The Catalan Numbers and Their Applications

A Math Exploration

18th April, 2012

A Math Exploration

The Catalan Numbers and Their Applications

Introduction

This exploration will talk about, as the title already implies, the Catalan numbers and about how they are applied in mathematics. I will also explore how they can be applied in our lives and in mine as an IB learner. I chose to do this topic, because I believe it is one the many interesting topics that aren't taught at school. I always strive for new things to learn and I thought this exploration would be the perfect opportunity to add something new to my knowledge of mathematics.

As an IB student, I do many sport activities for CAS, one of them being basketball. It is one of my favorite sports and it's very challenging and competitive. I personally love shooting baskets and my favorite game is the one-on-one drill. I wondered if there was a way I could find a connection between my interest, basketball and this topic. So I changed the rules of the one-on-one game and came up with a problem that we could solve. In this game, you play alone against a single player on one basket. The goal is to reach a certain number of points by shooting and scoring before the other player does. If you score, you get the ball and it's your turn again. However, your opponent may steal the ball from you and shoot himself. If we assume that every scored basket is worth one point and we score every time we shoot (but not every time we're in possession of the ball), in how many possible ways am I never behind until I win, if I have to score n goals in order to win?

The answer is C_n, the ordinary Catalan numbers. We will prove this later after introducing the Catalan numbers. They are a sequence of natural numbers that appear in different counting problems. The first fifteen Catalan numbers are:

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900, 2674440$$

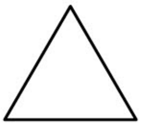
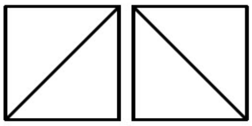
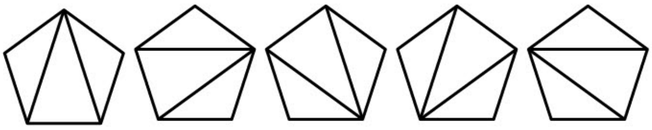
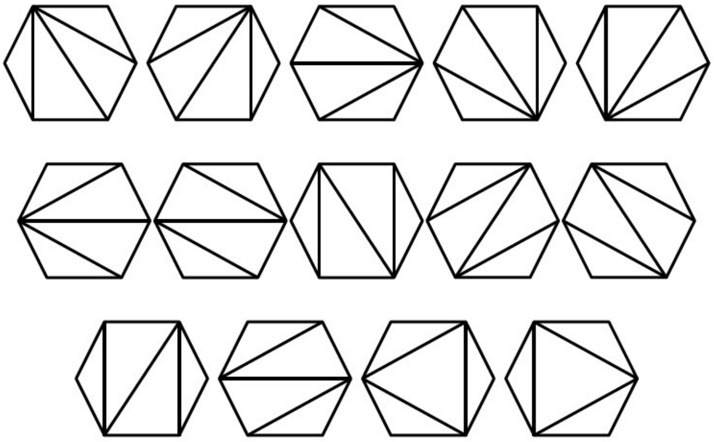
Although they seem to progress randomly, they do have a general formula by which they can be expressed:

$$C_n = \frac{1}{n+1}\binom{2n}{n}, \qquad n \in \mathbb{Z}^+$$ [1]

The interesting thing about these numbers is that they make up a single sequence which is based on a single formula, yet can be applied in so many ways. That's why in this exploration, I've decided to find the connection between the formula and the applications, and between themselves as well.

The Catalan numbers were named after Belgian mathematician Eugène Charles Catalan (1814–1894), although it was Leonhard Euler who first came up with the sequence in 1751 when trying to solve a problem by dividing a polygon into triangles. In fact, this problem Euler was trying to solve is one of the many interpretations of the Catalan numbers. It is said that the Catalan numbers represent the number of ways in which a regular polygon with $n+2$ sides can be divided into n triangles if different orientations are counted separately. For a clearer understanding, here is a table with illustrations.

Note: When $n=0$, there is one way to construct a triangle from the line, by not constructing it.

n	$n+2$	Polygon		C_n
0	2	Line		1
1	3	Equilateral triangle		1
2	4	Square		2
3	5	Regular pentagon		5
4	6	Regular hexagon		14

Other Interpretations

A few other interpretations are listed below and these are the ones we will be dealing with and we will try to prove.

- The number of ways you can form valid groupings of n pairs of parenthesis, where $n \in \mathbb{N}$. In other words each opening parenthesis **(** has a corresponding closing one **)**.

())(())(Non-valid grouping	(()()()) Valid grouping

n	**Parenthesis**	C_n
0	*	1
1	( )	2
2	( )( ), (( ))	2
3	( )( )( ), ( )(( )), (( ))( ), (( )( )), ((( )))	5
4	( )( )( )( ) (( ))( )( ) ( )(( ))( ) ( )( )(( )) (( )( ))( ) (( )( )( )) ( )(( )( )) (((( )))) ( )((( ))) (( )(( ))) ((( )( ))) ((( ))( )) ((( )))( ) (( ))(( ))	14

  - Note that here again, even if there are no parentheses, there is still one way in which you can arrange it, by not arranging it.

- the number of “mountain ranges” you can form with n up and down strokes, where $n \in \mathbb{N}$

$n = 0$

1 way

```
*
```

$n = 1$

1 way

```
/\
```

$n = 2$

2 ways

```
       /\
/\/\, /  \
```

$n = 3$

5 ways

```
                                  /\
           /\    /\      /\/\    /  \
/\/\/\, /\/  \, /  \/\, /    \, /    \
```

$n = 4$

14 ways

```
               /\,     /\      /\        /\  /\
/\/\/\/\, /\/\/  \, /\/  \/\, /  \/\/\, /  \/  \,

                                  /\      /\
 /\/\/\      /\/\    /\/\        /  \    /  \
/      \, /\/    \, /    \/\, /\/    \, /    \/\,

                                 /\
    /\      /\        /\/\      /  \
 /\/  \    /  \/\    /    \    /    \
/      \, /      \, /      \, /      \
```

  - Here again, even if there are no up and down strokes, there is still one way in which you can form a mountain range, by not forming it.

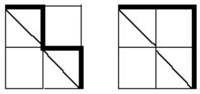
- The number of paths from one corner to the opposite corner of an $n \times n$ grid that doesn't cross the diagonal, where $n \in \mathbb{N}$

$n = 0$
1 way

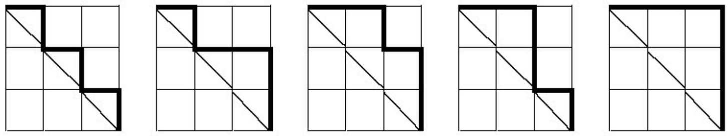
*

$n = 1$
1 way

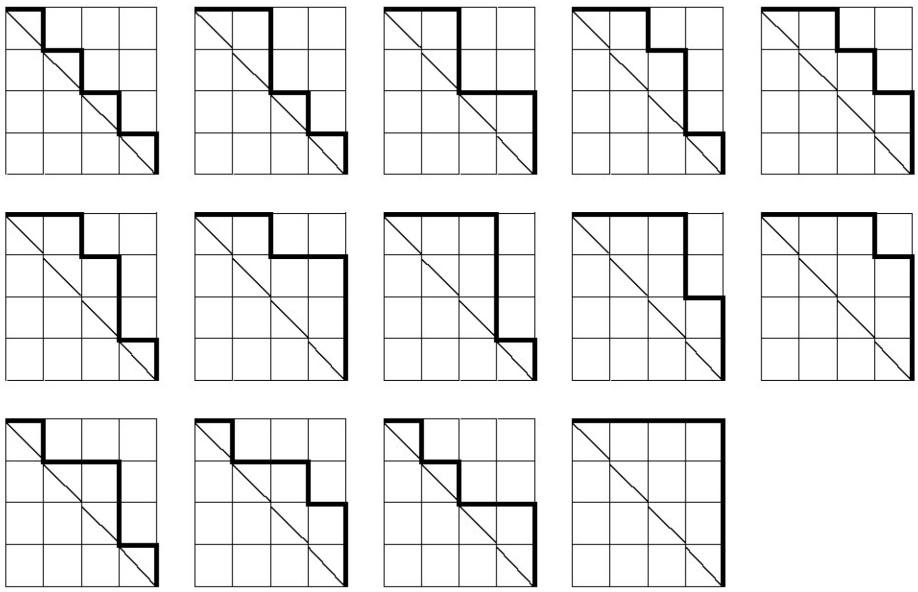
$n = 2$
2 ways



$n = 3$
5 ways



$n = 4$
14 ways

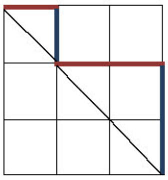


These are only a few of many other interpretations of the Catalan numbers.

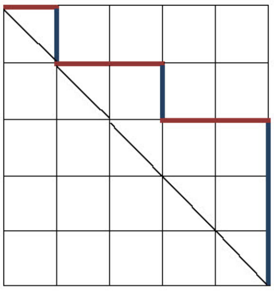
Proof

We can start by proving the grid method, that the number of paths from one corner to the opposite corner of an $n \times n$ grid that doesn't cross the diagonal is the Catalan number, where $n \in \mathbb{N}$.

We want to find the number of paths that don't cross the. Let's call these paths 'good' path (P_G) and the paths that do cross the diagonal 'bad' paths (P_B). If the number of total paths is P_T, we know that $P_G = P_T - P_B$. In order to find P_T, let's say R represents one step to the right and D represents one step downward. We know that since the grid is a square, there have to be the same number of R's and D's (n of each). For example:



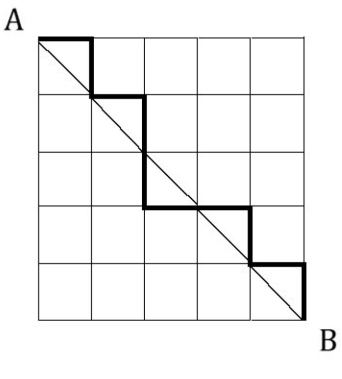
In this 3×3 grid, there are 3 steps to the right (red) and 3 steps downwards (blue).



This 5×5 grid has 5 steps to the right (red) and 5 steps downwards (blue).

Fig. 1

Therefore, for an $n \times n$ square, there are n D's and n R's, giving a total of $2n$ steps. We can write the path from A to B in terms of R's and D's as a sequence. For example:



$= RDRDDRRDRD$

Fig. 2

Therefore, P_T would be equal to the total number of possible arrangements of n Rs' and n Ds' in the sequence, which is:

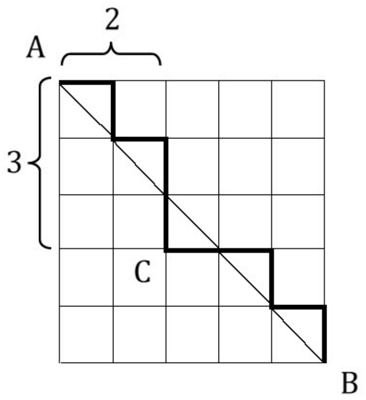
$$\binom{2n}{n} = \frac{2n!}{n!\,(2n-n)!}$$

For example, let $n = 5$.

$$\therefore P_T = \binom{10}{5} = \frac{10!}{5!\,5!} = 252$$

Every bad path will cross the diagonal at some point. Let's mark the point where the path ends as soon as it crosses the diagonal as C. Let the path cross the diagonal at the second R. This means that C is at the 2nd R and 3rd D (fig. 3).

The remaining path from C to B has 3 R's and 2 D's.



$R \rightarrow 2$
$D \rightarrow 3$

Fig. 3

Starting from the step after C, if we convert all R's to D and all D's to R, the new path which we'll call P' (the red path in fig. 4), will end at a new point. Let's call it B'.

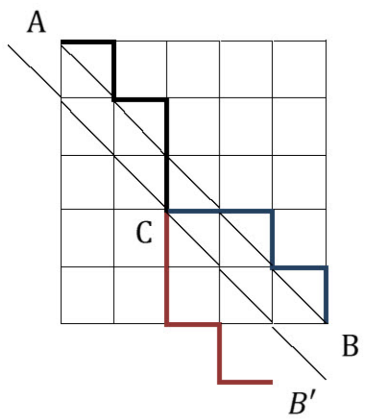
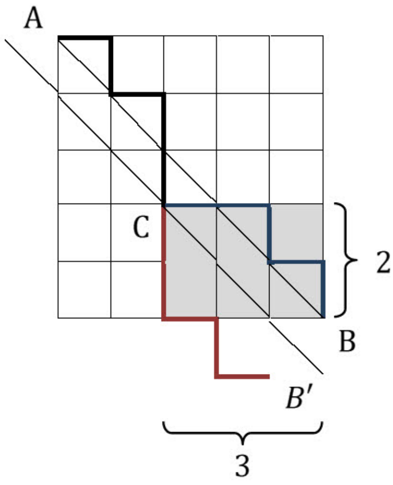


Fig. 4

Remaining grid from C to B is 3×2, where 3×2 is an ordered pair.



$R \rightarrow 5 - 2 = 3$
$D \rightarrow 5 - 3 = 2$

Fig. 5

The new remaining grid from C to B' is 2×3.

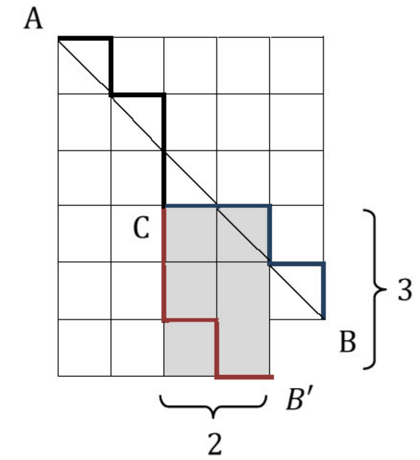


Fig. 6

Therefore the new full grid from A to B' is 4×6.

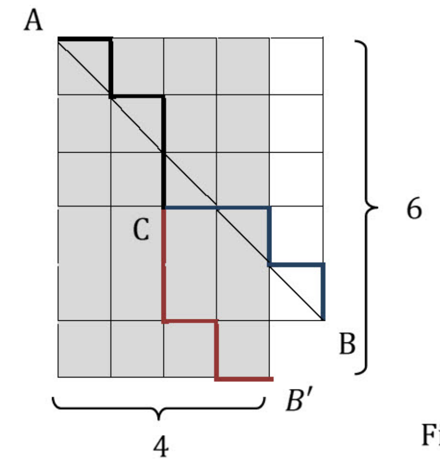


Fig. 7

Note that the new remaining grid has the same entries of the original grid, only their order has been changed. That is because by reflecting the remaining path, we also reflected the grid on F.

Our new path P', consists now of 4 R's and 6 D's and can be written as a sequence case $RDRDDDDRDR$. The total number of all possible arrangements of 4 Rs' and 6 Ds' in the sequence is:

$$\binom{6+4}{6} = \binom{10}{6} = 210$$

This is the total number of bad paths (P_B).

As mentioned before, the number of good paths $P_G = P_T - P_B$

$$\therefore P_G = \binom{10}{5} - \binom{10}{6} = 252 - 210 = 42$$

Conjecture

Now we want to find a conjecture for a general term. We want to find the number of paths that don't cross the diagonal from one corner of an $n \times n$ grid to the opposite corner, where $n \in \mathbb{N}$. We will illustrate the $n \times n$ grid by the grid in fig. 8, where the connecting dots represent all the possible values for k and n.

Assuming that the bad path crosses the diagonal at the k^{th} R, our point C will now be at the k^{th} R and $(k+1)^{th}$ D, where $k \in \mathbb{N}$, $k < n$ (fig. 8)

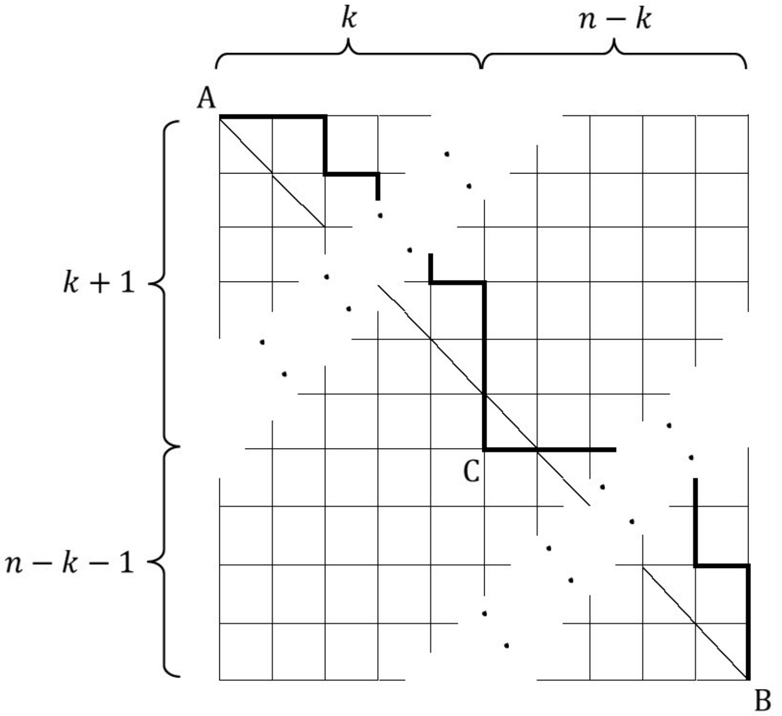


Fig. 8

Now, our remaining path from C to B has $(n-k)$ R's and $\left(n-(k+1)\right)$ D's.

Again, starting from the step after C, we convert all R's to D and all D's to R and gain the new path P' (the red path in fig. 9) which ends at B'.

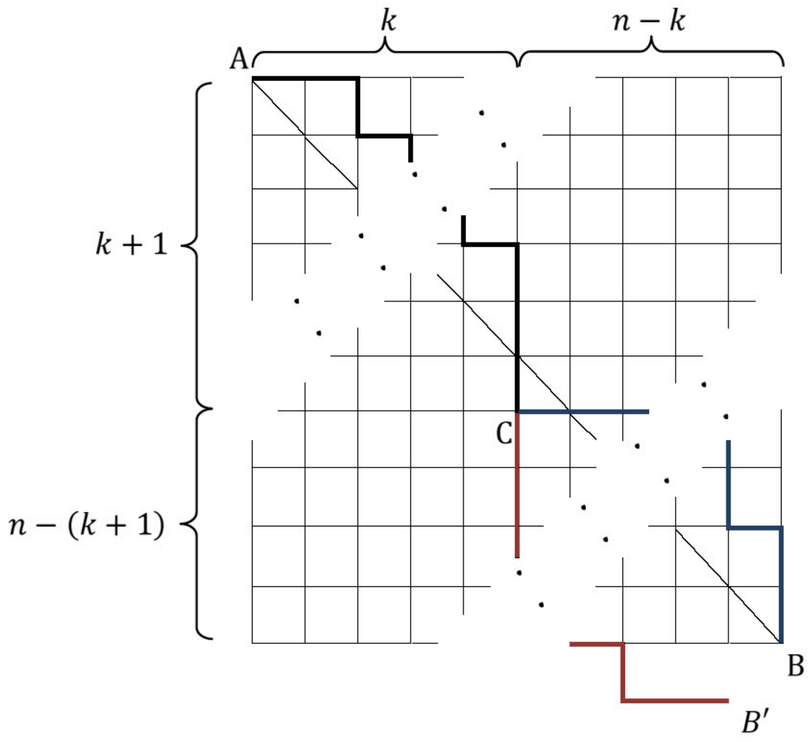


Fig. 9

Our remaining grid from C to B is $(n-k)\times(n-k-1)$, where $(n-k)\times(n-k-1)$ is an ordered pair. It is represented by the shaded area in fig. 10.

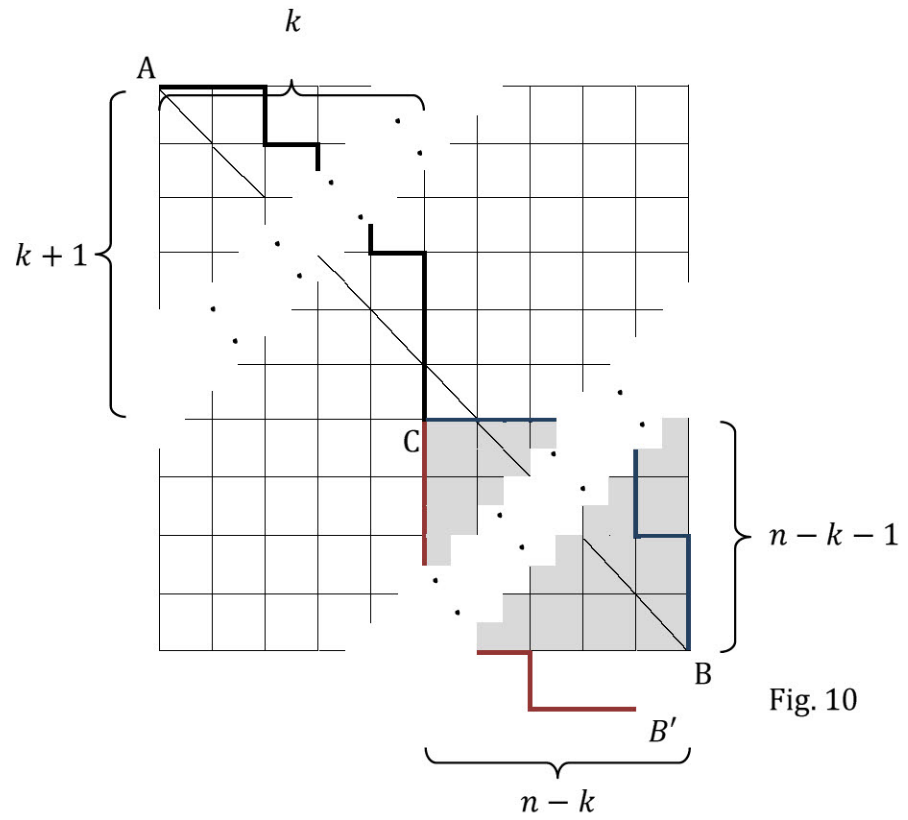
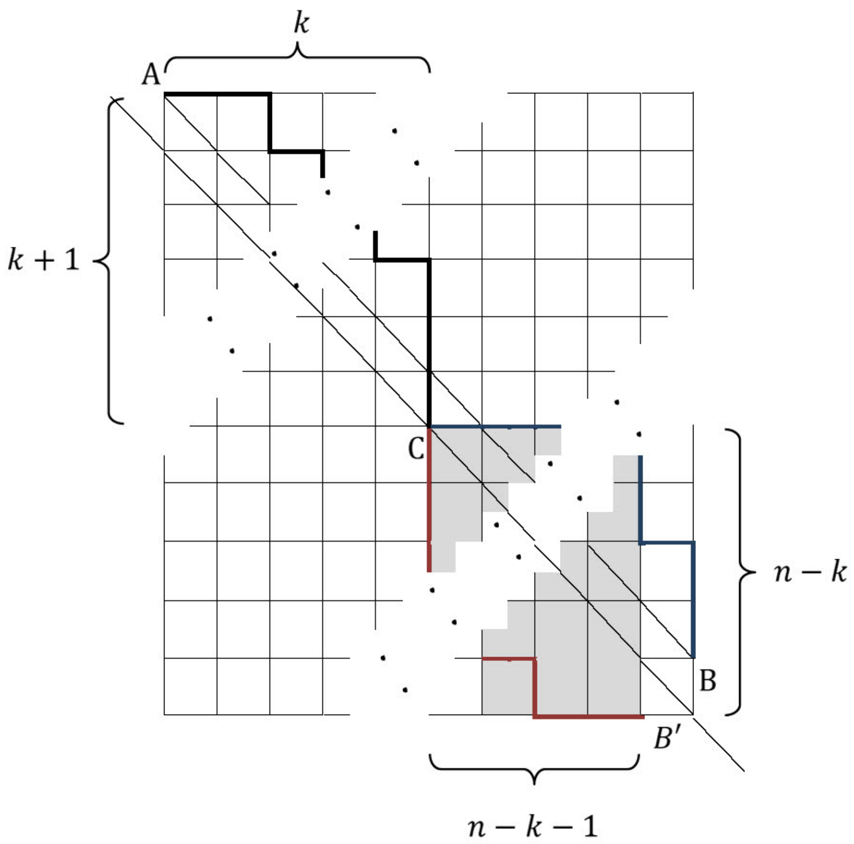


Fig. 10

Now, our new remaining grid from C to B' is $(n-k-1)\times(n-k)$.



\Fig. 11

The new full grid from A to B' is therefore:

$$k+(n-k-1)\times[(k+1)+(n-k)]=(n-1)\times(n+1).$$

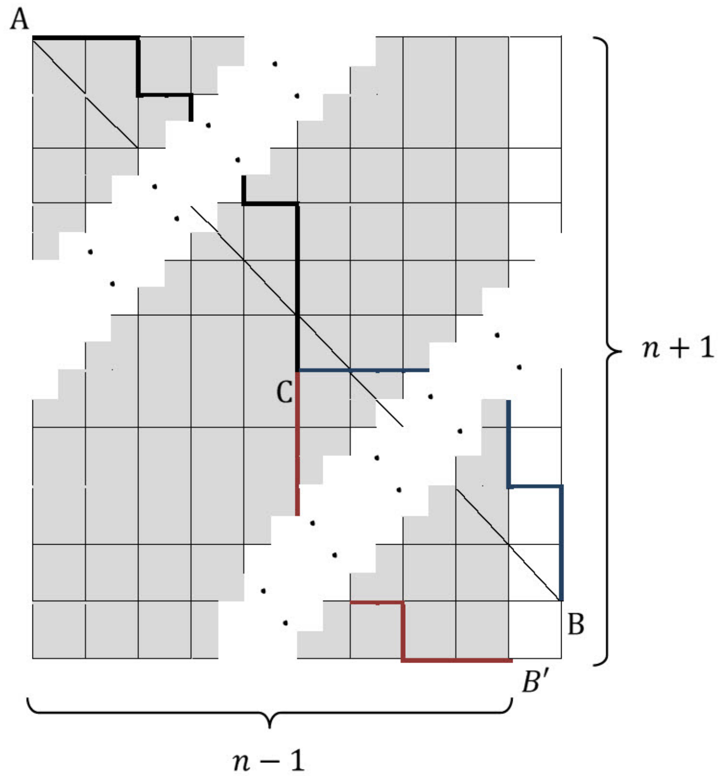


Fig. 6

Note that here, too, the new grid has the same entries of the original grid, and only their order has been changed.

This is true for all n and k, regardless of where C is.

Our new path P', consists now of $(n-1)$ R's and $(n+1)$ D's and can be written as a sequence. The total number of all possible arrangements of $(n-1)$ Rs' and $(n+1)$ Ds' in the sequence is:

$$\binom{(n+1)+(n-1)}{n+1} = \binom{2n}{n+1}$$

This is the total number of bad paths (P_B).

The number of good paths $P_G = P_T - P_B$

$$\begin{aligned}
\therefore P_G &= \binom{2n}{n} - \binom{2n}{n+1} \\
&= \binom{2n}{n} - \frac{(2n)!}{(n+1)!} \\
&= \binom{2n}{n} - \frac{n}{(n+1)}\binom{2n}{n} \\
&= \binom{2n}{n} \times 1 - \frac{n}{n+1} \\
&= \frac{1}{n+1}\binom{2n}{n}
\end{aligned}$$

Other Applications

Let's prove that Catalan numbers also apply to the other methods by finding a relation between them and the grid method we just proved. If we rotate the grid of a good path counterclockwise by 45°, we can spot the mountain ranges as well. An upstroke is equivalent to a step to the right and a down stroke is a step down.

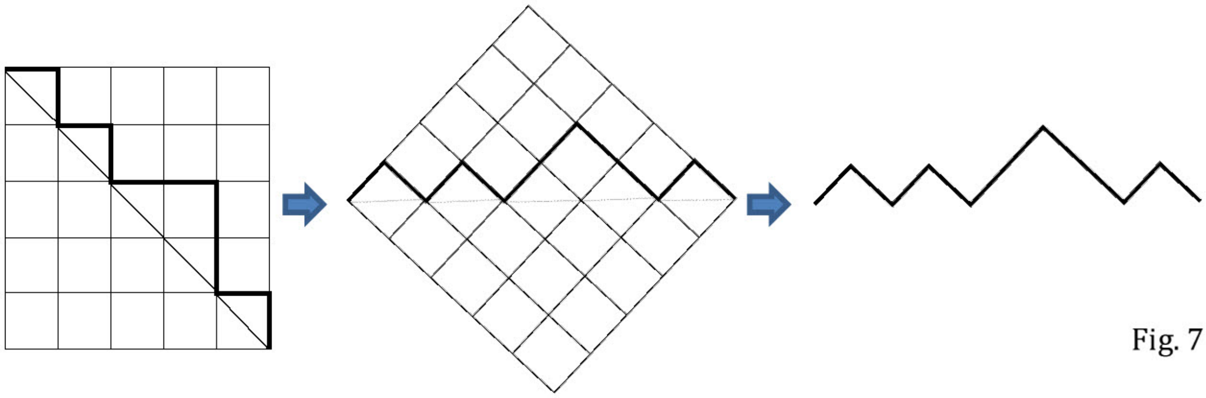


Fig. 7

As mentioned before, the parenthesis is also a similar interpretation. Every open and closed bracket is equivalent to an up and down stroke of the mountain range or a step to the right and down of the grid, respectively. Note how there can't, at any point, be more closed brackets than open ones, just like there can't be more steps downward than to the right in the grid, and like how there can't be more down strokes than upstrokes.

The Basketball Problem

Coming back to our basketball problem, let's now try to find a connection to our grid. To refresh our memories, here's the situation. One against one player on one basket, you shoot and if you score, you get a point. Then it's your turn again. However you might not always get a chance to shoot when you're in possession of the ball. Your opponent might steal it, and then it is his turn to shoot. We're also assuming that the chance of scoring every time is 100%. In other words, you don't always shoot when you are in possession of the ball, but when you do, you always score.

Let's say every time I score a point, it represents one step to the right (R) on our grid, and every time my opponent scores, we go one step down (D). In order for me to never be behind, my opponent must not, at any point, have scored more than me. In other words there can never, at a certain point, be more steps down (D) than to the right (R). Hence we can say that the good path in the grid represents all the ways we can score, with me never falling behind. That's because in every good path, D is never at any point more than R. Since the total number of good paths P_G of an $n \times n$ grid is the Catalan number C_n, that the total number of ways in which I am never losing when each of us have to score n points is C_n.

For example, if we had to score three points in order to win, there would be:

$$C_3 = \frac{1}{3+1}\binom{2\times 3}{3} = \frac{20}{4} = 5 \text{ ways}$$

No. of ways	**I score (s) or opponent scores (o)**					
1	s	s	s	o	o	o
2	s	s	o	s	o	o
3	s	s	o	o	s	o
4	s	o	s	s	o	o
5	s	o	s	o	s	o

After I score the nth point and win, it doesn't make a difference whether my opponent has to score their remaining points or not, because either way, I win.

Probability

If we want to find the probability $P(E)$ of me never losing in a game, we have to operate:

$$P(E) = \frac{n(E)}{n(S)}$$ [2]

Where the $n(E)$ is the number of occurrences of the event and $P(E)$ is the total number of possible outcomes.

In this case, where we have to score three points, $n(E)$ is the number of games in which I am never losing when playing for n points. This is the Catalan number $C_3 = 4$.

$n(S)$ is the total number of ways in which both of us can score n points, regardless of who is winning of losing. If we refer back to our path along the grid, we can see that P_T, the total number of paths from A to B is $n(S)$. When $n = 3$,

$$P_T = \binom{6}{3} = 20$$

$$\therefore P(E) = \frac{4}{20} = 20\%$$

The general term is simple to get, since we already know the general terms of C_n and P_T.

$$P(E) = \frac{n(E)}{n(S)} = \frac{C_n}{P_T} = \frac{\frac{1}{n+1}\binom{2n}{n}}{\binom{2n}{n}} = \frac{1}{n+1}$$

Conclusion

I really enjoyed working with the Catalan Numbers. I personally think it is incredible how one sequence of numbers can apply to various situations and I found it really interesting to try and find a way to prove it. I had many different approaches; I tried to use proof by induction, I also tried to find a general formula by finding a relationship between a number and its previous number. Once proven, it was really easy to find a link between the different interpretations and what I like most about these numbers was that I could find a connection to an everyday situations and problems, especially ones I'm interested in. In this way I was truly able to make this exploration personal and that's what I enjoyed the most.

---

[1] Roberts, Bill, and Sandy MacKenzie. "Sequences, Series and Binomial Theorem."*Mathematics Higher Bevel for the IB Diploma*. Oxford: Oxford UP, 2007. Print.

[2] Roberts, Bill, and Sandy MacKenzie. "Probabilty." *Mathematics Higher Bevel for the IB Diploma*. Oxford: Oxford UP, 2007. Print.