18.212 Algebraic Combinatorics

Spring 2021

Alex Postnikov

math.mit.edu/~apost/18.212/

Lecture 1. The Catalan Numbers

Definition. The n^{th} Catalan number C_n is the number of sequences $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{2n})$ with entries $\varepsilon_i = \pm 1$ such that

- there are n entries $\varepsilon_i = 1$ and there are n entries $\varepsilon_j = -1$.
- the partial sums $\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_i \geq 0 \quad \forall i$

Example. $n=3$. There are 5 such sequences:

$(1,1,1,-1,-1,-1)$, $(1,1,-1,1,-1,-1)$, $(1,1,-1,-1,1,-1)$

$(1,-1,1,1,-1,-1)$, $(1,-1,1,-1,1,-1)$.

So $C_3 = 5$.

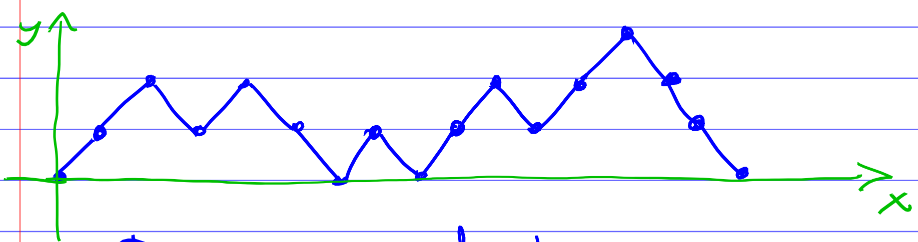
Graphically, such sequences $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{2n})$ are represented by Dyck paths so that

$\varepsilon_i = 1 \rightsquigarrow i^{th}$ step is "up"

$\varepsilon_j = -1 \rightsquigarrow j^{th}$ step is "down"

Example. $n=8$

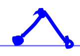
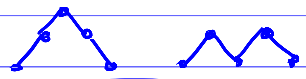
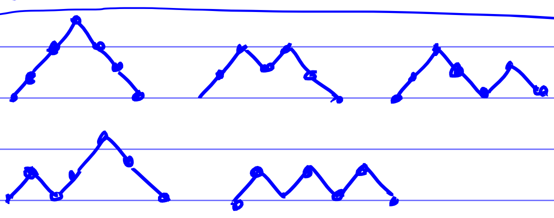
$(1,1,-1,1,-1,-1,1,-1,1,1,-1,1,1,-1,-1,-1)$



a Dyck path of length $2\cdot 8$

Each Dyck path

- starts at $(0,0)$ and ends at $(2n,0)$
- has n "up" steps and n "down" steps
- always stays weakly above the x-axis.

n	C_n	Dyck paths
0	1	•
1	1	
2	2	
3	5	
4	14	14 Dyck paths of length $2\cdot 4$
5	42	42 Dyck paths of length $2\cdot 5$

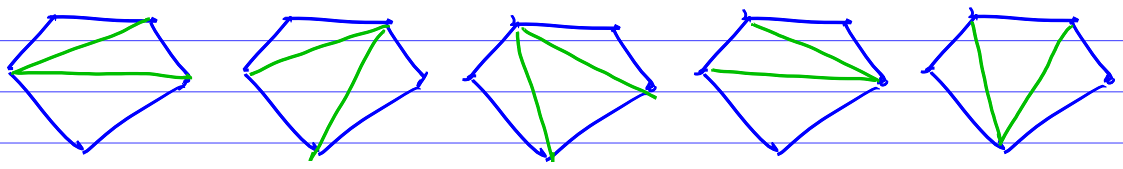
the Catalan numbers

There are many other combinatorial interpretations of the Catalan numbers C_n.

For example, C_n equals the number of triangulations of a convex $(n+2)$-gon.

A <u>triangulation</u> is a way to subdivide a polygon into triangles by non-crossing chords.

<u>Example</u>. 5 triangulations of a pentagon:



$$C_3 = 5$$

a bijection is a one-to-one correspondence

<u>Exercise</u>. Find a <u>bijection</u> between Dyck paths of length $2n$ and triangulations of an $(n+2)$-gon.

Is there an explicit formula for C_n?

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

Example. $C_4 = \frac{1}{5}\binom{8}{4} = \frac{8\cdot 7\cdot 6\cdot 5}{5\cdot 4\cdot 3\cdot 2\cdot 1}$

$$= 14.$$

There are many proofs of this formula for C_n.

We'll discuss several interesting proofs. But let me start with most straightforward way to prove it.

The following approach works for many different combinatorial sequences $A_0, A_1, A_2, \ldots$

- Find a recurrence relation for A_n, i.e. express A_n in terms of $A_{n-1}, A_{n-2}, \ldots, A_0$
- Use the ordinary generating funct.

$$A(x) := A_0 + A_1 x + A_2 x^2 + \ldots$$

(or the exponential generating funct. $A_0 + A_1 \frac{x}{1!} + A_2 \frac{x^2}{2!} + A_3 \frac{x^3}{3!} + \ldots$)

Rewrite the recurrence relation for A_n as an algebraic or differential equation for the generating function $A(x)$.

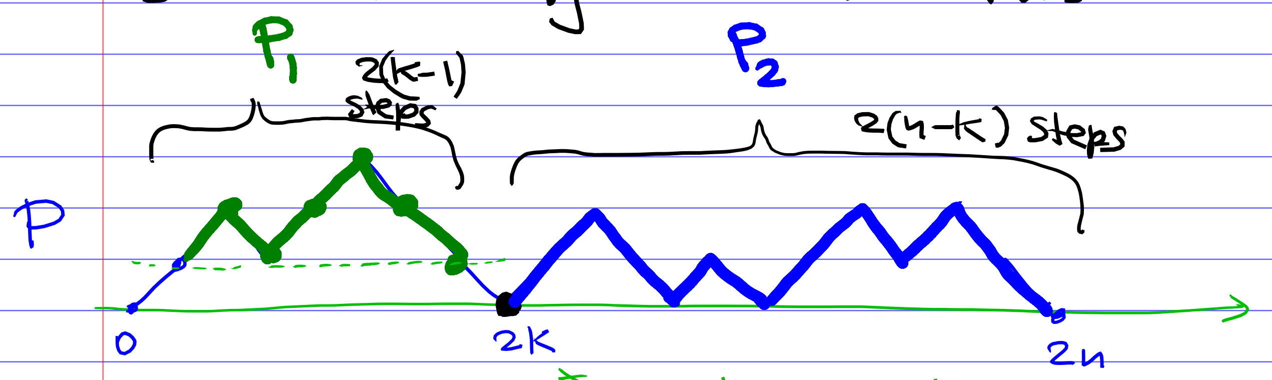
- Use algebra or analysis to solve the equation
- Deduce an explicit formula for A_n.

Usually, this is the most boring way to solve a combinatoria problem. But this is a very powerful techinque. So we'll discuss how it works for the Catalan numbers C_n.

- Recurrence relation

A recurrence relation describes a way to decompose some combinatorial objects into smaller objects of the same kind.

Let's see how this works for Dyck paths. We want to break a Dyck paths into smaller Dyck paths. Here is one way to do this



this is the first point on the x-axis where path P touches it.

$k \in \{1, 2, \dots, n\}$

$$P \rightsquigarrow P_1 \quad , \quad P_2$$

a Dyck path of length $2n$ — a Dyck path of length $2(k-1)$ — a Dyck path of length $2(n-k)$.

- k can be any number $1, 2, \dots, n$
- P_1 can be any Dyck path of length $2(k-1)$
- P_2 can be any Dyck path of length $2(n-k)$.

- This decomposition of a Dyck path P into two shorter Dyck paths P_1 and P_2 is unique.

So we get

Recurrence Relation for C_n:

$$C_n = \sum_{k=1}^{n} C_{k-1} \cdot C_{n-k} \quad \text{for } n \geq 1$$

$$C_0 = 1$$

Examples. $C_1 = C_0 \cdot C_0 = \boxed{1}$

$C_2 = C_0 \cdot C_1 + C_1 \cdot C_0 = \boxed{2}$

$C_3 = C_0 C_2 + C_1 \cdot C_1 + C_2 \cdot C_0$
$= 1 \cdot 2 + 1 \cdot 1 + 2 \cdot 1 = \boxed{5}$

$C_4 = C_0 C_3 + C_1 C_2 + C_2 C_1 + C_3 C_0$
$= 1 \cdot 5 + 1 \cdot 2 + 2 \cdot 1 + 5 \cdot 1$
$= \boxed{14}$

etc.

- Rewrite this as an algebraic equation for the ordinary generating funct.

$$C(x) := C_0 + C_1 x + C_2 x^2 + \dots$$

Remark. Here we ignore the question of convergence or divergence of this power series. We treat $C(x)$ as a formal power series (i.e an infinite series with some coeffs) We can define how to add & multiply such formal power series without worring about their convergence.

recurr. relation

$$C_n = \sum_{k=1}^{n} C_{k-1} \cdot C_{n-k}$$

multiply it by powers of x

$$C_n x^n = x \cdot \sum_{k=1}^{n} (C_{k-1} x^{k-1})(C_{n-k} x^{n-k})$$

then sum this over all $n \geq 1$ and all $k \in \{1, 2, \dots, n\}$.

$$LHS = C(x) - 1$$

the constant term is not included into the summation

$$RHS = x \cdot C(x) \cdot C(x)$$

We obtain

Proposition The generating function $C(x) = \sum_{n \geq 0} C_n x^n$ satisfies

$$C(x) - 1 = x\, C(x)^2$$

or

$$\boxed{x\, C(x)^2 - C(x) + 1 = 0}$$

$$x C^2 - C + 1 = 0$$

How to solve this equation?

Let us treat x as a parameter and C as an unknown variable.

Then we've got a quadratic equation for C.

$$a C^2 + b C + d = 0$$

where $a = x$, $b = -1$, $d = 1$

$$C = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{1 \pm \sqrt{1-4x}}{2x}$$

We obtain two possible solutions:

$$C(x) = \frac{1+\sqrt{1-4x}}{2x} \quad \text{and}$$

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

Question: which one is the correct solution?

If $C(x) = \dfrac{1+\sqrt{1-4x}}{2x}$. Then

as $x \to 0$, $C(x) \approx \dfrac{1+1}{2x} \to \infty$

However $C(0)$ should be $C_0 = 1 \neq \infty$.

So $C(x) = \dfrac{1-\sqrt{1-4x}}{2x}$

is the correct solution.

Proposition

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

What should we do next?

Binomial Formula:

$$(1+y)^n = \binom{n}{0} y^0 + \binom{n}{1} y^1 + \binom{n}{2} y^2 + \dots$$

binomial coefficients

This formula is true not only in the well-known case when n is a positive integer, but also for any real (and even complex) number n.

We just need to define the binomial coefficients $\binom{n}{k}$ when n is any complex number (but k is still non-negative integer).

$$\binom{n}{k} := \frac{n(n-1)(n-2)\dots(n-k+1)}{k!}$$

One can easily prove the binomial formula

$$(1+y)^n = \sum_{k\geq 0} \binom{n}{k} y^k$$

by taking the Taylor series for $(1+y)^n$.

Let's specialize it for $y = -4x$ and $n = \frac{1}{2}$.

$$\sqrt{1-4x} = (1-4x)^{1/2}$$

$$= \sum_{k \geq 0} \binom{1/2}{k} (-4x)^k.$$

Now $C(x) = \dfrac{1-\sqrt{1-4x}}{2x}$

Extracting the coefficient of x^n in the RHS, we obtain

$$C_n = -\frac{1}{2} \binom{1/2}{n+1} (-4)^{n+1}$$

$$= -\frac{1}{2} \frac{1/2 \cdot (\frac{1}{2}-1)(\frac{1}{2}-2) \cdots (\frac{1}{2}-n)}{(n+1)!} (-4)^{n+1}$$

$$= \cdots = \frac{1}{n+1} \binom{2n}{n},$$

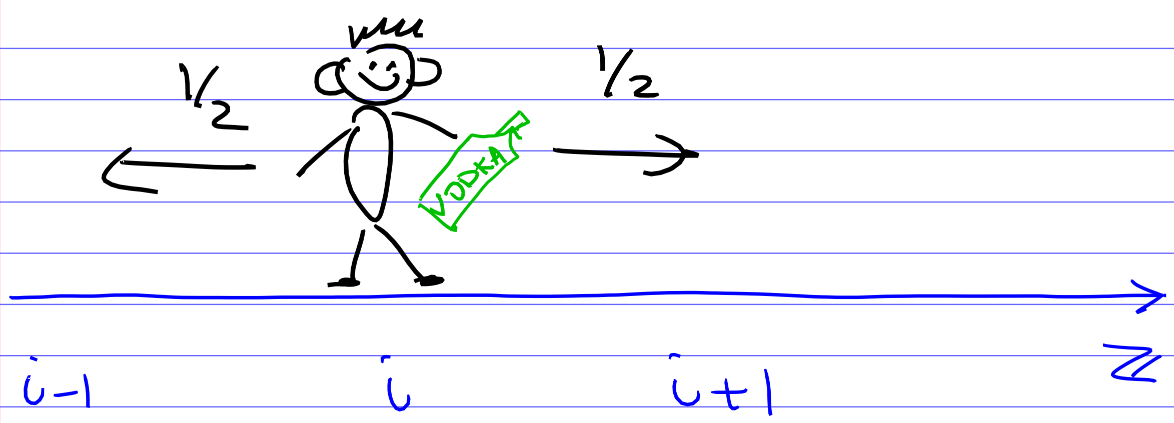
as needed.

Q.E.D.

<u>Remark</u> There are shorter and more interesting proofs of the formula $C_n = \frac{1}{n+1}\binom{2n}{n}$. We'll discuss this next time...

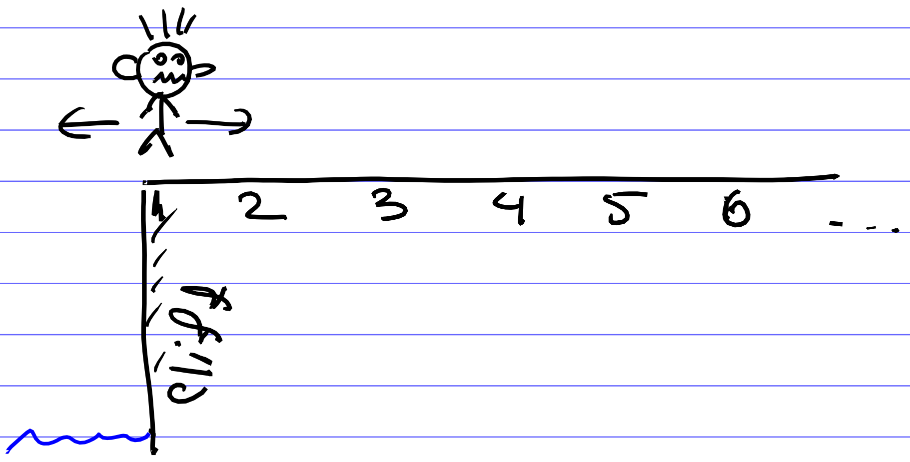
Drunkard's Walk

A simple random walk on the integer line



The drunkard makes a step to the right with probability $1/2$ or a step to the left with probability $1/2$, and continues randomly walking like this.

Now assume that there is a cliff at $x = 0$. Initially, the drunkard starts at $x = 1$ (right at the edge of the cliff).
If he steps off the cliff after some number of steps, he falls.



Problem. What is the probability that the drunkard does not fall off the cliff?

Let's find all possibilities for the drunkard to fall off the cliff

$$\text{Prob}\begin{pmatrix}\text{falls off}\\ \text{the cliff}\end{pmatrix} = \overset{\leftarrow}{\frac{1}{2}} +$$

$$+ \overset{\rightarrow}{\frac{1}{2}} \cdot \overset{\leftarrow}{\frac{1}{2}} \cdot \overset{\leftarrow}{\frac{1}{2}} + \overset{\rightarrow}{\frac{1}{2}} \overset{\leftarrow}{\frac{1}{2}} \cdot \overset{\rightarrow}{\frac{1}{2}} \cdot \overset{\leftarrow}{\frac{1}{2}} \overset{\leftarrow}{\frac{1}{2}}$$

$+ \dots$

all these possibilites correspond to Dyck paths with one extra left step in the end.

$$= \frac{1}{2} C_0 + \left(\frac{1}{2}\right)^3 C_1 + \left(\frac{1}{2}\right)^5 C_2 + \dots$$

the Catalan numbers

$$= \frac{1}{2} C\left(\frac{1}{4}\right).$$

Let us now plug this into the expression $C(x) = \frac{1-\sqrt{1-4x}}{2x}$

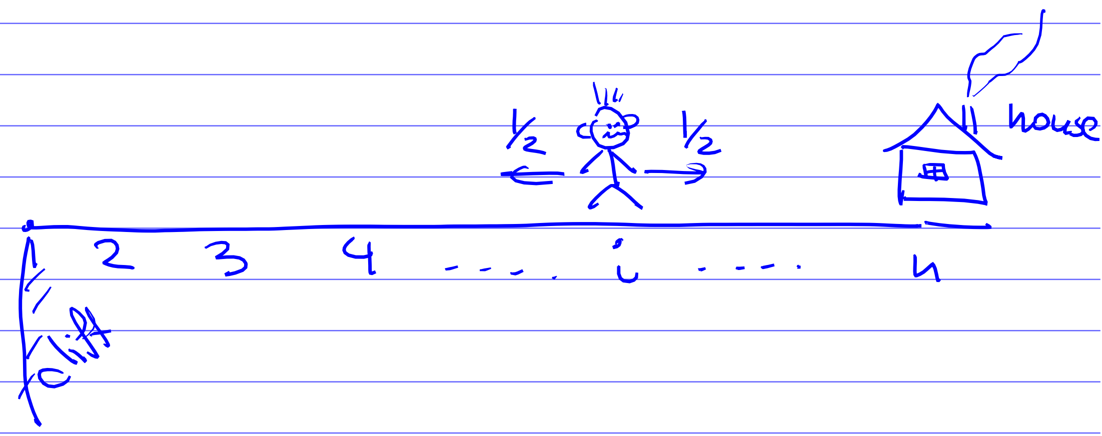
$$\text{Prob}\begin{pmatrix}\text{falls off}\\ \text{the cliff}\end{pmatrix} = \frac{1}{2} \frac{\left(1-\sqrt{1-\frac{4}{4}}\right)}{2 \cdot \frac{1}{4}} = \underline{1}$$

Not very good news for the drunkard.

<u>Proposition</u>. $\text{Prob}\begin{pmatrix}\text{drunkard}\\ \text{falls}\end{pmatrix} = 1.$

There is a simpler solution of the drunkard's walk problem that does not use the Catalan numbers

Let's give the drunkard some hope. Assume that his house is located at $x=n$ (and the cliff is still at $x=0$)



- the drunkard starts at $x=i \in \{0,1,2,\dots,n\}$.

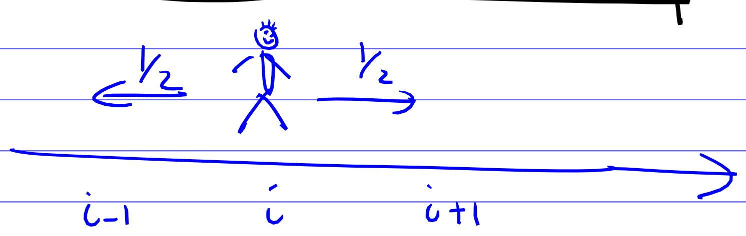
- If he first reaches $x=0$ (cliff) he falls
- If he first reaches $x=n$ (house), he goes to sleep.

Let p_i be the probability that the drunkard reaches the house before reaching the cliff.

The probabilities $p_0, p_1, \dots, p_n$ satisfy the relations:

- For $1 \leq i \leq n-1$,

$$\boxed{p_i = \frac{1}{2} p_{i-1} + \frac{1}{2} p_{i+1}}$$



- $\boxed{p_0 = 0}$ the drunkard starts at $x=0$ & falls right away
- $\boxed{p_n = 1}$ the drunkar starts at the house $(x=n)$ & goes to sleep right away

It is not hard to show that this system of linear equations for $p_0, p_1, \dots, p_n$ has a unique solution

$$p_i = \frac{i}{n} \quad \text{for } i = 0, 1, \dots, n.$$

In particular, $p_1 = \frac{1}{n}$

Taking the limit as $n \to \infty$ we obtain the solution of the original problem without the house.

$$\text{Prob}\begin{pmatrix}\text{does not}\\ \text{fall off}\\ \text{the cliff}\end{pmatrix} = \lim_{n\to\infty} p_1 =$$

$$= \lim_{n\to\infty} \left(\frac{1}{n}\right) = 0.$$