Lecture 1. Two proofs of the formula $C_n = \frac{1}{n}\binom{2n-2}{n-1}$ for the Catalan numbers

M. Klazar

February 23, 2024

Let $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \{0, 1 \dots\}$. For $n \in \mathbb{N}$, the n-th *Catalan number* $C_n \in \mathbb{N}$ is the cardinality of the set

$$\mathcal{D}_n = \{(u_1, \dots, u_{2n-2}) \in \{-1, 1\}^{2n-2} \mid \forall\, m\, :\, \textstyle\sum_{i=1}^m u_i \ge 0 \wedge \sum_{i=1}^{2n-2} u_i = 0\}\ .$$

Note that any word $u \in \mathcal{D}_n$ has length $|u| = 2n - 2$ and has $n - 1$ ones and $n - 1$ minus ones. We set $\mathcal{D}_1 = \{\emptyset\}$. For example, $C_4 = |\mathcal{D}_4| = 5$ because

$$\mathcal{D}_4 = \{111000,\ 110100,\ 101100,\ 110010,\ 101010\}$$

where we write for brevity 0 instead of -1 and omit commas and brackets. The elements of the sets $\mathcal{D}_n$ are called *Dyck words.*

Theorem 1 *For every $n \in \mathbb{N}$ we have $C_n = \frac{1}{n}\binom{2n-2}{n-1}$.*

The first proof of Theorem 1. We consider the generating function (GF) $C = C(x) = \sum_{n=1}^{\infty} C_n x^n \in \mathbb{R}[[x]]$. Every nonempty Dyck word u has a unique decomposition

$$u = 1\, v\, (-1)\, w$$

where v and w are possibly empty Dyck words and $1v(-1)$ is the shortest initial segment of u with sum 0. Restricting the map $u \mapsto (v, w)$ to $\mathcal{D}_n$ we get a bijection

$$\mathcal{D}_n \to \bigcup_{i=1}^{n-1} \mathcal{D}_i \times \mathcal{D}_{n-i},\ n \ge 2\ .$$

Thus we have the *basic recurrence* that $C_1 = 1$ and for $n \ge 2$,

$$C_n = \sum_{i=1}^{n-1} C_i C_{n-i}\ .$$

In terms of $C(x)$ it means that $C(x) = x + C(x)^2$ and $C^2 - C + x = 0$. The quadratic formula yields two solutions

$$C = C(x) = \tfrac{1}{2}\big(1 \pm \sqrt{1 - 4x}\big)\ .$$

Newton's binomial theorem says that for any $\alpha \in \mathbb{R}$,

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n \in \mathbb{R}[[x]]$$

where for $n \geq 1$

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!}\ ,$$

and $\binom{\alpha}{0} = 1$. For any $\alpha, \beta \in \mathbb{R}$ it holds more generally that$(1+\beta x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} \beta^n x^n$. Thus

$$\sqrt{1-4x} = (1+(-4)x)^{1/2} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-4)^n x^n = 1 - 2x - 2x^2 - \ldots\ .$$

Hence

$$C(x) = \frac{1}{2}\left(1 - \sum_{n=0}^{\infty} \binom{1/2}{n} (-4)^n x^n\right) = x + x^2 + 2x^3 + \ldots$$

and

$$C_n = \underbrace{[x^n]C(x)}_{\text{the coefficient of } x^n \text{ in } C(x)} = -\frac{1}{2} \cdot (-4)^n \cdot \binom{1/2}{n}\ .$$

Certainly $C_1 = 1 = \frac{1}{1}\binom{0}{0}$. For $n \geq 2$ the number C_n equals

$$\begin{aligned} -\frac{1}{2} \cdot (-4)^n \cdot \binom{1/2}{n} &= 2^{2n-1} \cdot \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{5}{2} \cdot \ldots \cdot \frac{2n-3}{2}}{n!} \\ &= \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-3) \cdot 2^{n-1} \cdot (n-1)!}{n! \cdot (n-1)!} \\ &= \frac{1}{n} \cdot \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots \cdot (2n-2)}{(n-1)! \cdot (n-1)!} \end{aligned}$$

and we get the stated formula $C_n = \frac{1}{n}\binom{2n-2}{n-1}$. □

But this computation contains a gap which borders on a logical fallacy. In the equality $\sqrt{1-4x} = (1-4x)^{1/2}$ we have on the left side a formal power series (FPS) in $\mathbb{R}[[x]]$ with constant term 1 and such that its square equals $1-4x$. On the right side we have the FPS $P(x) = \sum_{n=0}^{\infty} \binom{1/2}{n} (-4)^n x^n$. What is missing is the proof that really

$$P(x)^2 = 1 - 4x\ .$$

We see that we need to show the equality

$$P(x)^2 = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} \binom{1/2}{k} \binom{1/2}{n-k} \right) (-4)^n x^n = 1 - 4x\ .$$

We get it from the next identity.

Theorem 2 (the Vandermonde convolution) *In the ring $\mathbb{R}[x, y]$ of bivariate real polynomials, the identity*

$$\sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k} = \binom{x+y}{n}$$

holds for every $n \in \mathbb{N}_0$.

Now we view binomial coefficients as rational polynomials: for $k \in \mathbb{N}$,

$$\binom{x}{k} = \frac{x(x-1)\dots(x-k+1)}{k!} \in \mathbb{Q}[x]$$

and $\binom{x}{0} = 1$. By Theorem 2,

$$\sum_{k=0}^{n} \binom{1/2}{k}\binom{1/2}{n-k} = \binom{\frac{1}{2}+\frac{1}{2}}{n} = \binom{1}{n} = \begin{cases} 1 & \dots \quad n = 0,\, 1 \text{ and} \\ 0 & \dots \quad n \geq 2 \end{cases}$$

and the equality $P(x)^2 = 1 - 4x$ follows. We deduce Theorem 2 by means of the next theorem.

Theorem 3 *Let $d \in \mathbb{N}_0$, $X, Y \subset \mathbb{R}$ with $|X| = |Y| = d+1$ and $F \in \mathbb{R}[x, y]$ be a nonzero polynomial with degree at most d. Then*

$$\exists\, (u,\, v) \in X \times Y :\; F(u,\, v) \neq 0\;.$$

Proof. We write

$$F(x,\, y) = x^{n_1}p_1(y) + x^{n_2}p_2(y) + \dots + x^{n_k}p_k(y)$$

where $d \geq n_1 > n_2 > \dots > n_k \geq 0$, $k \in \mathbb{N}$, $p_i \in \mathbb{R}[y]$ and every p_i is a nonzero polynomial with degree at most d. Since every nonzero (univariate) polynomial over a field with degree at most d has at most d roots, there exists a $v \in Y$ such that $p_1(v) \neq 0$. Then $G(x) = F(x, v) \in \mathbb{R}[x]$ is a nonzero polynomial with degree at most d and there exists a $u \in X$ such that $G(u) = F(u, v) \neq 0$. □

Proof of Theorem 2. For any $n \in \mathbb{N}_0$ we set

$$F_n = F_n(x,\, y) = \sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k} - \binom{x+y}{n} \in \mathbb{R}[x,\, y]\;.$$

If F_n is a nonzero polynomial then $\deg F_n \leq n$. It is not hard to see that $F_n(x, y) = 0$ for every $x, y, n \in \mathbb{N}_0$: if A and B are disjoint sets with cardinalities $|A| = x$ and $|B| = y$, then by counting the sets $C \subset A \cup B$ with $|C| = n$ in two ways we get that (for $x, y \in \mathbb{N}_0$)

$$\sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k} = \binom{x+y}{n}\;.$$

Using Theorem 3 we see that $F_n(x,y)$ is a zero polynomial. $\square$

Only now is the first proof of the formula $C_n = \frac{1}{n}\binom{2n-2}{n-1}$ complete.

Theorem 3 inspired me to give in the second part of the course a survey of the proof of the following remarkable theorem. For $N \in \mathbb{N}$ we set $[N] = \{1, 2, \dots, N\}$.

Theorem 4 (Bombieri–Pila) *Suppose that $d \in \mathbb{N}$ and that $F \in \mathbb{R}[x, y]$ is a nonzero irreducible polynomial with degree d. Then for any $N \in \mathbb{N}$,*

$$|\{(u,\, v) \in [N]^2 \mid F(u,\, v) = 0\}| \le \log(N+2)^{O(d)} \cdot N^{1/d}$$

(the shift $N+2$ removes the inconvenient values $\log 1 = 0$ and $\log 2 < 1$). The polynomial $F(x, y) = y^d - x$ shows that up to the logarithmic factor the bound is tight. For $d = 1$ we have the simple bound $|\dots| \le N$. In [2] the theorem was proven with the bound $|\dots| \le N^{1/d+o(1)}$. In [4] the term $N^{o(1)}$ was improved to the polylogarithmic factor. I will follow the survey [1], or maybe not, but will begin with the pioneering 1926 result of V. Jarník (1897–1970) [3] that if $\Gamma \subset [0, N]^2$, $N \in \mathbb{N}$, is the graph of a monotone and strictly convex or strictly concave function $f\colon [0, N] \to [0, N]$, then

$$\max_{\Gamma} |\Gamma \cap \mathbb{Z}^2| = 3\pi^{-2/3} N^{2/3} + O(N^{1/3} \log N)\ .$$

The second proof of Theorem 1. $C_1 = 1 = \frac{1}{1}\binom{0}{0}$ is trivial and we assume that $n \ge 2$. Besides the set $\mathcal{D}_n$ of Dyck words with length $2n-2$ we consider the sets of words

$$\mathcal{A}_n = \{u \in \{-1,\, 1\}^{2n-2} \mid \textstyle\sum u_i = 0\}$$

and

$$\mathcal{B}_n = \{u \in \{-1,\, 1\}^{2n-2} \mid \textstyle\sum u_i = 2\}$$

with the same length $2n-2$. Clearly, $\mathcal{D}_n \subset \mathcal{A}_n$. The formula for the Catalan numbers follows from the next proposition which shows that $C_n = |\mathcal{D}_n| = |\mathcal{A}_n| - |\mathcal{A}_n \setminus \mathcal{D}_n|$ equals

$$|\mathcal{A}_n| - |\mathcal{B}_n| = \binom{2n-2}{n-1} - \binom{2n-2}{n} = \left(1 - \frac{n-1}{n}\right)\binom{2n-2}{n-1} = \frac{1}{n}\binom{2n-2}{n-1}\ .$$

$\square$

Thus bijective combinatorics easily trumps generating functions, at least in the case of the formula $C_n = \frac{1}{n}\binom{2n-2}{n-1}$.

Proposition 5 *Let $n \ge 2$. The map*

$$f\colon \mathcal{A}_n \setminus \mathcal{D}_n \to \mathcal{B}_n,\ f(u) = v\ ,$$

where v arises from u by changing signs in the shortest initial segment of u with sum -1, is a bijection.

Proof. Let $n \geq 2$ and $u \in \mathcal{A}_n \setminus \mathcal{D}_n$. Thus u has $n-1$ ones and $n-1$ minus ones and not all initial sums are nonnegative. It follows that there is the shortest initial segment u' of u with sum -1. It has one more -1 than 1's, and in the rest of u it is the other way around. Thus if we change signs in u' we get the word $v = f(u) \in \mathcal{B}_n$ and this transformation turns u' in v' which is the shortest initial segment of v with sum 1. We transform v' in the same way and get $w = g(v) \in \mathcal{A}_n \setminus \mathcal{D}_n$; this defines the map $g\colon \mathcal{B}_n \to \mathcal{A}_n \setminus \mathcal{D}_n$. Clearly, $w = u$. The maps f and g are inverses of one another and f is a bijection. $\square$

In [5] R. P. Stanley describes very many families of structures counted by the Catalan numbers. To conclude we mention one striking "Catalanian" result due to P. Valtr in [6]. A *convex chain* is a finite set of points in the plane $\mathbb{R}^2$ such that the points lie on the graph of a strictly convex function. For $n \in \mathbb{N}$, $n \geq 3$, let U_n be the event that n random and independent points selected in the unit square $[0,1]^2$ form a convex n-gon, and V_n be the event that they form a convex chain. Then, by [6],

$$\Pr(V_n \mid U_n) = \frac{1}{C_n} \,.$$

For $n = 3$ it is clear that the conditional probability equals $\frac{1}{2}$, but for $n > 3$ it is far from clear why we get the reciprocal of the n-th Catalan number.

References

[1] T. F. Bloom and J. D. Lichtman, The Bombieri–Pila determinant method, arXiv:2312.12890v1, 2023, 17 pp.

[2] E. Bombieri and J. Pila, The number of integral points on arcs and ovals, *Duke Math. J.*, **59** (1989) 337–357

[3] "Vojtěch Jarník in Göttingen", Über die Gitterpunkte auf konvexen Kurven, *Math. Z.* **24** (1926), 500–518

[4] J. Pila, Density of integer points on plane algebraic curves, *IMRN*, **18** (1996) 903–912

[5] R. P. Stanley, *Catalan Numbers*, Cambridge University Press, Cambridge, UK 2015

[6] P. Valtr, Catalan numbers via random planar point sets, 441–443. In: I. Bárány and K. Böröczky (eds.), *Intuitive Geometry (Budapest, 1995)*, János Bolyai Mathematical Society, Budapest 1997