TIKI TAKA WITH CATALAN NUMBERS

Bharath Krishna S

Integrated PhD 1st year

School of Mathematics

IISER Thiruvananthapuram

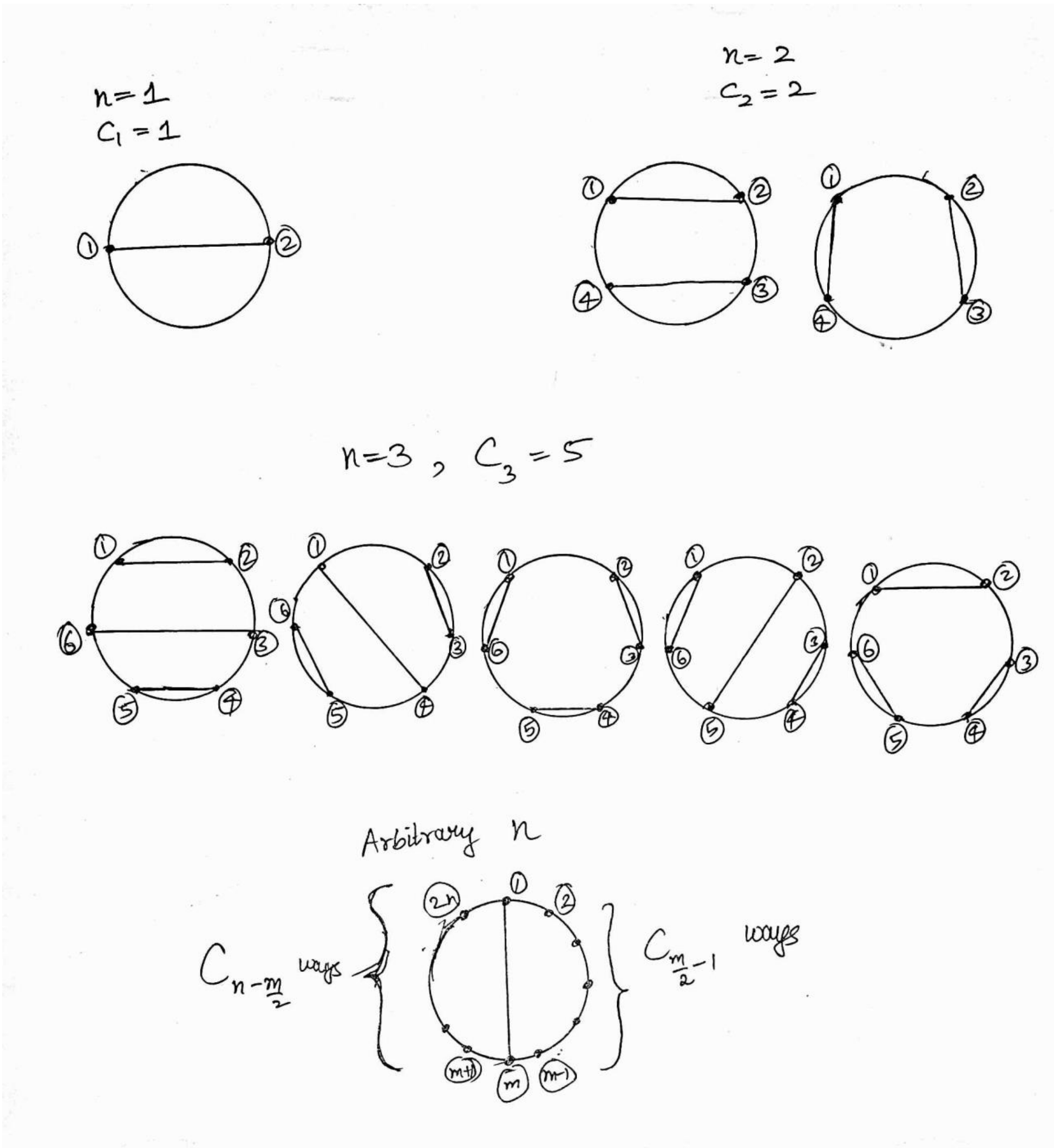
bharathmaths21@iisertvm.ac.in

Imagine solving an exercise problem in which there are 66 sub divisions. Yes,you read it right! That is exactly what Exercise 6.19 of Richard Stanley's Enumerative Combinatorics asks one to do. The beauty of the exercise is that, all those questions are related to one concept – Catalan numbers. Stanley then went on to write a book titled Catalan numbers which contained more than 200 applications of this famous sequence of numbers. What is so special about this sequence that it has a monologue dedicated to it? Let's explore and have fun.

We begin with a real life motivation. A group of friends (even number of them) meet in a get together event. They start playing a game. The game involves them to sit in a circular formation and shake hands. The main rule is that one person should be involved in only one handshake. Also all of them should shake hands(these two conditions forces us to have even number of people). One of those excited persons in that group,in an attempt to make the game interesting,actually brings another restriction to the game. They don't want two pairs of hands crossing each other,while shaking hands. The players now are interested in finding the number of ways in which,they can do this task. So, let us help them in this endeavor.

Let us assume that there are n pairs of players in the game,where n is a natural number. Let us name the players $1,2,3,\dots\dots,2n$. First note that,1 can't shake hands with an odd number. Doing this will either leave one person without a handshake or one person shaking hands with two people or leads to intersection of a pair of hands, all of which aren't desired due to the given conditions. In fact there is nothing special about 1. This observation shows that the parities of the two numbers involved in a handshake must be different. Let C_n denote the number of possible handshakes of above mentioned form,in a group of $2n$ people. Let us say 1 shakes hands with m, where m is even. Clearly, $2\ to\ m-1$ (case 1) have to shake hands among themselves, for if they don't,it will lead to crossing of handshakes. The same holds good for $m+1\ to\ 2n$ (case 2). By finding the number of pairs of hands in each case,it is clear that, there are $C_{\frac{m}{2}-1}$ possible handshakes in Case 1 and $C_{n-\frac{m}{2}}$ possible handshakes in Case 2. Thus the total number of handshakes in which 1 shakes hand with m is $C_{\frac{m}{2}-1}C_{n-\frac{m}{2}}$. But the m we chose was arbitrary and $m \in A =$

$\{2,4,6,\ldots,2n\}$. This gives the required number of handshakes to be $C_n = \sum_{m \in A} C_{\frac{m}{2}-1} C_{n-\frac{m}{2}}$. Many a times,a picture is worth thousand words,so the above idea is summarized in the following illustration.



The illustration also highlights another fact – the hand shake problem is equivalent to finding the number of chords in a circle containing $2n$ points in its boundary such that no two chords cross each other. As George Polya writes in "How to Solve It", "Analogy pervades all our thinking". This method showcases the power of analogies and geometric approach in problem solving.

One of the basic properties while defining structures like groups,rings,etc is associativity. The binary operation used to define them have to necessarily be associative. However,there do exists binary operations that are not associative – for example consider subtraction in $\mathbb{Z}$. When a binary operation * on a set A is associative, $a * (b * c) = (a * b) * c$, for all $a, b, c \in A$. If not, then these two values are different, for some $a, b, c \in A$. Now we are interested in finding the number of such distinct combinations possible in $n + 1$ symbols, when these symbols are not associative under *. We need $n \geq 2$ to deal with associativity. The case $n = 3$ gives 2 possible combinations namely $a * (b * c), (a * b) * c$. For $n = 4$, there are 5 possiblities – $(((a * b) * c) * d), ((a * (b * c)) * d), (a * ((b * c) * d)), (a * (b * (c * d))), ((a * b) * (c * d))$. Now,let us proceed to generalising things. Let C_n denote the required number of combinations on $n + 1$ symbols. Each bracketing can be split into two parts(the first bracket acts as a split)– one part contaning k and the other part containing $n - k + 1$ terms. By our notation, the number of bracketings in the first part is C_{k-1}, while that for the second part is C_{n-k}. For each bracketing in the first part,there are C_{n-k} bracketings in the second part and also using the fact that k is arbitrary(note that k can't be $0\ or\ n+1$ as such a split will not make sense of associativity), we get,

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

Now, we shall to find an explicit formula for C_n. A standard tool used in combinatorics in these types of situations is generating functions. As the name suggests, this method will help us to "generate" the n^{th} Catalan number. We won't go into the deep intricacies of this concept, but as far as this article is concerned,the following definition will suffice. Let $\{a_n\}$ be a sequence. Then $\sum_{n=0}^{\infty} a_n x^n$, is actually a polynomial in *x* and this is called the generating function of the sequence $\{a_n\}$. We will try to simplify the generating function of the sequence of Catalan numbers, in pursuit of the required formula. Let,

$$F(x) = \sum_{n=0}^{\infty} C_n x^n \qquad (2)$$

Intuitively, from the handshake problem, it is clear that $C_0 = 1$(only possibility is the empty handshake!) and $C_1 = 1$(refer illustration). So,we take these as the initial conditions of our generating function. Now,

$$
\begin{aligned}
\left(F(x)\right)^2 &= \left(\sum_{n=0}^{\infty} C_n x^n\right)\left(\sum_{n=0}^{\infty} C_n x^n\right) \\
&= (C_0 + C_1 x + C_2 x^2 + \cdots)(C_0 + C_1 x + C_2 x^2 + \cdots) \\
&= {C_0}^2 + (C_0 C_1 + C_1 C_0)x + \left(C_0 C_2 + {C_1}^2 + C_2 C_0\right)x^2 + \cdots \\
&= {C_0}^2 + C_2 x + C_3 x^2 + C_4 x^3 \ldots + C_{n+1} x^n + \cdots
\end{aligned}
$$

$$x\left(F(x)\right)^2 = F(x) - 1 \qquad (3)$$

The above simplification is done using (1) and (2) and has reduced the problem into a quadratic equation in the variable $F(x)$. Solving (3),we get

$$F(x) = \frac{1 \pm \sqrt{1-4x}}{2x} \qquad (4)$$

Here, we use the identity,

$$(1 + x)^{\alpha} = \sum_{n \geq 0} \binom{\alpha}{n} x^n \qquad (5)$$

where $\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\ldots\ldots\ldots(\alpha-n+1)}{n!}$ for any real number α. Note that this is a generalized version of binomial theorem. If $\alpha \in \mathbb{N} \cup \{0\}$, this reduces to the usual form of binomial theorem. Thus (4) becomes,

$$F(x) = \frac{1}{2x}(1 \pm \left(1 + \binom{1/2}{1}(-4x) + \binom{1/2}{2}(-4x)^2 + \binom{1/2}{3}(-4x)^3 + \cdots\right)) \qquad (6)$$

$F(x) = \frac{1+\sqrt{1-4x}}{2x}$ isn't the required solution as there is no $\frac{1}{x}$ term in (2), but it occurs in (6). So,

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

$$\therefore F(x) = \frac{1}{2x}\left(1 - \left(1 + \binom{1/2}{1}(-4x) + \binom{1/2}{2}(-4x)^2 + \binom{1/2}{3}(-4x)^3 + \cdots\right)\right) \quad (7)$$

We have constructed $F(x)$ such that, the coefficient of x^n gives C_n. Thus from (7), we get,

$$C_n = \frac{1}{2}\left(\frac{1}{2}\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)\left(\frac{5}{2}\right) \ldots \ldots \left(\frac{2n-3}{2}\right)\left(\frac{2n-1}{2}\right)\right)\frac{(4)^{n+1}}{(n+1)!}$$

$$= \frac{(1.3 \ldots \ldots (2n-1)(2)^n)}{(n+1)n!}$$

$$= \frac{(1.3.5 \ldots \ldots (2n-1).1.2.3 \ldots \ldots (n)(2)^n)}{(n+1)(n)!\,(n)!}$$

$$= \frac{(2n)!}{(n+1)(n)!\,(n)!}$$

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

We now have an explicit formula to calculate the nth Catalan number.

Necessity is the mother of all inventions. So what prompted Mathematicians to define this sequence? We dig deep into the time capsule and explore the history of this sequence. Mongolian mathematician Ming Antu has written some trigonometric identities involving Catalan numbers in his book that came out in 1730s. Then came Leonhard Euler’s correspondence with Christian Goldbach and Johann Segner.

While this article began with a real life motivation, the problem that motivated Euler to define this sequence was a one which is similar to the handshake problem. Euler was interested in finding the number of triangulations of a (n+2) sided polygon such that none of the diagonals intersect. This lead to a series of letters between Euler and Goldbach, which resulted in the generating function for C_n. But Euler couldn't proceed further. His work was simplified by his correspondence with Segner. Segner had found the formula in (1) by a combinatorial approach and used it to calculate C_n for $n \leq 18$. But he had made a mistake while computing C_{13}, which was spotted by Euler. Euler went on to summarise this work in a paper and gave credit to Segner. But some versions claim that, unlike his correspondence with Goldbach,Euler wasn't open to discussing his findings with Segner. Rather he just wanted to "test" Segner. Nevertheless,Segner responded to the challenge well. One must now be wondering how the sequence got its name. It is named after Eugene Catalan, who was the first person to give an explicit formula for C_n. This name was given by American Mathematician John Riordan. Before that, these were known as Segner numbers or Euler Segner numbers for obvious reasons.

Catalan numbers are used in various branches of Mathematics. Some of them are:

- The number of binary trees on n vertices is given by C_n.

- For finding the number of certain type of paths in a $\mathbb{N} \times \mathbb{N}$ grid including the origin(Refer exercises)
- Some counting problems involving sets,sequences,groups and posets.
- Counting problems related to Young Tableux, which is widely used in Representation theory.

Hope this was a beautiful exposition into the world of Catalan numbers. Before, we sign off, a few exercises for the readers to try and the references:

- ❖ Consider the $\mathbb{N} \times \mathbb{N}$ grid including the origin. What is the number of increasing paths in the grid from $(0,0)\ to\ (n,n)$ that doesn't cross the line $y = x$?
- ❖ Prove that the number of binary search trees on n vertices is C_n
- ❖ Let A=(a_{ij}) be a $n \times n$ matrix such that $(a_{ij}) = C_{i+j-2}$. What is the determinant of this matrix?

References:

1. "Catalan Numbers",Richard Stanley
2. "Enumerative Combinatorics", Richard Stanley
3. "Notes on Introductory Combinatorics",George Polya,Robert Tarjan and Donald Woods
4. "Catalan numbers for the Classroom?", *Elemente der Mathematik* ,Judita Cofman
5. Web references from YouTube, Wikipedia, Math Stackexchange and Brilliant.