Catalan number

In combinatorial mathematics, the **Catalan numbers** form a sequence of natural numbers that occur in various counting problems, often involving recursively defined objects. They are named for the Belgian mathematician Eugène Charles Catalan (1814–1894).

The n^{th} Catalan number is given directly in terms of binomial coefficients by

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!\,n!} \qquad \text{for } n \geq 0.$$

Properties

An alternative expression for C_n is

$$C_n = \binom{2n}{n} - \binom{2n}{n-1} \qquad \text{for } n \geq 1.$$

This shows that C_n is a natural number, which is not *a priori* obvious from the first formula given. This expression forms the basis for André's proof of the correctness of the formula (see below under second proof).

The Catalan numbers satisfy the recurrence relation

$$C_0 = 1 \quad \text{and} \quad C_{n+1} = \sum_{i=0}^{n} C_i\, C_{n-i} \quad \text{for } n \geq 0.$$

They also satisfy:

$$C_0 = 1 \quad \text{and} \quad C_{n+1} = \frac{2(2n+1)}{n+2} C_n,$$

which can be a more efficient way to calculate them.

Asymptotically, the Catalan numbers grow as

$$C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$$

in the sense that the quotient of the n^{th} Catalan number and the expression on the right tends towards 1 for $n \to \infty$. (This can be proved by using Stirling's approximation for $n!$.)

The only Catalan numbers C_n which are odd are those for which $n = 2^k - 1$. All others are even.

Applications in combinatorics

There are many counting problems in combinatorics whose solution is given by the Catalan numbers. The book *Enumerative Combinatorics: Volume 2* by combinatorialist Richard P. Stanley contains a set of exercises which describe 66 different interpretations of the Catalan numbers. Following are some examples, with illustrations of the cases $C_3 = 5$ and $C_4 = 14$.

- C_n is the number of **Dyck words** of length $2n$. A Dyck word is a string consisting of n X's and n Y's such that no initial segment of the string has more Y's than X's (see also Dyck language). For example, the following are the Dyck words of length 6:

  $$\text{XXXYYY} \quad \text{XYXXYY} \quad \text{XYXYXY} \quad \text{XXYYXY} \quad \text{XXYXYY}.$$

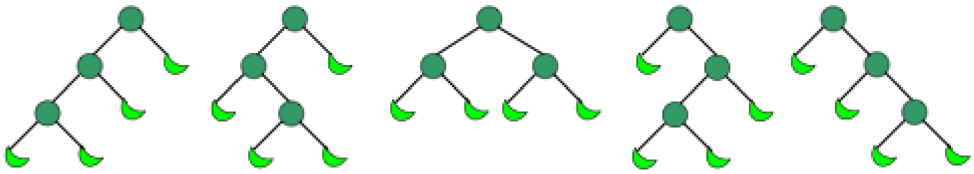
- Re-interpreting the symbol X as an open parenthesis and Y as a close parenthesis, C_n counts the number of expressions containing n pairs of parentheses which are correctly matched:

  $$((())) \quad ()(()) \quad ()()() \quad (())() \quad (()())$$

- C_n is the number of different ways $n + 1$ factors can be completely parenthesized (or the number of ways of associating n applications of a binary operator). For $n = 3$, for example, we have the following five different parenthesizations of four factors:

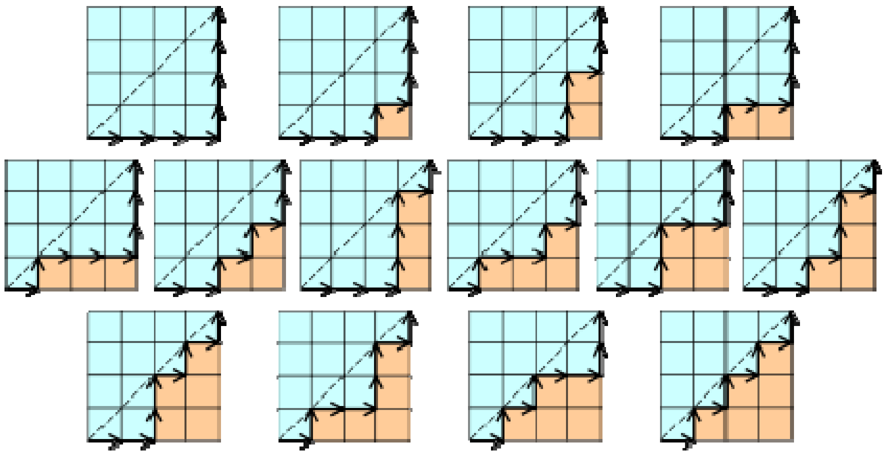
  $$((ab)c)d \quad (a(bc))d \quad (ab)(cd) \quad a((bc)d) \quad a(b(cd))$$

- Successive applications of a binary operator can be represented in terms of a binary tree. It follows that C_n is the number of rooted ordered binary trees with $n + 1$ leaves:

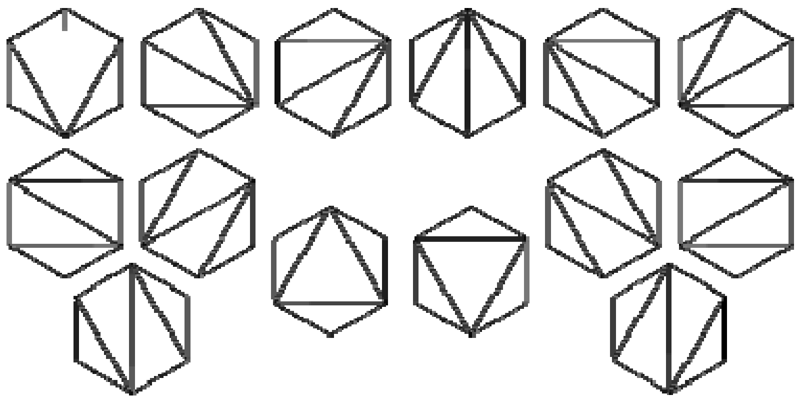


If the leaves are labelled, we have the quadruple factorial numbers.

- C_n is the number of non-isomorphic full binary trees with n vertices that have children, usually called internal vertices or branches. (A rooted binary tree is *full* if every vertex has either two children or no children.)
- C_n is the number of **monotonic paths** along the edges of a grid with $n \times n$ square cells, which do not pass above the diagonal. A monotonic path is one which starts in the lower left corner, finishes in the upper right corner, and consists entirely of edges pointing rightwards or upwards. Counting such paths is equivalent to counting Dyck words: X stands for "move right" and Y stands for "move up". The following diagrams show the case $n = 4$:

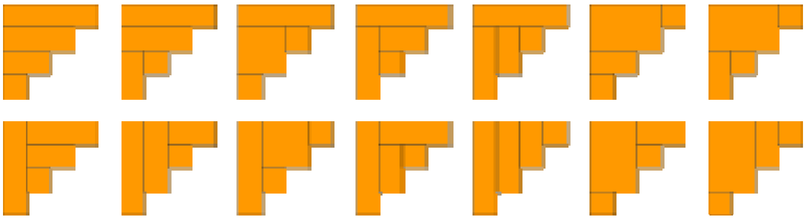


- C_n is the number of different ways a convex polygon with $n + 2$ sides can be cut into triangles by connecting vertices with straight lines. The following hexagons illustrate the case $n = 4$:



- C_n is the number of stack-sortable permutations of {1, ..., n}. A permutation w is called **stack-sortable** if $S(w) = (1, ..., n)$, where $S(w)$ is defined recursively as follows: write $w = unv$ where n is the largest element in w and u and v are shorter sequences, and set $S(w) = S(u)S(v)n$, with S being the identity for one-element sequences.
- C_n is the number of noncrossing partitions of the set {1, ..., n}. *A fortiori*, C_n never exceeds the nth Bell number. C_n is also the number of noncrossing partitions of the set {1, ..., 2n} in which every block is of size 2. The conjunction of these two facts may be used in a proof by mathematical induction that all of the *free* cumulants of degree more than 2 of the Wigner semicircle law are zero. This law is important in free probability theory and the theory of random matrices.

- C_n is the number of ways to tile a stairstep shape of height *n* with *n* rectangles. The following figure illustrates the case *n* = 4:



- C_n is the number of [Young tableaux](https://) whose diagram is a 2 by n rectangle. In other words, it is the number ways the numbers 1, 2, ... 2n can be arranged in a 2 by n rectangle so that each row and each column is increasing. As such, the formula can be derived as a special case of the hook formula.
- C_n is the number of ways that the vertices of a convex 2n-gon can be paired so that the line segments joining paired vertices do not intersect.

Proof of the formula

There are several ways of explaining why the formula

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

solves the combinatorial problems listed above. The first proof below uses a generating function. The second and third proofs are examples of bijective proofs; they involve literally counting a collection of some kind of object to arrive at the correct formula.

First proof

We first observe that many of the combinatorial problems listed above satisfy the recurrence relation

$$C_0 = 1 \quad \text{and} \quad C_{n+1} = \sum_{i=0}^{n} C_i\, C_{n-i} \quad \text{for } n \ge 0.$$

For example, every Dyck word *w* of length ≥ 2 can be written in a unique way in the form

$w = Xw_1Yw_2$

with (possibly empty) Dyck words w_1 and w_2.

The generating function for the Catalan numbers is defined by

$$c(x) = \sum_{n=0}^{\infty} C_n x^n.$$

The two recurrence relations together can then be summarized in generating function form by the relation

$$c(x) = 1 + xc(x)^2;$$

in other words, this equation follows from the recurrence relations by expanding both sides into power series. On the one hand, the recurrence relations uniquely determine the Catalan numbers; on the other hand, the generating function solution

$$c(x) = \frac{1-\sqrt{1-4x}}{2x}$$

has a power series at 0 and its coefficients must therefore be the Catalan numbers. (Since the other solution has a pole at'0, this reasoning doesn't apply to it.)

The square root term can be expanded as a power series using the identity

$$\sqrt{1+y} = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} y^n = 1 - 2\sum_{n=1}^{\infty} \binom{2n-2}{n-1} \left(\frac{-1}{4}\right)^n \frac{y^n}{n}.$$

This is a special case of Newton's generalized binomial theorem; as with the general theorem, it can be proved by computing derivatives to produce its Taylor series. Setting $y = -4x$ and substituting this power series into the expression for *c(x)* and shifting the summation index *n* by 1, the expansion simplifies to

$$c(x) = \sum_{n=0}^{\infty} \binom{2n}{n} \frac{x^n}{n+1}.$$

The coefficients are now the desired formula for C_n.

Second proof

This proof depends on a trick known as André's reflection method (not to be confused with the Schwarz reflection principle in complex analysis), which was originally used in connection with Bertrand's ballot theorem. The reflection principle has been widely attributed to Désiré André, but his method did not actually use reflections; and the reflection method is a variation due to Aebly and Mirimanoff[1]. It is most easily expressed in terms of the "monotonic paths which do not cross the diagonal" problem (see above).

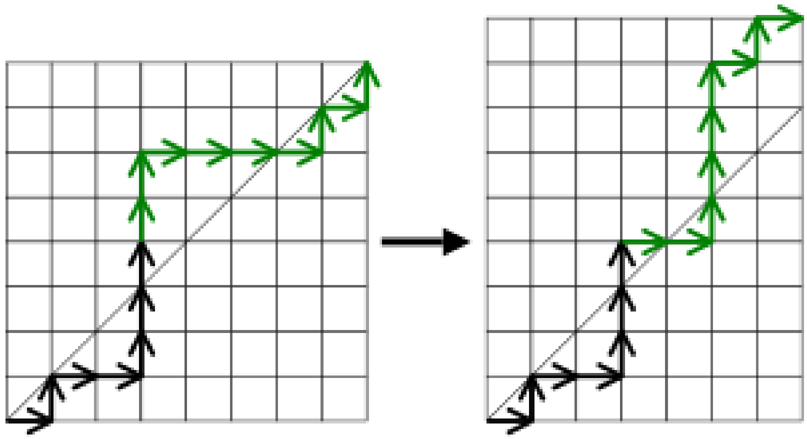


Figure 1. The green portion of the path is flipped.

Suppose we are given a monotonic path in an $n \times n$ grid that *does* cross the diagonal. Find the first edge in the path that lies above the diagonal, and *flip* the portion of the path occurring after that edge, along a line parallel to the diagonal. (In terms of Dyck words, we are starting with a sequence of n X's and n Y's which is *not* a Dyck word, and exchanging all X's with Y's after the first Y that violates the Dyck condition.) The resulting path is a monotonic path in an $(n - 1) \times (n + 1)$ grid. Figure 1 illustrates this procedure; the green portion of the path is the portion being flipped.

Since every monotonic path in the $(n - 1) \times (n + 1)$ grid must cross the diagonal at some point, every such path can be obtained in this fashion in precisely one way. The number of these paths is equal to

$$\binom{2n}{n-1}.$$

Therefore, to calculate the number of monotonic $n \times n$ paths which do *not* cross the diagonal, we need to subtract this from the *total* number of monotonic $n \times n$ paths, so we finally obtain

$$\binom{2n}{n} - \binom{2n}{n-1}$$

which is the nth Catalan number C_n.

Third proof

The following bijective proof, while being more involved than the previous one, provides a more natural explanation for the term $n + 1$ appearing in the denominator of the formula for C_n.

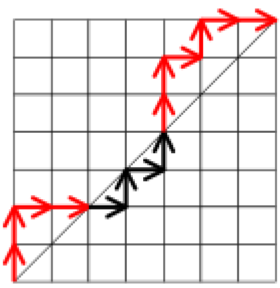


Figure 2. A path with exceedance 5.

Suppose we are given a monotonic path, which may happen to cross the diagonal. The **exceedance** of the path is defined to be the number of pairs of edges which lie *above* the diagonal. For example, in Figure 2, the edges lying above the diagonal are marked in red, so the exceedance of the path is 5.

Now, if we are given a monotonic path whose exceedance is not zero, then we may apply the following algorithm to construct a new path whose exceedance is one less than the one we started with.

- Starting from the bottom left, follow the path until it first travels above the diagonal.
- Continue to follow the path until it *touches* the diagonal again. Denote by *X* the first such edge that is reached.
- Swap the portion of the path occurring before *X* with the portion occurring after *X*.

The following example should make this clearer. In Figure 3, the black circle indicates the point where the path first crosses the diagonal. The black edge is *X*, and we swap the red portion with the green portion to make a new path, shown in the second diagram.

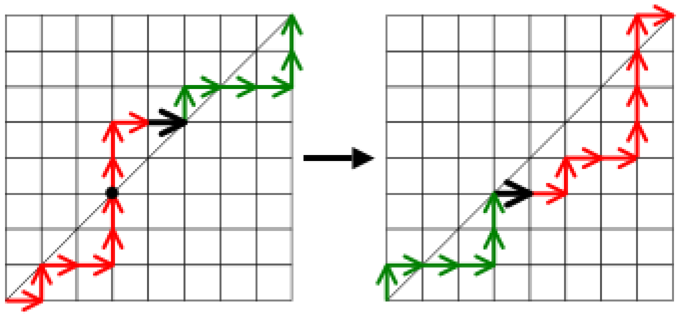


Figure 3. The green and red portions are being exchanged.

Notice that the exceedance has dropped from three to two. In fact, the algorithm will cause the exceedance to decrease by one, for any path that we feed it.

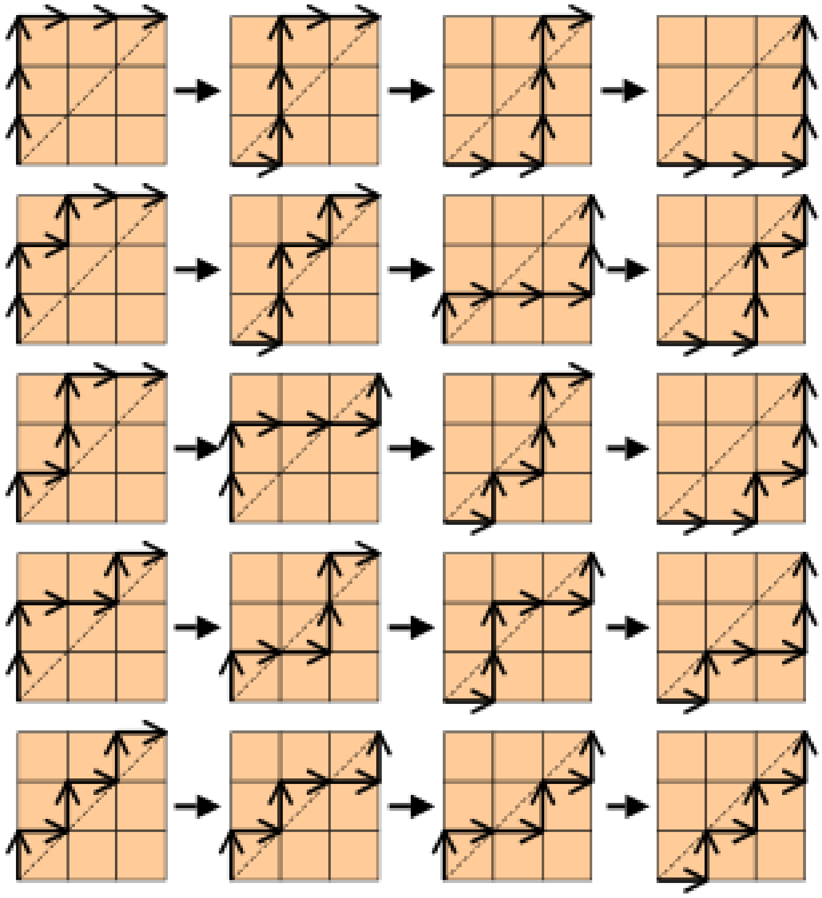


Figure 4. All monotonic paths in a 3×3 grid, illustrating the exceedance-decreasing algorithm.

It is also not difficult to see that this process is *reversible*: given any path *P* whose exceedance is less than *n*, there is exactly one path which yields *P* when the algorithm is applied to it.

This implies that the number of paths of exceedance *n* is equal to the number of paths of exceedance $n - 1$, which is equal to the number of paths of exceedance $n - 2$, and so on, down to zero. In other words, we have split up the set of *all* monotonic paths into $n + 1$ equally sized classes, corresponding to the possible exceedances between 0 and *n*. Since there are

$$\binom{2n}{n}$$

monotonic paths, we obtain the desired formula

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

Figure 4 illustrates the situation for $n = 3$. Each of the 20 possible monotonic paths appears somewhere in the table. The first column shows all paths of exceedance three, which lie entirely above the diagonal. The columns to the right show the result of successive applications of the algorithm, with the exceedance decreasing one unit at a time. Since there are five rows, $C_3 = 5$.

Fourth proof

This proof uses the triangulation definition of Catalan numbers to establish a relation between C_n and C_{n+1}. Given a polygon *P* with $n+ 2$ sides, first mark one of its sides as the base. If *P* is then triangulated, we can further choose and orient one of its $2n+1$ edges. There are $(4n+2)C_n$ such decorated triangulations. Now given a polygon *Q* with $n+3$ sides, again mark one of its sides as the base. If *Q* is triangulated, we can further mark one of the sides other than the base side. There are $(n+2)C_{n+1}$ such decorated triangulations. Then there is a simple bijection between these two kinds of decorated triangulations: We can either collapse the triangle in *Q* whose side is marked, or in reverse expand the oriented edge in *P* to a triangle and mark its new side. Thus

$$(4n + 2)C_n = (n + 2)C_{n+1}.$$

The binomial formula for C_n follows immediately from this relation and the initial condition $C_1 = 1$.