $(7\cdot 5\cdot 3\cdot 1)\cdot 2^4 4! = 8!$. 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

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

5.1 A Strange Geometric Result



Consider the figure above. Assume that the circle has radius r, so $OR = r$ and that the length of RP is 1. What is the length of QP? We know $QP = r - OQ$ and $OQ = \sqrt{r^2 - 1}$ so:

$$QP = r - \sqrt{r^2 - 1}.$$

Let $r = 5$ and we obtain:

$$QP = 0.101020514433643803\ldots$$

Notice that the catalan numbers appear in the decimal expansion; we can see $1, 1, 2, 5, 14$ and "almost" 42. What's going on?

What's even stranger is if we let $r = 5,000,000$. We obtain:

$$\begin{aligned} QP \quad = \quad & .00000010000000000000100000000000002000000000000050000000000001400000 \\ & 00000004200000000000132000000000004290000000000143000000000048620000 \\ & 00000016796000000000587860000000020801200000000742900000000267444000 \\ & 00000096948450000003535767000001296447900000477638700000176726319 \\ & 00000656412042000024466267020000914825636400034305961365001289904147 \\ & 32404861946401452183673530721526953355091600663747951750370022422166 \\ & 514061498650209244944636039227464340648770503213613041225123944042\ldots \end{aligned}$$

where we find the first 25 Catalan numbers surrounded by various numbers of zeroes.

Hint: Look at the generating function for the Catalan numbers.

6 Catalan's Triangle

In this section we will consider a triangle somewhat akin to Pascal's triangle that will provide a nice method to generate the Catalan numbers.

1	6	20	48	90	132	132
1	5	14	28	42	42	
1	4	9	14	14		
1	3	5	5			
1	2	2				
1	1					
1						

Ignoring the dots for a moment, the rule is simple: As we build up, each row contains one more number than the previous. The first row consists of a single 1. Once a row is complete construct the next row up beginning over the left-most element of the row below and the number placed there is the sum of the number directly below it and the number directly to the left. If there is no number in the slot below or in the slot to the left, just use zero. To make sure you understand the rule, construct the eighth row and make sure that you obtain: $1, 7, 27, 75, 165, 297, 429, 429$.

Notice that the numbers running up the diagonal of this triangle, $1, 1, 2, 5, 14, 42, 132$, are the Catalan numbers. Why is this? You may wish to experiment a little before reading on.

In Section 1.3 we saw that the number of diagonal-avoiding paths is counted by the Catalan numbers. For any point in the triangle above, and consider the problem of counting the number of paths from that point to the bottom-most point where the only allowable moves are one step to the left or one step down, where you are constrained to remain on the lattice points of the triangle.

If we begin at the lowest one, we are already there, so there is only one path; namely, the empty path: don't do anything and you're done. The two 1's in the next row make sense, too, since there's only one path to the bottom from each. (One of the paths is two steps long and one is only one step long, but we are not counting the number of steps, but the number of paths, and there is only one path from each of the points.)

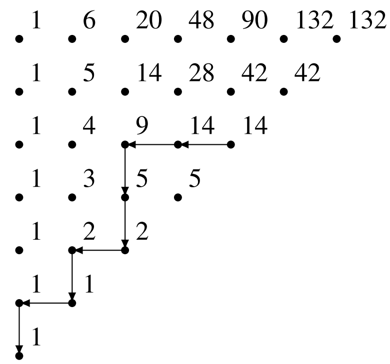
Now start with any other point in the triangle. If that point is on the left of the row, the only possible path is straight down, so there's only one path. If the point is on the right-most end of the row, the only move you can make is to the left, so there are the same number of paths from there to the bottom as there are from the point immediately to your left. For any other point, it is possible to make the initial move to the left or down, in which case the number of paths to the bottom is the sum of the number of paths to the bottom from the point below you and from the point immediately to your left. Notice that this is exactly the same rule we used to generate the numbers in the grid, so every number in the triangle represents the number of paths to the bottom from the point immediately below and to the left of that number.

The paths that start on the diagonal are the paths we counted when we were generating the Catalan numbers, so the diagonal numbers are the Catalan numbers.

6.1 Enumerating all Diagonal-Avoiding Paths

A very interesting observation that was pointed out to me by Patrick Labarque is the following. Suppose we'd like to number all the diagonal-avoiding paths, giving the empty path number 0, The unique path on the triangle with three dots is numbered 1, the two paths on the next larger triangle with six dots get numbers 2 and 3, and so on.

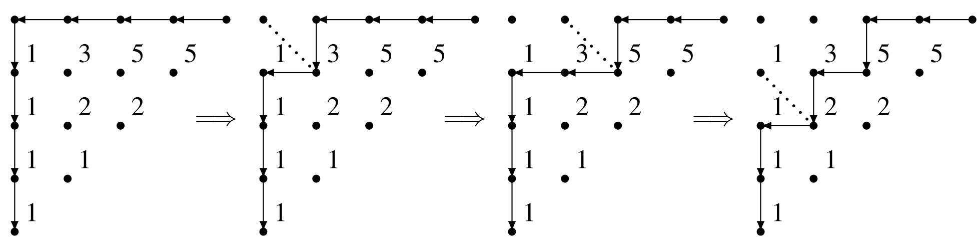
The enumeration can be done by drawing the path as in the sample illustration below:



The number to assign to each path is the sum of the numbers under and to the right of the path; in this case, $1 + 1 + 2 + 5 + 5 = 14$.

It is not too hard to prove that every path from a diagonal point to the bottom-most point has a unique sum of the grid numbers below it and that there are no omitted numbers, assuming that the null path is assigned number 0.

First notice a few things. If we add all the numbers in a particular row, their sum is the Catalan number on the right end of the row above. For example, in the fourth row up, we have $1+3+5+5 = 14$ which is the right-most number in the fifth row up. This makes sense, since if we consider paths starting on the row above, we will move left at least once, but perhaps all the way to the left before taking our first step down. When we take the first downward step, the number below is the number of ways to complete a path where that is the first step down. The sum of all of them, or in other words, the sum of the numbers in the row, is the total number of paths, which is the Catalan number.



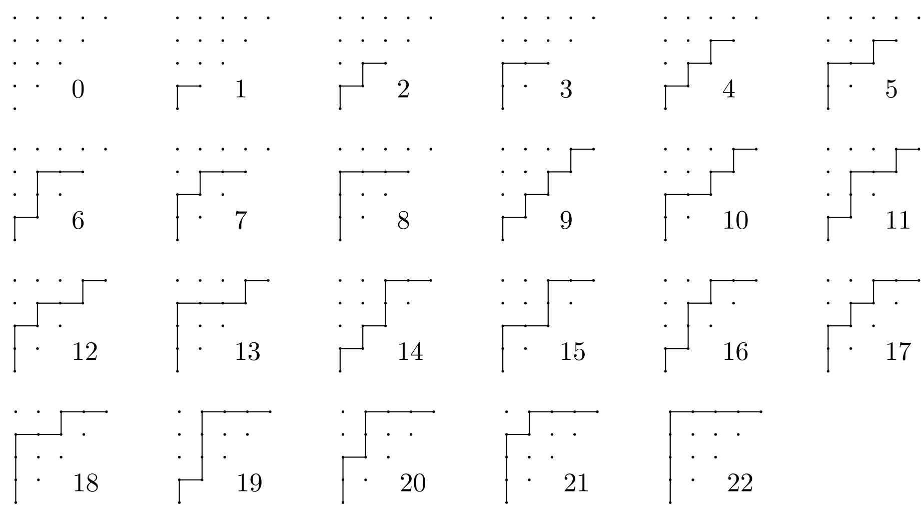
To show why the enumeration works, let's assume that all the paths are correctly enumerated for the "mountain range paths" for diagonals of length $0, 1, 2$ and 3. The empty path is assigned the number 0, the unique path that takes one step on the diagonal is assigned 1, the next two are assigned 2 and 3, and the next five are assigned $4, 5, 6, 7$ and 8.

We now would like to make sure that the next fourteen paths are assigned to the numbers $9, 10, \ldots, 22$. If we begin by looking at the path that encloses all the numbers as in the figure above, the top row sums to 14, and that added to the numbers below that sum to 8 yields 22.

We can obtain any other path by doing successive "foldings" of the original path, removing one number at a time and always folding from upper left to lower right. In the figure above, three steps of unfolding are illustrated, and the dotted paths show the motion of each fold.

If any folding occurs, the first one has to be from the upper left corner. This produces a path with a 1 removed from below it, and it eliminates exactly one path, so this path will be assigned $22 - 1 = 21$. In the illustration, the next fold eliminates three possible paths. Remember that the 3 that was exposed refers to the number of paths from the dot to its lower left to the bottom. When we do that fold, since we've now skipped three paths, we need to subtract 3 from 21, yielding path number 18, et cetera. You can also see that when the path is completely folded down to the minimum path covering the four diagonal steps, the only remaining numbers are $1 + 1 + 2 + 5 = 9$: the first available number after the $0, 1, \ldots, 8$ that were used to enumerate all the shorter paths.

Following is a list of all the paths from number 0 to number 22. The first one (or zeroth one, if you prefer), of course is the empty path. It is a worthwhile exercise to check that the sum of the numbers under at least a few of these paths is equal to the path number.



7 More Examples Without Proof

Here are some more counting problems whose answer is "the Catalan numbers". You can use these as exercises.

7.1 Permutations avoiding 123

A permutation of n numbers consists of a rearrangement of those n numbers. Without any constraints, there are $n!$ permutations. To completely define a permutation, all that is required is an n-tuple of the numbers $\{1, 2, \ldots, n\}$ with the following interpretation: The n-tuple $(p_1, p_2, \ldots, p_n)$ is the permutation that takes 1 to p_1, 2 to p_2, ..., n to p_n. All the p_i in such an n-tuple are distinct. For example, here is a list of all the permutations of the set $\{1, 2, 3\}$:

$$(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1). \tag{11}$$

We will say that a permutation "avoids 123" if in the n-tuple as described above, it is impossible to find three numbers i, j and k such that $i < j < k$ and $p_i < p_j < p_k$. In other words, the n-tuple contains no subsequence of length three that is increasing. For example $(4, 1, 2, 3)$ because you can find 1, 2 and 3 in order. The permutation $(2, 1, 3, 4)$ fails because of the subsequence 2, 3 and 4

(and also the subsequence 1, 3 and 4). As an example of one that works, the permutation $(4, 1, 3, 2)$ avoids 123.

By definition, all the permutations of 0, 1 or 2 elements avoid 123, since we can't find three different numbers in the n-tuples. There are 1, 1 and 2 such permutations. For $n = 3$, there is only one permutation from the list in List 11 that fails: $(1, 2, 3)$, leaving 5 permutations that avoid 123.

For $n = 4$, there are 24 permutations, 14 of which avoid 123 and 10 of which that do not. Here is a list of the 14:

$$(1,4,3,2),(2,1,4,3),(2,4,1,3),(2,4,3,1),(3,1,4,2),$$
$$(3,2,1,4),(3,2,4,1),(3,4,1,2),(3,4,2,1),(4,1,3,2),$$
$$(4,2,1,3),(4,2,3,1),(4,3,1,2),(4,3,2,1).$$

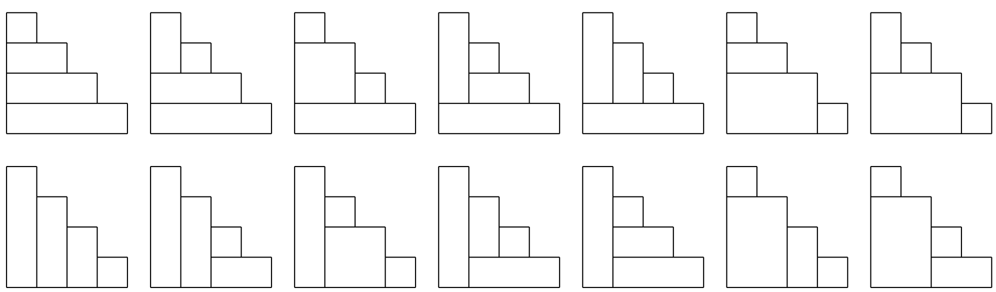
Here is a list of the 10 that do not avoid 123:

$$(1,2,3,4),(1,2,4,3),(1,3,2,4),(1,3,4,2),(1,4,2,3),$$
$$(2,1,3,4),(2,3,1,4),(2,3,4,1),(3,1,2,4),(4,1,2,3).$$

The number of permutations of n elements that avoid 123 is C_n.

7.2 Tiling with Rectangles

Given a "triangular" region composed of n blocks on a side, in how many different ways can the region be tiled with exactly n rectangles? The illustration below shows that for $n = 4$ there are exactly $C_4 = 14$ tilings.



8 Some Interesting Matrices

An interesting property of the Catalan numbers is that if you form either of the following Hankel matrices, their determinants are as follows. These results (and others) are nicely proved using arguments related to counting paths in the following paper by Mays and Wojciechowski:

`http://www.math.wvu.edu/ jerzy/research/21catalan.pdf`

It is also true that the only sequence $C_0, C_1, C_2, \ldots$ that satisfies the following equations is the Catalan numbers.

$$Det \begin{pmatrix} C_0 & C_1 & C_2 & \dots & C_n \\ C_1 & C_2 & C_3 & \dots & C_{n+1} \\ C_2 & C_3 & C_4 & \dots & C_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_n & C_{n+1} & C_{n+2} & \dots & C_{2n} \end{pmatrix} = 1.$$

$$Det \begin{pmatrix} C_1 & C_2 & C_3 & \dots & C_n \\ C_2 & C_3 & C_4 & \dots & C_{n+1} \\ C_3 & C_4 & C_5 & \dots & C_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_n & C_{n+1} & C_{n+2} & \dots & C_{2n-1} \end{pmatrix} = 1.$$

$$Det \begin{pmatrix} C_2 & C_3 & C_4 & \dots & C_n \\ C_3 & C_4 & C_5 & \dots & C_{n+1} \\ C_4 & C_5 & C_6 & \dots & C_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_n & C_{n+1} & C_{n+2} & \dots & C_{2n-2} \end{pmatrix} = n.$$

It is also very interesting to note that the determinants of the following Hankel matrices (and their obvious generalizations) are very tiny:

$$Det \begin{pmatrix} 1/1 & 1/1 & 1/2 & 1/5 & 1/14 \\ 1/1 & 1/2 & 1/5 & 1/14 & 0 \\ 1/2 & 1/5 & 1/14 & 0 & 0 \\ 1/5 & 1/14 & 0 & 0 & 0 \\ 1/14 & 0 & 0 & 0 & 0 \end{pmatrix} = \frac{1}{537824}.$$

$$Det \begin{pmatrix} 1/1 & 1/1 & 1/2 & 1/5 & 1/14 \\ 1/1 & 1/2 & 1/5 & 1/14 & 1/42 \\ 1/2 & 1/5 & 1/14 & 1/42 & 1/132 \\ 1/5 & 1/14 & 1/42 & 1/132 & 1/429 \\ 1/14 & 1/42 & 1/132 & 1/429 & 1/1430 \end{pmatrix} = -\frac{1}{122489812645200000}.$$