Catalan Numbers (short)

DBW

December 22, 2016

Let us define the Catalan numbers C_n be the following recursion formula

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

with initial value $C_0 = 1$. We thus find $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, $C_4 = 14$, $C_5 = 42$, $C_6 = 132$ and $C_7 = 429$.

Parenthesis. If we have n pairs of parenthesis (), in how many ways can we form a chain of these brackets, such that we have an allowed chain? In order that these brackets form an allowed chain, then, reading the chain from left to right, there may not be at any step more)-brackets than (-brackets. Conversely, if this constraint is satisfied, the chain is allowed; indeed the first) has to combine with the last (appearing before him, so that these two can then be removed and as long as the constraint is satisfied we can keep on removing parenthesis until we have the empty chain - which is allowed.

Counting the number of these parenthesis words is thus equivalent to the following; among all chains with n times $+1$ and n times -1, how many such chains have intermediate nonnegative sums, when adding from left to right?

Mountains ranges. The last interpretation can be thought of as walking on a vertical axis; with each step we go either one up or one down. The question is, how many ways are there to make $2n$ steps such, that one ends at the same height as where one has started, but never goes below the initial height.

The vertical steps can be made into a mountain range; additionally to the vertical move, we make one move to the right at each step, thus making a path in a two-dimensional grid. Then such a path corresponds to a sequence of steps of the form / and \, where one never crosses the horizontal line at which one starts and ends.

Counting of votes leading to a draw. Suppose we have two candidates in a voting and that both receive equal number of votes. How many ways can the ballots be counted, such that one candidate never leads in the counting? This problem obviously has the same solution as the previous.

Choosing the order to calculate a product. Suppose we have $n + 1$ numbers that are to be multiplied. How many ways, without changing the order, are there to do this multiplication? Since we keep the order fixed, we might as well consider $n + 1$ matrices that need to be multiplied together.

We can use parenthesis to denote the way is done the multiplication: for example $((AB)(CD))$ means, first calculate AB, then calculate CD and then multiply these two results together. At any step we replace two objects by one, so with $n+1$ objects, we need n steps to reduce the string to a single object. Hence we need n parenthesis pairs. So we are led to an allowed chain of parenthesis. Conversely, if we have an allowed chain with n parenthesis pairs, we can fill it up and we have a multiplication order. Indeed, each combination () means a product of two objects XY, resulting in a new object Z, so we can first remove thesse pairs, and look for new () combinations. Erasing the () subsequently one ends with an empty string in just a few steps, and then working your way back through the steps one can insert objects.

The recursive formula

Let us consider the ways to count the sequences of $2n$ vertical moves along an axis that end where they started and never go below their starting point. Let C_n be the number of such sequences. Suppose we know C_k for $k = 0, 1, 2, \ldots, n$ and want to use this to find C_{n+1}. Let $2l$ be the first time after the start, where our path returns to start. Clearly, l is at least 1. The first $2l$ steps are such that first one goes one up, then makes $2l-2$ steps such that one never goes below height 1, and then goed to level 0 again. There are C_{l-1} such paths. Figuratively, the first $2l$ steps can be changed by removing the first and the last of these $2l$ steps. Then one still has an allowed path. For the remaining $2(n+1-l)$ there are C_{n+1-l} ways to do the steps. Thus there are $C_{l-1}C_{n+1-l}$ paths that first touch level 0 after $2l$ steps. If $l = n+1$, we can leave out the first and the last step of the whole sequence and obtain a legal path of length $2n$. The number l thus takes values $1, 2, \ldots, n+1$. Thus there are $C_0C_n + C_1C_{n-1} + \ldots + C_nC_0$ ways to make a legal path of length $2n+2$.

There is another way to motivate the recursive formula. Suppose we have a legal chain X of $n+1$ pairs of brackets. The first bracket is (, and this one pairs up with a). Then our chain X must be of the form $X = (A)B$, where A and B are allowed chains with a total of n pairs of brackets. If A has l pairs, then B has $n-l$ pairs, thus there are C_lC_{n-l} numbers of legal chains of $n+1$ pairs, where the first bracket pairs up with the bracket at place $2l+2$. Clearly, l can be 0, leaving A void, and up to n, leaving B void. The counting results in the same recursive formula.

An explicit formula

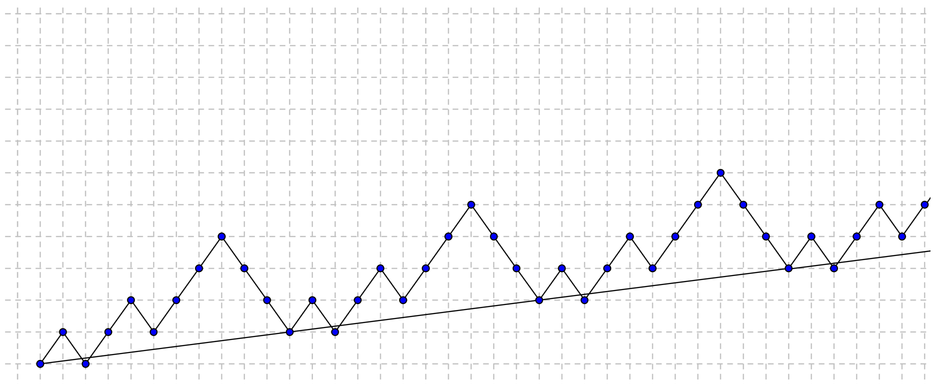
In this section we want to motivate $C_n = \frac{1}{n+1}\binom{2n}{n}$. Of course, one could show that the recursive formula works using an induction argument. We prefer a counting argument.

We consider the mountain ranges, i.e. paths of the fom //\/\\. In fact, we will start with a sequence that is not balanced, but goes up in total. We thus consider paths with $n+1$ / and n \. There are a total of $\binom{2n+1}{n}$ such paths. We first consider any such path. We denote our points on the grid x_0, x_1 and so on, up to x_{2n+1}.

We periodically continue our path on the grid, and continue the labels of the points. Our pattern has period $2n+1$. The line L_0 connecting x_0, x_{2n+1}, x_{4n+2}, and so on has slope $\frac{1}{2n+1}$ and thus only touches a grid point at precisely those points. In particular, it touches no other x_i for any i.

We call C the curve that connects the lattice points of our path in natural order. Thus C contains all segments $x_i x_{i+1}$. If C is such that C only touches L_0 at $x_{k(2n+1)}$ for $k = 0, 1, 2, \ldots$ we have an allowed sequence by removing the first /. Indeed, the line connecting x_1 and x_{2n+1} is horizontal and C never crosses (or touches) L_0 and thus never crosses this horizontal before x_{2n+1}.

Below is a periodically repeated sequence of 12 points, connected by a 6 / and 5 \. The line C crosses L_0 and thus the path connecting x_1 with x_{12} is not an allowed mountain range.



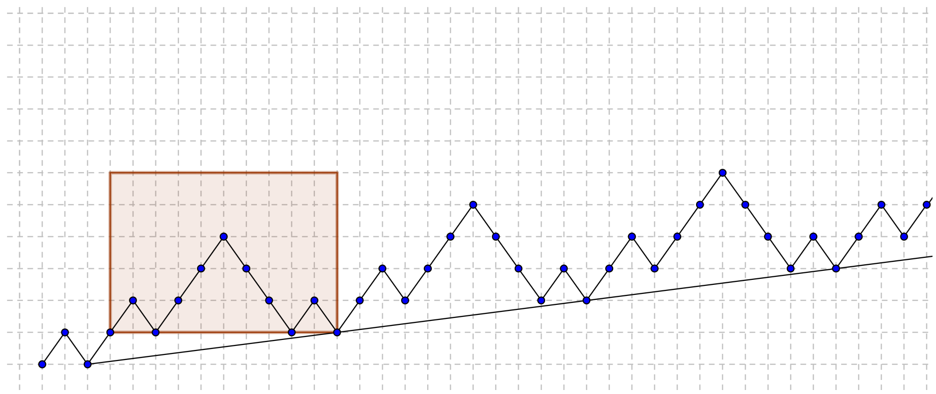
We now want to proof the following statement: For any chain of $n+1$ / and n \there exists precisely one cyclic permutation of the symbols that result in an allowed path of length $2n$ by removing the first step.

If we cyclically permute the / and \, this corresponds to relabeling the indices on our points in a cyclic way, thus choosing another starting point.

Now consider the lines L_i connecting x_i with $x_{i+k(2n+1)}$ with k any integer and $0 \leq i \leq 2n$. All these lines are parallel. But then there is a lowest line L_a for $0 \leq a \leq 2n$. The term "lowest" should be clear from the context; if we have any vertical line V, then the lowest line is the one crossing V at the lowest (smallest) ordinate. But then, the sequence of C starting at x_{a+1} and ending at x_{a+2n+1} is an allowed sequence of points in the sense that the / and \connecting them is an allowed mountain range.

Thus starting with any chain of $n+1$ / and n \, there is precisely one resulting in an allowed mountain range.

Below is the same sequence of points as in the picture above, but now we start the sequence at x_2. The allowed mountain range is now given in the shaded box.



We can then consider all chains of $n+1$ / and n \and group them in such a way that two chains are in the same group if and only if they are cyclic permutations of another. Each group contains $2n+1$ elements. And each group results in one allowed mountain range. The allowed number of mountain ranges is thus

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