Lecture 7: Aprill 11, 2011

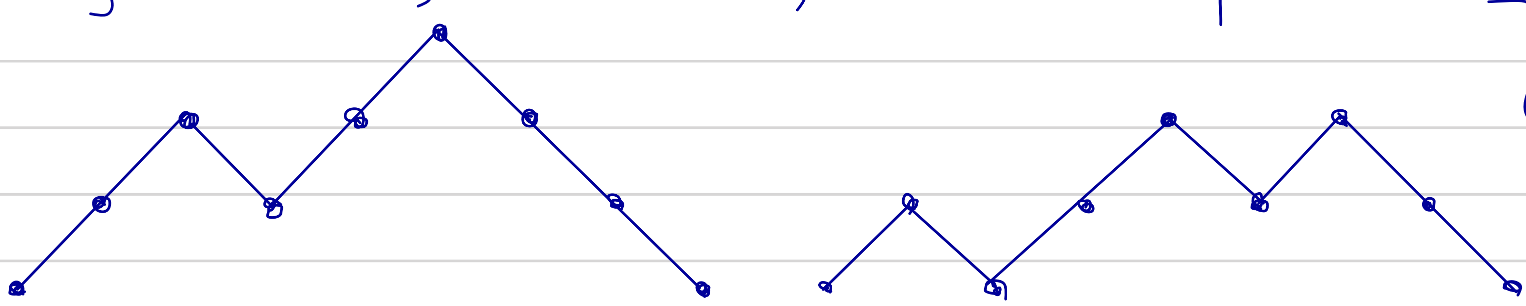
Def: A **Dyck Path** is a finite sequence $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \{\pm 1\}^n$ with the properties that:

(a) For $1 \leq k \leq n$, $\varepsilon_1 + \cdots + \varepsilon_k \geq 0$

(b) $\varepsilon_1 + \cdots + \varepsilon_n = 0$.

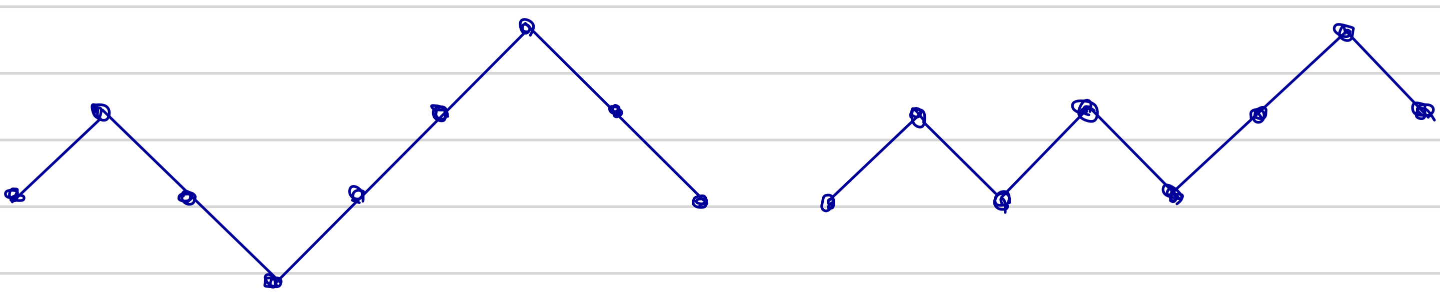
Since ± 1 are odd, $\varepsilon_1 + \cdots + \varepsilon_n$ has the same parity as n. Thus, condition (b) requires n to be even.

They are called paths since they can be represented nicely as lattice paths: the vertices are the points $(k, \varepsilon_1 + \cdots + \varepsilon_k)$ for $1 \leq k \leq n$ in addition to $(0,0)$. Thus, a Dyck path is a slope ± 1 lattice path that begins and ends on the x-axis, and never drops below it.



$+ + - + + - - -$ $\qquad$ $+ - + + - + - -$

Dyck Paths



$+ - - + + + - -$ $\qquad$ $+ - + - + + -$

Not Dyck Paths

From the path representation, we see something that wasn't obvious from the definition:

$(\varepsilon_1, \ldots, \varepsilon_n)$ is a Dyck path $\Longleftrightarrow$ $(\varepsilon_n, \ldots, \varepsilon_1)$ is a Dyck path.

Now, as we saw last day, if S is the unilateral shift operator on $\ell^2(\mathbb{N})$, $\mathfrak{A} = \langle S \rangle_*$, and $\varphi_0(A) = \langle A e_0, e_0 \rangle$, then in the NCPS $(\mathfrak{A}, \varphi_0)$, the $*$-distribution of S is described by

$$\varphi_0(S^{\varepsilon_1} \cdots S^{\varepsilon_n}) = \begin{cases} 1, & \text{if } (\varepsilon_n, \ldots, \varepsilon_1) \text{ is a Dyck path} \\ 0, & \text{otherwise.} \end{cases}$$

Remarks: (1) no need to reverse ε by the above observation.

(2) So the odd joint moments of S, S^* are all 0.

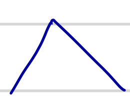
Now, consider the self-adjoint element $X = S + S^*$. We can use the $*$-distribution of S to compute the moments of X.

$$\varphi_0(X^n) = \varphi_0\left((S+S^*)^n\right)$$

$$(S+S^*)^n = \sum_{\varepsilon \in \{1,*\}^n} S^{\varepsilon_1} S^{\varepsilon_2} \cdots S^{\varepsilon_n}$$

$$\therefore \varphi_0(X^n) = \sum_{\varepsilon \in \{1,*\}^n} \underbrace{\varphi_0(S^{\varepsilon_1} \cdots S^{\varepsilon_n})}_{\substack{=1 \text{ if } \varepsilon \text{ is a Dyck path,} \\ =0 \text{ otherwise}}} = \#\text{Dyck paths of ``length'' } n.$$

So immediately we see that $\varphi_0(X^n) = 0$ if n is odd. For the even moments, we must enumerate the set of Dyck paths. Let's look at a few examples, for small n.

Eg. $n=2$:  $\# = 1$

$n=4$:  $\# = 2$

$n=6$:  $\# = 5$

If you plug 1, 2, 5 into the Sloan database, the first result turns out to be the right answer: these are the Catalan #s.

Here we will present two ways of enumerating Dyck paths: one will give a(n important) recurrence, the other will give a closed formula.

For $n \geq 1$, let $C_n = \#$Dyck paths of length $2n$. $C_1 = 1, C_2 = 2, C_3 = 5, \ldots$

Enumeration 1

First observation: any Dyck path must start with +, and (by reflective symmetry) end in −. So we can start by throwing these away:

$$(+, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_{n-2}, \varepsilon_{n-1}, -) \iff (\underbrace{\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_{n-2}, \varepsilon_{n-1}}_{\text{another lattice path.}})$$

<u>Question</u>: when is this length $2(n-1)$ lattice path <u>also</u> a Dyck path? <u>Answer</u>: when the original path never touches the x-axis except @ $(0,0)$ and $(2n,0)$.

<u>Def</u>: a Dyck path is <u>irreducible</u> if it only touches the x-axis at the beginning and the end. Denote

$$C_n' = \#\text{irreducible Dyck paths of length } n.$$

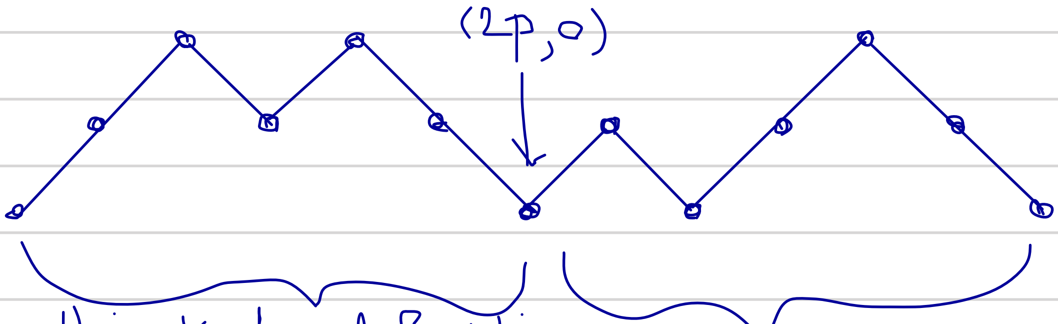
The above observation gives a bijective proof of the fact that

$$C_n' = C_{n-1}. \Leftarrow \text{ even makes sense for } n=1, \text{ as } C_0 = 1 \text{ is natural.}$$

Now we can setup our recursion. Let $C_n^{(p)} = \#$Dyck paths of length $2n$ that first touch the x-axis at the point $(2p,0)$, for $1 \leq p \leq n$; so the irreducible paths are exactly those counded by $C_n^{(n)}$. Then

$$C_n = C_n^{(1)} + C_n^{(2)} + \cdots + C_n^{(n)}.$$

But $C_n^{(p)}$ factors in a nice way.



The behaviour of the path after the first x-touch is independent of the path before.

this is, by definition, an irreducible Dyck path of length $2p$.

This is another (not necessarily irreducible) Dyck path of length $2(n-p)$

Thus, $C_n^{(p)} = C_p' \cdot C_{n-p}$.

$$\therefore\ C_n = C_n^{(1)} + C_n^{(2)} + \cdots + C_n^{(n)} = C_1' C_{n-1} + C_2' C_{n-2} + \cdots + C_{n-1}' C_1 + C_n'$$

Using the fact that $C_k' = C_{k-1}$ for $k \geq 1$, and $C_0 = 1$, we have proved the following.

<u>Prop</u>: $\{C_n : n \geq 0\}$ satisfy $C_0 = 1$ and

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + \cdots + C_{n-1} C_0$$
$$= \sum_{k=1}^{n} C_{k-1} C_{n-k}.$$

This is called the <u>Catalan recurrence</u>, and indeed generates the Catalan numbers. In fact, there is a way (without an ansatz/induction) of deriving the formula for C_n from here. We will address that approach, using generating functions, in a bit; first I want to present a second, very different enumeration trick, due to the 19th Century French combinatorialist Desiré André.

<u>Enumeration 2: André's Reflection Trick</u>

First, let's ignore the constraints of a Dyck path, and just look at general lattice paths. We can count the # of lattice paths (with slopes ± 1) that begin at $(0,0)$ and end at $(2n, 2m)$.

Say such a path has u $+$'s and v $-$'s; then $u+v=2n$ and $u-v=2m$; that is, $u=n+m$ and $v=n-m$. To specify a path, we must select which of the n arcs are $+$ (or $-$), so the # of paths is:

$\#\pm$paths from $(0,0)\to(2n,2m)$ $= \binom{2n}{n+m} = \binom{2n}{n-m}$.

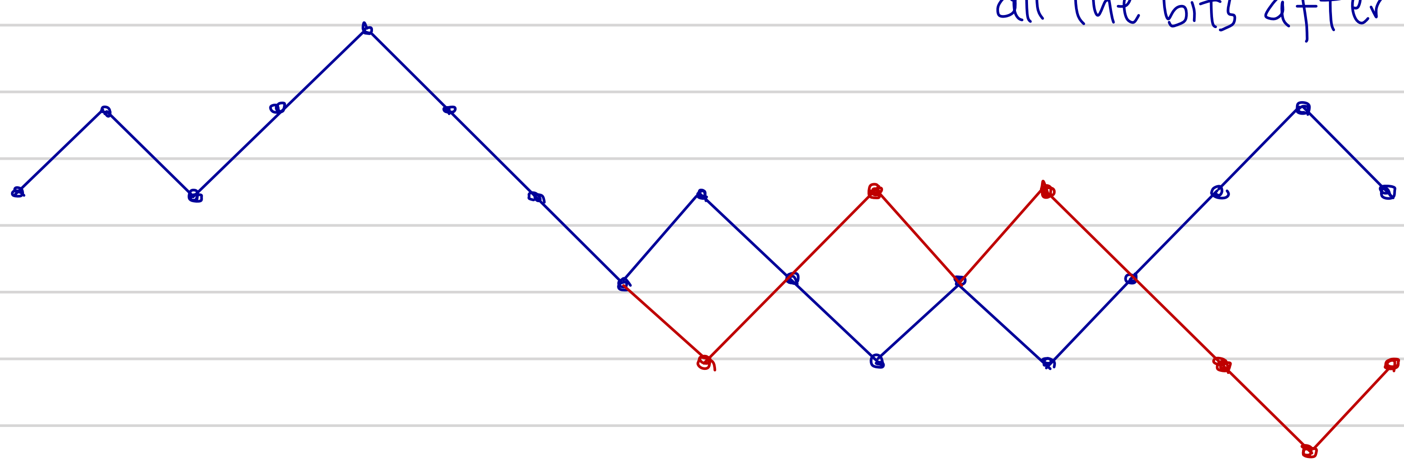
In particular, with $m=0$, $\pm$paths from $(0,0)\to(2n,0)$ number $\binom{2n}{n}$.

Now, which among these are <u>not</u> Dyck paths?

↳ If ε is not a Dyck path, there must be a first time the path drops below the x-axis; so let p be minimal so that $(p,-1)$ is on the path. Then break ε up at p:

$$\begin{array}{l} + - + + - - \overset{\downarrow}{-} + - - + - + + + - \\ + - + + - - - \color{red}{- + + - + - - - +} \end{array}$$

Create a new path $\color{red}{R(\varepsilon)}$ by reversing all the bits after p.



Since the sum of all the bits up to p is -1, and since ε ends at $(0,0)$, the sum of the bits after p is $+1$; reversing means $R(\varepsilon)$ ends at $(2n,-2)$.

So, $R: \{\text{non-Dyck paths } (0,0)\to(2n,0)\} \longrightarrow \{\text{lattice paths } (0,0)\to(2n,-2)\}$

<u>Claim</u>: R is a bijection.

Pf. If η is any lattice path $(0,0)\to(2n,-2)$, there must be a (minimal) $p<2n$ with $(p,-1)\in\eta$. Then perform the same reflection $R(\eta)$ as above; the same analysis shows that $R(\eta)$ ends at $(0,0)$; it is not a Dyck path, since $(p,-1)\in R(\eta)$. Since $R^2=\mathrm{id}$, this concludes the proof. ///

Hence, $C_n = \#\{(0,0)\to(2n,0)\} - \#\{\text{non-Dyck } (0,0)\to(2n,0)\}$

$= \#\{(0,0)\to(2n,0)\} - \#\{(0,0)\to(2n,-2)\}$

$= \binom{2n}{n} - \binom{2n}{n-1} \overset{\text{easy}}{\underset{\text{calculation}}{=}} \underline{\frac{1}{n+1}\binom{2n}{n}}$ ///

Any way you slice it, we now know how to count Dyck paths.

Thus, going back to the shift operator S, and $X = S + S^*$, we have

$$\varphi_0(X^n) = \begin{cases} C_n = \frac{1}{n+1}\binom{2n}{n}, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

Now, the question before us is: what measure (on $\mathbb{R}$) has these moments? I could give another ansatz that could be easily checked — but it is more instructive to go about this by introducing yet <u>another</u> method for finding a formula for C_n:

Enumeration 3: Generating Function.

Given the Catalan #s C_n, form the power series

$$\mathcal{C}(z) = \sum_{n=0}^{\infty} C_n z^n \leftarrow \text{ordinary generating function.}$$

As combinatorialists, we treat this as a formal power-series. As analysts, we wonder where (if) it converges. For the moment, we ignore the formula we developed for C_n, and rely on the recurrence relation

$$C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k}. \quad (*)$$

Note: by definition, C_n counts Dyck paths, a subset of $\{1,-1\}^{2n}$, which has size $2^{2n} = 4^n$; so $C_n < 4^n$.

<u>Exercise</u>: show from $(*)$ directly that, with $C_0 = 1$, $C_n \le 4^n$.

So, we can be assured that $\mathcal{C}(z)$ converges on $\{|z| < \frac{1}{4}\}$ at least. Usual theory of power series justifies all the following calculations on this domain.

$$\mathcal{C}(z) = \sum_{n=0}^{\infty} C_n z^n = 1 + \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} C_{k-1} C_{n-k} \right) z^n$$

$$= 1 + z \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} C_{k-1} z^{k-1} \cdot C_{n-k} z^{n-k} \right)$$

$$= 1 + z \sum_{n=1}^{\infty} \sum_{\ell=0}^{n-1} C_\ell z^\ell \cdot C_{n-(\ell+1)} z^{n-(\ell+1)})$$

$$= 1 + z \sum_{m=0}^{\infty} \sum_{\ell=0}^{m} C_\ell z^\ell \cdot C_{m-\ell} z^{m-\ell}$$

$$= 1 + z \left(\sum_{m=0}^{\infty} C_m z^m \right)^2 = 1 + z\,\mathcal{C}(z)^2$$

Thus $z\,\mathcal{C}(z)^2 - \mathcal{C}(z) + 1 = 0$. It follows that

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

must be $-$, b/c $\mathcal{C}(0) = 1$ is finite.

<u>Exercise</u>: Use the binomial theorem (for non-integer powers) to develop the Taylor series of $\mathcal{C}(z)$:

$$\mathcal{C}(z) = \frac{1 - \sqrt{1-4z}}{2z} = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} z^n$$

providing another derivation of the formula for the Catalan number C_n.

As we will see next time, having an explicit formula for the generating function immediately allows us to find the measure from its moments.