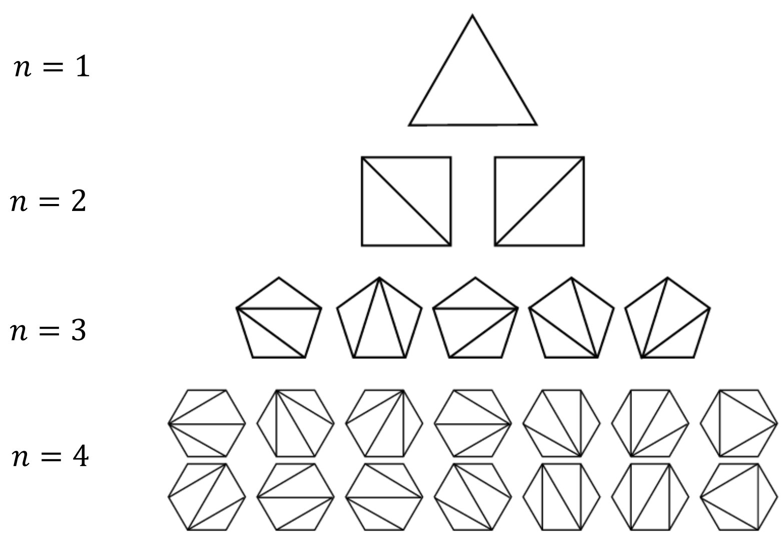
Gen Lipkin

Catalan Numbers and Combinatorial Applications

In a letter to C. Goldbach in 1751, Leonhard Euler posed the question of how to find the number of ways to subdivide a convex n-gon into triangles by drawing nonintersecting diagonals in its interior. This problem would become the foundation for the Catalan numbers.

Let us take Euler's problem and represent it using variables. Let C_n be the number of different triangulations of a $(n+2)$-gon, where C_n and n are nonnegative integers. If we were to just use trial and error, we could determine the triangulations of several polygons as seen in Figure 1. We would see that a triangle could be triangulated one way – itself, a quadrilateral could be divided into triangles using 2 different diagonals, a pentagon could be triangulated 5 ways, and so on.

Figure 1



n	**1**	**2**	**3**	**4**
$\boldsymbol{C}$	1	2	5	14

These numbers, C_n, are what we refer to as the Catalan numbers.

Although interesting, it is not practical to draw out each polygon and look for each distinct triangulation. Thankfully, scientist Johann Segner came to the rescue in 1758. He created a formula to solve for C_n, given n. His formula is below, given that $C_0 = 1$.

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-k-1}$$

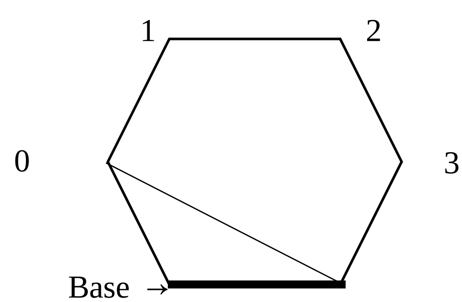
Let's try out his formula and see if it works for a 6-gon, since we already know that C_6 should be 14. First, n + 2 = 6 so n = 4.

$$C_6 = C_0C_3 + C_1C_2 + C_2C_1 + C_3C_0 = 15 + 12 + 21 + 51 = 5 + 2 + 2 + 5 = 14$$

We arrive at $C_6 = 14$, which is correct. As we saw while solving the above example, Segner's formula is recursive and requires knowing all the terms before C_n in order to solve for C_n itself.

Let us refer back to our example and arbitrarily pick a side of the 6-gon to be the base and label the points not on the base 0 through 3. We will start off with drawing our first diagonal so that it creates a triangle with its base as the one we chose and its vertex at Point 0, as seen in Figure 2.

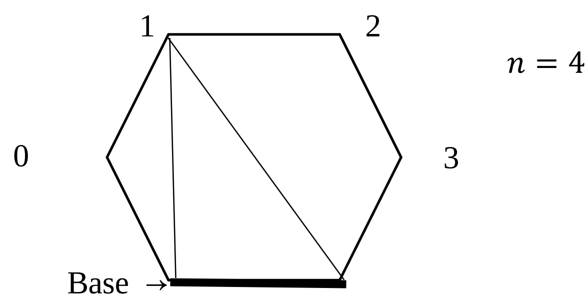
Figure 2



This diagonal divides the polygon into an angle and a 5-gon; we do not count the base as a side. Recall that the number of triangulations for an angle (C_0) is 1 and the number of triangulations for a 5-gon (C_3) is 5. This is how we get the term C_0C_3, and what it represents – the product of the number of triangulations of the shapes our diagonal made.

With our designated base, we create another triangle with a vertex at Point 1, as seen in Figure 3.

Figure 3



These diagonals divide the polygon into two shapes, excluding the triangle we created since we don't count the base as a side. To the left, the triangle created has 1 triangulation (C_1). To the right, the 5-gon has 5 triangulations (C_5). This is how we arrive at the term C_1C_5. This pattern continues with the triangle's vertex as Point 2 and then finally Point 3, which is how we arrive at the terms C_2C_1, and C_3C_0, respectively, from Segner's formula.

While Segner's formula allows us to find C_n far more easily than tediously drawing and counting triangulations, it requires us to do so by finding all the Catalan terms prior to C_n.

Euler created a formula that wasn't recursive – all you need is the value of n to solve for C_n. His formula was

$$C_n \quad = \qquad \frac{2 * 6 * 10 * \ldots * (4n - 2)}{(n + 1)!}$$

in which you multiply by increasing intervals of 4 until you reach the number $4n - 2$ and then divide the product by the factorial $n + 1$. Let's try the formula with our previous example of $n = 4$.

$$C_n = \frac{2 * 6 * 10 * 14}{5!} = \frac{1{,}650}{120} = 140$$

We see that Euler's formula works and requires less information than Segner's. This formula, however, can get quite laborious when n equals a large number (for example, $n = 240$). You'd need to keep adding 4 to 10 all the way until you reached 958. Talk about a lot of work to solve a problem!

In 1838, French mathematician Joseph Liouville was asked for a simple way to derive Euler's formula from Segner's recurrence. Another French mathematician Gabrielle Lame wrote to Liouville to tell him he had found a solution. That solution was published in the *Journal de Matȟ ematiques Pures et Appliquees*, which was a journal founded by Liouville.

Lame counted the number of triangulations An of a (n+2)-gon with one of its (n-1) diagonals aligned. "On the one hand, An = 2(n − 1)Cn. On the other hand, by summing over all possible directed diagonals we have An = n C1Cn−1 +C2Cn−2 +...+CnC1 " (Pak, *History of Catalan Numbers).* Combining the two above formulas with Segner's recurrence gives us Euler's formula.

In 1839, a Belgian mathematician Eugene Catalan, a former student of the Liouville, was studying at Ecole Polytechnique. Having seen Lame's work, he was inspired to obtain the now standard formulas

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

He is, of course, who the number sequence is named after. The above formulas provide a direct way to arrive at the nth Catalan number.

We will now go a hundred years in time and look at the work of H. Urban, who built upon Segner's formula. In 1941, H. Urban noticed a pattern for *Cn*. Using Segner's formula, he saw that:

[IMAGE ON NEXT PAGE]

$$
\begin{aligned}
\frac{C_1}{C_0} &= \frac{1}{1} = \frac{2}{2} = \frac{4 \cdot 1 - 2}{1 + 1} \\
\frac{C_2}{C_1} &= \frac{2}{1} = \frac{6}{3} = \frac{4 \cdot 2 - 2}{2 + 1} \\
\frac{C_3}{C_2} &= \frac{5}{2} = \frac{10}{4} = \frac{4 \cdot 3 - 2}{3 + 1} \\
\frac{C_4}{C_3} &= \frac{14}{5} = \frac{14}{5} = \frac{4 \cdot 4 - 2}{4 + 1} \\
\frac{C_5}{C_4} &= \frac{42}{14} = \frac{18}{6} = \frac{4 \cdot 5 - 2}{5 + 1}
\end{aligned}
$$

Urban, noticing a clear pattern, inferred that $\frac{C_n}{C_{n-1}} = \frac{4n - 2}{n-1}$, where n ≥ 1 since we can't have C_{-1}. It's a given that $C_0 = 1$. If we multiply C_{n-1} on both sides, we get that $C_n = \frac{4n - 2}{n-1} * C_{n-1}$. Urban's formula requires a lot less work than Euler's and Segner's. It is recursive but only requires knowing the one term directly before C_n, not all of them like in Segner's formula.

As it has become custom, let's try out this new formula with $n = 4$.

$$C_4 = \frac{4(4) - 2}{4-1} * C_3 = \frac{14}{5} * 5 = 14.$$

Once again, the newly introduced formula works!

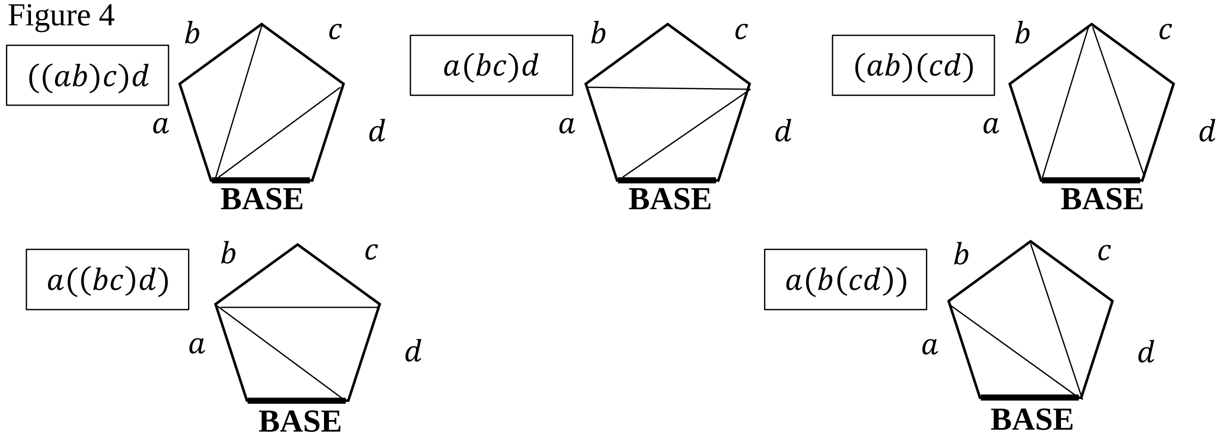
What else can these Catalan numbers do? The Catalan numbers C_n have many different= combinatorial applications. One of which is calculating the number of ways C_n of parenthesizing applications of a binary operator. A binary operator is an operation in which you can only perform on two numbers at a time. In multiplication, for example, you can only multiply 2 numbers at one time.

Say we wanted to find the number of ways of parenthesizing $a * b * c * d$, while still keeping that order of variables; we would find 5 ways. One way would be $((ab)c)d$ in which we multiply a and b, then multiply their product by c, then finally multiply *that* product by d. We are still only multiplying 2 numbers at one time. The order of parentheses doesn't matter because the order in which we multiply the numbers doesn't matter due to the associative property.

The five ways of parenthesizing $a * b * c * d$ are $((ab)c)d$, $a(bc)d$, $(ab)(cd)$, $a((bc)d)$, and $a(b(cd))$. We can actually represent these ways as triangulated polygons. Let us create a 5-gon with an arbitrarily designated base and 4 sides labeled $a, b, c,$ and d.

First we connect the sides of the two numbers we're multiplying together first. For $((ab)c)d$, we first connect sides a and b by drawing a diagonal because we multiplied a and b first. This creates a triangle. Then we draw a diagonal so that a shape is created that includes side c and sides a and b because we then multiplied c with the product of a and b (Figure 4).

Figure 4



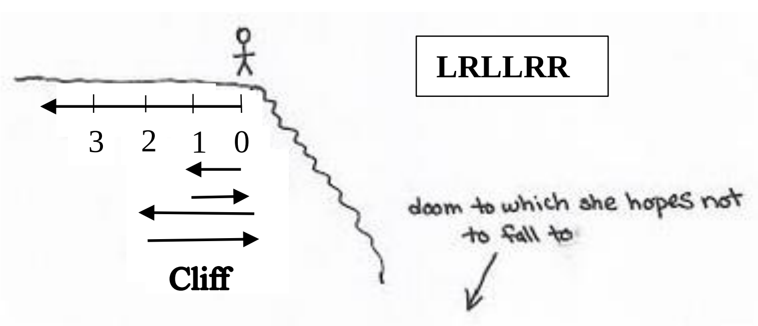
The polygon we created has 5 sides, so there are C_3 different ways of triangulating it (Recall that a 5-gon has $n + 2$ sides, so in this case $n = 3$.) and therefore C_3 different ways of parenthesizing $a * b * c * d$. We can conclude that the number of ways of parenthesizing a binary operator is equal to C_{r-1}, where r is the number of terms. For $a * b * c * d$, there are 4 terms so $r = 4$; the number of ways of parenthesizing $a * b * c * d$ is C_3, which is 5.

There is yet another application of the Catalan numbers. A Dyck word is a word that contains an equal amount of two different "letters;" they can be anything from numbers to letters to parentheses. In a Dyck word, the number of the second letter can never be greater than that of the first letter in an initial string. An initial string is a part of the word that must never have more second letters grouped together than previous first letters.

Let's look at a concrete example. Let's arbitrarily pick L and R to be our two letters in the Dyck words. We want to find C_n, the number of Dyck words we can form with an n amount of L's and an n amount of R's. Let's have $n = 3$. There are five different Dyck words we can write because we know that $C_3 = 5$. They are LLLRRR, LLRLRR, LRLLRR, LRLRLR, and LLRRLR.

The Dyck words can be graphed as directions. Say there is a person standing on a cliff, and we're giving her directions. She can never move more to the right in one move than to the left or she'll fall, as seen in Figure 5.

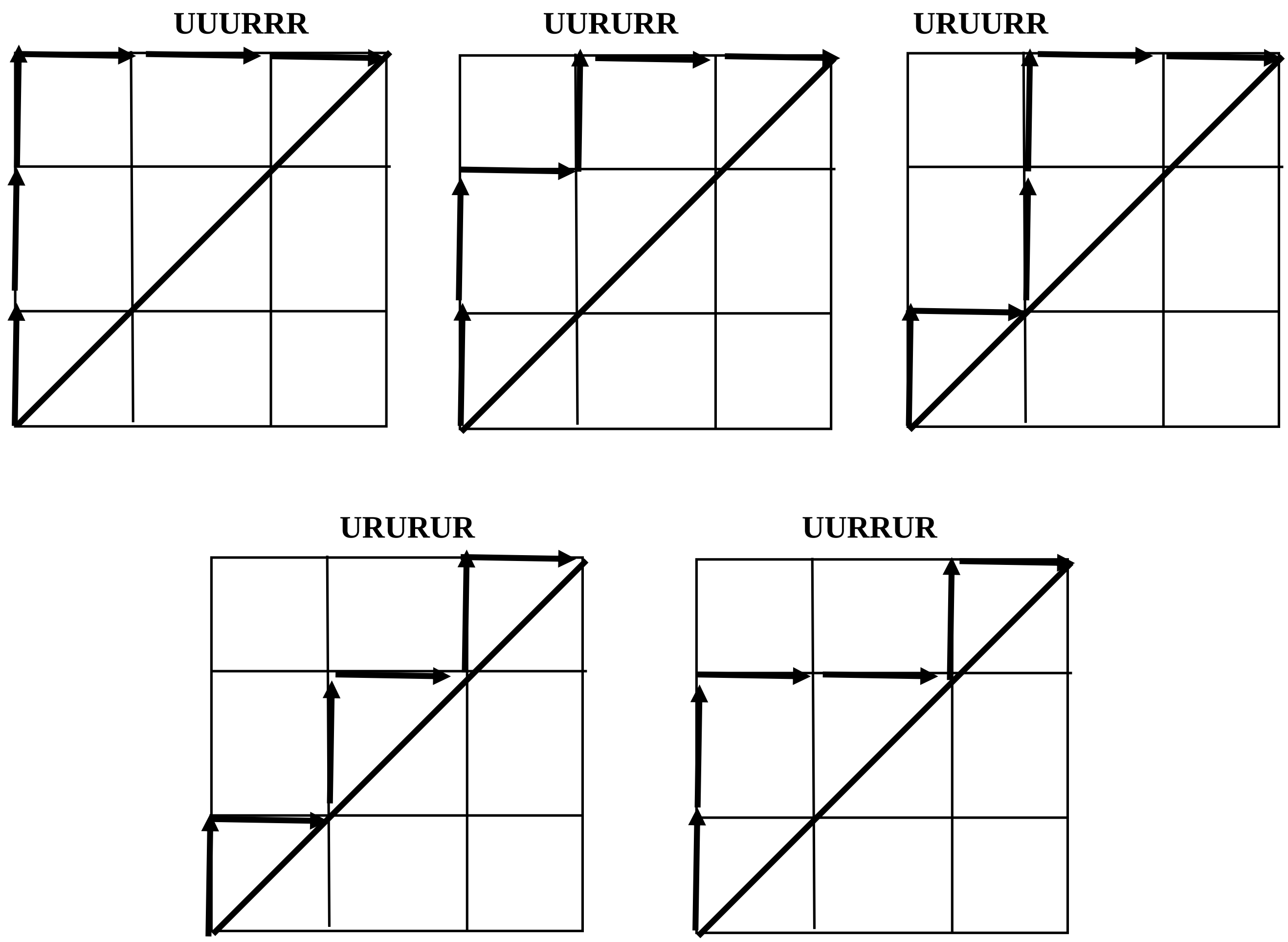
Figure 5



The R's, correspondingly, can never be grouped together more than the following L's.

The Dyck words can also be graphed in 2 dimensions. If we replace our L's (lefts) with U's (ups), we can graph a 2-D path. We can use this modified example to illustrate a different scenario. If we draw a square and start from the bottom left corner and travel to the upper right corner, we can count our paths using Dyck words. Since there will never be more R's than U's in an initial string, we know that we will never cross the diagonal of the square that connects our starting and finishing points, as seen in Figure 6.

Figure 6



The number of paths is still 5 because we did not change the number of U's and R's.

If we were curious as to the total number of paths we could take if weren't concerned with passing the diagonal (thus the number of R's would be able to surpass the number of U's), we could calculate it by $\binom{q}{w}$, where q is the total number of letters and w is the number of each letter. In our examples, the total number of paths would be $\binom{6}{3}$, or 20.

And that's it! You have now become acquainted with the Catalan numbers and several of their combinatorial applications. This paper serves mainly as an introduction to wet your palate. If you'd like to delve deeper, there are a great deal more applications than those listed here.

Works Cited

http://www.math.ucla.edu/~pak/lectures/Cat/Larcombe-The_18th_century_Chinese_discovery_of_the_Catalan_numbers.pdf

http://www.youtube.com/watch?v=VqJLED3KT4E

Catalan Numbers with Applications by Thomas Koshy

http://en.wikipedia.org/wiki/Catalan_number#Applications_in_combinatorics

http://www.loria.fr/cortier/Slides-GTVerif/Genestier.pdf

http://oeis.org/A000108

http://www.ijera.com/papers/Vol2_issue2/AA22161166.pdf