CATALAN OBJECTS

AMRIT KANDASAMY, NEIL IYER, ALEKSEI LOPATIN

Abstract. This expository paper surveys Catalan numbers from both algebraic and combinatorial perspectives. After presenting their explicit formula and main recurrences, we examine several combinatorial problems whose solutions are counted by Catalan numbers. In each case, we either construct a bijection with a standard Catalan object or verify that the problem satisfies the Catalan recurrence relation and initial conditions. We then turn to the generating function for the Catalan sequence and use it to study the growth rate of these numbers. We present each use both algebraically and combinatorially. We end with the generating function of Catalan Numbers, and how they can be used to quickly derive that of other sequences.

1. Introduction

The Catalan numbers are from the integer sequence

$$C_n = \frac{1}{n+1}\binom{2n}{n}, n \geq 0,$$

whose first few values are 1, 1, 2, 5, 14, 42, 132, ... At first, they look like some family of binomial coefficients, but they turn out to count a variety of combinatorical structures. A *Catalan object* is any family of objects $\mathcal{F}_n$ indexed by $n \geq 0$ such that

$$|\mathcal{F}_n| = C_n$$

for all n, where n is some parameter (number of steps, number of vertices, number of pairs of parentheses, and so on).

1.1. **Early European appearances, Euler and Segner.** One of the earliest known appearances of the Catalan Sequence in Europe is in the work of Leohnard Euler. Consider a convex polygon with $n+2$ vertices. A triangulation of this polygon is designed as the maximal set of noncrossing diagonals that partitions the polygons into triangles. Let T_n denote the number of triangulations of a convex polygon with $n+2$ vertices. For small n we can compute directly:

$$T_0 = 1, T_1 = 1, T_2 = 2, T_3 = 5, T_4 = 14.$$

These are already the first few Catalan numbers. Euler studied this problem in the 1750s and found a closed product formula for what we now write as

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

At that time, the connection between the formula and polygon triangulations was still being studied. The main method that eventually organized everything was a recurrence relation due to Johann Andreas von Segner.

Date: December 8, 2025.

Segner observed that any triangulation of a convex $(n+2)$-gon can be decomposed by looking at the unique triangle that contains a fixed base edge. Suppose we label the vertices of the polygon $0, 1, \ldots, n+1$ around the boundary and focus on the base edge $(0, n+1)$. In any triangulation there is a unique vertex k with $1 \leq k \leq n$ such that $(0, k)$ and $(k, n+1)$ are both diagonals. These two diagonals split the polygon into two smaller convex polygons with vertices

$$0, 1, \ldots, k \quad \text{and} \quad k, \ldots, n+1.$$

Every triangulation is obtained by choosing a triangulation of each smaller polygon. If we let T_n denote the number of triangulations of an $(n+2)$-gon, then this decomposition implies the recurrence

$$T_0 = 1, \qquad T_n = \sum_{i=0}^{n-1} T_i \, T_{n-1-i} \quad \text{for } n \geq 1.$$

We will see this recurrence again in later sections, it is the standard Catalan recurrence. In modern notation, we simply write $T_n = C_n$.

The recurrence was published by Segner in 1758. Euler had already identified the same numbers through analytical methods. Later, Lamé gave the first proof that the recurrence and the closed form $\frac{1}{n+1}\binom{2n}{n}$ match. But, Euler's and Senger's work shows that Catalan numbers encode recursive triangulations long before the numbers were named or recognized as a sequence.

1.2. **Mingantu and divisions of a circle.** The history of Catalan numbers extends beyond Europe. Around the same period, the Mongolian mathematician Mingantu, worked at the Qing court in China. He studied problems about dividing a circle and expressing cord lengths in terms of arcs.

Fix a circle of radius 1 and consider an arc that is divided into n equal sub-arcs. Let x be the chord length corresponding to a small sub-arc and let y_n be the chord length corresponding to the whole arc with n segments. For a given n, Mingantu built models, geometric ones, that relate y_n to x. He then converted those relationships into algebraic expressions.

For example, when $n = 2$ he obtained an expression of the form

$$y_2 = 2x - (\text{higher order terms in } x),$$

and for $n = 4$ he derived an infinite series

$$y_4 = 4x - \frac{10}{4}x^3 + \frac{14}{4^3}x^5 - \frac{12}{4^5}x^7 + \cdots .$$

Interestingly, we can rewrite these series using trigonometric functions. If we interpret $x = \sin(\alpha)$ and $y_n = \sin(n\alpha)$, Mingantu's calculations lead to expansions of the form

$$\sin(2\alpha) = 2\sin\alpha - \sum_{n\geq 1} a_n \sin^{2n+1}\alpha$$

and similar formulas for $\sin(4\alpha)$ and higher multiples. The coefficients a_n in these expansions can be expressed in terms of Catalan numbers. In particular, you can get an identity of the shape

$$\sin(2\alpha) = 2\sin\alpha - \sum_{n\geq 1} C_n \frac{\sin^{2n+1}\alpha}{4^{n-1}},$$

where $C_n = \frac{1}{n+1}\binom{2n}{n}$. The coefficients, which we now understand as Catalan numbers, already were present in Mingantu's work on trigonometric series and circle divisions.

It is important to note that this perspective was made clear much later, where math historians reinterpreted Mingantu's constructions using modern notation. Today, his work is cited as a discovery of the Catalan sequence, obtained by analyzing power series rather than counting combinatorial objects.

1.3. **Eugène Catalan and parenthesizations.** The actual sequence as we know it today is named after Eugène Charles Catalan (1814 – 1894), a Belgian mathematician who worked in France and Belgium in the nineteenth century. In a paper written in 1838, Catalan studied the following problem: given a product of $n+1$ factors

$$x_0 x_1 x_2 \cdots x_n,$$

in how many ways can one insert parentheses so that the product is well-defined if multiplication is associative but not commutative? Each paranthesized expression corresponds to a binary tree structure on the $n+1$ factors. For example, when $n+1=4$ the five possible scenarios are

$$(((x_0x_1)x_2)x_3), \quad ((x_0(x_1x_2))x_3), \quad ((x_0x_1)(x_2x_3)), \quad (x_0((x_1x_2)x_3)), \quad (x_0(x_1(x_2x_3))).$$

Catalan derived the formula

$$\#\{\text{parenthesizations on } n+1 \text{ factors}\} = \frac{1}{n+1}\binom{2n}{n}.$$

He wrote the expression in factored form and analyzed it in terms of non associative products. In other words, he showed that C_n counts the number of ways to fully parenthesize a product of $n+1$ letters. This interpretation is the most standard combinatorical models of Catalan numbers, and is an example of a Catalan object.

The same sequence therefore appears in two very different settings by the mid-nineteenth century. At this stage, however, the numbers were not treated as a unified object *yet*, but they were made gradually over the following decades.

1.4. **Generalizations and a modern view.** One early extension is now called the *Fuss–Catalan* family. Fix an integer $m \geq 1$ and look at a convex polygon with $mn+2$ vertices. Instead of triangulating it, one can ask for the number of ways to cut it into $(m+2)$-gons using noncrossing diagonals. The answer is

$$F_n^{(m)} = \frac{1}{(m-1)n+1}\binom{mn}{n},$$

and for $m=2$ this reduces to the usual Catalan number C_n. So, all of the classical polygon triangulations and parenthesization are $m=2$ case of a broader pattern where binary structure is replaced by $(m+2)$-ary structure.

2. Definitions and Notation

Definition 2.1. Let $n \geq 3$. A *convex n-gon* is a polygon in the plane with n vertices in convex position, labeled cyclically $v_1, v_2, \ldots, v_n$.

A *diagonal* of the n-gon is a segment joining two non-adjacent vertices.

A *triangulation* of a convex n-gon is a maximal collection of pairwise noncrossing diagonals. Equivalently, a triangulation is a decomposition of the polygon into triangles whose vertices are the original vertices.

Definition 2.2. The multinomial coefficient $\binom{n}{k_1,\ldots,k_d}$ is the number of ways to color k_i elements for $1 \le i \le d$ with the color c_i where there are d colors called $c_1, \ldots, c_d$. Obviously, we have the condition on the size of the set being $k_1 + \cdots + k_d = n$.

For calculation purposes, we have the following proposition.

Proposition 2.3.

$$\binom{n}{k_1, \ldots, k_d} = \frac{n!}{k_1! \cdots k_d!}$$

This is the number of ways to color the numbers from 1 to n c_d colors. We can color k_1 of the numbers c_1, then k_2 of the numbers c_2, and so on. This results in the RHS.

3. Dyck Paths, Ballot Sequences, and the Open Parentheses Problem

With our built up knowledge, we introduce lattice paths from $(0, 0)$ to (n, n).

Definition 3.1. A lattice path from $(0, 0)$ to (n, n) is a path with steps going north or east.

We have a way to calculate the number of such paths. We make a bijection from the number of paths in a lattice to the number of ways to arrange $RR \cdots UU$ where there are n Rs and n Us. An R corresponds to an east step and a U corresponds to a north step. Notice that the number of ways to calculate this is $\frac{(2n)!}{n!n!}$.

We can work with multinomial coefficients in lattice paths too.

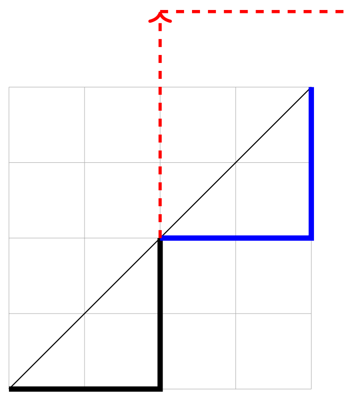
Proposition 3.2. *The number of paths from* $(0, 0, \ldots, 0)$ *to* $(n_1, \ldots, n_k)$ *taking steps of the form* $(0, \ldots, 0, 1, 0, \ldots, 0)$ *-i.e. moving parallel to one of the coordinate axes in the positive direction - is equal to the multinomial coefficient* $\binom{n_1+\cdots+n_k}{n_1,\ldots,n_k}$.

Catalan Numbers appear in Dyck paths in the following theorem.

Theorem 3.3. *The number of paths from* $(0, 0)$ *to* (n, n) *taking steps to the north or east and never going above the line* $y = x$ *is equal to* $\frac{1}{n+1}\binom{2n}{n}$.

Proof. There are many proofs of this theorem with similar approaches. Consider a path from $(0, 0)$ to (n, n) using north and east steps. Essentially, it either never crosses $y = x$ or it does at some point (k, k) to $(k, k + 1)$. Analyzing this path, it is made up of k east steps and $k + 1$ north steps and $n - k$ east steps and $n - k - 1$ north steps. We now reflect the part of the path after $(k, k + 1)$. The modification is that after $(k, k + 1)$, we convert every north step to an east step and every east step to a north step. An example is shown below.

This results in a total of $k + (n - k - 1) = n - 1$ east steps and $k + 1 + n - k = n + 1$ north steps, therefore the path ends at $(n - 1, n + 1)$. These "bad" paths sum to $\binom{2n}{n-1}$ possible ways. By complementary counting, there are $\binom{2n}{n} - \binom{2n}{n-1}$ good paths. ∎



It is also possible to generalize to an (m, n) rational Catalan number. This is defined as the number of paths from $(0, 0)$ to (m, n) not going above the line from $(0, 0)$ to (m, n). A similar argument verifies that $C_{m,n} = \frac{1}{m+n}\binom{m+n}{n}$.

We now present one proof of the recursion for the Catalan Numbers through Dyck paths.

Theorem 3.4. *For each $n \geq 0$, we have*

$$C_{n+1} = C_0C_n + C_1C_{n-1} + C_2C_{n-2} + \cdots + C_{n-1}C_1 + C_nC_0 = \sum_{k=0}^{n} C_kC_{n-k}$$

Proof. Consider a Dyck path from $(0, 0)$ to $(n + 1, n + 1)$, and specifically the first time when it hits the line $y = x$ (disregarding $(0, 0)$). Without loss of generalization, suppose it hits at $(k + 1, k + 1)$ for a $k \in [0, n]$. Look at the subpath of the Dyck path from $(0, 0)$ to $(k + 1, k + 1)$. This path begins with an east step and finishes with a north step. This subpath containes a path from $(1, 0)$ to $(k + 1, k)$.

As we assumed that $(k + 1, k + 1)$ was the first time that the path touched $y = x$, the path from $(1, 0)$ to $(k + 1, k)$ stays below $y = x$. Essentially, this is a path from $(1, 0)$ to $(k + 1, k)$ never crossing $y = x - 1$. But, this is a Dyck path of length k, and there are C_k such subpaths.

After $(k+1, k+1)$, we have a path from $(k+1, k+1)$ to $n+1, n+1$ never crossing $y = x$. This is a shifted version of a path from $(0, 0)$ to $(n - k, n - k)$ not crossing $y = x$, so there are C_{n-k} of those. Therefore, there are C_kC_{n-k} Dyck paths from $(0, 0)$ to $(n + 1, n + 1)$. We sum from $k = 0$ to n, which gives us the desired result. ∎

We now dive into the problem of paranthesizations of an expression with symbols, which has a bijection with Dyck paths. Consider the number of ways to paranthesize the product *abcd* for $n = 3$ (we typically consider paranthesizations of $n + 1$ symbols). For example, here are some possibilities:

$$(((ab)c)d), (a(b(cd))), (((ab)c)d)$$

Theorem 3.5. *There are C_n ways of performing parenthesizations on $n + 1$ symbols so that the expressions are parenthesized in pairs.*

Proof. We aim to make a bijection with Dyck paths from $(0, 0)$ to (n, n). Let us begin with a parenthesized product and delete all the right parentheses and the rightmost variable. For $n = 3$, this means deleting d. Examples are $(((abc$ and $((ab(c$. Notice that all of these modifications are distinct and we can get the fully parenthesized expression by placing the last variable at the end, and left from right, inserting a parenthesis when we have two consecutive symbols or groups of symbols.

We can now get a bijection with Dyck paths. From a Dyck path, replace every east step with a left parenthesis and every north step with a symbol. Then perform the modification to reinsert the right parentheses and the final variable. This results in an invertible process. We start with a parenthesized expression and retain a Dyck path, so this is a bijection. ∎

We can further emphasize how this modification works with a slightly larger example. Consider the string $(((a(bc((de(fg$. We insert h in at the end (we are forced to) to get $(((a(bc((de(fgh$. Working from the left, we close off bc like $(((a(bc)((de(fgh$. Since (bc) is a group and we have a outside, we parenthesize as follows $(((a(bc))((de(fgh$. Furthermore, the next pair if (de), which we cap off as $(((a(bc))((de)(fgh$. The last consecutive symbols are fg, so $(((a(bc))((de)(fg)h$. Since we have the adjacent blocks $(a(bc))$ and $((de)(fg))$, we group them $(((a(bc))((de)(fg)))h$. The h is grouped with everything else, so we end with $(((a(bc))((de)(fg)))h)$. It can be verified that going backward and deleting h and the right parentheses we have $(((a(bc((de(fg$.

Dyck paths are very useful in making bijections to other sets that involve the Catalan Numbers. Consider the following set: sequences of length $2n$ consisting of n a's and n b's, such that for each $1 \leq k \leq 2n$, the number of a's among the first k letters is at least as large as the number of b's among the first k letters.

For example for $n = 2$, $abab$ works, but $abba$ doesn't.

Theorem 3.6. *The cardinality of this set is C_n.*

Proof. We make a bijection between these elements to Dyck paths. We biject each a to an east step and each b to a north step. Therefore, this is just the number of Dyck paths from $(0, 0)$ to (n, n) where an a corresponds to $(1, 0)$ and a b corresponds to $(0, 1)$.

We can verify that this is a bijection through the same proof of why there are C_n Dyck paths from $(0, 0)$ to (n, n) with the same "reflection" idea. It is also sufficient to note that the condition "the number of a's among the first k letters is at least as large as the number of b's among the first k letters" means that there are at least as many east steps as north steps. But, this is equivalent to saying that we never cross $y = x$. ∎

This is closely related to the Ballot Problem. Joseph Bertrand proposed the following problem in 1887.

Suppose that in an election, candidate A receives a votes and candidate B receives b votes where $a \geq b$. Compute the number of ways the ballots can be ordered so that A maintains at least as many votes as B throughout the counting of the ballots.

Bertrand showed the number of ways is $\frac{a-b}{a+b}\binom{a+b}{a}$. We now consider a more general problem where $a \geq kb$ for a positive integer k. Note the remaining details still apply.

Theorem 3.7. *The number of ways to do so for the modified Ballot Problem is $\frac{a-kb}{a+b}\binom{a+b}{b}$.*

Proof. We can represent a ballot permutation as a lattice path from $(0, 0)$ where the votes for A are upsteps $(1, 1)$ and votes for B are downsteps $(1, -k)$. We wish to find the number of paths with a upsteps and b downsteps where no step ends on or below the x-axis. Paths that are above the x-axis are *good*. Paths that end on or below the x-axis are *bad*. A downstep that starts above the x-axis and ends on or below it is called a *bad step*.

For $i \in [0, k]$ let $\boldsymbol{S}_i$ denote the set of all bad paths whose first bad step ends i units below the x-axis. Obviously, these $k + 1$ sets are disjoint and their union is the set of all bad paths.

Note that the paths in $\boldsymbol{S}_k$ are those paths that start with a downstep. Therefore $|\boldsymbol{S}_k| = \binom{a+b-1}{a}$ as we choose a upsteps from a total of $a+b-1$ steps since the first step is forced to be a downstep.

We claim that $|\boldsymbol{S}_i| = |\boldsymbol{S}_k|$ for $i \neq k$.

Consider a path in $\boldsymbol{S}_i$ and specifically the first step of that path that ends i units below the x-axis. Let X be the initial segment of that path that ends with that step and call that path XY. Let X' denote the path that results from rotating X by 180 degrees, which swaps its endpoints. Since X ends with a downstep, X' starts with a downstep, and $X'Y \in \boldsymbol{B}_k$.

This same process converts a path in $\boldsymbol{B}_k$ into a path in $\boldsymbol{B}_i$ for $i \neq k$. If the original path is in $\boldsymbol{B}_k$, then take the first step that ends i units below the x-axis. Let X be the initial part of that path that ends with that step and call said path XY. Since X has to end with an upstep, we have $X'Y \in \boldsymbol{B}_i$.

Consequently, each of the $k+1$ sets $\boldsymbol{B}_i$ have cardinality $\binom{a+b-1}{a}$, and the number of good paths is (by complementary counting):

$$\binom{a+b}{a} - (k+1)\binom{a+b-1}{a} = \frac{a-kb}{a+b}\binom{a+b}{b}$$

■

4. Q-Analogues and Catalan Polynomials

This section aims at understanding how Catalan Numbers are related to q-analogues. The field of q-analogues has many definitions, and we will choose only the ones pertinent to describing the Catalan Numbers.

A q-analogue of a counting function is defined as a polynomial in q that reduces to the function when $q = 1$. Beyond that value, it satisfies other algebraic properties such as recursions.

Definition 4.1. The q-analogue of the real number n, denoted as $[n]$ is

$$[n] = \frac{1-q^n}{1-q}$$

It can be verified by L'Hopital's rule that $[n] \to n$ as $q \to 1^-$. We can describe $n!$ as a q-analogue in the following way

$$[n]! = \prod_{i=1}^{n} [i] = (1+q)(1+q+q^2)\cdots(1+q+\cdots+q^{n-1})$$

We now define several terms to understand q-analogues.

Definition 4.2. A statistic on a set S is a combinatorial rule that maps an element of N to each element of S. A permutation statistic is a statistic on the symmetric group S_n. Clearly $|S_n| = n!$.

Definition 4.3. We define a word $\sigma_1\sigma_2\sigma_2\cdots\sigma_n$ as a linear list of elements of a multiset of nonnegative integers.

Definition 4.4. We define an inversion of a word σ as a pair (i,j) for $1 \leq i < j \leq n$ such that $\sigma_i > \sigma_j$.

We define a descent as an integer i for $i \in [1, n-1]$ such that $\sigma_i > \sigma_{i+1}$.

For example, consider the permutation 261543. The pair $(1, 3)$ is an inversion as $2 > 1$. Additionally, $i = 2$ is a descent as $6 > 1$. We can then define the inversion statistic and major index statistic.

Definition 4.5. The inversion statistic $\mathrm{Inv}(\sigma)$ is the number of inversions of σ:

$$\mathrm{Inv}(\sigma) = \sum_{i<j, \sigma_i > \sigma_j} 1$$

The major index statistic $\mathrm{Maj}(\sigma)$ is the sum of the descents of σ:

$$\mathrm{Maj}(\sigma) = \sum_{i, \sigma_i > \sigma_{i+1}} i$$

Definition 4.6. A permutation statistic is said to be Mahonian if its distribution over S_n is $[n]!$.

It turns out that both Inv and Maj are Mahonian.

Theorem 4.7.

$$\sum_{\sigma \in S_n} q^{\mathrm{Inv}(\sigma)} = \sum_{\sigma \in S_n} q^{\mathrm{Maj}(\sigma)} = [n]!$$

Proof. We give a proof for the case of the inversion statistic. The major index statistic has a similar idea, but it is outside the scope of the paper.

Consider a permutation $A \in S_{n-1}$, and define the operation $A(k)$ for the permutation in S_n to insert n between the $(k-1)$th element and the kth element of A. For instance, if $A = 1342$, then $A(2) = 15342$. This operation allows us to relate the inversion of $A(k)$. Notice that when n is inserted, it is clearly larger than the rest of the $n - k$ elements, so there are $n - k$ new pairs added.

Therefore, $\mathrm{Inv}(A(k)) = \mathrm{Inv}(A) + n - k$. Therefore, we can write that:

$$\sum_{\sigma \in S_n} q^{\mathrm{Inv}(\sigma)} = \sum_{A \in S_{n-1}} (1 + q + q^2 + \cdots + q^{n-1}) q^{\mathrm{Inv}(A)}$$

By induction, we can show that Inv is Mahonian. ■

At this point, we might ask how these q-analogues can relate to Catalan Numbers. To do so, let us introduce some new information.

Definition 4.8. The gaussian polynomial is for $n, k \in \mathbb{N}$:

$$\begin{bmatrix} n \\ k \end{bmatrix} = \frac{[n]!}{[k]![n-k]!} = \frac{(1-q^n)(1-q^{n-1} \cdots (1-q^{n-k+1})}{(1-q^k)(1-q^{k-1}) \cdots (1-q)}$$

Let $B_{n,m}$ denote a lattice board of width n and height m. Given $\pi \in B_{n,m}$, we define $\sigma(\pi)$ to be an element in $S_{n,m}$ from the algorithmn: we initialize toward an empty string. For every north step, we place a 0 to the end. For every east step, we place a 1 to the end. We define $\pi(\sigma)$ to be the inverse of the following algorithmn.

We let $a_i(\pi)$ be the number of complete squares, in the ith row from the bottom of π, which are to the right of π and to the left of the line $y = x$ (this should sound familiar to the typical definition of the Catalan numbers with Dyck paths). We denote $a_i(\pi)$ to be the *length* of the ith row of π. We then define the *area vector* of *pi* to be $(a_1(\pi), a_2(\pi), a_3(\pi), \cdots, a_n(\pi))$. We set $\mathrm{Area}(\pi) = \sum_i a_i(\pi)$.

We define $B^+_{n,m}$ to be the paths characterized by the property that there are at least as many 0's as 1's. We then have the following by MacMahon as a possible characterization of the q-analogue of C_n:

Theorem 4.9.

$$\sum_{\pi \in B^+_{n,n}} q^{\text{Maj}(\sigma(\pi))} = \frac{1}{[n+1]} \begin{bmatrix} 2n \\ n \end{bmatrix}$$

The proof of this requires techniques beyond the scope of the paper, but it does exist. The second natural q-analogue of C_n was studied by Carlitz and Riordan who defined it as the following.

Definition 4.10.

$$C_n(q) = \sum_{\pi \in B^+_{n,n}} q^{\text{Area}(\pi)}$$

Notice that we can also transform the recurrence for Catalan Numbers into q-Catalan numbers recursion. Recall the the normal recursion is $C_n = \sum_{k=0}^{n} C_k C_{n-k}$.

We define the q-Catalan Numbers $C_n = C_n(q)$ by

$$z = \sum_{n=1}^{\infty} C_{n-1} z^n (1-z)(1-qz)\cdots(1-q^{n-1}z)$$

We present a sketch. Notice that

$$z^2 = \sum_{k \geq 1} C_{k-1} z^k (z)_k q^{-k} q^k z =$$

$$= \sum_{k \geq 1} C_{k-1} z^k (z)_k q^{-k} \sum_{i \geq 1} C_{i-1} (q^k z)^i (q^k z)_i =$$

$$= \sum_{n \geq 2} (\sum_{k+i=n} C_{k-1} C_{i-1} q^{k(i-1)}) z^n (z)_n$$

We can therefore rewrite the expression for z as:

$$z = C_0 z(1-z) + \sum_{n \geq 2} C_{n-1} z^n (z)_n$$

By analyzing the coefficients, we find that

$$C_{n-1} = \sum_{k+i=n-2} C_k C_i q^{(k+1)i}$$

We can re-index as following:

$$C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k} q^{(k+1)(n-k)}$$

Letting $C_n(q)' = q^{\binom{n}{2}} C_n(q^{-1})$:

$$C_{n+1}(q)' = \sum_{k=0}^{n} q^k C'_k C'_{n-k}$$

The first few values are

$$C_0 = C_1 = 1$$

$$C_2 = 1 + q$$
$$C_3 = 1 + q + 2q + q^3$$
$$C_4 = 1 + q + 2q^2 + 3q^3 + 3q^4 + 3q^5 + q^6$$

5. Catalan Numbers in Sets and Pattern Avoidance

In this section, we will study three families of Catalan objects arising from sets and permutations:

(1) stack-sortable permutations of S_n;
(2) noncrossing partitions of $[n] = \{1, 2, \ldots, n\}$;
(3) permutations avoiding the pattern 123.

We show that each of these families is counted by the Catalan number

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

, and then we will analyze the structure in detail.

Let S_n denote the symmetric group on $\{1, 2, \ldots, n\}$ written in a one-line notation. Given a permutation $\pi = \pi_1\pi_2\ldots\pi_n \in S_n$ and a pattern $\sigma \in S_k$, we say that π *contains the pattern* σ if there exists indices

$$1 \leq i_1 < i_2 < \cdots < i_k \leq n$$

such that the subsequence $\pi_{i_1}\pi_{i_2}\ldots\pi_{i_k}$ is order-isomorphic to σ, that is,

$$\pi_{i_a} < \pi_{i_b} \quad \Longleftrightarrow \quad \sigma_a < \sigma_b$$

for all $1 \leq a, b \leq k$. If π does not contain σ, we say that π *avoids* σ For a set of patterns B, we write

$$\mathrm{Av}_n(B) = \{\pi \in S_n : \pi \text{ avoids every } \beta \in B\}.$$

A theorem states that for every pattern $\sigma \in S_3$ and every $n \geq 0$ one has

$$|\operatorname{Av}_n(\sigma)| = C_n.$$

The six patterns of length 3 fall into orbits under symmetries of S_n (reversal, complement, inverse), and these symmetries imply equalities of avoidance numbers within each orbit. The fact that *all* six patterns give Catalan numbers requires additional arguments. In this section we will provide enumerations (Catalan) for $\sigma = 231$ and $\sigma = 123$; the other four patterns then follow by composing with these symmetries.

5.1. **Stack-sorting and 231-avoiding permutations.**

5.1.1. *The stack-sorting procedure.* We consider the classical model of a single stack between an input and an output queue. The input is the permutation

$$1, 2, \ldots, n$$

in increasing order. The stack initially is empty, and the output is initially empty. The allowed operations are:

- push: move the first element of the input onto the top of the stack;
- pop: move the top element of the stack to the end of the output.

Our requirement is that every element must be pushed exactly once and popped exactly once, and we must never pop from an empty stack.

Given a permutation $\pi = \pi_1\pi_2 \dots \pi_n \in S_n$, we say that π *is stack-sortable* if there exists some legal sequence of push and pop operations that produces π as the output sequence.

It is convenient to encode an execution of the algorithm by a word w over the alphabet $\{P, Q\}$, where P means "push" and Q means "pop". Any legal execution on input $1, 2, \dots, n$ has exactly n pushes and n pops, so w has length $2n$ with n copies of P and n copies of Q, and the number of Q's never exceeds the number of P's in any prefix of w. Thus w can be interpreted as a Dyck path of semilength n by reading P as an up-step and Q as a down-step.

Definition 5.1. Let $\mathcal{S}_n$ denote the set of stack-sortable permutations in S_n, and let $s_n = |\mathcal{S}_n|$.

Our goal is to show that $s_n = C_n$.

5.1.2. *Characterization via pattern avoidance.* The key structural result is the following.

[Knuth] A permutation $\pi \in S_n$ is stack-sortable if and only if it avoids the pattern 231. Equivalently,

$$\mathcal{S}_n = \mathrm{Av}_n(231).$$

Proof. (*Only if.*) Suppose π contains a 231-pattern. Then there exist indices $i < j < k$ such that

$$\pi_j < \pi_k < \pi_i.$$

Consider any legal stack execution that produces π on output. When the input element π_i is read, both π_j and π_k have not yet been output, so at that moment at least one of them is in the stack (possibly both). Since the input is $1, 2, \dots, n$ in increasing order, the relative positions of π_j, π_k, π_i in the stack force π_i to sit above π_j. This makes it impossible to output π_j before π_i, contradicting the assumption that the output order is π. A standard local configuration argument (see any detailed exposition of Knuth's theorem) turns this into a formal proof that no permutation containing 231 can be obtained.

(*If.*) Now suppose π avoids 231. We must show there is some legal execution that outputs π. Consider the greedy algorithm: whenever possible, pop the stack if the top of the stack is the next desired output element of π; otherwise, push the next input element. Formally:

- If the stack is nonempty and its top equals the next element we want in π, pop it to output.
- Otherwise, push the next input element (if any remain).

One checks that this algorithm never pops from an empty stack and eventually performs exactly n pushes and n pops.

Assume for contradiction that π avoids 231 but the greedy algorithm fails to output π. Then there is a smallest index k such that the k-th symbol output by the algorithm does not equal π_k. Immediately before the k-th output step, all elements $\pi_1, \dots, \pi_{k-1}$ have been output in the correct order, and the top element x of the stack is not equal to π_k. Let y be the next unread input element, if any.

A careful case analysis of the relative order of x, y, π_k in π shows that these three elements must form a 231-pattern, contradicting the assumption that π avoids 231. Hence, if π avoids 231, the greedy algorithm succeeds, and π is stack-sortable. ∎

5.1.3. *Enumeration via the position of n.* We now count the 231-avoiding permutations. Let

$$a_n = |\operatorname{Av}_n(231)| = |\mathcal{S}_n|.$$

We derive a recurrence for a_n by looking at the position of the largest element n in the permutation.

Let $\pi \in \operatorname{Av}_n(231)$ and write

$$\pi = \sigma\, n\, \tau$$

where σ is the subsequence of elements before n and τ is the subsequence after n. Let $k = |\sigma|$, so $0 \leq k \leq n-1$ and $|\tau| = n-1-k$.

Lemma 5.2. *For any $\pi \in \operatorname{Av}_n(231)$ written as $\sigma\, n\, \tau$ with $|\sigma| = k$, all elements of σ are* smaller *than all elements of τ. Moreover, after standardizing the values in σ and τ to obtain permutations in S_k and S_{n-1-k}, we get permutations in $\operatorname{Av}_k(231)$ and $\operatorname{Av}_{n-1-k}(231)$ respectively.*

Proof. Suppose, for contradiction, that there exist $x \in \sigma$ and $y \in \tau$ with $x > y$. Then the subsequence x, n, y satisfies

$$\operatorname{pos}(x) < \operatorname{pos}(n) < \operatorname{pos}(y) \quad \text{and} \quad n > x > y,$$

so x, n, y are order-isomorphic to $2, 3, 1$. Thus x, n, y form a 231-pattern in π, contradicting 231-avoidance. Hence every element in σ is smaller than every element in τ.

Standardizing the letters means replacing the smallest letter in σ by 1, the next smallest by 2, and so on, to obtain a permutation $\sigma' \in S_k$. Similarly, we standardize τ to $\tau' \in S_{n-1-k}$. If σ' contained a 231-pattern, then the same indices would give a 231-pattern in σ (and hence in π), which is impossible. Thus $\sigma' \in \operatorname{Av}_k(231)$. The same argument applies to τ'. ■

Conversely, given any pair of 231-avoiding permutations $\alpha \in S_k$ and $\beta \in S_{n-1-k}$, we can construct a unique 231-avoiding permutation $\pi \in S_n$ with $|\sigma| = k$ as follows. Think of α as a permutation on the set $\{1, \ldots, k\}$ and β as a permutation on the set $\{1, \ldots, n-1-k\}$. Embed them into $\{1, \ldots, n-1\}$ by letting α' act on the smallest k values $\{1, \ldots, k\}$ and letting β' act on the largest $n-1-k$ values $\{k+1, \ldots, n-1\}$. Then all entries of α' are smaller than all entries of β', and we set

$$\pi = \alpha'\, n\, \beta'.$$

By construction, π has $|\sigma| = k$, and any 231-pattern would have to use n as its largest element. But for any triple (x, n, y) with $x \in \sigma$ and $y \in \tau$ we have $x < y < n$, so (x, n, y) is of type 132, not 231. Inside σ and inside τ there are no 231-patterns by assumption, so π avoids 231.

This construction is inverse to the standardization procedure discussed in the lemma, so we obtain a bijection between

$$\operatorname{Av}_k(231) \times \operatorname{Av}_{n-1-k}(231) \quad \text{and} \quad \{\pi \in \operatorname{Av}_n(231) : |\sigma| = k\}.$$

It follows that

$$a_n = \sum_{k=0}^{n-1} a_k\, a_{n-1-k}, \qquad n \geq 1,$$

with initial condition $a_0 = 1$ (the empty permutation). This is exactly the Catalan recurrence

$$C_0 = 1, \qquad C_n = \sum_{k=0}^{n-1} C_k C_{n-1-k}.$$

Theorem 5.3. *For all $n \geq 0$,*

$$|\mathcal{S}_n| = |\operatorname{Av}_n(231)| = C_n.$$

Proof. Both sequences (a_n) and (C_n) satisfy the same recurrence and initial condition, and the Catalan sequence is uniquely determined by these. Therefore $a_n = C_n$ for all n. ■

Thus, stack-sortable permutations form a Catalan family. The push–pop words of stack-sorting executions give an explicit bijection between $\operatorname{Av}_n(231)$ and Dyck paths of semilength n, but the recurrence already shows why Catalan numbers appear.

5.2. **Noncrossing partitions of** $[n]$.

5.2.1. *Definitions.* A *set partition* of $[n] = \{1, 2, \dots, n\}$ is a collection of nonempty, pairwise disjoint subsets (called *blocks*) whose union is $[n]$. For example,

$$\pi = \{\{1, 4, 5\}, \{2, 3\}, \{6\}\}$$

is a partition of $[6]$.

We represent a partition by placing the numbers $1, 2, \dots, n$ in order on a horizontal line and, for each block, drawing arcs connecting consecutive elements of the block. For the example above we draw arcs $(1, 4)$, $(4, 5)$, and $(2, 3)$ over the line.

Definition 5.4. A partition π of $[n]$ is *noncrossing* if in this diagram there are no crossing arcs, that is, there do not exist $a < b < c < d$ such that a is connected to c and b is connected to d by arcs in the same partition. Otherwise, π is *crossing*.

Let NC_n denote the set of noncrossing partitions of $[n]$, and let $p_n = |\mathrm{NC}_n|$. We show that $p_n = C_n$.

5.2.2. *Structure via the block containing* 1. Consider a noncrossing partition $\pi \in \mathrm{NC}_n$. Let B be the block containing 1, and write

$$B = \{1 = b_1 < b_2 < \cdots < b_k\}$$

with $k \geq 1$. Since we have drawn arcs between consecutive elements of the block, we have arcs (b_1, b_2), (b_2, b_3), …, (b_{k-1}, b_k).

Between b_j and b_{j+1} (for $1 \leq j < k$) there may be other points of $[n]$ belonging to other blocks. Define the intervals

$$I_j = \{b_j + 1, b_j + 2, \dots, b_{j+1} - 1\}, \quad 1 \leq j < k,$$

and also

$$I_0 = \{2, 3, \dots, b_1 - 1\}, \qquad I_k = \{b_k + 1, b_k + 2, \dots, n\}.$$

The noncrossing condition implies that each block of π (other than B) is entirely contained in exactly one of the I_j's. In other words, once we know B, the rest of the partition splits into $(k + 1)$ independent noncrossing partitions, one on each interval I_j.

Let $|I_j| = n_j$ so that

$$n_0 + n_1 + \cdots + n_k = n - k.$$

For each j, the restriction of π to I_j is a noncrossing partition of a set of size n_j, which can be canonically identified with a noncrossing partition of $[n_j]$. Conversely, given B and a choice of noncrossing partitions on each I_j, we can reconstruct π. This decomposition is actually quite standard when considering noncrossing partitions. It shows that NC_n admits a recursive description in terms of smaller noncrossing partitions glued around the distinguished block B. More concretely, if $p_n = |\mathrm{NC}_n|$ and $p_0 = 1$, then for $n \geq 1$ we obtain the recurrence

$$p_n = \sum_{k\geq 1} \sum_{\substack{n_1+\cdots+n_k=n-k \\ n_j\geq 0}} p_{n_1} p_{n_2} \cdots p_{n_k}. \tag{5.1}$$

Here $k = |B|$ is the size of the block containing 1, and $n_j = |I_j|$ is the size of the j-th interval between consecutive elements of B (including the final interval after the largest element of B); these sizes satisfy

$$n_1 + \cdots + n_k = n - k.$$

This decomposition is quite standard when considering noncrossing partitions. It shows that NC_n admits a recursive description in terms of smaller noncrossing partitions glued around the distinguished block B. We will later show how to translate this into generating functions, and yield the Catalan functional equation

$$P(x) = \sum_{n\geq 0} p_n x^n = 1 + xP(x)^2,$$

which is the same equation satisfied by the Catalan generating function

$$C(x) = \sum_{n\geq 0} C_n x^n.$$

We refer to standard references for the full derivation of this functional equation from the decomposition above.

Theorem 5.5.

For all $n \geq 0$,

$$p_n = |\mathrm{NC}_n| = C_n.$$

Proof. The power series $P(x)$ and $C(x)$ satisfy the same functional equation

$$F(x) = 1 + xF(x)^2$$

and the same initial condition $F(0) = 1$. This equation has a unique formal power series solution, so $P(x) = C(x)$ and hence $p_n = C_n$ for all n. ■

Thus, noncrossing partitions of $[n]$ form another Catalan family. There are explicit bijections between NC_n, triangulations of an $(n + 2)$-gon, plane trees, and Dyck paths.

5.3. **123-avoiding permutations.** We now show that the number of 123-avoiding permutations in S_n is also C_n. This gives a third Catalan family, this time directly in terms of pattern avoidance.

Let

$$b_n = |\,\mathrm{Av}_n(123)|.$$

5.3.1. *Recurrence via the position of n.* Take $\pi \in \mathrm{Av}_n(123)$ and write

$$\pi = \sigma\, n\, \tau$$

with $|\sigma| = k$, $|\tau| = n-1-k$. We derive constraints on σ and τ.

Lemma 5.6. *For any $\pi \in \mathrm{Av}_n(123)$ written as $\sigma\, n\, \tau$ with $|\sigma| = k$, all elements in σ are greater than all elements in τ. Moreover, after standardizing, σ is a 123-avoiding permutation in S_k and τ is a 123-avoiding permutation in S_{n-1-k}.*

Proof. Suppose there exist $x \in \sigma$ and $y \in \tau$ with $x < y$. Then the triple x, y, n forms a 123-pattern in π: we have

$$\mathrm{pos}(x) < \mathrm{pos}(n) < \mathrm{pos}(y)$$

and values

$$x < y < n,$$

so x, y, n are order-isomorphic to $1, 2, 3$ in increasing positions. This contradicts 123-avoidance. Thus every element of σ is greater than every element of τ.

Standardizing σ and τ as before, any 123-pattern in the standardization would correspond to a 123-pattern in π, so both standardizations avoid 123. ■

Conversely, given $\alpha \in \mathrm{Av}_k(123)$ and $\beta \in \mathrm{Av}_{n-1-k}(123)$, we can reconstruct a unique 123-avoiding permutation in S_n with $|\sigma| = k$ by letting the letters of α be the largest k values less than n, the letters of β be the smallest $n-1-k$ values, and then inserting them as σ, τ in $\sigma\, n\, \tau$.

This gives a bijection

$$\mathrm{Av}_k(123) \times \mathrm{Av}_{n-1-k}(123) \longleftrightarrow \{\pi \in \mathrm{Av}_n(123) : |\sigma| = k\}.$$

Summing over all k we obtain:

Proposition 5.7. *The numbers $b_n = |\mathrm{Av}_n(123)|$ satisfy*

$$b_0 = 1, \qquad b_n = \sum_{k=0}^{n-1} b_k\, b_{n-1-k} \quad \text{for } n \geq 1.$$

This is again the Catalan recurrence. Therefore:

Theorem 5.8. *For all $n \geq 0$,*

$$|\mathrm{Av}_n(123)| = C_n.$$

By the symmetries of permutation patterns of length 3, this implies:

Corollary 5.9. *For any $\sigma \in S_3$ and any $n \geq 0$,*

$$|\mathrm{Av}_n(\sigma)| = C_n.$$

The Catalan enumeration of 123-avoiding permutations does admit a lot of refinements. A descent of a permutation $\pi \in S_n$ is an index i with $1 \leq i \leq n-1$ such that $\pi_i > \pi_{i+1}$. Let $d(\pi)$ be the number of descents of π.

Though we will discuss this greater in-depth later, the *Narayana numbers* $N(n,k)$ are defined by

$$N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}, \quad 1 \leq k \leq n.$$

They satisfy

$$\sum_{k=1}^{n} N(n,k) = C_n.$$

It is a nontrivial theorem that $N(n,k)$ counts the number of 123-avoiding permutations in S_n with exactly $k-1$ descents. In other words,

$$|\{\pi \in \mathrm{Av}_n(123) : d(\pi) = k-1\}| = N(n,k).$$

There are bijections from $\mathrm{Av}_n(123)$ to Dyck paths of semilength n that carry descents to peaks, and the same Narayana numbers count Dyck paths with a fixed number of peaks.

6. Triangulation

In this section we show that the number of ways to triangulate a convex polygon is counted by the Catalan numbers.

Lemma 6.1. *Any triangulation of a convex n-gon uses exactly $n-3$ diagonals and produces exactly $n-2$ triangles.*

Proof. Let d be the number of diagonals in a triangulation of an n-gon, and f the number of resulting triangular faces. Consider the planar graph whose vertices are the n vertices of the polygon and whose edges are the n sides of the polygon together with the d diagonals.

This graph has

$$V = n, \qquad E = n + d, \qquad F = f + 1$$

faces (where the extra face is the exterior). By Euler's formula for planar graphs,

$$V - E + F = 2 \quad \Rightarrow \quad n - (n+d) + (f+1) = 2,$$

which simplifies to

$$-f + 1 - d = 2 - 2n \quad \Rightarrow \quad f = d + 1.$$

On the other hand, each triangular face has 3 edges, and each edge lies on at most two faces. Counting incidences of edges and faces gives

$$3f = 2d + n,$$

because every diagonal is counted twice (it borders two triangles) and every side of the polygon is counted once (it borders exactly one triangle). Substituting $f = d + 1$ into $3f = 2d + n$ yields

$$3(d+1) = 2d + n \quad \Rightarrow \quad 3d + 3 = 2d + n \quad \Rightarrow \quad d = n - 3, \quad f = d + 1 = n - 2.$$

■

For $n \geq 0$, let T_n denote the number of triangulations of a convex $(n+2)$-gon. Thus $T_0 = 1$ (there is exactly one way to triangulate a triangle).

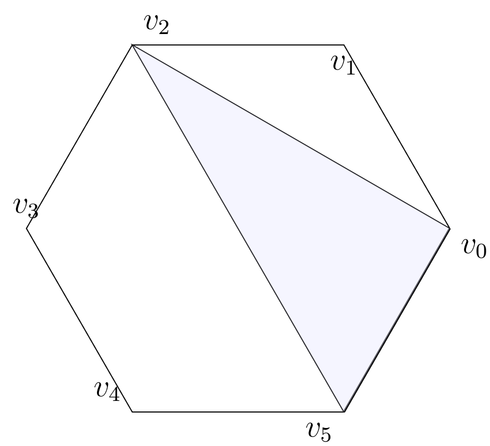
To derive a recurrence for T_n, we fix a distinguished side and look at the unique triangle that contains it.

Lemma 6.2. *For $n \geq 1$, the numbers T_n satisfy*

$$T_n = \sum_{k=0}^{n-1} T_k \, T_{n-1-k}.$$

Proof. Consider a convex $(n+2)$-gon with vertices labeled $v_0, v_1, \ldots, v_{n+1}$ in counterclockwise order. Fix the base edge $v_0 v_{n+1}$.

In any triangulation, the edge $v_0 v_{n+1}$ belongs to exactly one triangle, whose third vertex is some v_j with $1 \le j \le n$. This triangle (v_0, v_j, v_{n+1}) splits the polygon into two smaller convex polygons:



- a left polygon with vertices $v_0, v_1, \ldots, v_j$, which has $j+1$ vertices and hence $j-1$ diagonals in any triangulation;
- a right polygon with vertices $v_j, v_{j+1}, \ldots, v_{n+1}$, which has $(n+2-j)$ vertices and hence $(n-j-1)$ diagonals in any triangulation.

By definition of T_n, the left polygon has T_{j-1} triangulations and the right polygon has T_{n-j} triangulations. For a fixed j, these choices are independent, so the number of triangulations in which the third vertex is v_j is

$$T_{j-1}\, T_{n-j}.$$

Summing over all $j = 1, 2, \ldots, n$ gives

$$T_n = \sum_{j=1}^{n} T_{j-1}\, T_{n-j},$$

which is the same as

$$T_n = \sum_{k=0}^{n-1} T_k\, T_{n-1-k}.$$

■

Theorem 6.3. *For all $n \ge 0$, the number of triangulations of a convex $(n+2)$-gon is the nth Catalan number:*

$$T_n = C_n.$$

Proof. We have shown that (T_n) satisfies

$$T_0 = 1, \qquad T_n = \sum_{k=0}^{n-1} T_k\, T_{n-1-k} \quad (n \ge 1).$$

The Catalan numbers (C_n) satisfy the same recurrence with the same initial condition. By induction on n, a sequence determined by this recurrence and initial value is unique, so $T_n = C_n$ for all n. ■

Thus one of the standard interpretations of C_n is

> C_n is the number of triangulations of a convex $(n+2)$-gon by noncrossing diagonals.

Triangulations also encode ways to split a polygon recursively into smaller pieces.

Definition 6.4. Fix a convex $(n+2)$-gon P. A *binary splitting scheme* for P is a rooted full binary tree whose

- root is labeled by P,
- each internal node is labeled by a convex subpolygon of P with at least 4 vertices, and has two children labeled by the two subpolygons obtained by drawing a diagonal that splits it into two smaller convex polygons,
- each leaf is labeled by a triangle.

Two splitting schemes are considered distinct if their associated binary trees differ as rooted labeled trees.

Intuitively, a splitting scheme records the history of repeatedly drawing one diagonal at a time until only triangles remain.

Corollary 6.5. *For $n \geq 0$, the number of binary splitting schemes of a convex $(n+2)$-gon is also C_n.*

Proof. Given a binary splitting scheme, take the union of all diagonals drawn during the splitting process. Because each step uses a diagonal that lies entirely inside the polygon and never crosses previous diagonals, the final union of diagonals is a triangulation of the polygon.

Conversely, given a triangulation, we can recover a binary splitting scheme by recursively "undoing" diagonals. For example, fix a base edge v_0v_{n+1} and proceed as follows:

- Let (v_0, v_j, v_{n+1}) be the unique triangle in the triangulation containing the base edge. First split along the diagonal v_0v_j or v_jv_{n+1} to produce two smaller polygons.
- Recursively apply the same procedure inside each subpolygon (using the induced triangulation) until only triangles remain.

This produces a binary splitting scheme whose set of diagonals is exactly the original triangulation. The two constructions are inverse to each other, yielding a bijection between triangulations and binary splitting schemes.

Since we already know there are C_n triangulations of a convex $(n+2)$-gon, there are also C_n binary splitting schemes. ■

Triangulations already give one geometric realization of the Catalan numbers. Another very clean euclidean picture comes from noncrossing matchings drawn as chords of a circle.

Definition 6.6. Place $2n$ labeled points in convex position on a circle, labeled $1, 2, \dots, 2n$ in clockwise order. A *noncrossing perfect matching* is a collection of n disjoint chords such that

- each point is an endpoint of exactly one chord, and
- no two chords intersect in the interior of the circle.

Let M_n denote the number of noncrossing perfect matchings on $2n$ points.

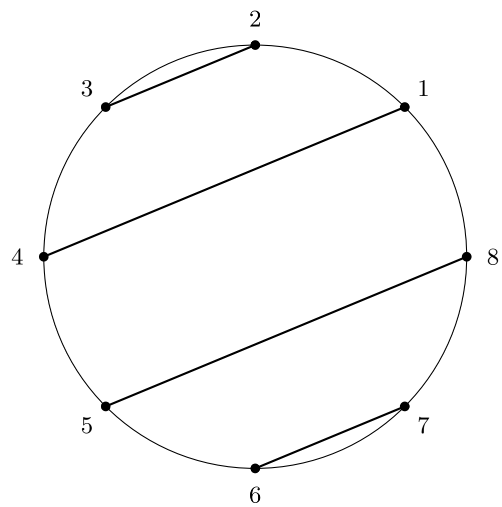


Figure. A noncrossing perfect matching on $2n = 8$ points on a circle.

Lemma 6.7. *The sequence $(M_n)_{n\geq 0}$ satisfies*

$$M_0 = 1, \qquad M_n = \sum_{k=0}^{n-1} M_k\, M_{n-1-k} \text{ for } n \geq 1.$$

Proof. The case $n = 0$ corresponds to the empty matching. For $n \geq 1$, fix the point labeled 1. In any noncrossing perfect matching, point 1 is joined by a chord to exactly one other point, say point $2j$ for some $1 \leq j \leq n$.

The chord $(1, 2j)$ divides the remaining $2n - 2$ points into two blocks:

- the $2(j-1)$ points strictly between 1 and $2j$ in clockwise order,
- the $2(n-j)$ points strictly between $2j$ and 1 around the other side.

Because the matching is noncrossing, these two blocks are completely independent: every chord whose one endpoint lies in one block must have its other endpoint in the same block, otherwise it would cross the chord $(1, 2j)$.

Thus, for fixed j, the number of noncrossing matchings where 1 is matched with $2j$ is

$$M_{j-1}\, M_{n-j},$$

where M_{j-1} counts matchings on the first block of $2(j-1)$ points and M_{n-j} counts matchings on the second block of $2(n-j)$ points.

Summing over $j = 1, 2, \ldots, n$ gives

$$M_n = \sum_{j=1}^{n} M_{j-1}\, M_{n-j} = \sum_{k=0}^{n-1} M_k\, M_{n-1-k},$$

as claimed. ■

Theorem 6.8. *For all $n \geq 0$ we have*

$$M_n = C_n,$$

where C_n is the nth Catalan number.

Proof. The Catalan numbers (C_n) satisfy the same recurrence

$$C_0 = 1, \qquad C_n = \sum_{k=0}^{n-1} C_k\, C_{n-1-k} \text{ for } n \geq 1.$$

Since the recurrence together with $M_0 = C_0 = 1$ uniquely determines the sequence, induction on n shows $M_n = C_n$ for all n. ■

Thus the Catalan number C_n is also the number of noncrossing perfect matchings of $2n$ points on a circle. This is another very geometric model: the Catalan structure is literally encoded by nonintersecting chords in the plane.

6.1. **Noncrossing partitions and chord diagrams.** We can generalize the previous picture by allowing groups of more than two points to be connected together via chords.

Definition 6.9. Let $[n] = \{1, 2, \ldots, n\}$. A set partition π of $[n]$ is *noncrossing* if, when the numbers $1, 2, \ldots, n$ are placed on a circle in clockwise order and the elements of each block of π are joined by chords inside the circle, no two chords cross.

Let $\mathrm{NC}(n)$ denote the set of noncrossing partitions of $[n]$.

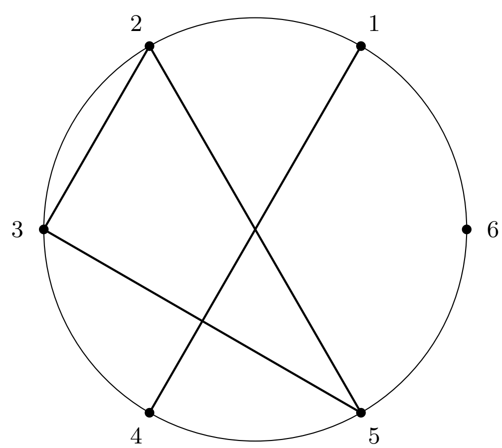


Figure. A noncrossing partition of [6] with blocks $\{1,4\}$, $\{2,3,5\}$, and $\{6\}$.

A classical result states that

$$|\mathrm{NC}(n+1)| = C_n.$$

We will not give the full proof here, but the idea is to encode a noncrossing partition via a recursive decomposition very similar to the noncrossing matching argument: picking the block containing 1 and looking at the induced subpartitions between its elements yields the same Catalan recurrence.

Geometrically, this says that Catalan numbers count all ways to decompose a convex polygon with labeled vertices into "clusters" of vertices joined by nonintersecting chords, without insisting that each cluster is just a pair.

6.2. **Higher polygon dissections and Fuss–Catalan numbers.** Triangulations are just one kind of polygon dissection. We can also ask for dissections into larger polygons, and this leads to a natural generalization of the Catalan numbers.

Definition 6.10. Fix an integer $m \geq 1$. A *dissection of type* m of a convex polygon is a way to subdivide it into smaller polygons, each having exactly $m+2$ sides, using noncrossing diagonals.

Let $D_n^{(m)}$ denote the number of dissections of a convex polygon with

$$N = (m+1)n + 2$$

vertices into $(m+2)$-gons.

For $m = 1$, $D_n^{(1)}$ counts triangulations of an $(n+2)$-gon, so $D_n^{(1)} = C_n$.

Lemma 6.11. *For fixed $m \geq 1$, the numbers $D_n^{(m)}$ satisfy*

$$D_0^{(m)} = 1, \qquad D_n^{(m)} = \sum_{\substack{i_1+\cdots+i_{m+1}=n-1 \\ i_j \geq 0}} D_{i_1}^{(m)} \cdots D_{i_{m+1}}^{(m)} \text{ for } n \geq 1.$$

Proof. Consider a convex polygon with $N = (m+1)n + 2$ vertices, labeled $v_0, v_1, \ldots, v_{N-1}$ in order. Fix the base edge $v_0 v_{N-1}$.

In any dissection of type m, there is a unique $(m+2)$-gon containing the base edge. Suppose its other vertices are

$$v_{j_1}, v_{j_2}, \ldots, v_{j_m}$$

with

$$0 < j_1 < j_2 < \cdots < j_m < N - 1.$$

These m internal vertices, together with v_0 and v_{N-1}, cut the large polygon into $m+1$ subpolygons, each of which must itself be dissected into $(m+2)$-gons. If the rth subpolygon has

$$(m+1)i_r + 2$$

vertices, then dissecting it contributes a factor $D_{i_r}^{(m)}$, and the condition that the total number of vertices add up correctly forces

$$i_1 + \cdots + i_{m+1} = n - 1.$$

Conversely, any choice of nonnegative integers $i_1, \ldots, i_{m+1}$ summing to $n-1$, together with dissections of the corresponding subpolygons, glue together to give a valid dissection of the original polygon.

Summing over all compositions $n - 1 = i_1 + \cdots + i_{m+1}$ yields the recurrence. ∎

Theorem 6.12. *For all $n \geq 0$ and $m \geq 1$,*

$$D_n^{(m)} = \frac{1}{mn+1} \binom{(m+1)n}{n}.$$

In particular, when $m = 1$ this reduces to

$$D_n^{(1)} = \frac{1}{n+1} \binom{2n}{n} = C_n.$$

Idea of proof. The numbers

$$\mathrm{Cat}_n^{(m)} = \frac{1}{mn+1} \binom{(m+1)n}{n}$$

are called the *Fuss–Catalan numbers.* One checks directly that

$$\mathrm{Cat}_0^{(m)} = 1$$

and that they satisfy exactly the same $(m+1)$-fold convolution recurrence as the $D_n^{(m)}$:

$$\mathrm{Cat}_n^{(m)} = \sum_{i_1+\cdots+i_{m+1}=n-1} \mathrm{Cat}_{i_1}^{(m)} \cdots \mathrm{Cat}_{i_{m+1}}^{(m)} .$$

Since the recurrence with the given initial condition determines the sequence uniquely, we must have $D_n^{(m)} = \mathrm{Cat}_n^{(m)}$ for all n.

■

7. Generating Functions of Catalan Numbers

Let

$$C(z) = \sum_{n\geq 0} C_z z^n$$

be the generating function of the Catalan numbers. Using our recurrence on C_n, we derive a functional equation for $C(z)$.

Lemma 7.1. *We claim that*

$$C(z) = 1 + zC(z)^2$$

Proof. By (insert recurrence ref),

$$C(z) = C_0 + \sum_{n\geq 0} C_{n+1} z^{n+1} = 1 + \sum_{n\geq 0} (\sum_{i=0}^{n} C_i C_{n-i}) z^{n+1}$$

Factoring out z, we get a Cauchy Product

$$C(z) = 1 + z \sum_{n\geq 0} \left(\sum_{i=0}^{n} C_i \, C_{n-i} \right) z^n = 1 + z \left(\sum_{i\geq 0} C_i z^i \right) \left(\sum_{j\geq 0} C_j z^j \right) = 1 + z\, C(z)^2.$$

So we are done. ■

Lemma 7.2. *The functional equation*

$$C(z) = 1 + zC(z)^2$$

has unique solution $C(z) \in \mathbb{C}[[z]]$ *with* $C(0) = 1$*, namely*

$$C(z) = \frac{1 - \sqrt{1-4z}}{2z}.$$

Proof. By quadratic formula,

$$zC(z)^2 - C(z) + 1 = 0 \implies C(z) = \frac{1 \pm \sqrt{1-4z}}{2z}.$$

Expanding $\sqrt{1-4z}$ as a formal power series around $z = 0$,

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

Then

$$\frac{1 + \sqrt{1-4z}}{2z} = \frac{1 + (1 - 2z + O(z^2))}{2z} = \frac{2 - 2z + O(z^2)}{2z} = \frac{1}{z} - 1 + O(z),$$

which is not a formal power series since it has a pole at $z = 0$.

The other branch is

$$\frac{1-\sqrt{1-4z}}{2z} = \frac{1-(1-2z+O(z^2))}{2z} = \frac{2z+O(z^2)}{2z} = 1+O(z),$$

which is a genuine power series with constant term 1. Hence it is the unique solution in $\mathbb{C}[[z]]$ with $C(0) = 1$. ■

Catalan numbers are not the only sequence whose generating function satisfies a simple algebraic equation. A closely related family is given by the Motzkin numbers, which interpolate between Catalan numbers and central binomial coefficients.

Definition 7.3. A *Motzkin path* of length n is a lattice path from $(0, 0)$ to $(n, 0)$ using steps

$$U = (1, 1), H = (1, 0), D = (1, -1),$$

that never goes below the x-axis. Let M_n be the number of Motzkin paths of length n, and define the generating function

$$M(z) = \sum_{n \geq 0} M_n z^n.$$

Lemma 7.4. *The generating function $M(z)$ satisfies*

$$M(z) = 1 + z\,M(z) + z^2 M(z)^2.$$

Proof. Consider a nonempty Motzkin path and inspect its first step.

If the first step is H, then the rest is an arbitrary Motzkin path. This contributes $zM(z)$. The factor z records the H, and $M(z)$ encodes the remaining stuff.

If the first step is U, then there is a unique first return to the x-axis via a matching D. Between U and D we have a possibly empty path staying strictly above the axis. Lowering it by one unit gives another Motzkin path. After D comes a second Motzkin path. This contributes $z^2 M(z)^2$. One factor of z for U, one for D, and a factor of $M(z)$ for each subpath.

Adding the empty path counted by 1, gives

$$M(z) = 1 + zM(z) + z^2 M(z)^2.$$

as desired ■

Solving this similarly to the Catalan GF, we get

$$M(z) = \frac{1 - z\sqrt{1 - 2z - 3z^2}}{2z^2}.$$

We can also derive this via its relation to the Catalan Numbers.

Theorem 7.5. *We claim that*

$$\sum_{k=0}^{n} C_k = 1 + \sum_{k=1}^{n} \binom{n}{k} M_{k-1}$$

Proof. We expand the square root from $M(z)$ as a power series. Note that

$$1 - 2z - 3z^2 = (1-z)^2 - 4z^2.$$

Thus

$$\sqrt{1-2z-3z^2} = (1-z)\sqrt{1 - \frac{4z^2}{(1-z)^2}} = (1-z)\sqrt{1-4u}, u = \frac{z^2}{(1-z)^2}.$$

Using the Catalan expansion

$$\sqrt{1-4u} = 1 - 2\sum_{k\geq 1} C_k u^k,$$

we obtain

$$\sqrt{1-2z-3z^2} = (1-z)\left(1 - 2\sum_{k\geq 1} C_k \left(\frac{z^2}{(1-z)^2}\right)^k\right) = (1-z) - 2\sum_{k\geq 1} C_k \frac{z^{2k}}{(1-z)^{2k-1}}.$$

Substitute this back into the closed form for $M(z)$ to get

$$M(z) = \frac{1 - z - \left[(1-z) - 2\sum_{k\geq 1} C_k \frac{z^{2k}}{(1-z)^{2k-1}}\right]}{2z^2} = \sum_{k\geq 1} C_k \frac{z^{2k}}{z^2(1-z)^{2k-1}}.$$

Hence

$$M(z) = \sum_{k\geq 0} C_k \frac{z^{2k}}{(1-z)^{2k-1}}.$$

Finally, expand the geometric factor to get

$$\frac{1}{(1-z)^{2k-1}} = \sum_{m\geq 0} \binom{m+2k-2}{2k-2} z^m.$$

Therefore,

$$M(z) = \sum_{k\geq 0} C_k \sum_{m\geq 0} \binom{m+2k-2}{2k-2} z^{m+2k} = \sum_{n\geq 0}\left(\sum_{k=0}^{\lfloor n/2\rfloor} \binom{n}{2k} C_k\right) z^n.$$

Extracting coefficients of z^n gives

$$M_n = \sum_{k=0}^{\lfloor n/2\rfloor} \binom{n}{2k} C_k,$$

From this, we inversely see that

$$C_{n+1} = \sum_{k=0}^{n} \binom{n}{k} M_k$$

which gives

$$\sum_{k=0}^{n} C_k = 1 + \sum_{k=1}^{n} \binom{n}{k} M_{k-1}$$

as desired.

■

We also discuss a related type of numbers, namely the Narayana numbers, and their algebraic relation to Catalan.

Definition 7.6. The Narayana Numbers $N(n,k)$ refine the Catalan numbers by counting Dyck paths of semilength n with exactly k peaks. Equivalently,

$$N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}, 1 \le k \le n.$$

They satisfy

$$\sum_{k=1}^{n} N(n,k) = C_n,$$

where C_n is the nth Catalan number.

Proposition 7.7. *Let $N(n,k)$ be the Narayana numbers, i.e. the number of Dyck paths of semilength n with exactly k peaks. Then the bivariate generating function*

$$S(z,t) = \sum_{n=1}^{\infty} \sum_{k=1}^{n} N(n,k)\, z^n\, t^{k-1}$$

is given by

$$S(z,t) = \frac{1 - z(t+1) - \sqrt{1 - 2z(t+1) + z^2(t-1)^2}}{2tz}.$$

Moreover, it can be expressed as

$$S(z,t) = \frac{z}{1 - z(t+1)}\, C\left(\frac{z^2 t}{\left(1 - z(t+1)\right)^2}\right).$$

Proof. Let $\mathcal{D}$ be the set of all Dyck paths including the empty path. For a path $P \in \mathcal{D}$, let $|P|$ be its semilength and $\mathrm{pk}(P)$ its number of peaks. Define the bivariate generating function

$$F(z,t) = \sum_{P \in \mathcal{D}} z^{|P|} t^{\mathrm{pk}(P)} = 1 + \sum_{n \ge 1} \sum_{k=1}^{n} N(n,k)\, z^n t^k.$$

Then

$$S(z,t) = \sum_{n \ge 1} \sum_{k=1}^{n} N(n,k) z^n t^{k-1} = \frac{F(z,t) - 1}{t}.$$

Every nonempty Dyck path P has a unique decomposition

$$P = U\, A\, D\, B,$$

where $A, B \in \mathcal{D}$ are Dyck paths possibly empty, and $U = (1,1)$, $D = (1,-1)$.

We split into two cases

Case 1: A is empty Then $P = UDB$. The two steps UD form a peak at height 1, so

$$\mathrm{pk}(P) = 1 + \mathrm{pk}(B), |P| = 1 + |B|.$$

The contribution of all such paths to F is

$$\sum_{B \in \mathcal{D}} z^{1+|B|} t^{1+\mathrm{pk}(B)} = zt \sum_{B \in \mathcal{D}} z^{|B|} t^{\mathrm{pk}(B)} = zt\, F(z,t).$$

Case 2: A is nonempty Then $P = UADB$ with $A \neq \emptyset$. The subpath A is a Dyck path that starts at height 1 and returns to height 1; all its peaks are at height at least 2, but they are still peaks of P. The outer U and D do not create a new peak at height 1 in this case. Thus

$$\mathrm{pk}(P) = \mathrm{pk}(A) + \mathrm{pk}(B), |P| = 1 + |A| + |B|.$$

The contribution of these paths is

$$\sum_{\substack{A\in\mathcal{D}\\ A\neq\emptyset}} \sum_{B\in\mathcal{D}} z^{1+|A|+|B|} t^{\mathrm{pk}(A)+\mathrm{pk}(B)} = z \left(\sum_{\substack{A\in\mathcal{D}\\ A\neq\emptyset}} z^{|A|} t^{\mathrm{pk}(A)} \right) \left(\sum_{B\in\mathcal{D}} z^{|B|} t^{\mathrm{pk}(B)} \right) = z\,(F(z,t)-1)\,F(z,t).$$

Adding the empty path (counted by 1), we obtain the functional equation

$$F(z,t) = 1 + ztF(z,t) + z(F(z,t)-1)F(z,t) = 1 + zF(z,t)^2 + z(t-1)F(z,t).$$

Rewrite the equation as a quadratic in F

$$zF^2 + z(t-1)F - F + 1 = 0.$$

Solving for F gives

$$F(z,t) = \frac{1 - z(t-1) \pm \sqrt{1 - 2z(t+1) + z^2(t-1)^2}}{2z}.$$

At $z = 0$ we must have $F(0,t) = 1$ (only the empty path), so we choose the branch with $F(z,t) = 1 + O(z)$, namely

$$F(z,t) = \frac{1 - z(t+1) - \sqrt{1 - 2z(t+1) + z^2(t-1)^2}}{2z}.$$

Since $S(z,t) = \dfrac{F(z,t)-1}{t}$, we obtain

$$S(z,t) = \frac{1}{t}\left(\frac{1 - z(t+1) - \sqrt{1 - 2z(t+1) + z^2(t-1)^2}}{2z} - 1 \right)$$
$$= \frac{1 - z(t+1) - \sqrt{1 - 2z(t+1) + z^2(t-1)^2}}{2tz},$$

which is exactly the claimed closed form.

Set

$$\Delta(z,t) = 1 - 2z(t+1) + z^2(t-1)^2 = \big(1 - z(t+1)\big)^2 - 4z^2t.$$

Define

$$w = \frac{z^2t}{\big(1 - z(t+1)\big)^2}.$$

Then

$$1 - 4w = 1 - \frac{4z^2t}{\big(1 - z(t+1)\big)^2} = \frac{\Delta(z,t)}{\big(1 - z(t+1)\big)^2},$$

so

$$\sqrt{1-4w} = \frac{\sqrt{\Delta(z,t)}}{1 - z(t+1)}.$$

Hence the Catalan generating function satisfihaes

$$C(w) = \frac{1-\sqrt{1-4w}}{2w} = \frac{1-\sqrt{\Delta(z,t)}/(1-z(t+1))}{2w}.$$

Now rewrite $S(z,t)$

$$\begin{aligned} S(z,t) &= \frac{1-z(t+1)-\sqrt{\Delta(z,t)}}{2tz} \\ &= \frac{1-z(t+1)}{2tz}\left(1-\frac{\sqrt{\Delta(z,t)}}{1-z(t+1)}\right) \\ &= \frac{1-z(t+1)}{2tz}\left(1-\sqrt{1-4w}\right). \end{aligned}$$

Since

$$1-\sqrt{1-4w} = 2w\,C(w)$$

and

$$w = \frac{z^2t}{\left(1-z(t+1)\right)^2},$$

we get

$$\begin{aligned} S(z,t) &= \frac{1-z(t+1)}{2tz}\cdot 2w\,C(w) = \frac{1-z(t+1)}{tz}\cdot\frac{z^2t}{\left(1-z(t+1)\right)^2}\,C(w) \\ &= \frac{z}{1-z(t+1)}\,C\left(\frac{z^2t}{\left(1-z(t+1)\right)^2}\right), \end{aligned}$$

as desired ■

The GenFunc for Catalan numbers in itself does not seem to have many other applications.

As we have seen, though, it is often useful in parameterizing other sequences.

8. References

Email address: `Emails`