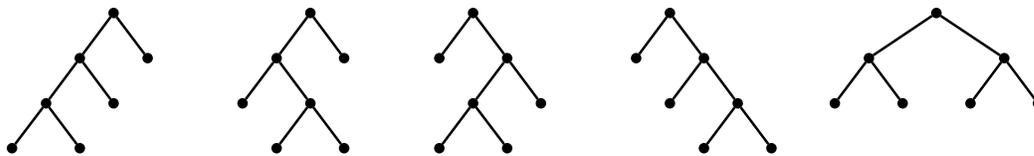


CATALAN NUMBERS

Catalan numbers arise from many counting problems. Here is one that is perhaps easiest to state from a typesetting point of view: suppose you have an operation $*$ that isn't associative. In how many ways can you “associate” a product of n things (kept on a fixed order)? That is, how many ways can you parenthesize a product of n letters? The cases $n = 1$ and $n = 2$ are trivial: there is only one choice (a and $a * b$ respectively). In the case $n = 3$ you have the two choices $(a * b) * c$ and $a * (b * c)$. For $n = 4$ there are five choices:

$$((a * b) * c) * d, \quad (a * (b * c)) * d, \quad a * ((b * c) * d), \quad a * (b * (c * d)), \quad (a * b) * (c * d)$$

Each of these choices can be represented by a tree diagram:



Technically, these are known as planar binary rooted trees. Here a *tree* is a graph with no circuits, a *rooted tree* is a tree with a distinguished vertex (the *root*, which is at the top in the examples above), and a *planar tree* is one embedded in the plane. In a rooted tree it is possible to assign a direction to each edge (down the page in the examples above) so the directions go away from the root and each vertex has one incoming edge. The vertices with no outgoing edges are called *terminal*. If every non-terminal vertex has exactly two outgoing edges, the tree is called *binary*.

Let's write P_n for the number of ways of parenthesizing n letters, or equivalently the number of planar binary rooted trees that have n terminal vertices. Let

$$P(t) = P_1 t + P_2 t^2 + P_3 t^3 + \cdots = \sum_{n \geq 1} P_n t^n$$

be the corresponding generating function. We claim that

$$(1) \quad P_n = P_1 P_{n-1} + P_2 P_{n-2} + \cdots + P_{n-2} P_2 + P_{n-1} P_1 = \sum_{k=1}^{n-1} P_k P_{n-k}.$$

To see this, consider a parenthesized product of n letters. The last operation done combines a parenthesization of (say) k letters on left with $n - k$ letters on the right. This can be done in $P_k P_{n-k}$ ways. Consideration of all possible values of k gives equation (1).

Now as a recurrence, equation (1) is much worse than anything we've dealt with so far: it has $n - 1$ terms! But generating functions make it manageable. Note that

$$\begin{aligned} P(t)^2 &= (P_1t + P_2t^2 + P_3t^3 + \dots)^2 \\ &= P_1^2t^2 + (P_1P_2 + P_2P_1)t^3 + (P_1P_3 + P_2^2 + P_3P_1)t^4 + \dots \\ &= P_2t^2 + P_3t^3 + P_4t^4 + \dots \quad (\text{using equation (1)}) \\ &= P(t) - t. \end{aligned}$$

Hence $P(t)$ satisfies the quadratic equation

$$(2) \quad P(t)^2 - P(t) + t = 0.$$

Solving equation (2) by the quadratic formula gives

$$P(t) = \frac{1 \pm \sqrt{1 - 4t}}{2},$$

and we choose the negative square root to have $P(0) = 0$. Thus

$$(3) \quad P(t) = \frac{1 - \sqrt{1 - 4t}}{2} = \frac{1}{2} - \frac{1}{2}(1 - 4t)^{\frac{1}{2}}.$$

Expanding out equation (3) using the binomial formula gives

$$P(t) = \frac{1}{2} - \frac{1}{2} \sum_{i \geq 0} (-4t)^i \binom{\frac{1}{2}}{i} t^i = \sum_{i \geq 1} 2(-4)^{i-1} \binom{\frac{1}{2}}{i} t^i = \sum_{i \geq 1} (-1)^{i-1} 2^{2i-1} \binom{\frac{1}{2}}{i} t^i.$$

Now

$$\begin{aligned} \binom{\frac{1}{2}}{i} &= \frac{(\frac{1}{2})(\frac{1}{2} - 1)(\frac{1}{2} - 2) \dots (\frac{1}{2} - i + 1)}{i!} \\ &= \frac{(1)(-1)(-3) \dots (-2i + 3)}{2^i i!} \\ &= \frac{(-1)^{i-1} (2i - 3)(2i - 5) \dots (3)(1)}{2^i i!} \\ &= \frac{(-1)^{i-1} (2i - 2)!}{(2i - 2)(2i - 4) \dots (2) 2^i i!} \\ &= \frac{(-1)^{i-1} (2i - 2)!}{2^{i-1} (i - 1)! 2^i i!} \\ &= \frac{(-1)^{i-1}}{2^{2i-1} i} \binom{2i - 2}{i - 1}. \end{aligned}$$

Hence

$$P(t) = \sum_{i \geq 1} \frac{1}{i} \binom{2i - 2}{i - 1} t^i,$$

or

$$P_n = \frac{1}{n} \binom{2n-2}{n-1}, \quad n \geq 1.$$

The *Catalan numbers* are generally defined by $C_n = P_{n+1}$ for $n \geq 0$. Thus the Catalan numbers have generating function

$$C(t) = C_0 + C_1t + C_2t^2 + \cdots = P_1 + P_2t + P_3t^2 + \cdots = \frac{1 - \sqrt{1-4t}}{2t}$$

and closed form

$$(4) \quad C_n = \frac{1}{n+1} \binom{2n}{n}, \quad n \geq 0.$$

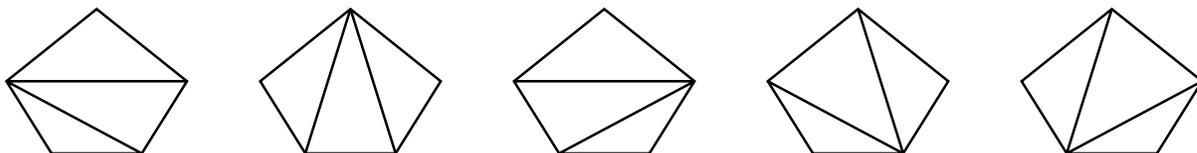
Exercise 1. Using equation (4), show that Catalan numbers satisfy the recurrence

$$C_{n+1} = \frac{2(2n+1)}{n+2} C_n.$$

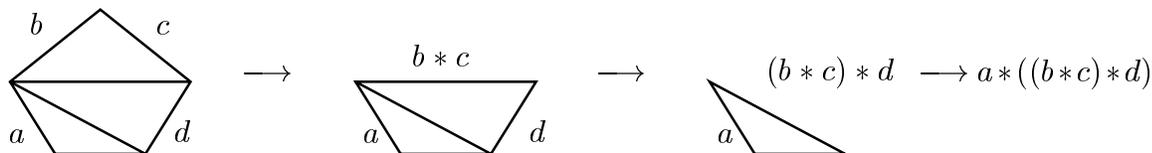
So the Catalan number C_n counts parenthesizations of $(n+1)$ -fold products, and planar binary rooted trees with $n+1$ terminal vertices. But C_n counts lots of other things. For example, consider the possible ways to “triangulate” (cut into triangles) a regular polygon with $n+2$ sides, using noncrossing diagonals. There are two ways to do this if $n=2$:



In the case $n=3$, there are five ways:



What does this have to do with parenthesizing $(n+1)$ -fold products? Label all but the bottom side of the $(n+2)$ -sided polygon with a letter. Then combine each pair of labels belonging to the same triangle, and after “collapsing” these pairs repeat the process until a parenthesized product is obtained. Here is an example.



Exercise 2. Work out the parenthesized products represented by the other four triangulations of the pentagon.

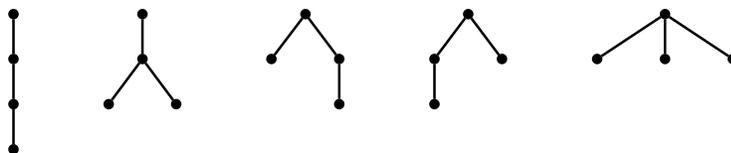
Catalan numbers also count what are known as balanced bracket arrangements (BBAs). A BBA of weight n is a sequence of left brackets \langle and right brackets \rangle so that:

1. there are a total of n left brackets and n right brackets; and
2. reading left to right, the running total of right brackets never exceeds the running total of left brackets.

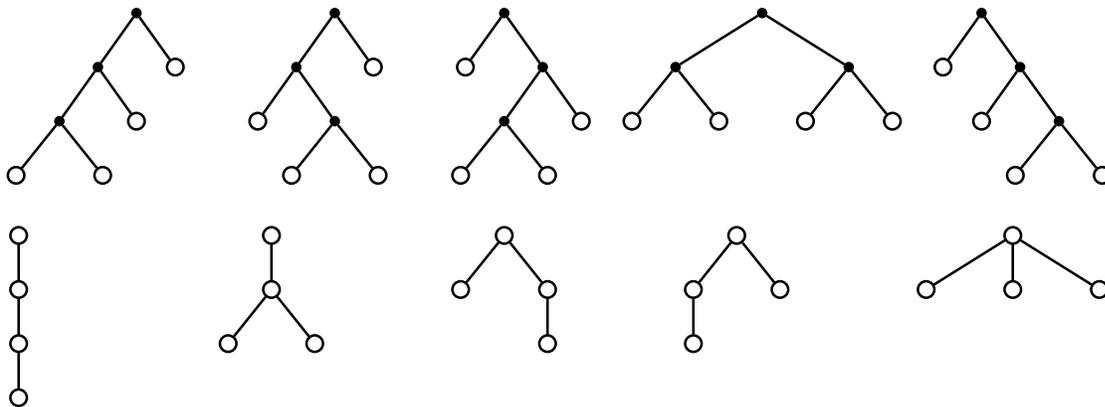
For example, here are the five BBA's of weight three:

$\langle\langle\langle\rangle\rangle\rangle$ $\langle\langle\rangle\rangle\rangle$ $\langle\rangle\langle\langle\rangle\rangle$ $\langle\langle\rangle\rangle\langle\rangle$ $\langle\rangle\langle\rangle\langle\rangle$

We claim that there are C_n BBA's of weight n . To see this, we first note that each BBA above can be associated with a planar rooted tree (not necessarily binary) that has four vertices:



But aren't these the wrong kind of trees? Yes they are, but we now show that there is a one-to-one correspondence between the sets of *planar binary rooted trees with n terminal vertices* and of *planar rooted trees with n total vertices*. To see this, start with a planar binary rooted tree with n terminal vertices and draw it so that all its edges run either southwest or southeast. Now collapse all the southeast-running edges: the result is a planar rooted tree with n total vertices. To illustrate, here's the case $n = 4$, with each planar binary rooted tree (terminal vertices with open circles) above its collapsed version:



It may seem that we've lost information collapsing all those edges, but if you look at the examples carefully you'll see that the process is reversible. The edges of the collapsed tree correspond to the southwest-running edges of the binary tree, while the vertices of collapsed tree correspond to chains of southeast-running edges of the binary tree. The number of outgoing edges of a vertex v of the collapsed tree equals the number of non-terminal vertices in the chain of southeast-running edges of the binary tree that was collapsed into v .

Thus, we have isomorphisms of the following sets:

$$\begin{aligned} \{\text{BBA's of weight } n\} &\cong \{\text{planar rooted trees with } n + 1 \text{ vertices}\} \cong \\ &\quad \{\text{planar binary rooted trees with } n + 1 \text{ terminal vertices}\} \cong \\ \{\text{parenthesized products of } n + 1 \text{ letters}\} &\cong \{\text{triangulations of the } (n + 2)\text{-gon}\} \end{aligned}$$

and all have C_n elements.

Exercise 3. Using these isomorphisms, find the BBA of weight four that corresponds to the parenthesized product $(a * b) * ((c * d) * e)$.

We have only scratched the surface of the subject here: Richard Stanley has compiled a list of 100 types of objects that are counted by Catalan numbers! (The first 66 occur as an exercise in his book *Enumerative Combinatorics*.)