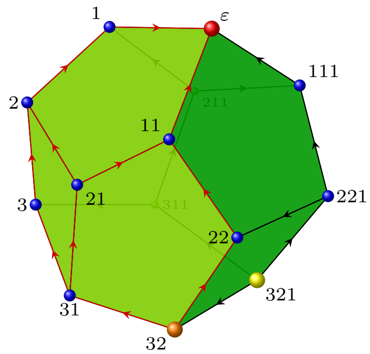
CATALAN COMBINATORICS



FRANÇOIS BERGERON

Contents

Introduction 1
1. A very brief history 2
2. Dealing with Catalan numbers 3
3. More combinatorial structures counted by Catalan numbers 3
4. Polynomial count 6
Appendix A. Integer Sequences 7
Appendix B. Partitions and Diagrams 7
Appendix C. q-Analogs 8
Appendix D. Generalizations 9
Appendix E. Problems 13
References 14

Introduction

The study of the interplay between Catalan combinatorics and abstract algebra has interesting ties with many subjects including: Computer Science, Algebraic Geometry, Knot Theory, and Quantum Mechanics. On the combinatorial side, the story involves many classical structures including triangulation of polygons, binary trees, Dyck paths, partitions contained in a staircase, well-parenthesized expressions, two-row standard Young tableaux, and pattern avoiding permutations; to mention but a few. The enumeration of all of these objects involves Catalan numbers $1/(n+1)\binom{2n}{n}$. Several interesting families of polynomials arise when one considers weighted enumerations of the combinatorial objects considered in this broad subject. As a matter of fact, since they arise in so many areas, several generalizations of these families of object have been introduced over the years. On top of this rich combinatorics, ties with abstract algebra also abound. Typically, one sees the Catalan numbers appear as the dimensions of interesting vector spaces (for more on this, see [1, 2, 3]).

Date: June 16, 2021.

The purpose of these notes is to give a small first taste of this vast subject, as well as some indications on where to find further material.

1. A very brief history

The sequence of **Catalan**[1] **numbers**:

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, \ldots, \qquad C_n := \frac{1}{n+1}\binom{2\,n}{n}, \tag{1.1}$$

already occurs explicitly in the work of Leonhard Euler (1707–1783) as the number of triangulations of n-polygons. It had also been used by Minggatu (c.1692–c.1763), to relate the series expansion of $\sin(2x)$ to that of $\sin(x)$, as well as to calculate the ratio of division of a circle. In between Euler-Minggatu's era and that of Catalan, several others[2] studied properties, recurrences and formulas for these numbers. A generalization was also proposed by Nicolas Fuss (1755-1826), who worked for 10 years as Euler's assistant. For more on the history of Catalan numbers, and a discussion of why they are called as they are, see Igor Pak's text: *History of Catalan Numbers*, which appeared as an appendix of Richard Stanley's book on Catalan Numbers [8]. See also [6] for an accessible book on many subjects related to Catalan numbers. For a broad source of material on occurrences of Catalan numbers in research papers, see entry A000108 of the Online Encyclopedia of Integer Sequences (*a.k.a.* OEIS).

In Euler's interpretation of Catalan numbers, one considers all the ways that a n-sided polygon may be decomposed into triangles by joining pairs of vertices by non-crossing segments. This is illustrated for the hexagon in Figure 1.1. Another instance of combinatorial structures enumerated by

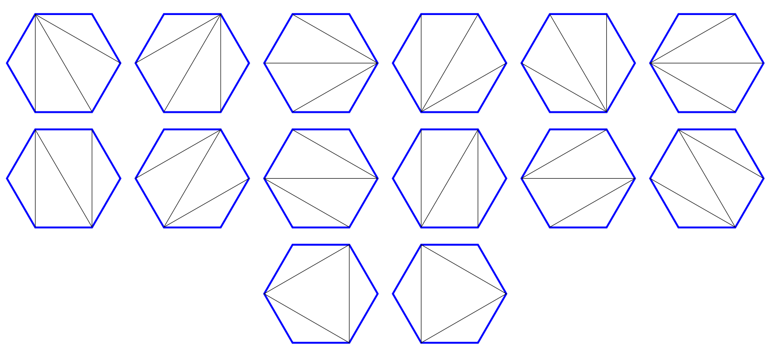


Fig. 1.1. The 14 triangulations of the hexagon.

the Catalan numbers, is the set of partitions[3] contained in a staircase partition $(n-1, n-2, \ldots 2, 1)$. For $n = 4$, there are 14 such partitions:

$$\{321, 32, 311, 31, 3, 221, 22, 211, 21, 2, 111, 11, 1, \varepsilon\},$$

where ε stands for the empty partition. As we have already mentioned, there are many families $(\mathcal{F}_n)_{n\in\mathbb{N}}$ of "nice" combinatorial structures that are counted by the Catalan numbers: *i.e.*

$$\mathrm{card}(\mathcal{F}_n) = \frac{1}{n+1}\binom{2n}{n}, \qquad \text{for all } n \in \mathbb{N}.$$

[1] For Eugène Charles Catalan (1814–1894).
[2] Such as Johann Andreas Segner (1704–1777).
[3] See the relevant section for definitions.

For any other family $(\mathcal{G}_n)_{n\in\mathbb{N}}$ having the same number of elements, one may certainly find bijections[4] between the sets $\mathcal{F}_n$ and $\mathcal{G}_n$. However, finding an explicit description of such a bijection is often an interesting challenge (see Problem 2), which may prove to be very hard in some instances.

2. Dealing with Catalan numbers

When studying sequences of numbers, one looks for interesting ways to describe them. Different descriptions often add interesting new perspectives. Some classical ways of calculating Catalan numbers include:

$$C_n = \prod_{k=2}^{n} \frac{n+k}{k}, \tag{2.1}$$

$$= \sum_{k=0}^{n-1} C_k\, C_{n-1-k}, \qquad C_0 := 1, \tag{2.2}$$

$$= \frac{2\,(2n-1)}{n+1} C_{n-1}; \tag{2.3}$$

as well as via their **generating series**

$$\mathcal{C}(x) := \sum_{0}^{\infty} C_n\, x^n \tag{2.4}$$

$$= 1 + x + 2x^2 + 5x^3 + 14x^4 + 42x^5 + 132x^6 + 429x^7 + 1430x^8 + \dots,$$

which is such that

$$\mathcal{C}(x) = 1 + x\, C(x)^2; \tag{2.5}$$

as may be show using Equation 2.2. It follows that

$$\mathcal{C}(x) = \frac{1 - x\sqrt{1-4\,x}}{2\,x}. \tag{2.6}$$

Using this, one may prove that Catalan numbers grow asymptotically as

$$C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}. \tag{2.7}$$

Equation 2.5 may also be exploited to calculate $\mathcal{C}(x)$ as the limit of "iterations" of $f(z) = z^2 + x$:

$$f(0),\ f(f(0)),\ f(f(f(0))),\ \dots f^{<n>}(0), \dots$$

This is closely linked to the definition of the Mandelbrot[5] set (see Figure 2.1).

3. More combinatorial structures counted by Catalan numbers

In his book *Catalan Numbers*, Richard Stanley presents 214 different families of structures counted by Catalan numbers. On top of those already mentioned, some of the best known are discussed below.

[4] Recall that two sets have the same number of elements if and only if there exists a bijection between them.

[5] Benoit Mandelbrot (1924–2010).

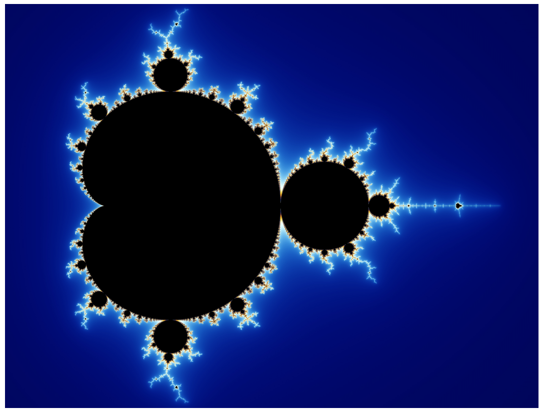


FIG. 2.1. The Mandelbrot set.

3.1. **Dyck paths.** A **southeast lattice path** is a sequence of points $(a_k, b_k) \in \mathbb{N}^2$, for $0 \leqslant k \leqslant \ell$, such that

$$(a_k, b_k) = \begin{cases} (a_{k-1}, b_{k-1}) + (0, -1) & \text{or,} \\ (a_{k-1}, b_{k-1}) + (1, 0). \end{cases}$$

The path is said to start at (a_0, b_0) and end at (a_ℓ, b_ℓ). A **Dyck path** of size n is southeast path going from $(0, n)$ to $(n, 0)$, that stays below the **diagonal** joining $(0, n)$ to $(n, 0)$. In other terms, for every point on the path, we have $a_k + b_k \leqslant n$. The number of size n Dyck paths is the Catalan number. The 14 Dyck paths of size 4 appear in Figure 3.1.

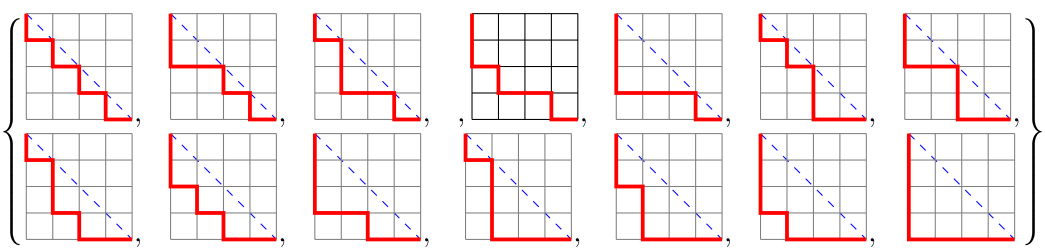


FIG. 3.1. Dyck paths of size 4

3.2. **Balanced words.** A length n **word** $\alpha = a_1 a_2 \cdots a_n$, is a sequence of **letters**. These letters are simply the name that is given to elements of a given set A, which is called the **alphabet**. We consider here the alphabet $A = \{0, 1\}$. We write $|\alpha|_0$ (resp. $|\alpha|_1$) for the number of copies of the letter 0 (resp.1) in a word α. For $k \leqslant n$, we denote by α_k the **initial factor** of length k of α, *i.e.* $\alpha_k = a_1 a_2 \cdots a_k$. With these notations at hand, we may define the notion of **balanced word**. This is a word α in which there are as many 0's and 1's, and such that

$$|\alpha_k|_0 \geqslant |\alpha_k|_1, \qquad \text{for all} \qquad 1 \leqslant k \leqslant n.$$

The number of balanced words of length $2n$ is the Catalan number. For example, here are the 14 balanced words of length 8:

$$\{01010101,\ 00110101,\ 01001101, 00101101, 00011101, 01010011, 00110011,$$
$$01001011, 00101011, 00011011, 01000111, 00100111, 00010111, 00001111\}.$$

3.3. **Binary trees.** Recall that a simple **graph** is a pair of sets (V, E), with V any finite set whose elements are called **vertices**, and $E \subset \{\{a, b\} \mid a, b \in V,\ a \neq b\}$. The elements of E, which are called **edges**, are thus some of the two-element subsets of V. If $\{a, b\}$ is an edge, then a and b are said to be **neighbours**. A length n **path** in a graph is a sequence of vertices $(a_0, a_1, \ldots, a_n)$, such that (a_k, a_{k+1}) belongs to E. The graph (V, E) is said to be **connected** if there is a path going from any vertex to any other. It is said to be **acyclic** if there is at most one such **strict**[6] path between any two vertices.

A **Tree** is an acyclic and connected graph. It is said to be **rooted**, if one of its vertices (also called **nodes**) is "selected", and called the **root**. Hence, in a rooted tree there is one and only one (strict) path linking any vertex to the root. The length of this path is the **distance** of the vertex to the root. Thus, in a rooted tree, one may naturally orient from a to b any edge $\{a, b\}$, if b sits farther from the root than a. One says that b is a **child** of a.

A **labeled binary tree** is a rooted tree with the further condition that any node has at most two children. Furthermore, there is a left-right order on these children, and for nodes with only one child, this child may either be a left child or a right child. The family $\mathcal{B}(U)$ of binary trees with vertex set V may be recursively described as follows

- if $U = \varnothing$, there is but one binary tree which is denoted by 0;
- if not, elements of $\mathcal{B}(U)$ are of the form (r, G, D), with
  i) $r \in U$, which is called the root,
  ii) $G \in \mathcal{B}(V)$, for some subset V of $U \backslash \{r\}$,
  iii) $D \in \mathcal{B}(W)$, with W the remaining vertices, hence $U = \{r\} \cup V \cup W$ (disjoint union).

A size n **(unlabelled) binary tree** is the "shape" of such a tree on a set of n vertices. Informally, all labels are replaced by a $\bullet$. The number of unlabeled binary tree is the Catalan number. For instance, Figure 3.2 illustrates the 14 binary trees with 4 nodes, with the root being the topmost node (in red).

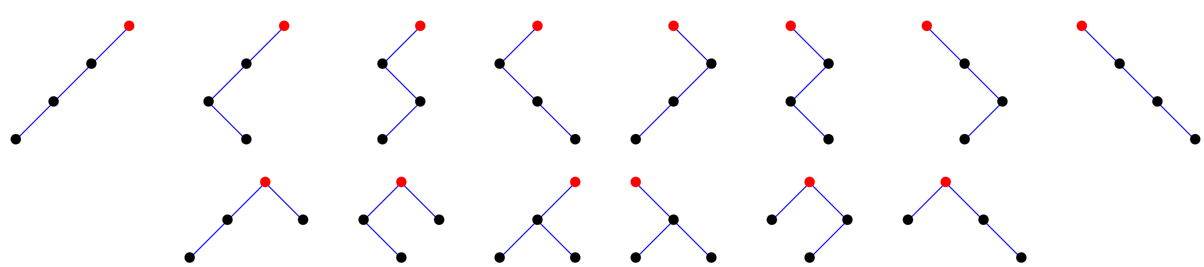


FIG. 3.2. The 14 binary trees of size 4.

3.4. **Excluded pattern permutations.** A **permutation** of the set $[n] = \{1, 2, \ldots, n\}$ may be described (in one line notation) as a rearrangement $a_1 a_2 \cdots a_n$ of the numbers from 1 to n. Hence, the two sets $\{1, 2, \ldots, n\}$ and $\{a_1, a_2, \ldots, a_n\}$ are the same. A permutation is said to **avoid the pattern** 312, if there does not exist $i < j < k$ such that $a_j < a_k < a_i$. The number of permutations of $[n]$ that avoid the pattern 312 is the Catalan number. The 14 such permutations for $n = 4$ are:

$$1234, 1243, 1324, 1342, 1432, 2134, 2143, 2314, 2341, 2431, 3214, 3241, 3421, 4321.$$

[6] One in which all vertices are different.

This is also true for permutations that avoid any length 3 pattern. However, the number of permutations that avoid a pattern of length 4 may depend on which permutation of 4 is chosen as a pattern (see wiki/Permutation_pattern).

3.5. **Standard Young tableaux.** A **two-row standard tableau** may be described as a $(2 \times n)$ matrix:

$$\begin{pmatrix} b_1 & b_2 & \cdots & b_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix},$$

such that $a_1 < a_2 < \ldots < a_n$, $b_1 < b_2 < \cdots < b_n$, and $a_k < b_k$ for all $1 \leqslant k \leqslant n$. Moreover, one requires that all the number from 1 to $2n$ occur exactly once in the matrix. The number of two-row standard tableaux with rows of length n is equal to the Catalan number. The 14 two-row standard tableaux for $n = 4$ are:

$$\begin{pmatrix} 2&4&6&8\\1&3&5&7 \end{pmatrix}, \begin{pmatrix} 3&4&6&8\\1&2&5&7 \end{pmatrix}, \begin{pmatrix} 2&5&6&8\\1&3&4&7 \end{pmatrix}, \begin{pmatrix} 3&5&6&8\\1&2&4&7 \end{pmatrix}, \begin{pmatrix} 4&5&6&8\\1&2&3&7 \end{pmatrix}, \begin{pmatrix} 2&4&7&8\\1&3&5&6 \end{pmatrix}, \begin{pmatrix} 3&4&7&8\\1&2&5&6 \end{pmatrix}$$

$$\begin{pmatrix} 2&5&7&8\\1&3&4&6 \end{pmatrix}, \begin{pmatrix} 3&5&7&8\\1&2&4&6 \end{pmatrix}, \begin{pmatrix} 4&5&7&8\\1&2&3&6 \end{pmatrix}, \begin{pmatrix} 2&6&7&8\\1&3&4&5 \end{pmatrix}, \begin{pmatrix} 3&6&7&8\\1&2&4&5 \end{pmatrix}, \begin{pmatrix} 4&6&7&8\\1&2&3&5 \end{pmatrix}, \begin{pmatrix} 5&6&7&8\\1&2&3&4 \end{pmatrix}$$

4. Polynomial count

One may shed more light on specific properties of elements of a family $\mathcal{F}_n$, via a "weighted" enumeration involving some associated **measure**, $\omega : \mathcal{F}_n \to \mathbb{N}$, on the combinatorial structures considered. As it happens, the distribution of this measure may be nicely described via the polynomial

$$F_n(x) := \sum_{f \in \mathcal{F}_n} x^{\omega(f)}, \tag{4.1}$$

in a variable x. Clearly, the coefficient of x^k is the number of elements in $\mathcal{F}_n$ having ω-measure equal to k. Setting $x = 1$, we evidently get back the cardinality of $\mathcal{F}_n$, *i.e.* $F_n(1) = \mathrm{card}(\mathcal{F}_n)$. Observe in particular that[7]

$$\frac{F_n'(1)}{F_n(1)} = \frac{1}{\mathrm{card}(\mathcal{F}_n)} \sum_{f \in \mathcal{F}_n} \omega(f),$$

gives the **mean** value of $\omega(f)$, as f runs over all elements of $\mathcal{F}_n$. One may also calculate the **variance** of ω in a similar manner (with a formula that also involves the second derivative of $F_n(x)$).

Let us consider the family $\mathcal{F}_n$ of Ferrers' diagrams (see Appendix B) of partitions contained in the staircase δ_n. For instance, we have the cardinality 14 set of Ferrers' diagrams: For a given

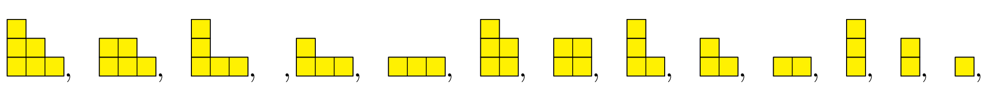
$\mathcal{F}_4 := \{$  $\varnothing. \}.$

Fig. 4.1. Ferrers' diagram of size 4 partitions.

Ferrers' diagram μ, considered as a sub-diagram of δ_n, the **area** of μ is by definition:

$$\mathrm{area}(\mu) := |\delta_n| - |\mu|,$$

which is simply the number of boxes of δ_n which lie outside μ. We then consider the q-polynomial:

$$C_n(q) := \sum_{\mu \in \mathcal{C}_n} q^{\mathrm{area}(\mu)} \tag{4.2}$$

[7]As is usual, $F_n'(x)$ stands for the derivative of $F_n(x)$.

In this way, we get a **q-analog**[8]. The first of these are:

$$
\begin{aligned}
C_0(q) &= 1,\\
C_1(q) &= 1,\\
C_2(q) &= q+1,\\
C_3(q) &= q^3+q^2+2q+1,\\
C_4(q) &= q^6+q^5+2q^4+3q^3+3q^2+3q+1,\\
C_5(q) &= q^{10}+q^9+2q^8+3q^7+5q^6+5q^5+7q^4+7q^3+6q^2+4q+1,\\
C_6(q) &= q^{15}+q^{14}+2q^{13}+3q^{12}+5q^{11}+7q^{10}+9q^9+11q^8+14q^7+16q^6\\
&\qquad +16q^5+17q^4+14q^3+10q^2+5q+1.
\end{aligned}
$$

One can prove that these polynomials satisfy the recurrence:

$$C_n(q) = \sum_{k=0}^{n-1} q^k C_k(q) C_{n-1-k}(q), \qquad \text{with} \qquad C_0(q) = 1. \tag{4.3}$$

An alternate q-analog[9] for Catalan numbers is

$$B_n(q) := \frac{1}{[n+1]_q} \begin{bmatrix} 2n \\ n \end{bmatrix}. \tag{4.4}$$

Checking that this is always a polynomial, and even that its coefficients are all positive integers, is not immediate. See Appendix C for notation.

Appendix A. Integer Sequences

Not only do Integer sequences play an interesting role in research, they may also serve as an interesting tools to find hard to reach information, and unexpected new links between subjects. A nice paper in which these notions are explored is *Fingerprint Databases for Theorems* by Sara C. Billey and Bridget E. Tenner.

Probably the first to explicitly exploit this idea was Neil Sloane in his book *Handbook of Integer Sequences* published in 1973. It compiled 2300 sequences that had occurred in scientific literature, together with detailed references to the papers in which these sequences had appeared. It rapidly became an essential research tool for anyone interested in integer sequences. In 1995, with the help of Simon Plouffe, Sloane published a greatly expanded version called the *Encyclopedia of Integer Sequences*, containing 5487 sequences. Since this coincided with the beginning of web based data banks, it was natural to turn the book into an online version, now well known as the *Online Encyclopedia of Integer Sequences*, *a.k.a.* OEIS.

In parallel, tools to automatise the discovery of formulas or recurrences for sequences became a stable of computer algebra systema. Among the early ones was "GFUN", included as a part of Maple, which grew out of the paper [4]. See: here for the current version. Other systems, such as Findstat have expanded these ideas in other direction, including tools for exploring and discovering bijections between families of combinatorial structures.

Appendix B. Partitions and Diagrams

[8] A polynomial in q, see Appendix C of the Catalan number

[9] There may be several interesting different q-analogs for a given sequence of numbers.

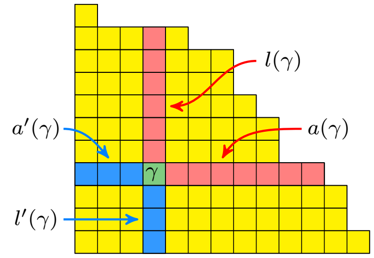


FIG. B.1. Arm, co-arm, leg, and co-leg.

As usual **integers partitions** (or simply **partitions**) are decreasing sequences $\mu = \mu_1\mu_2\cdots\mu_\ell$ of integers $\mu_i \geqslant 0$. We denote by ε, the unique **empty** partition of 0. The μ_i's are the **parts** of μ, and $|\mu| := \mu_1 + \mu_2 + \ldots + \mu_\ell$ is its **size**. When $|\mu| = m$, μ is said to be a partition of m, and this is denoted by $\mu \vdash m$. The **length** $\ell(\mu) = \ell$ of μ is the number of its non-zero parts. We write $k \in \mu$ when $k = \mu_i$ is a part of μ. One denotes by $\eta(\mu)$ the integer $\sum_j (j-1)\mu_j$. A **diagram**[10] is any finite subset $\boldsymbol{d}$ of $\mathbb{N}^+ \times \mathbb{N}^+$. Its elements are called **cells**, hence the usual set notation $\gamma \in \boldsymbol{d}$ states that γ is a cell of $\boldsymbol{d}$. The **content** of a cell $\gamma = (i,j)$, is the difference $c(\gamma) := i - j$. The **(Ferrers) diagram** of a partition μ is the set of cells (i,j), for which $1 \leqslant j \leqslant k$, and $1 \leqslant i \leqslant \mu_j$. Most often, we denote the same way both a partition and its diagram. The cell $\gamma = (i,j)$ is said to sit in **column** i and **row** j of (the diagram of) μ. Clearly partitions of size m have a total of m cells in their diagram. The **arm** (resp. **co-arm**) of a cell $\gamma = (i,j)$ is the number $a(\gamma) = a_\mu(\gamma) := \mu_j - i$ of cells that sit to the right of γ (resp. left) on the same row (resp. $a'(\gamma) = a'_\mu(\gamma) := i - 1$). The **leg** (resp. **co-leg**) of γ is the number $l(\gamma) = l_\mu(\gamma) := \mu'_i - j$ of cells that sit above γ (resp. below) in the same column (resp. $l'(\gamma) = l'_\mu(\gamma) := j - 1$). Assuming that partitions are padded with infinitely many 0-parts, we sometimes extend the notion of arm and leg to cells lying outside the partition, in which case they take negative values. The **hook length**, of a cell $\gamma = (a,b)$ in μ, is defined as $\varepsilon(\gamma) = \varepsilon_\mu(\gamma) := 1 + a_\mu(\gamma) + l_\mu(\gamma)$. The **conjugate** μ', of a partition μ, is the partition whose diagram is $\mu' = \{(j,i) \,|\, (i,j) \in \mu\}$. We sometimes say that the cell (j,i) is the conjugate of the cell (i,j). Considering cells as "vectors" in $\mathbb{N}\times\mathbb{N}$, we set $(\eta(\mu), \eta(\mu')) = \sum_{(i,j)\in\mu}(i,j)$. In particular, we have the **staircase partition** $\delta_n = (n-1, n-2, \ldots, 2, 1)$.

APPENDIX C. q-ANALOGS

The notion of **q-analog** goes back to the 19th-century. The idea is to replace integers by polynomials in q, that play a role that is "analogous" to the a corresponding numbers. By assumption the polynomials have positive integer coefficient, and specialize to the corresponding number when the variable q is set to 1. One of the aim is to find formulas/identities for the polynomials obtained, that "mimick" formulas/identities for the corresponding numbers. Note that a number may have different analogs, depending on the "role" it plays. For instance, the number 6 affords (at least) the two different q-analogs:

- $q^5 + q^4 + q^3 + q^2 + q + 1$, and
- $q^3 + 2q^2 + 2q + 1 = (q^2 + q + 1)(q + 1)$.

A first systematic way to get a q-analog of an integer n is to consider the polynomial

$$q^{n-1} + q^{n-1} + \ldots + q + 1,$$

which is usually denoted by $[n]_q$. For sure $[1]_q = 1$, and one sets $[0]_q := 0$. Observe that,

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

A simplistic identity is then

$$[n+k]_q = [n]_q + q^n [k]_q.$$

[10]Naturally drawn using Cartesian coordinates, rather than the slightly awkward matrix coordinates. As it should $\mathbb{N}$ contains 0, and $\mathbb{N}^+$ does not.

One may introduce a q-analog for $n!$ by setting

$$n!_q := [n]_q\,[n-1]_q\,\cdots[2]_q\,[1]_q,$$

with $0!_q := 1$. With these at hand, the next step is to "construct" q-analogs of binomial coefficients by setting:

$$\begin{bmatrix} n \\ k \end{bmatrix} := \frac{n!_q}{k!_q(n-k)!_q}.$$

Although it is true, it is not clear at first glance that these are indeed polynomials, and that they have positive integer coefficients. One way to see this is to show that they satisfy a q-version of the Pascal triangle rule:

$$\begin{bmatrix} n \\ k \end{bmatrix} = q^k \begin{bmatrix} n-1 \\ k \end{bmatrix} + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}, \qquad \text{with} \qquad \begin{bmatrix} n \\ n \end{bmatrix} = \begin{bmatrix} n \\ 0 \end{bmatrix} = 1.$$

There are many other interesting formulas of this nature.

One way to deal with q-analogs[11] is to exploit the decomposition of $[n]_q$ into "irreducible" polynomials, called **cyclotomic** and denoted $\varphi_n(q)$ (there is exactly one such for each integer $n \geqslant 2$). The cyclotomic polynomials play a role analogous to prime numbers. However, one must be careful here because they do not have all of their coefficients positive. Their "defining" property is that

$$[n]_q = \prod_{d\,|\,n} \varphi_d(q), \tag{C.1}$$

where d runs over the divisors of n (different from 1, but including n itself). For instance,

$$[6]_q = \varphi_2(q)\varphi_3(q)\varphi_6(q). \tag{C.2}$$

As it happens, one may recursively calculate $\varphi_n(q)$ using Equation C.1, as

$$\varphi_n(q) = \frac{[n]_q}{\prod_{d\,|\,n\ d \neq n} \varphi_d(q)}. \tag{C.3}$$

The first few values (which are also already defined in Sage) of the cyclotomic polynomials are:

$$\begin{aligned}
\varphi_2(q) &= q + 1, \\
\varphi_3(q) &= q^2 + q + 1, \\
\varphi_4(q) &= q^2 + 1, \\
\varphi_5(q) &= q^4 + q^3 + q^2 + q + 1, \\
\varphi_6(q) &= q^2 - q + 1, \\
\varphi_7(q) &= q^6 + q^5 + q^4 + q^3 + q^2 + q + 1, \\
\varphi_8(q) &= q^4 + 1, \\
\varphi_9(q) &= q^6 + q^3 + 1, \\
\varphi_{10}(q) &= q^4 - q^3 + q^2 - q + 1.
\end{aligned}$$

In particular, one may directly check that Equation C.2 holds. As is apparent here, it follows from the definition that for n a prime number, the cyclotomic polynomial is simply $[n]_q$ itself.

Appendix D. Generalizations

[11] Which may actually be efficiently exploited to find formulas with a computer.

For any pair of positive real numbers (r, s), consider the partition $\tau_{rs} := t_1 t_2 \cdots t_n$ having parts equal to $t_j := \lfloor r - j\, r/s \rfloor$. Then, τ_{rs} is said to be a **triangular partition**. In other words, the $\triangle$-partition τ_{rs} contains all the cells $\gamma = (i, j)$ lying below the line joining $(0, s)$ to $(r, 0)$, hence

$$(i, j) \in \tau_{r,s} \qquad \text{iff} \qquad \frac{i}{r} + \frac{j}{s} < 1.$$

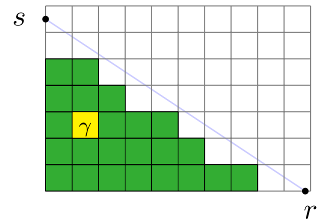


Fig. D.1. Triangular partition τ_{rs}.

Many different pairs (r, s) may give rise to a given $\triangle$-partition τ. Indeed, as illustrated in Figure D.2, we have an equality $\tau = \tau_{rs} = \tau_{r's'}$ when there are no positive integer coordinate points outside the partition[12] between the lines $x/r + y/s = 1$ and $x/r' + y/s' = 1$. Sometimes, but not always, there exist integers m and n in $\mathbb{N}$ such $\tau = \tau_{mn}$.

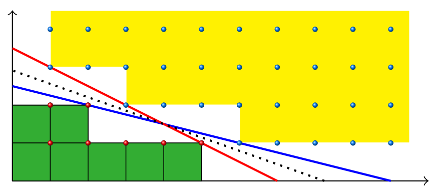


Fig. D.2. Many lines characterize the same triangular partition.

The beginning of the sequence of number of triangular partitions of size n is

$$1, 1, 2, 3, 4, 6, 7, 8, 10, 12, 13, 16, 16, 18, 20, 23, \ldots$$

for $m \geqslant 0$. Table D.1 displays all triangular partitions of $m \leqslant 6$.

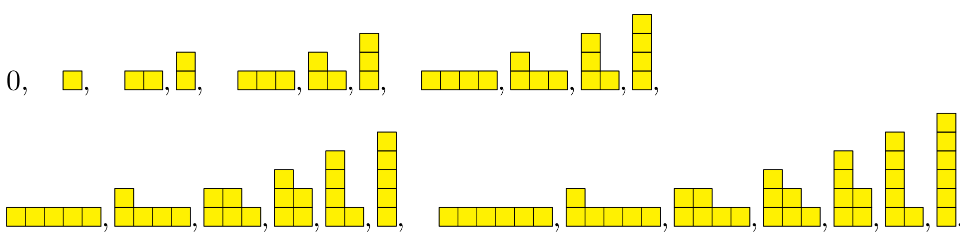


Table D.1. All $\triangle$-partitions, for $n \leqslant 6$.

D.1. **Triangular Dyck Paths.** For any triangular partition τ, we define the set of **τ-Dyck paths** as $\mathcal{D}_\tau := \{\alpha \mid \alpha \subseteq \tau\}$. Observe that conjugation gives a bijection between $\mathcal{D}_\tau$ and $\mathcal{D}_{\tau'}$. We further write $\mathcal{D}_{(m,n)}$, when $\tau = \tau_{m,n}$ with m and n integers. As we have already mentioned, classical Dyck paths correspond to the case $\tau = \delta_n$, hence when $m = n$. As we will see, the case when the greatest common divisor of m and n is 1 is of particular interest. We then say that m and n are **coprime**. The **ν-area** of a τ-Dyck path α is the number of cells lying in the skew shape ν/α, hence $\mathrm{area}_\tau(\alpha) := |\tau| - |\alpha|$.

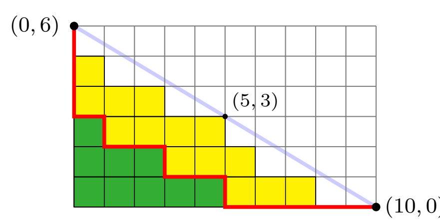


Fig. D.3. The $\tau_{(10,6)}$-Dyck path 531000.

[12]Including points that could lie on the lines.

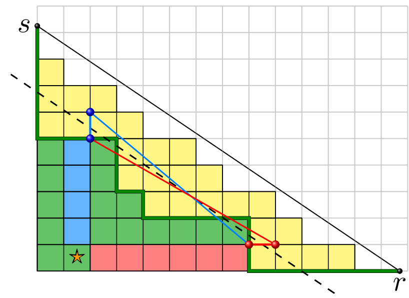


FIG. D.4. A dinv cell

For a τ-Dyck path α, we set $\mathrm{dinv}_\tau(\alpha)$ to be equal to the cardinality of the set of **τ-dinv-cells** (or simply dinv-cells) of α:

$$\mathrm{Dinv}_\tau(\alpha) := \left\{\gamma \in \alpha \,\middle|\, \tfrac{l(\gamma)}{a(\gamma)+1} < \overline{\sigma}_\tau + \varepsilon < \tfrac{l(\gamma)+1}{a(\gamma)}\right\},$$

where any division by 0 gives $+\infty$. Here $\varepsilon > 0$ is some small (enough) irrational value, that removes possible ambiguities in inequalities. The respective dinv-cells of the partitions contained in $\tau = 32$ are marked by stars in Figure D.5.

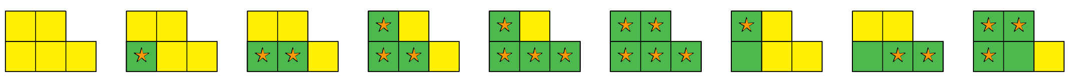


FIG. D.5. Dinv-cells for subpartitions of 32. The first 6 subpartitions have full-dinv.

D.2. **Counting Triangular Dyck Paths.** The enumeration of triangular Dyck paths has an interesting (ongoing) history. The simple counting in the case of any integer pairs (m, n) had already been worked out in the 1950s (see [**?**]). However, it is only rather recently that the overall enumerative combinatorics community has become aware of that fact. For a long time, only the "coprime case" was deemed really understood. Sample values for the integral case of **(m, n)-Catalan numbers** $\mathcal{C}_{mn}$ are given in Table D.2. These include the classical **Fuss-Catalan** numbers when

m/n	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	2	2	3	3	4	4
3	1	2	5	5	7	12	12
4	1	3	5	14	14	23	30
5	1	3	7	14	42	42	66
6	1	4	12	23	42	132	132
7	1	4	12	30	66	132	429

TABLE D.2. Rectangular Catalan numbers $\mathcal{C}_{rs}$.

$m = kn$ (or equivalently when $m = kn + 1$): $\mathcal{C}_{kn,n} = \mathcal{C}_{kn+1,n} = \frac{1}{kn+1}\binom{(k+1)n}{n}$. This formula is a special case of the following lemma. Indeed, it is a direct extension of the classical "cycling" argument proving that the number Dyck paths corresponds to the Catalan numbers.

Lemma D.1. *For m and n coprime integers, we have* $\mathcal{C}_{mn} = \frac{1}{m+n}\binom{m+n}{n}$.

k/n	0	1	2	3	4	5	6	7	8	9
1	1	1	2	5	14	42	132	429	1430	4862
2	1	1	3	12	55	273	1428	7752	43263	246675
3	1	1	4	22	140	969	7084	53820	420732	3362260
4	1	1	5	35	285	2530	23751	231880	2330445	23950355
5	1	1	6	51	506	5481	62832	749398	9203634	115607310
6	1	1	7	70	819	10472	141778	1997688	28989675	430321633

TABLE D.3. Fuss-Catalan numbers $\mathcal{C}_{kn,n}$.

1												
2												
1^2	2											
3	3											
1^3	21	3										
4	5	4										
1^4	21^2	31	4									
5	7	7	5									
1^5	21^3	2^21	32	41	5							
6	9	9	9	9	6							
1^6	21^4	2^21^2	321	42	51	6						
7	11	12	14	12	11	7						
1^7	21^5	2^21^3	3211	421	52	61	7					
8	13	15	19	19	15	13	8					
1^8	21^6	2^21^4	2^31^2	32^21	431	53	62	71	8			
9	15	18	18	23	23	18	18	15	9			
1^9	21^7	2^21^5	2^31^3	32^21^2	3^221	432	531	63	72	81	9	
10	17	21	2^2	30	28	28	30	2^2	21	17	10	
1^{10}	21^8	2^21^6	2^31^4	32^21^3	3^221^2	4321	532	631	73	82	91	(10)
11	19	24	26	37	37	42	37	37	26	24	19	11

TABLE D.4. Number of τ-Dyck paths

For all triangular partitions τ of size at most 10, the number of τ-Dyck paths are given in the Table D.4. A formula for the non-coprime case of τ_{mn} is described in the next subsection.

D.3. **Bizley formula.** In the general "integral" situation, the enumeration formula of $(m \times n)$-Dyck paths takes the form of a sum of terms indexed by partitions of the greatest common divisor of m and n. This is what makes it harder to "guess" a formula[13], since the numbers obtained do not factor nicely in general, even if they are effectively sums of nicely factorized numbers. The **Grossman-Bizley** formula (see [**?**]) is:

$$\mathcal{C}_{mn} := \sum_{\mu \vdash d} \frac{1}{z_\mu} \prod_{k \in \nu} \frac{1}{a+b} \binom{k(a+b)}{ka}, \tag{D.1}$$

where $(m, n) = (ad, bd)$ with a and b coprime, so that $d = \gcd(m, n)$. It is worth recalling that $d!/z_\mu$ is the number of size d permutations of cycle type μ, where

$$z_\mu := \prod_k k^{d_k}\, d_k!.$$

[13]Most guessing approaches rely (directly or indirectly) on the fact the numbers considered have nice factorization in small prime numbers.

Here, μ has d_i parts of size i. Specific examples of (D.1) are:

$$\begin{aligned}\mathcal{C}_{2a,2b} &= \tfrac{1}{2}\left(\tfrac{1}{a+b}\tbinom{a+b}{a}\right)^2 + \tfrac{1}{2}\left(\tfrac{1}{a+b}\tbinom{2a+2b}{2a}\right),\\ \mathcal{C}_{3a,3b} &= \tfrac{1}{6}\left(\tfrac{1}{a+b}\tbinom{a+b}{a}\right)^3 + \tfrac{1}{2}\left(\tfrac{1}{a+b}\tbinom{a+b}{a}\right)\left(\tfrac{1}{a+b}\tbinom{2a+2b}{2a}\right) + \tfrac{1}{3}\left(\tfrac{1}{a+b}\tbinom{3a+3b}{3a}\right).\end{aligned} \tag{D.2}$$

Observe that, for fixed coprime numbers a and b, all the formulas for $(m, n) = (ad, bd)$, with $0 \leqslant d$, may be presented in the form of the generating function:

$$\sum_{d=1}^{\infty} \mathcal{C}_{ad,bd}\, x^d = \exp\Big(\sum_{k\geqslant 1} \tfrac{1}{a+b}\tbinom{k(a+b)}{ka}\, x^k/k\Big). \tag{D.3}$$

D.4. **Area enumerator.** For any triangular partition τ, the q-area enumerator of τ-Dyck paths, is

$$\mathcal{C}_\tau(q) := \sum_{\alpha\subseteq\tau} q^{\mathrm{area}_\tau(\alpha)}. \tag{D.4}$$

Since conjugation is an area-preserving bijection between the set of τ-Dyck paths and the set of τ'-Dyck paths, we clearly have $\mathcal{C}_\tau(q) = \mathcal{C}_{\tau'}(q)$. The q-area enumerators for all triangular partitions of size at most 6 are as follows (avoiding repetitions for conjugate partitions):

$$\begin{array}{llll}
\mathcal{C}_1 = & q+1, & \mathcal{C}_2 = & q^2+q+1,\\
\mathcal{C}_{21} = & q^3+q^2+2q+1, & \mathcal{C}_3 = & q^3+q^2+q+1,\\
\mathcal{C}_{31} = & q^4+q^3+2q^2+2q+1, & \mathcal{C}_4 = & q^4+q^3+q^2+q+1,\\
\mathcal{C}_{32} = & q^5+q^4+2q^3+2q^2+2q+1, & \mathcal{C}_{41} = & q^5+q^4+2q^3+2q^2+2q+1,\\
\mathcal{C}_5 = & q^5+q^4+q^3+q^2+q+1, & \mathcal{C}_{321} = & q^6+q^5+2q^4+3q^3+3q^2+3q+1,\\
\mathcal{C}_{42} = & q^6+q^5+2q^4+2q^3+3q^2+2q+1, & \mathcal{C}_{51} = & q^6+q^5+2q^4+2q^3+2q^2+2q+1,\\
\mathcal{C}_6 = & q^6+q^5+q^4+q^3+q^2+q+1. & &
\end{array}$$

D.5. **Counting by Area and Dinv.** We now consider enumeration of τ-Dyck paths with respect to both statistics: "area" and "dinv", in the triangular case. The resulting (q, t)-polynomials:

$$\mathcal{C}_\tau(q,t) := \sum_{\alpha\subseteq\tau} q^{\mathrm{area}_\tau(\alpha)} t^{\mathrm{dinv}_\tau(\alpha)}, \tag{D.5}$$

plays a central role in a wide range of subjects. Finding formulas for these is not easy, and many properties are still being studied. In the case of τ_{ab}, with a and b coprime integers, we have the nice q-analog (obtained via setting $t = 1/q$)

$$q^{|\tau_{ab}|}\mathcal{C}_{\tau_{ab}}(q, 1/q) = \frac{1}{[a+b]_q}\begin{bmatrix} a+b \\ a \end{bmatrix}. \tag{D.6}$$

Recall that, for all k and n we have $\tau_{kn,n} = \tau_{kn+1,n}$, hence the above also covers such cases.

Appendix E. Problems

Problem 1. Directly prove that at least one of the following family of structures is counted by Catalan numbers: triangulations of n-gons, subpartitions of the staircase δ_n, Dyck paths, balanced words, binary trees, permutations avoiding 321, two equal row standard tableaux. ■

Problem 2. Give explicit bijections between size n structures of the following Catalan numbers enumerated structures: triangulations of n-gons, subpartitions of the staircase δ_n, Dyck paths, balanced words, binary trees, permutations avoiding 321, two equal row standard tableaux. Conclude that all of them are counted by the Catalan numbers. ■

Problem 3. Prove that the recurrence of Equation 2.2 holds, as well as that of Equation 4.3. Exploiting the fact that two series are equal if and only their coefficients are the same, prove Equation 2.5 by comparing coefficients on both sides of the equation. Deduce from this Formula (2.6) (being careful to give an explanation for the sign). ■

Problem 4. In each case, directly prove that the number of following family of structures satisfies the recurrence of Equation 2.2: triangulations of n-gons, subpartitions of the staircase δ_n, Dyck paths, balanced words, binary trees, permutations avoiding 321, two equal row standard tableaux. Conclude that all of them are counted by the Catalan numbers. ■

Problem 5. Assuming that there exist integer coefficient polynomials $\varphi_d(q)$ such that

$$q^n - 1 = \prod_{d\,|\,n} \varphi_d(q),$$

for all n, check for small values that $B_n(q)$ defined by Equation 4.4 is indeed a polynomial. Using the values provided for these in Appendix C, also verify that the result is indeed a positive integer polynomial. Generalize if you can. ■

Problem 6. As a challenge, prove Lemma D.1. ■

References

[1] **Advanced Book**: *Algebraic Combinatorics and Coinvariant Spaces*, F. Bergeron, CMS Treatise in Mathematics, CMS and A.K.Peters, Monograph, 2009. (On page 1)

[2] **Survey**: *Combinatorics of r-Dyck paths, r-Parking functions, and the r-Tamari lattices*, F. Bergeron, 2012, 36 pages. arxiv.org/abs/1202.6269. (On page 1)

[3] **Advanced Lecture Notes**: *Symmetric Functions and Combinatorics*, F. Bergeron, 65 pages, 2018. http://bergeron.math.uqam.ca/wp-content/uploads/2018/12/Symmetric-Functions.pdf (On page 1)

[4] **Paper**: *Computing the generating function of a series given its first few terms*, F.Bergeron and S. Plouffe, Experimental Mathematics, Vol. 1, no. 4, (1992) 307-312. (On page 7)

[5] **Paper**: *Fingerprint Databases for Theorems*, S. C. Billey and B. E. Tenner, Notices of the AMS Volume 60, Number 8, 1034–1039, 2013.

[6] **Book**: *Catalan Numbers with Applications*, T. Koshi, Oxford University Press, 2009. (On page 2)

[7] **Book**: *Encyclopedia of Integer Sequences*, N. Sloane and S. Plouffe, Academic Press, 1995, 587 pp.

[8] **Book**: *Catalan Numbers*, R.P. Stanley, Cambridge University Press, April 2015. doi.org/10.1017/CBO9781139871495 (On page 2)