Catalan Numbers and Random Matrices

Lawrence Ip
lip@eecs.berkeley.edu
Department of Electrical Engineering
and Computer Sciences
University of California, Berkeley
supervised by Alan Edelman

May 19 1999

1 Introduction

Catalan numbers, the number of ways of pairing n brackets, arise naturally in some random matrix results in the computation of moments of eigenvalue distributions.

In Section 2 we show that Catalan numbers can be represented as the number of Dyck paths. In Section 3 we show that this representation can be used to drastically simplify some of the calculations that Wigner used for his derivation of the semi-circle law and in Section 4 we show that Dyck paths can be used to interpret an expression combinatorially to prove its validity. This last result is the main result of this project.

2 Catalan Numbers and Dyck Paths

The most common definition of the nth Catalan number is "the number of ways of pairing n brackets". So we have $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, $C_4 = 14$, $C_5 = 42$, $\ldots$. The standard way to find a general formula for C_n is to derive a recurrence relation for it and then solve this using generating functions.

However, this is not the only representation. In his new book, Stanley [2], lists 66 combinatorial interpretations of Catalan numbers. We show that C_n is also the number of Dyck paths of length $2n$.

A Dyck path is a path on the square lattice with steps $(1,1)$ and $(1,-1)$ from $(0,0)$ to $(2n,0)$ that never falls below the x-axis. Consider a pairing of n brackets. At each step, let δ_i be the number of unmatched brackets to the left of the ith bracket. So $(()())$ would be encoded 0,1,2,1,2,1,0. If we now interpret the δ_i as the heights of a Dyck path, we see that there is a one to one correspondence between pairings of n brackets and Dyck paths of length $2n$.

We give a bijective proof of a formula for the number of Dyck paths and hence a formula for C_n. Consider Dyck paths of length $2n$. Now add a downward step at the end of the Dyck path. There are now $n+1$ downward steps. Choose one of these steps to be special. Then we have $(n+1)C_n$ possible configurations. Now cut the Dyck path into two parts, the part before the special step and the part after it (including the special step). Move the right part so that it starts from $(0,0)$ and append the left part to the right part. We now have a special step going from $(0,0)$ to $(1,-1)$ and a path with steps $(1,1)$ and $(1,-1)$ going from $(1,-1)$ to $(2n+1,-1)$. We can recover the Dyck path with the extra downward step and special step by finding the leftmost minimum of the new path, cutting it into two parts and reassembling. Thus this operation is a bijection. See Figure 1 for an example.

Clearly the number of paths from $(1,-1)$ to $(2n+1,-1)$ in is $\binom{2n}{n}$. So we find that

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

3 Wigner's Semi-circle Law

Wigner's semi-circle law essentially says that the histogram of the eigenvalues of a random symmetric matrix are approaches a semi-circle. Wigner [3] first derived this for a special case where he allowed the entries to be $\pm v$.

In his derivation he computed the asymptotics of the moments of the distribution by considering the trace of H^k and used this to calculate the resulting characteristic function (Fourier transform of the distribution).

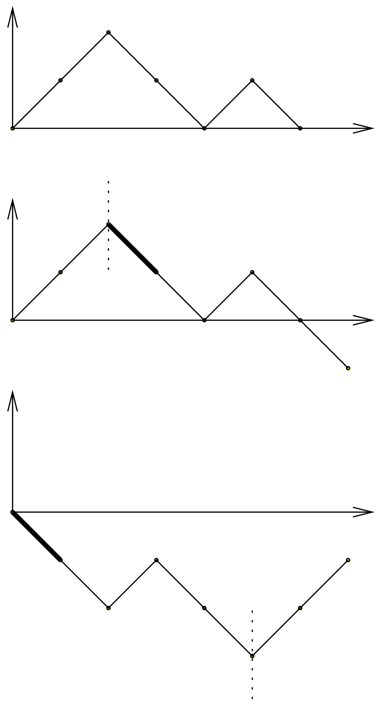

Figure 1: Bijection in derivation of formula for C_n.

Due to the symmetry of the distribution, the terms in the expression for the νth moment

$$M_\nu = \mathcal{R}^{-1} \sum_{\text{set}} \sum_{i_1,\dots,i_{\nu-1}} v_{0i_1} \cdots v_{i_{\nu-1}0},$$

are non-zero only if $0, i_1, i_2, \dots, i_{\nu-1}, 0$ is a valid sequence. A valid sequence, is a sequence of the form $0, i_1, i_2, \dots, i_{2\nu-1}, 0$, where no two successive members of the sequence are equal and each step jl and lj occurs an even number of times.

Wigner then shows that the number of valid sequences of length $2\nu+1$ is asymptotically $t_\nu(2N)^\nu v^{2\nu}$, where t_ν is the number of type sequences with length $2\nu+1$.

A type sequence of length $2\nu+1$ is a sequence that starts at 1 and ends at 0, with successive members differing by ± 1 and all members non-negative. By interpreting the type sequence as the heights of a Dyck path we see that the number of type sequences t_ν is the νth Catalan number.

4 Random Regular Graphs

McKay [1] derived the eigenvalue distribution of a large random regular graph. Again the approach is to calculate moments of the distribution. Here a diagonal entry of A^k has the interpretation as the number of closed walks of length k.

McKay gave a derivation for the exact number of closed walks by using combinatorial arguments, arriving at the following result.

Let ν_0 be a vertex of a graph X. A closed walk of length $r \geq 0$ starting at ν_0 is a sequence $\nu_0, \nu_1, \nu_2, \dots, \nu_r$ of vertices of X such that $\nu_r = \nu_0$, and ν_{i-1} is adjacent to ν_i for $1 \leq i \leq r$.

Lemma 2.1. *Suppose X is regular of degree v. Let ν_0 be a vertex of X, and suppose that the subgraph of X induced by the vertices at distance at most $r/2$ from ν_0 is acyclic. Then the number of closed walks of length r in X*

starting at ν_0 is $\theta(r)$, where $\theta(r) = 0$ if r is odd, and

$$\begin{aligned}
\theta(2s) &= \sum_{k=1}^{s} \binom{2s-k}{s} \frac{k}{2s-k} v^k (v-1)^{s-k} \\
&= v \sum_{k=0}^{s-1} \binom{2s}{k} \frac{s-k}{s} (v-1)^k \\
&= \sum_{k=1}^{s} \binom{2s}{k} \frac{2s-2k+1}{2s-k+1} (v-1)^k.
\end{aligned}$$

Edelman and Strang later independently derived an expression for the number of closed walks by calculating the first few values and then passing it to the superseeker (`http://www.research.att.com/~njas/sequences/-eisonline.html`) to guess the general formula

$$m_n(z) = (1+z)^{2n+1} - \sum_{j=0}^{n-1} (1+z)^{2j+1} z^{1+n-j} C_{n-j}.$$

In the notation of McKay's paper, $z = v-1$ and $s = n+1$, and the expression becomes

$$\theta(2s) = v^{2s-1} - \sum_{j=1}^{s-1} v^{2j-1} (v-1)^{s-j+1} C_{s-j}. \tag{1}$$

This formula also involves Catalan numbers and it turns out that it has a straightforward combinatorial interpretation.

Let $\nu = \nu_0, \nu_1, \ldots, \nu_r$ be a walk of length r. We can think of this as a traversal of the vertices of the graph. Let δ_i be the depth of the traversal at step i. We decrease δ_i by one if we "backtrack" and increases δ_i by one otherwise. More formally,

$$m(d, i) = \max\{j : j < i, \delta_j = d\}$$

then

- $\delta_0 = 0$,
- $\delta_{i+1} = \delta_i - 1$ if $\nu_{m(\delta_i - 1, i)} = \nu_{i+1}$,

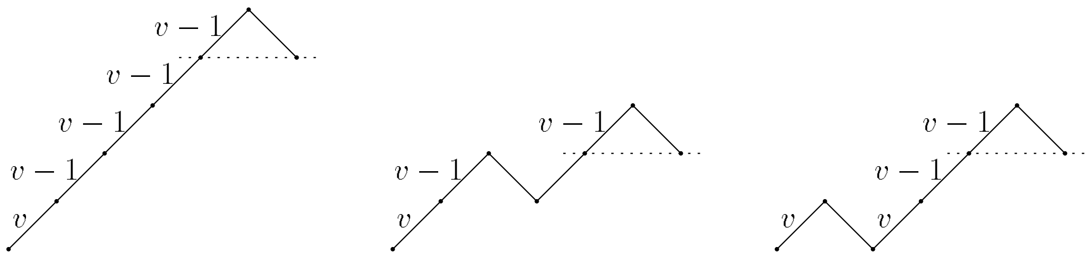

Figure 2: Plot of δ_i with $s = 3$, $j = 2$.

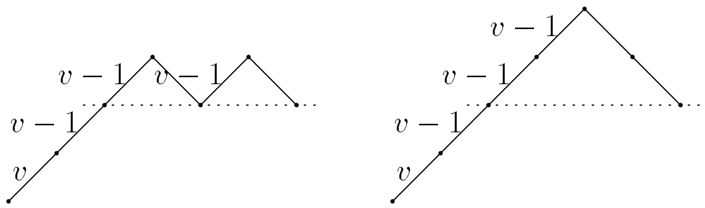

Figure 3: Plot of δ_i with $s = 3$, $j = 1$.

- $\delta_{i+1} = \delta_i + 1$ if $\nu_{m(\delta_i - 1, i)} \neq \nu_{i+1}$.

Note that for all *closed* walks in the setting of Lemma 2.1 this definition is equivalent to the definition given in McKay's paper because of the acyclic condition of the subgraph. Thus the number of closed walks is given by the number of walks with $\delta_{2s} = 0$.

Now clearly all closed walks must end with a backtrack. The number of walks that end with a backtrack is thus v^{2s-1}, the first term of Equation 1.

Now if one of these walks is not closed, then $\delta_{2s} = 2k$, where $k \geq 1$, because of parity considerations. In particular, $\delta_{2s} \geq 2$. Then let $2j - 1 = m(\delta_{2s} - 1, 2s)$. For any given j there are v^{2j-1} ways to traverse the first $2j - 1$ steps. The remaining $2s - 2j + 1$ steps must have $\delta_i > \delta_{2j-1}$ for all $i > 2j - 1$. Plotting the values of δ_i we can think of this as one "upward" step followed by a Dyck path of length $2(s-j)$. Now each upward step can occur in $v - 1$ ways and each "downward" step (corresponding to a backtrack), can occur in only one way. So the number of ways of doing this will be $(v-1)(v-1)^{s-j} C_{s-j}$. Summing over j gives us Equation 1.

5 Conclusion

The representation of Catalan numbers as the number of pairings of n pairs of brackets is the most well known but is not necessarily the most useful one. In this project, we showed two applications where using the Dyck path representation made it easier for the author to find direct combinatorial interpretations of various formulae. The reason Dyck paths may be easier to work with is the graphical nature of their representation.

6 Acknowledgments

To Frank Calegari for recalling the key idea for the bijective proof of the formula for the number of Dyck paths.

References

[1] B. D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.

[2] R. P. Stanley. *Enumerative Combinatorics*, volume 2. Cambridge University Press, 1999.

[3] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, November 1955.