# Decoding Catalan Numbers

- Honey Harshit Srivastava

(Last edited: 11th April 2026)

Dedicated to my loving parents

Shri Saroj K. Srivastava (father) and Smt. Vijai Raje Srivastava (mother)

# Table of Contents

## 1.1 Introduction

**"Mathematics rightly viewed, possesses not only truth, but supreme beauty."**

**-Bertrand Russell**

We have all studied mathematics at some stage in our lives. Many of us love it more than anything while some of us cannot stand the subject for several reasons. Today however, under the scope of this literary work, I will make a genuine attempt to show that love it or hate it mathematics is indeed supremely beautiful.

To prove the point, we will take a deep dive into the branch of combinatorics. Combinatorics is the field of mathematics that focuses on counting, arranging, and selecting finite discrete structures without any operating all of them (mathematical out of counting).

**LET'S GET COUNTING!!**

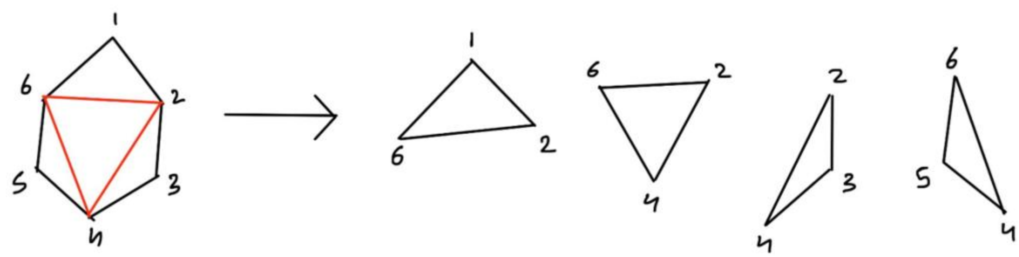
## 1.2 Motivation (family of counting problems)

We will now discuss some combinatorial problems that lead to Catalan numbers.

## Problem 1: Triangulation of a polygon

**Triangulation:** Polygon triangulation is a process of dividing a polygon into sets of triangles using their non intersecting diagonals. It is trivial that in this case, vertices of each triangle must also be the vertices of the original polygon (as we are using non intersecting diagonals only). For this matter we will talk exclusively about convex polygons.

Example:

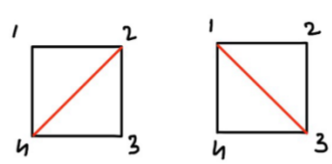


The above figure shows one triangulation of a hexagon using diagonals (2,4) (6,2) and (4,6).

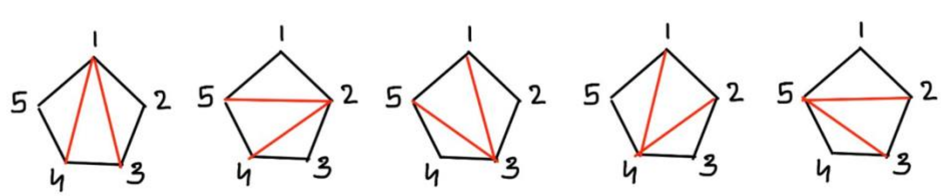
It is also worth observing that to triangulate a convex polygon of m sides, we need (m-3) non intersecting diagonals and this process yields (m-2) triangles.

**Number of triangulations of an (n+2) gon**

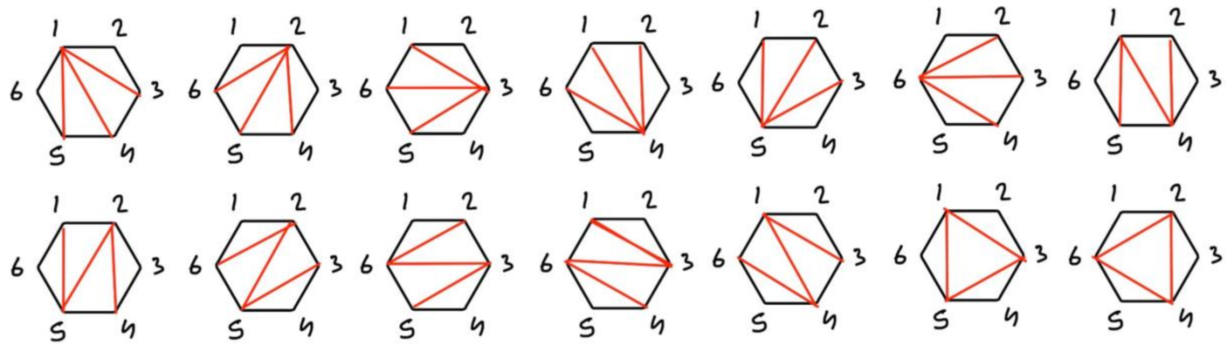
Each polygon can be triangulated in multiple ways. But in how many ways exactly can we triangulate a polygon with (n+2) sides? Let's look at few examples.



2 ways to triangulate a quadrilateral.



5 ways to triangulate a pentagon.



14 ways to triangulate a hexagon.

So, what is the pattern here? How do we find out the exact number of triangulations for an (n+2) gon? We will discuss more about this in the next section (1.3).

## Problem 2: Dyck words and Dyck paths

**Dyck words:** Dyck words are sequences made up of two types of characters (X and Y) that satisfy the conditions mentioned below:

a) Total number of X's = Total number of Y's (say n)

b) At any point in the sequence, the number of X's is greater than or equal to the number of Y's (#X>=#Y)

**XYXXYXYY:** a dyck word of n=3

**XYYXYXXY:** not a dyck word as #X>#Y after 3 characters

In a variation of the same problem, we construct a valid parenthesis using "("and ")". It is structurally identical to the Dyck word problem as each opening bracket "(" has a corresponding closing one ")" in a valid parenthesis.

**XYXXYXYY: ( ) ( ( ) ( ) )**

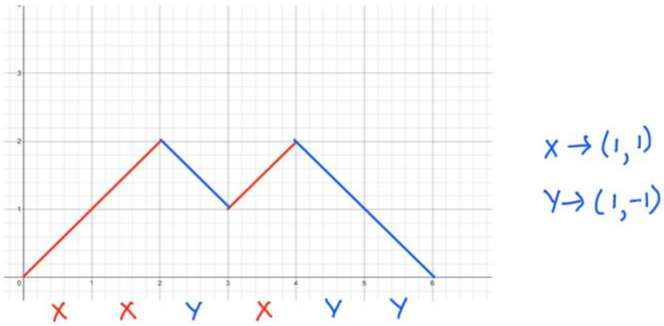
**Dyck paths:** Another problem similar to Dyck words is that of Dyck paths. A Dyck path of semi-length "n" is a path that satisfies the following conditions:

a) Starts at origin (0,0)

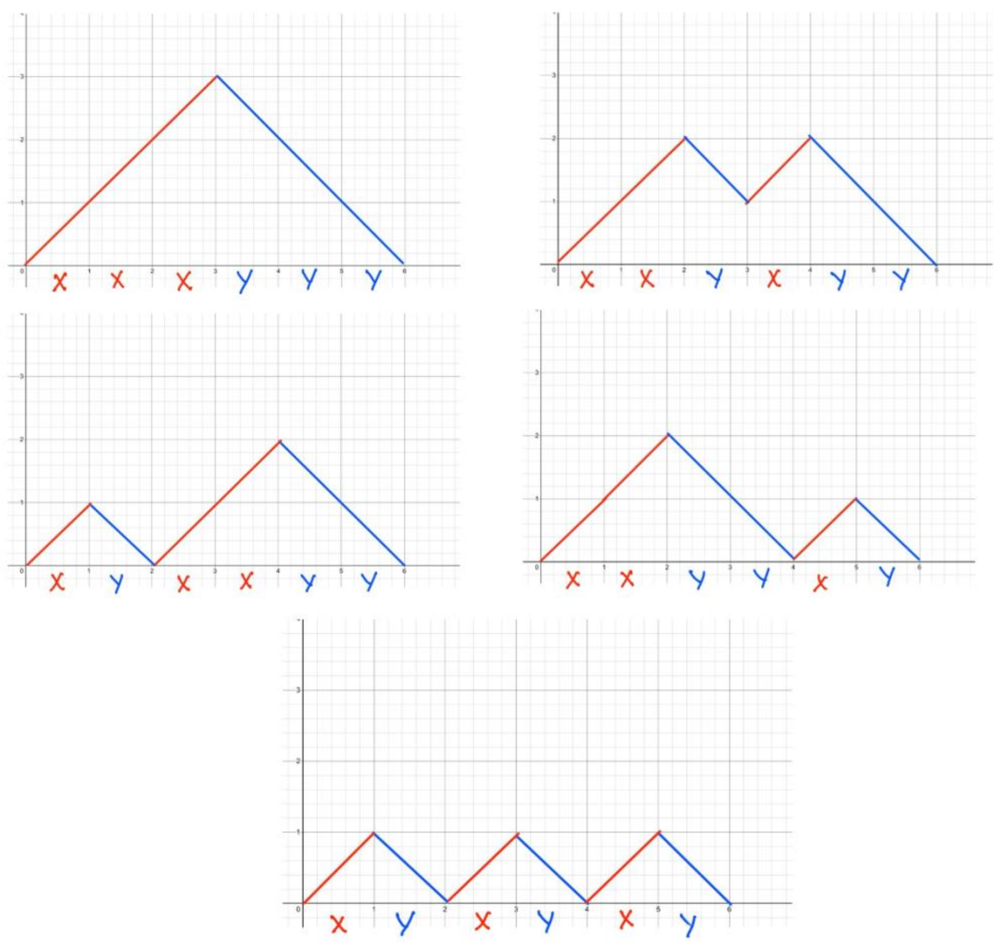
b) Ends at (2n,0)

c) Uses one two types of moves (up and down)

d) Never goes below X-axis




It is trivial that each Dyck word can also yield a Dyck path. The figure given below shows Dyck words of n=3 and their corresponding Dyck paths:



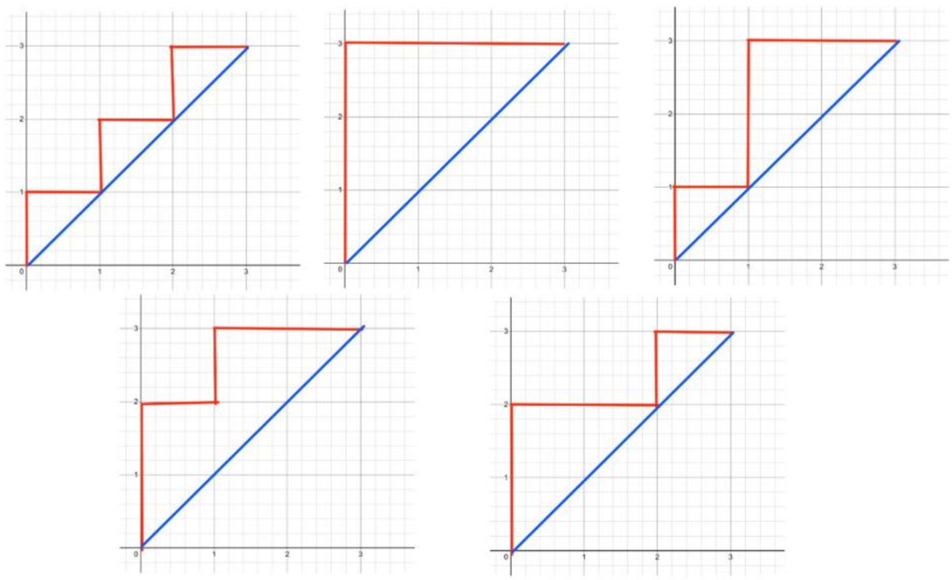
We meet the same question once again, how many Dyck words/Dyck paths of semi-length “n” can exist? We will come back to this in section 1.3.

### Problem 3: Lattice paths

In the lattice paths problem, we count the number of paths from (0,0) to (n,n) in a lattice that follow the given conditions:

a) Only use type types of moves (up and right)

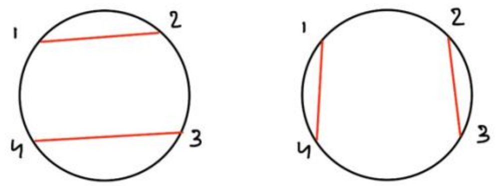
b) Don’t cross the diagonal (stay on or above the diagonal)



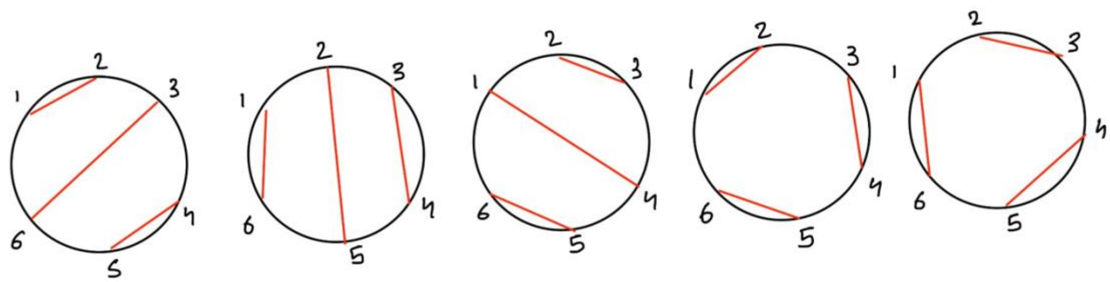
The figure shows the five possible lattice paths from (0,0) to (3,3). We will discuss about the number of lattice paths from (0,0) to (n,n) in the next section.

### Problem 4: Non intersecting curves and round table handshakes

In a round table conference, there are 2n people sitting around a circular table. We want to find the number of ways in which they can shake hands with each other such that hands don't cross each other. Each person must participate and can only shake one hand at a time. Let's look at a few examples:

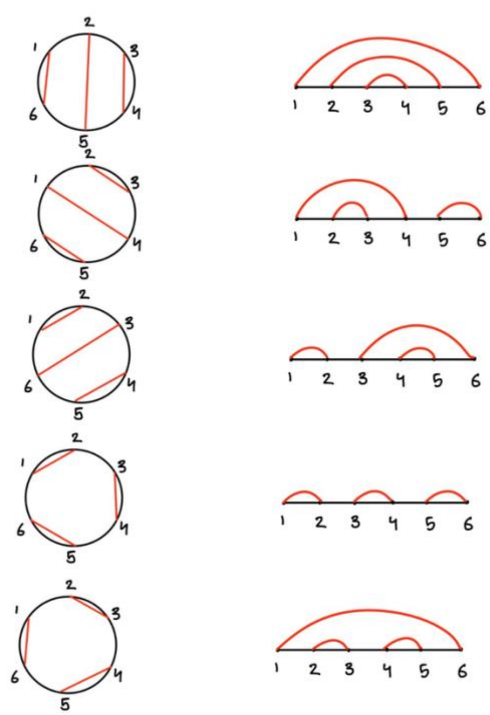


Two ways in which 4 people can shake hands. (n=2)



Five ways in which 6 people can shake hands. (n=3)

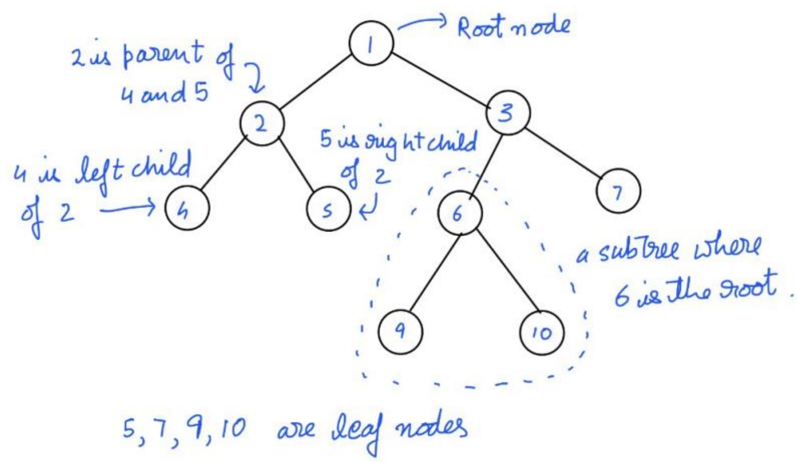
Another way to present the same question is by the method of non intersecting curves on a line with 2n points.



We will come back to this problem in the next section.

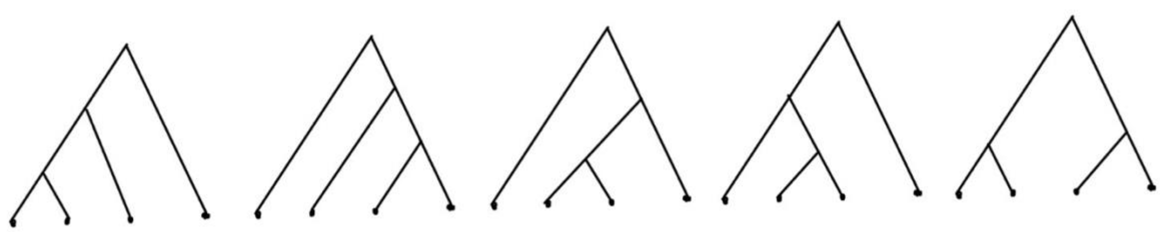
## Problem 5: Counting Binary Trees

A binary tree is a type of tree data structure where each node can have at most 2 children (left and right child). It would be helpful if we learn some basic terms related to a binary tree.




In a rooted binary tree, the nodes that have children are termed as "intermediate nodes" whereas the ones that don't have children are termed as "leaf nodes".

It is trivial that a rooted binary tree with n leaf nodes will have (n-1) intermediate nodes (each node has 0 or 2 children). Given below are the binary trees having 4 leaf nodes:



We can draw five binary trees that have 4 leaf nodes. But how do we count the number of trees having (n+1) leaf nodes? You guessed it, we will come back to this problem in section (1.3).

## 1.3 Self Similarity and Recursion

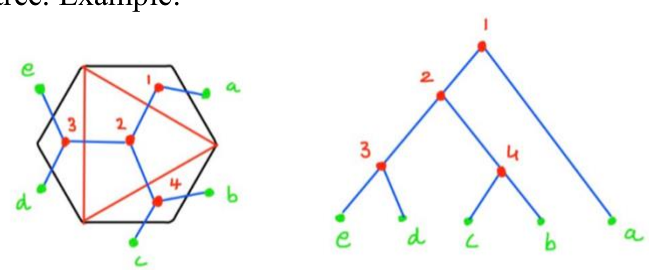
In this section, we will take a deep dive into the structure of these problems and study the hidden patterns at work. But first, let's study how these problems are all similar to each other.

**Triangulation of (n+2) gon and Counting Binary Trees**

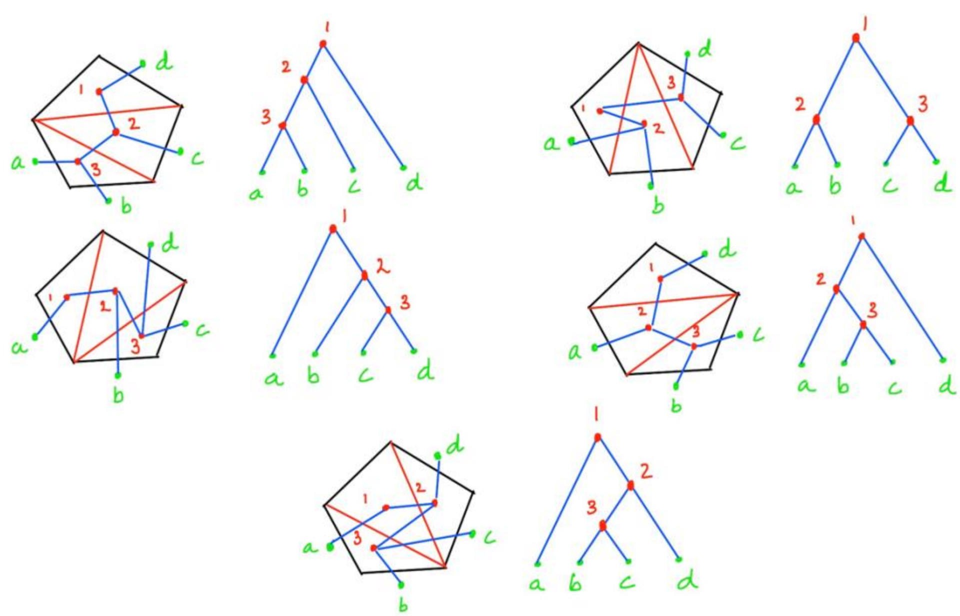
Consider a polygon of (n+2) sides can be triangulated in $T_{n+2}$ ways.

As for the number of rooted binary trees with (n+1) leaf nodes, we will represent it by $B_{n+1}$.

Triangulation of an (n+2) gon produces n triangles and uses (n+1) non intersecting diagonals in the process. Also, we know that binary tree having (n+1) leaves has n intermediate nodes. We can thus find a bijection between the two problems by considering each triangle to be an internal node of the corresponding tree. Example:



The above figure shows one possible triangulation of a hexagon and a corresponding binary tree.



The above figure shows the five triangulations of a pentagon and their corresponding binary trees having 4 leaf nodes. This bijection shows that for each triangulation of an (n+2) gon, there is a corresponding binary tree with (n+1) leaf nodes.

$$T_{n+2}=B_{n+1}$$

**Rooted Binary Trees and Dyck Words**

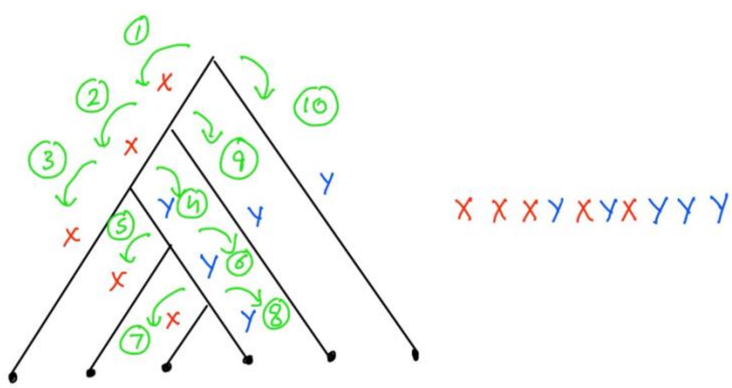
We will consider the number of Dyck words of semi-length n to be $D_{n.}$

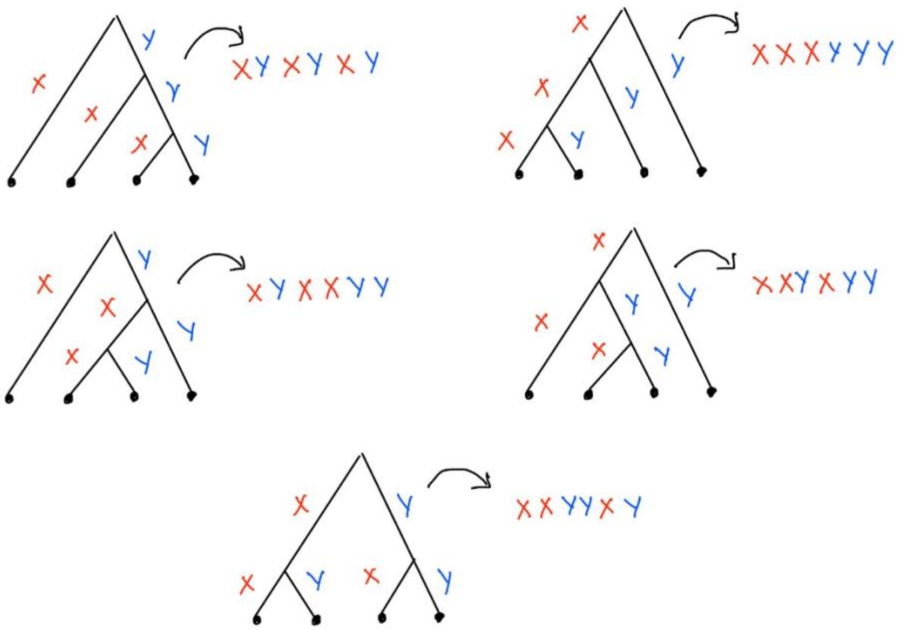
$B_{n+1}$ = number of binary trees with (n+1) leaf nodes.

There exists another beautiful bijection between the Dyck words and Binary trees. We already know that a binary tree having (n+1) leaf nodes has n intermediate nodes (each intermediate node has 2 children). Therefore, we can conclude that such a binary tree consists of (2n) branches. In a binary tree, we name the left branches "X" and right branches as "Y". We now begin a process that will convert our binary tree into a Dyck word.

1) We start at the root node and take the first step as left (X)
2) If we reach an intermediate node, we will take left again and will continue to do so until we reach a leaf node.
3) Once we reach the leaf node, we come back to the root of the tree and take the alternative right paths and continue to do so until we complete this process on all subtrees.

The process may sound a bit confusing, let's look at an example for better understanding.






The above image shows Binary Trees of 4 leaf nodes and their corresponding Dyck words. Therefore, we can conclude that for every Binary tree with (n+1) nodes, we can form a corresponding Dyck word of semi-length n.

$$D_n = B_{n+1}$$

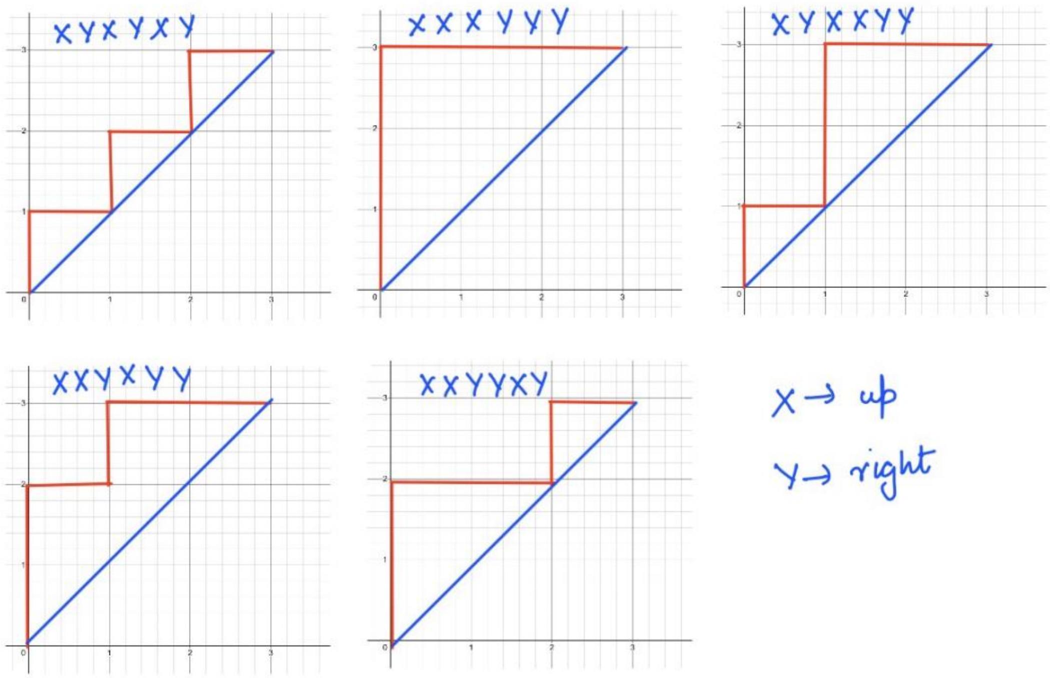
**Dyck words and Lattice Paths**

We consider the number of lattice paths from (0,0) to (n,n) to be $L_{n.}$

It is relatively easy to connect the concepts of Dyck words and lattice paths.

We know that a lattice path doesn't cross the diagonal at any stage (the number of "up" moves is always greater than or equal to the number of "right" moves in the path). Therefore, each Dyck word of semi-length n can yield a lattice path from (0,0) to (n,n).

$$D_n = L_n$$

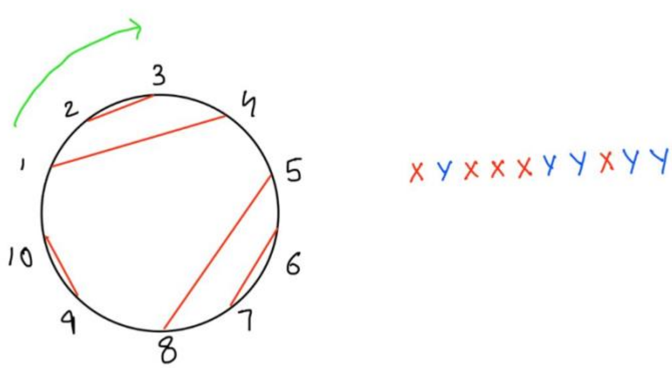



Lattice paths from (0,0) to (3,3) and corresponding Dyck words.

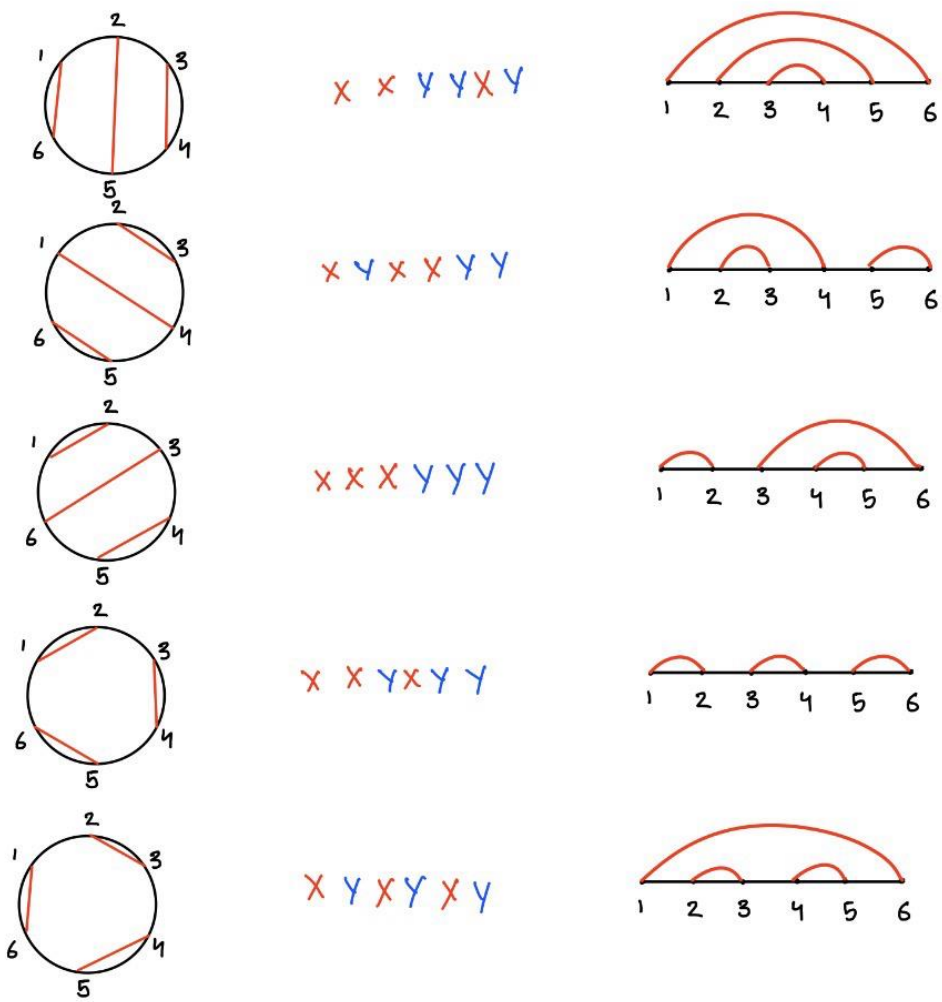
**Dyck Words and Round Table handshakes**

Consider the number of possible handshakes to be $R_n$. It is also relatively straightforward to connect the two problems. We follow the given process:

1) Start at person 1 on the table, moving in clockwise direction.
2) When you meet the first person involved in a handshake, mark him as X
3) When you meet the other person involved in the handshake, mark him as Y



In the figure given below, we see the handshakes of 6 people and their corresponding Dyck words.



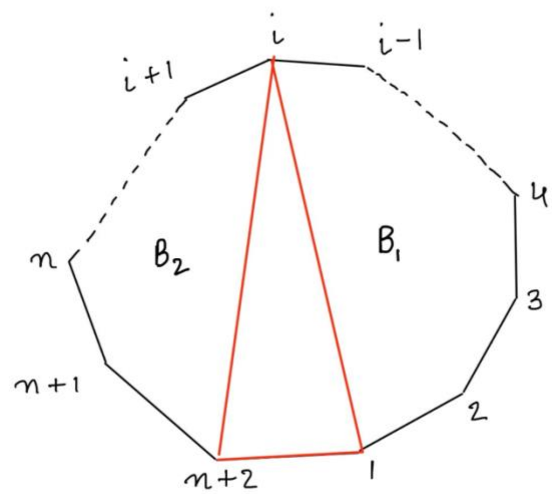

We have now connected the given problems. Let's take a look at the recursion now.

Suppose $P_1$ is a polygon having (n+2) sides. We fix the side [1, (n+2)] as our base.

We join our base side to a vertex "i" to form a triangle whose vertices are 1, (n+2) and i.

This triangle divides the polygon into two different polygons $B_1$ and $B_2$

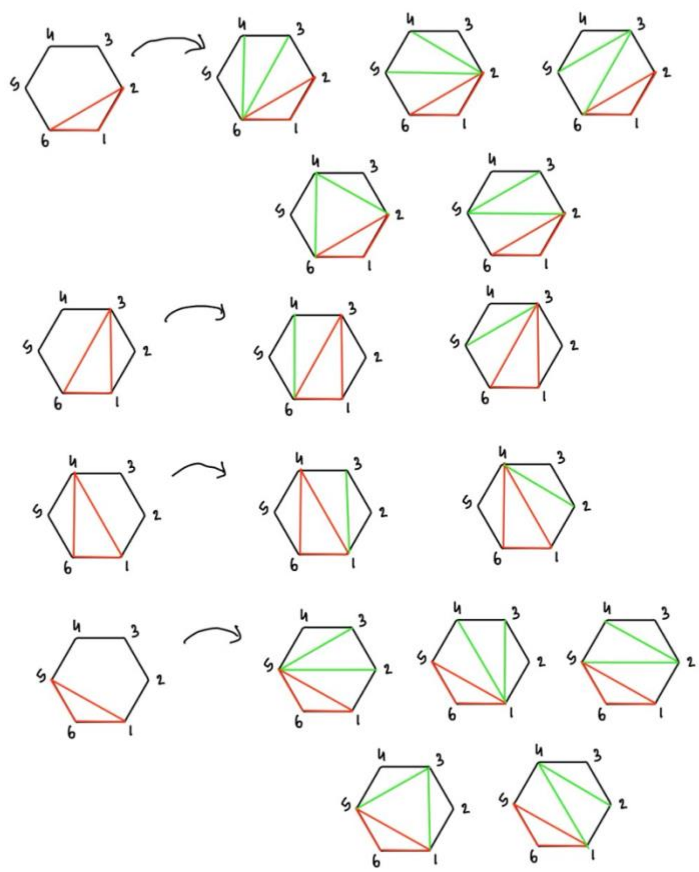
It is obvious that $B_1$ has "i" sides while $B_2$ has (n+3-i) sides.




The number of triangulations that have the triangle of vertices 1, i, (n+2) =

(Triangulations of $B_1$)*(Triangulations of $B_2$)

To be clearer, let's take a look at a hexagon



$$T_6 = T_2 \cdot T_5 + T_3 \cdot T_4 + T_4 \cdot T_3 + T_5 \cdot T_2$$

Therefore, the total number of triangulations of $T_{n+2}$=

$$T_{n+2} = T_2 T_{n+1} + T_3 T_n + T_4 T_{n-1} \cdots\cdots$$

$$T_{n+2} = \sum_{k=2}^{n+1} T_k T_{n+3-k}$$

We know the number of triangulations of (n+2) gon is counted by the n th Catalan number, the recurring relation of $C_n$ becomes:

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + C_2 C_{n-3} \cdots\cdots C_{n-1} C_0$$

$$C_n = \sum_{i=0}^{n-1} C_i C_{n-1-i}$$

This is because we know $T_{n+2} = C_n$ (shown previously)

## 1.4 General formula for n th Catalan Number

To find the general formula for the n th Catalan Number, we will use some calculus.

Consider a generating function G(x) for Catalan numbers

$$G(x) = \sum_{n \geq 0} C_n x^n$$

Using the recursive relation, we derived earlier,

$$C_n = \sum_{i=0}^{n-1} C_i \cdot C_{n-1-i} \quad \text{or} \quad C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$$

We take summation of both sides for all positive integer values of n

$$C_{n+1} x^n = \left( \sum_{i=0}^{n-1} C_i C_{n-i} \right) x^n$$

$$\sum_{n \geq 0} C_{n+1} x^n = \sum_{n \geq 0} \left( \sum_{i=0}^{n-1} C_i C_{n-i} \right) x^n$$

On the right-hand side, we have a Cauchy sum

$$\sum_{n\geq 0} C_{n+1}x^n = \sum_{n\geq 0}\left(\sum_{i=0}^{n-1} C_i\, C_{n-i}\right)x^n$$

$$\sum_{n\geq 0} C_{n+1}x^n = \left(\sum_{n\geq 0} C_n x^n\right)\left(\sum_{n\geq 0} C_n x^n\right)$$

$$\frac{1}{x}\left(\sum_{n\geq 0} C_{n+1}x^{n+1}\right) = G(x)\cdot G(x)$$

$$\frac{1}{x}\left[G(x) - C_0\right] = G^2(x)$$

$$xG^2(x) - G(x) + 1 = 0$$

On solving for G(x), we get:

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

we know $G(0)=1$ as $C_0 = 1$

$\therefore$ we take $\boxed{G(x) = \frac{1-\sqrt{1-4x}}{2x}}$

We use the binomial theorem to conclude that

$$(1+a)^n = \sum_{r=0}^{n}\binom{n}{r}a^r$$

$$\binom{1/2}{r} = \frac{(-1)^{r-1}\cdot 2\cdot (2r-2)!}{r\cdot 4^r\cdot (r-1)!\,(r-1)!}$$

$$(1-4x)^{1/2} = -\sum_{r\geq 0}\frac{2}{r}\binom{2r-2}{r-1}x^r$$

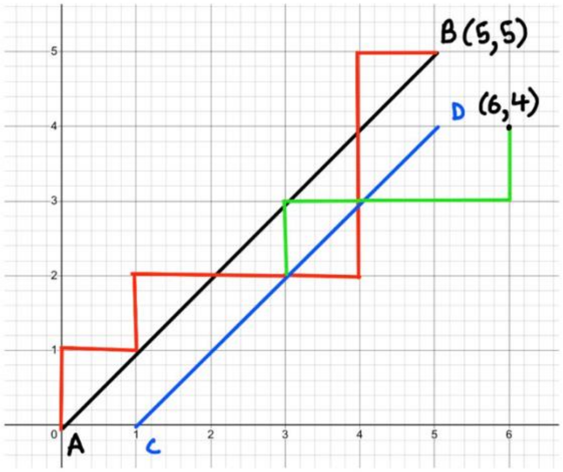
Therefore, the coefficient of $x^{r-1}$ in G(x) is

$$C_{r-1} = \frac{1}{r}\binom{2(r-1)}{(r-1)}$$

or

$$\boxed{C_r = \frac{1}{r+1}\binom{2r}{r}}$$

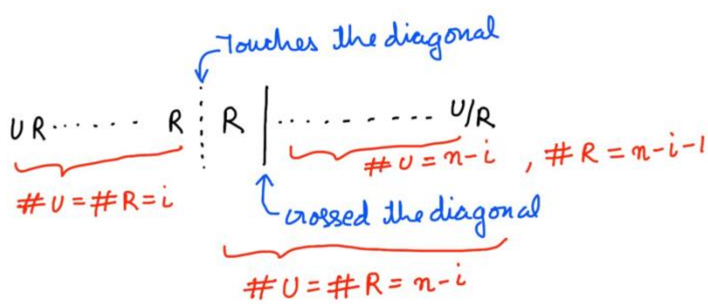
We have an alternate method to give the general formula for Catalan numbers. In this method, we will take the approach of lattice paths.




We know that the total number of paths from (0,0) to (n, n)

$$\text{Total no. of paths} = \binom{2n}{n}$$

We will subtract the illegal paths (ones crossing the diagonal). Suppose the path crosses the diagonal at (i, i). It is trivial that the path can only cross the diagonal by taking a right move.




If we reflect the later section along CD, the number of Rs and Us will interchange

$$\text{(before)} \quad \begin{aligned} \#U &= n-i \longrightarrow \#U = n-i-1 \\ \#R &= n-i-1 \longrightarrow \#R = n-i \end{aligned} \quad \text{(after)}$$

In the new string thus formed after reflection,

$$\#U = (n-i-1)+i = n-1$$

$$\#R = (n-i)+(i+1) = n+1$$

Therefore, this new reflected path leads to (n+1,n-1). This is true for all in valid paths. We can thus count the number of invalid paths.

Total no. of valid paths = Total no. of paths − Total no. of invalid paths

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

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

We have now reached the end of our essay, and I would like to express my sincere gratitude to everyone who supported and inspired me during the course of writing this essay. The beauty of mathematics often reveals itself through patients and curiosity, and this work is a reflection of that journey. I hope this essay not only clarifies the ideas surrounding Catalan numbers but also sparks for the curiosity about elegance hidden within mathematics.

(1973 words approximately)