The Catalan Numbers

Example 1. A sequence of parentheses is said to be well-formed (or legal) of there are an equal number of left and right parentheses and, when reading the string from left to right, the number of right parentheses never exceeds the number of left parentheses. For example, the strings below are well formed.

$$()(()); ()()((())); ()()()((()))(()())$$

For $n > 0$, let S_n be the set of all legal strings of n pairs of parentheses. For example,

$$S_3 = \{()()(), (())(), ()(()), ((())), (()())\}$$

since these are the only legal strings of three pairs of parentheses. Now let $c_0 = 1$ and for $n > 0$, let $c_n = |S_n|$.

Now for each $w \in S_n$, we scan the string from left to right and stop once we reach the *first* legal substring. Call this substring the **prefix** and define the **index** k of a given string to be the number of pairs of parentheses in its prefix. Finally, call a string **primitive** if $k = n$. It should be clear that for any given string, $1 \leq k \leq n$.

For example, if $w = (()())()() \in S_5$ then its prefix is the legal string preceding the vertical bar $\overbrace{(()())}^{\text{3 pairs}}|()()$ so $k = 3$. The indices associated with S_3 are 1, 2, 1, 3, 3, respectively. Notice that the last two strings are primitive.

For each $n > 0$, how many *primitive* strings in S_n are there?

Proposition. For $n > 0$, let p_n count the number of primitive strings of length n. Then

$$p_n = c_{n-1} \tag{1}$$

Proof: Let $w \in S_n$ be a primitive string. Then removing the first and last parentheses yields a legal string of length $n - 1$. On the other hand, if $w \in S_{n-1}$ is any legal string, then (w) is a primitive string in S_n. □

Now let $w \in S_n$ with index k. The remaining $n - k$ pairs of parentheses can be formed in c_{n-k} ways. So by the product rule there are $p_k c_{n-k}$ ways to form a string (from S_n) with index k. Summing over all $1 \leq k \leq n$ we have

$$c_n = \sum_{k=1}^{n} p_k c_{n-k} \tag{2}$$

Now (2) suggests that we consider ordinary power series generating functions. So let $C \overset{\text{ogf}}{\longleftrightarrow} \{c_n\}_{n\geq 0}$. Then

$$\begin{aligned} C(x) - 1 &= \sum_{n\geq 1} c_n x^n = \sum_{n\geq 1}\sum_{k=1}^{n} p_k c_{n-k}\, x^n \\ &= \sum_{n\geq 1} p_n x^n \sum_{n\geq 0} c_n x^n = \sum_{n\geq 1} c_{n-1} x^n \sum_{n\geq 0} c_n x^n \\ &= x \sum_{n\geq 0} c_n x^n \sum_{n\geq 0} c_n x^n = xC(x)C(x) \end{aligned}$$

In other words,

$$C(x) = 1 + xC(x)^2 \tag{3}$$

Now let $\alpha = C(x)$. Rearranging yields the following quadratic equation (in α)

$$0 = x\alpha^2 - \alpha + 1$$

So that

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

Now which should we choose? That is, should we let

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

One can easily confirm that

$$\lim_{x\to 0} \frac{1+\sqrt{1-4x}}{2x} = \infty$$

On the other hand,

$$\begin{aligned}
\lim_{x\to 0} \frac{1-\sqrt{1-4x}}{2x} &= \lim_{x\to 0} \frac{1-\sqrt{1-4x}}{2x} \frac{1+\sqrt{1-4x}}{1+\sqrt{1-4x}} \\
&= \lim_{x\to 0} \frac{4x}{2x(1+\sqrt{1-4x})} \\
&= \lim_{x\to 0} \frac{2}{1+\sqrt{1-4x}} = 1
\end{aligned}$$

Since $c_0 = 1$ implies that $C(0) = 1$, then

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

For an alternative derivation of $C(x)$ see section 2.6.6 of the HHM text.

The numbers $\{c_n\}_{n\geq 0}$ are called the **Catalan** numbers and they are among the more important sequences in all of combinatorics. There first 10 Catalan numbers are

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, \ldots \tag{5}$$

According to the site maintainers, the Catalan number reference is probably the most extensive entry at the OEIS. R. Stanley has identified 214 different kind of objects that can be counted using these numbers.

We will have more to say about this important sequence and its ordinary power series generating function in the Example 2 below.

Example 2. Find a closed formula for the Catalan numbers c_n defined in Example 1 (see (2)).

So by (4) we have

$$c_n = [x^n]C(x) = [x^n]\frac{1-\sqrt{1-4x}}{2x} \tag{6}$$

Recall the generalized Binomial theorem from second semester calculus,

$$\begin{aligned}
1-\sqrt{1-4x} = 1-(1-4x)^{1/2} &= -\sum_{n\geq 1}\binom{1/2}{n}(-4)^n x^n \\
&= -\sum_{n\geq 1}\binom{n-1/2-1}{n}(-1)^n(-4)^n x^n \\
&= -\sum_{n\geq 1}\binom{n-3/2}{n}4^n x^n \\
&= -4x\sum_{n\geq 0}\binom{n-1/2}{n+1}4^n x^n
\end{aligned}$$

Thus

$$\frac{1-\sqrt{1-4x}}{2x} = -2\sum_{n\geq 0}\binom{n-1/2}{n+1}4^n x^n$$

so that

$$c_n = [x^n]C(x) = -\binom{n-1/2}{n+1}2^{2n+1} \tag{7}$$

For example,

$$c_5 = -\binom{9/2}{6}2^{11} = 42$$

as expected. In the exercises we ask you to show that (7) has the more convenient form

$$c_n = \frac{1}{n+1}\binom{2n}{n}$$