Elemente der Mathematik

Catalan Numbers for the Classroom?

Judita Cofman

Judita Cofman teaches didactics of mathematics for prospective mathematics teachers at the University of Erlangen-Nürnberg. In her spare time she conducts "Maths-Workshops" for 13–19 years old youngsters.

Introduction

In a stimulating article P.J. Hilton and J. Pedersen [1] recommend the teaching of Catalan numbers to secondary school pupils. Their recommendation is backed by several arguments:

- Catalan numbers appear in various mathematical disciplines. By showing pupils different interpretations of the Catalan numbers one can make them aware of links between seemingly unrelated phenomena.

- The basic study of Catalan numbers is accessible at pre-university level; it requires no knowledge beyond elementary mathematics.

- In the process of discovering properties and generalizations of Catalan numbers, new techniques and methods of mathematical reasoning can be acquired.

Binomialkoeffizienten, Fibonaccizahlen, Catalanzahlen sind durch ganz einfache kombinatorische Probleme definiert. Trotzdem weisen sie eine überraschende Vielfalt von Strukturen auf; das Pascalsche Dreieck der Binomialzahlen ist dafür ein bekanntes und schönes Beispiel. Da viele dieser Strukturaussagen ohne besondere mathematische Vorbildung entdeckt werden können, bietet es sich geradezu an, Schülerinnen und Schüler im Bereich dieser speziellen Zahlen selbständig *mathematische Entdeckungen* machen zu lassen. Eindrücklich kann hier erfahren werden, wie eine einfache *Umformulierung* neue Einsichten und oft sogar die Lösung eines Problems bringt. Und schliesslich drängen sich sofort *Verallgemeinerungen* auf; an diesen wird exemplarisch klar, wie jede mathematische Erkenntnis zu immer neuen Fragen führt. – Der Beitrag von Judita Cofman geht zurück auf die "Advanced Royal Institution Master Classes for young people", in London, an denen die Autorin 1995 mit jungen Schülerinnen und Schülern über Catalanzahlen gearbeitet hat. *ust*

The authors describe various representations of the Catalan numbers and derive the formula $C_n = \dfrac{1}{n+1}\dbinom{2n}{n}$ for the n-th Catalan number C_n by counting two-dimensional lattice paths. The latter can be done by following the elegant method of D. André, a nineteenth century French mathematician.

The contents of the Hilton-Pedersen paper can be presented – with slight alterations – to advanced pupils in their final years at secondary school. My intention was to find out whether it made sense to talk to younger pupils – aged 13–14 – about Catalan numbers and to determine to what extent the youngsters could benefit from the underlying ideas.

Last summer, at the "Advanced Royal Institution Master Classes for young people" in London, I had the opportunity of presenting Catalan numbers to young teenagers, keen to study mathematics. I am greatly indebted to the organizers, Joan Glastonbury, Terry Heard and Martin Perkins for inviting me to participate in this event.

Attempts to simplify the approach to Catalan numbers led to a method for finding a formula for C_n which is different from André's, and can be extended to generalized Catalan numbers. This method will be described in Sections 2 and 3. Section 1 outlines the introductory work session with the pupils, aimed at raising their interest in the subject. Section 4 contains suggestions for further studies and comments on our work.

1 How to motivate the study of Catalan numbers

The youngsters were told that they are going to learn about a sequence of numbers which played an important role in the history of mathematics and continues to turn up in different areas of modern research. The sequence is named after E.C. Catalan, a 19-th century Belgian mathematician. It was discovered in the 18-th century by L. Euler in connection with the following problem:

In how many ways can a convex n-gon be divided into triangles by its diagonals such that no two of them intersect inside the n-gon?

Let E_n be the number of different partitions of a convex n-gon into triangles by means of non-intersecting diagonals. Euler succeeded in finding a formula for E_n for any natural number $n \geq 3$, but thought that his method was rather cumbersome.

It was suggested that, later on, interested pupils try to rediscover Euler's formula. Some pupils asked:

Euler was one of the greatest mathematicians of all times. Can we hope to solve a problem which he himself found difficult?

The pupils were reassured by pointing out that since the 18-th century mathematics has developed in many directions. In particular, the Catalan numbers have appeared in various areas of mathematics and some of their new interpretations have provided easier methods for their calculation.

We considered next a special case, namely the number E_5. By solving a series of exercises the pupils saw that:

- Each partition of a convex pentagon into triangles by nonintersecting diagonals leads to a tree-diagram, each tree-diagram leads to a code, and each code to a lattice path from $(0,0)$ to $(3,3)$ which lies below the line $y = x$ (see Figure 1, Steps 1–4).
- Conversely, each lattice path from $(0,0)$ to $(3,3)$ below the line $y = x$ leads to a code, each code to a tree-diagram, and each tree-diagram to a partition of a convex pentagon into triangles by non-intersecting diagonals. This is seen by reversing the steps in the constructions in Figure 1.
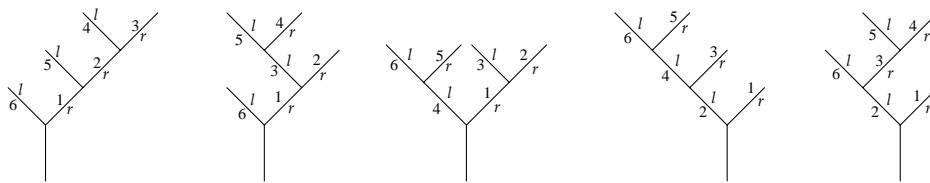
Step 1: The partitions of a convex pentagon into triangles
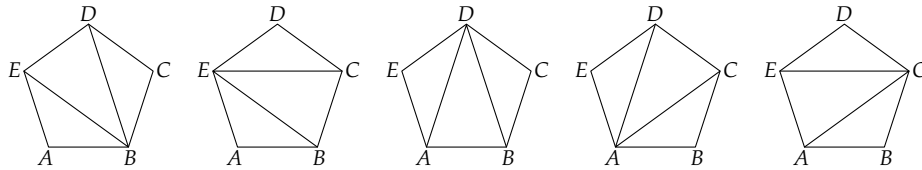
Step 2: a) The construction of the tree-diagrams, corresponding to the partitions
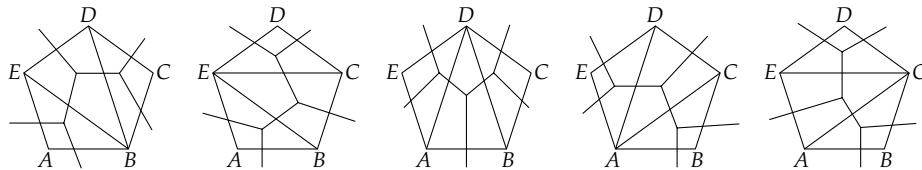
Step 2: b) The labelling of the branches of the tree-diagrams

Step 3: The codes, derived from the labelled tree-diagrams

$$r\ r\ r\ l\ l\ l \qquad r\ r\ l\ r\ l\ l \qquad r\ r\ l\ l\ r\ l \qquad r\ l\ r\ l\ r\ l \qquad r\ l\ r\ r\ l\ l$$
$$(\dot{1}\ \dot{2}\ \dot{3}\ \dot{4}\ \dot{5}\ \dot{6}) \quad (\dot{1}\ \dot{2}\ \dot{3}\ \dot{4}\ \dot{5}\ \dot{6}) \quad (\dot{1}\ \dot{2}\ \dot{3}\ \dot{4}\ \dot{5}\ \dot{6}) \quad (\dot{1}\ \dot{2}\ \dot{3}\ \dot{4}\ \dot{5}\ \dot{6}) \quad (\dot{1}\ \dot{2}\ \dot{3}\ \dot{4}\ \dot{5}\ \dot{6})$$
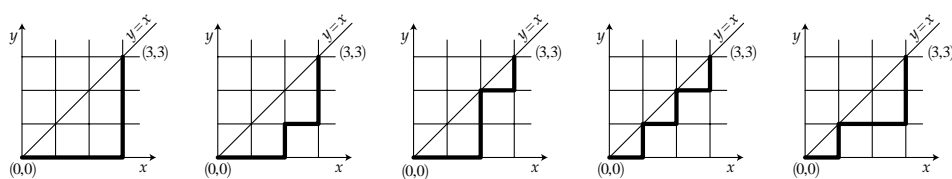
Step 4: The lattice paths, obtained from the codes

Fig. 1

In this way we found four different interpretations of E_5, related to geometry, graphs, codes and lattice paths respectively.

The number of lattice paths from $(0,0)$ to (n,n) below the line $y = x$ is usually denoted by C_n and is called the n-th Catalan number. From Figure 1 the pupils concluded that $E_5 = C_3$, and formulated the following

Hypothesis 1: $E_n = C_{n-2}$ for $n \geq 3$.

In the ensuing discussion it was pointed out that the above hypothesis could be verified by proving the following two statements:

Theorem 1: *The different partitions of a convex n-gon into triangles by non-intersecting diagonals correspond to different lattice paths from $(0,0)$ to $(n-2, n-2)$, lying below the line $y = x$.*

Theorem 2: *The different lattice paths from $(0,0)$ to $(n-2, n-2)$, below the line $y = x$ correspond to different partitions of a convex n-gon into triangles by non-intersecting diagonals.*

Proving Theorem 1 would imply that $E_n \leq C_{n-2}$ and proving Theorem 2 would show that $C_{n-2} \leq E_n$. From these relations it would follow that $E_n = C_{n-2}$.

The pupils understood intuitively how the proofs of Theorems 1 and 2 would work for an arbitrary n. Consequently, they accepted the truth of the hypothesis. I told them that exact proofs of Theorems 1 and 2 involve various technical details, and that, at this stage, it would not be appropriate to dwell on them.

Finally, I made two remarks:

The first remark explained the connection between Catalan and the numbers named after him. Catalan stated and solved the following algebraic problem:

The product $a_1 \cdot a_2 \cdot a_3 \cdots a_n$ of n numbers $a_1, a_2, a_3, \ldots, a_n$ is computed step by step, so that at each stage two of the numbers are multiplied together. At no stage can the order of the factor be altered. In how many ways can this be done?

The youngsters found the answer for $n = 4$ and were shown that the different ways of computing a product of four numbers (say $2 \cdot 3 \cdot 5 \cdot 10$) are in one-to-one correspondence with the tree-diagrams in Figure 1 (see Figure 2).
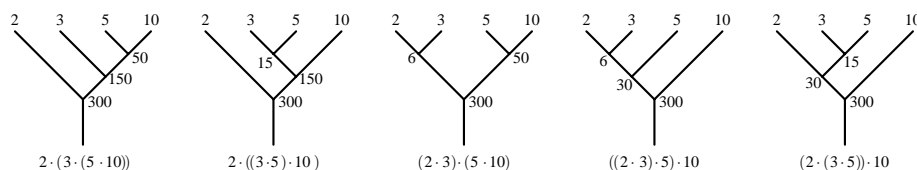

Fig. 2

Catalan's solution of the general case gave the number C_{n-1}. His problem provides yet another interpretation of the numbers E_n.

The second remark stressed the importance of lattice paths in modern mathematics: Many problems in combinatorics and in the theory of probability can be interpreted in

the terms of lattice paths. The solution of such problems is often simplified thanks to various established methods for counting lattice paths. – We could try to solve Euler's problem by counting the lattice paths from $(0,0)$ to (n,n), $n \geq 0$, underneath the line $y = x$.

2 How to guess a formula for C_n

From a group of 22 youngsters four volunteered to embark on a project on Catalan numbers. They produced good work. I shall mention their names and thank them for inspiring me to think about further problems related to this topic. The pupils were: Edmund Harrison (13 years old), and Maria Russell, Douglas Shaw, and Peter Zimmermann (all 14 years old).

The first task of the project was to try to find a formula for the numbers C_n of lattice paths from $(0,0)$ to (n,n) underneath the line $y = x$. This could have been done by following the well-known, ingenious method of André, which involves reflections of parts of lattice paths in the line $y = x$ (see e.g. [3]). Although this method is beautiful, I decided to adopt a different approach, enabling the pupils to guess what C_n is:

The youngsters first determined C_n for a sequence of initial values of n. This was done by completing, step by step, the number pattern in Fig. 3a. In this pattern each point is labelled with the number of those lattice paths connecting it to $(0,0)$ which pass below the line $y = x$. (Hence each number inside the pattern is the sum of the number below it and the number on its left). The labels of the points on the line $y = x$ were the first terms of the Catalan sequence:

$$1, 1, 2, 5, 14, 42, \ldots$$

The pupils were unable to provide a clue for continuing the sequence. It was decided to compare the Catalan numbers found so far with the numbers of all lattice paths from $(0,0)$ to (n,n), for $n = 0, 1, 2, 3, 4, 5$. Therefore another number pattern was constructed, in which the points were labelled with the numbers of all lattice paths connecting them with $(0,0)$ (see Fig. 3b).
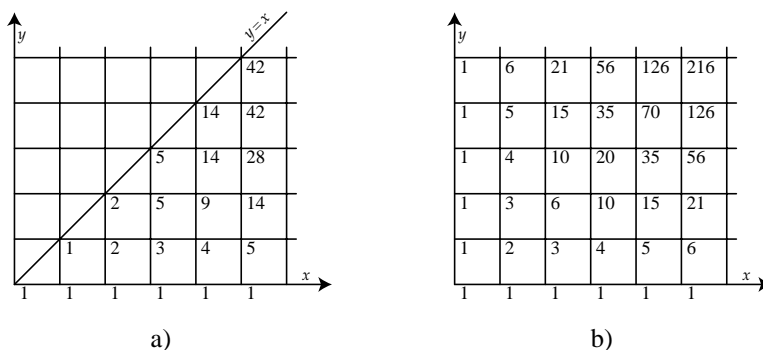

a)   b)

Fig. 3

n	0 1 2 3 4 5
C_n	1 1 2 5 14 42
number of all lattice paths from $(0,0)$ to (n,n)	1 2 6 20 70 252

Fig. 4

The numbers on the line $y = x$ in Fig. 3b were compared to the corresponding numbers in Fig. 3a. The table in Figure 4 was revealing; it led to

Hypothesis 2: C_n *multiplied by* $n + 1$ *gives the number of all lattice paths from* $(0,0)$ *to* (n,n).

Some of the youngsters recognized the number pattern in Fig. 3b as Pascals' triangle and reformulated Hypothesis 2, suggesting a formula for C_n:

Hypothesis 2′: $C_n = \dfrac{1}{n+1}\dbinom{2n}{n}$ *for* $n \geq 0$.

We were left with the problem of verifying Hypothesis 2. To do this, we agreed to try to partition the set of all lattice paths from $(0,0)$ to (n,n) into disjoint subsets of $n + 1$ elements each, such that each subset contained exactly one lattice path below the line $y = x$. Carrying out such a partition would prove Hypothesis 2.

The next section outlines a method for partitioning all lattice paths from $(0,0)$ to (n,n) subject to the conditions stated above.

3 A method for proving Hypothesis 2

Denote by S_n the set of all lattice paths from $(0,0)$ to (n,n) and by U_n the subset of S_n consisting of the lattice paths below the line $y = x$. Any lattice path $l \in S_n$ contains n horizontal steps. Denote the successive steps from $(0,0)$ to (n,n) by $h_1, h_2, \ldots, h_n$. Put $h_i = k$ if and only if the step h_i lies on the line $y = k$.

In this way any lattice path $l \in S_n$ can be represented by an ordered n-tuple $(h_1, h_2, \ldots, h_n)$. Since the heights of the horizontal steps of l form a non-decreasing sequence, it follows that

$$0 \leq h_1 \leq h_2 \leq \ldots \leq h_n \leq n. \tag{1}$$

Conversely, any ordered n-tuple of integers satisfying (1) corresponds to a lattice path $l \in S_n$.

The lattice paths in U_n are represented by ordered n-tuples which, in addition to (1), satisfy the inequalities

$$h_i \leq i - 1 \quad \text{for} \quad i = 1, 2, \ldots, n. \tag{2}$$

Representing lattice paths by ordered n-tuples makes it possible to explain ideas about partitioning S_n in terms of algebra. One has only to understand addition modulo $n + 1$.

$i \in \{0, 1, 2, 3\}, \oplus$ Denotes addition mod 4;
after addition mod 4 the entries in each triple
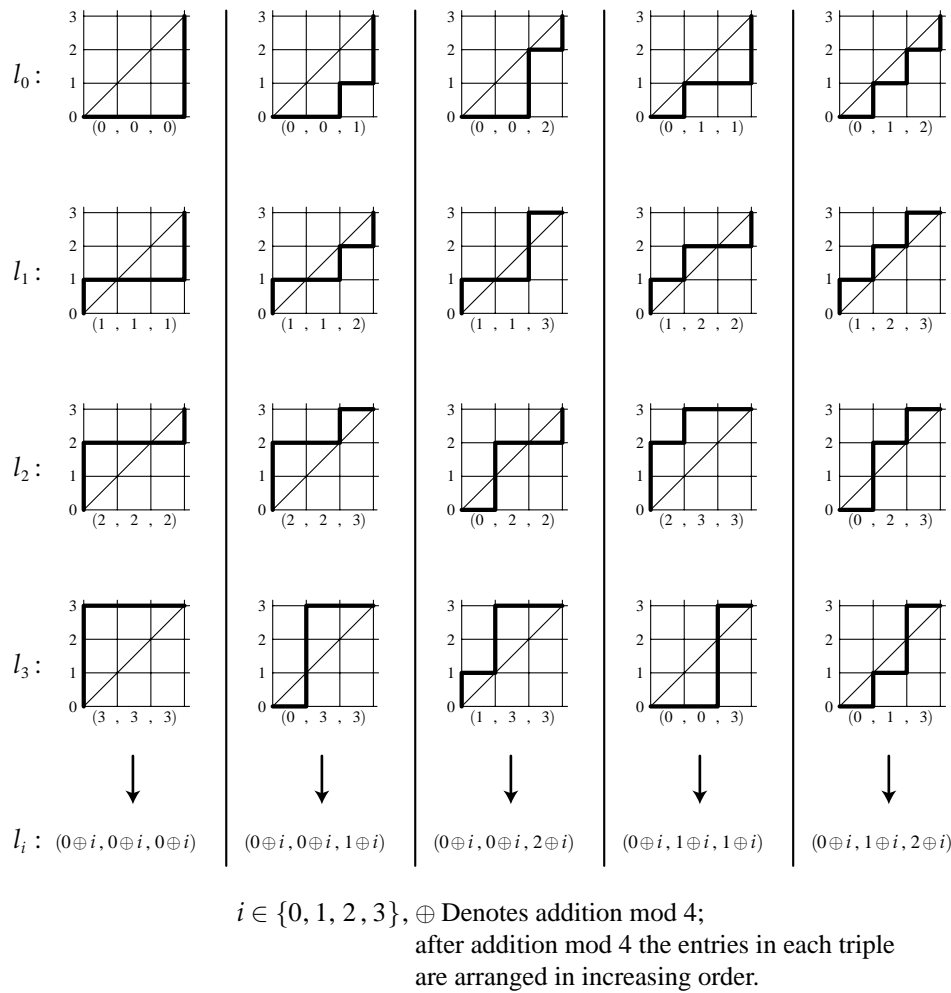are arranged in increasing order.

Fig. 5

To the youngsters the explanations were given for small values of n, accompanied by
diagrams (see Figures 5 and 6). Here we shall describe the procedure in general terms.
Take any $l_0 \in U_n$. From its corresponding n-tuple $(h_1, h_2, \ldots, h_n)$ form the n-tuples

$$(h_1 \oplus k, h_2 \oplus k, \ldots, h_n \oplus k) \quad \text{for} \quad k = 0, 1, \ldots, n,$$

where $\oplus$ denotes addition modulo $n + 1$. If in the n-tuple $(h_1 \oplus k, h_2 \oplus k, \ldots, h_n \oplus k)$
smaller numbers turn up after larger ones, rearrange them, so that they appear in non-
decreasing order. This associates a lattice path l_k with a (possibly rearranged) n-tuple
$(h_1 \oplus k, h_2 \oplus k, \ldots, h_n \oplus k)$. Denote the collection $l_0, l_1, \ldots, l_n$ of paths by L.

In order to verify Hypothesis 2 we have to prove

Theorem 3: *For any $l_0 \in U_n$ we have:*

(a) $l_k \notin U_n$ *for* $k = 1, \ldots, n$.

(b) $l_0, l_1, \ldots, l_n$ *are distinct.*

(c) *For any two distinct paths* $l_0, l_0' \in U_n$ *the corresponding sets* $L = \{l_0, l_1, \ldots, l_n\}$ *and* $L' = \{l_0', l_1', \ldots, l_n'\}$ *are disjoint.*

(d) *For any path* $l \in S_n$ *there exists a path* $l_0 \in U_n$ *such that* $l \in \{l_0, l_1, \ldots, l_n\}$.

The proofs of (a), (b) and (c) are straightforward. They are based on properties of addition mod $(n + 1)$ and on the relations (1) and (2).

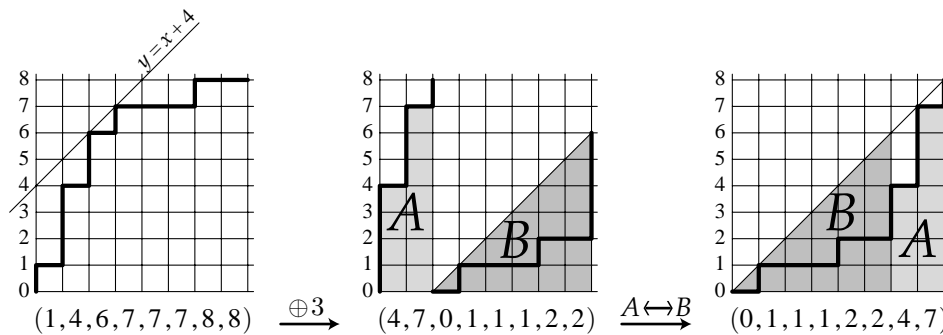The proof of (d) requires some thought; it is illustrated graphically in Figure 6.

$$(1,4,6,7,7,7,8,8) \quad \xrightarrow{\oplus 3} \quad (4,7,0,1,1,1,2,2) \quad \xrightarrow{A \leftrightarrow B} \quad (0,1,1,1,2,2,4,7)$$

Fig. 6

Let $l \leftrightarrow (h_1, h_2, \ldots, h_n)$ be a path in $S_n \setminus U_n$. Let m be the smallest integer such that l lies below the line $y = x + m$, and suppose that the first horizontal segment of the path from $(0,0)$ to (n,n) which touches the line $y = x + m$ is the r-th. Denote by h_r' the inverse of h_r with respect to addition mod $(n + 1)$ (that is, the number $n + 1 - h_r$), and construct the n - tuple

$$(h_0 \oplus h_r', h_1 \oplus h_r', \ldots, h_n \oplus h_r').$$

By rearranging the above n-tuple in non-decreasing order, we obtain an n-tuple which represents a lattice path. It is left to the reader to verify that this lattice path is a path $l_0 \in U_n$ and that the path l_r coincides with l in the corresponding set $L = \{l_0, l_1, \ldots, l_n\}$.

The proof of the Theorem 3 implies that $C_n = \dfrac{1}{n+1}\dbinom{2n}{n}$.

4 Suggestions for further study – for pupils and "grown-ups"

In the final part of the project the pupils studied questions related to the Catalan numbers:

(a): It was pointed out that the method for finding the formula for C_n can be adapted for calculating the number of lattice paths below the line $y = kx$ for any integer $k \geq 1$. It can be shown that for $k \geq 1$ the number of the lattice paths from $(0,0)$ to (n, kn)

which lie below the line with equation $y = kx$ is $kn + 1$ times smaller than the number of all lattice paths from $(0,0)$ to (n, kn).

(b): Maria studied three-dimensional lattice paths below the diagonal plane $y = z$. She formulated the Hypothesis:

Hypothesis 3: *The number $D_n^{(3)}$ of those lattice paths from $(0,0,0)$ to (n,n,n) which lie below the plane $z = y$ is $n + 1$ times smaller than the number of all lattice paths from $(0,0,0)$ to (n,n,n).*

Maria's hypothesis can be verified. This made me think of a problem for "grown-ups":

The "proper way" to generalize C_n in space is to consider the number $C_n^{(3)}$ of the lattice paths from $(0,0,0)$ to (n,n,n) which pass through the lattice points (x, y, z) such that $x \geq y \geq z$. Is it true that

Hypothesis 4: $C_n^{(3)} = \dfrac{1}{\binom{n+2}{2}} D_n^{(3)}$, *that is*

$$C_n^{(3)} = \frac{1}{\binom{n+2}{2}} \cdot \frac{1}{\binom{n+1}{1}} \cdot \frac{(3n)!}{n!n!n!} \; ? \tag{3}$$

The idea for stating Hypothesis 4 arose from considerations indicated in Figure 7.
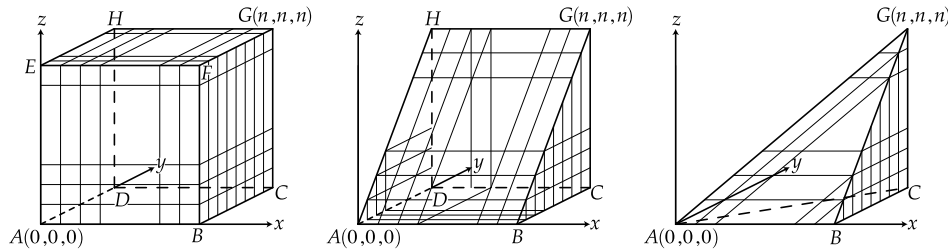
The lattice paths from $(0,0,0)$ to (n,n,n) are contained in a cube $ABCDEFGH$.

Their number is $\dfrac{(3n)!}{n!n!n!}$.

The lattice paths from $(0,0,0)$ to (n,n,n) through the lattice points (x,y,z) with $y \geq z$ are contained in the wedge $ABCDGH$.

Their number is $D_n^{(3)}$.

The lattice paths from $(0,0,0)$ to (n,n,n) through the lattice points (x,y,z) with $x \geq y \geq z$ are contained in the pyramid $ABCG$.

Their number is $C_n^{(3)}$.

The number of the lattice points on the edge BG of $ABCDGH$ is $n + 1$, and $D_n^{(3)} = \dfrac{(3n)!}{n!n!n!} \cdot \dfrac{1}{n+1}$.

The number of the lattice points on the face ABG of $ABCDGH$ is $1 + 2 + 3 + \cdots + (n+1) = \binom{n+2}{2}$.

Is it true that $C_n^{(3)} = D_n^{(3)} \cdot \dfrac{1}{\binom{n+2}{2}}$?

Fig. 7

To my astonishment, Hypothesis 4 turned out to be true. I verified it by using a recurrence relation for $C_n^{(3)}$. Formula (3) is known; it can be found, e.g. in an equivalent form in Macmahon's beautiful book [2]. Moreover, the formula can be further generalized, for arbitrary, k - dimensional spaces (see [3]):

The number of those k-dimensional lattice paths from $(0, 0, \ldots, 0)$ to $(n, n, \ldots, n)$ which pass through lattice points $(x_1, x_2, \ldots, x_k)$ such that $x_1 \geq x_2 \geq \ldots \geq x_k$ is given by the formula

$$C_n^{(k)} = \frac{1}{\binom{n+1}{1}} \cdot \frac{1}{\binom{n+2}{2}} \cdots \frac{1}{\binom{n+k-1}{k-1}} \cdot \frac{(kn)!}{n!n!\cdots n!}.$$

I wish to conclude this article by remarks on classroom teaching provoked by our activities. Watching pupils at work confirmed my belief, formed on various occasions, that solving simple combinatorial problems plays an important role in developing thought processes, and should be practised throughout secondary school. Many questions on counting, selecting and ordering are accessible at an early age. There are many topics in the syllabus, whose teaching can profit from the inclusion of a selection of carefully formulated problems of combinatorial nature. In many schools Fibonacci numbers have already entered the classroom. Should we introduce Catalan numbers to young teenagers? Probably not, but we should definitely teach them some of the ideas underlying their study.

References

[1] P.J. Hilton and Jean Pedersen, Catalan – Zahlen und Wege in einem ganzzahligen Gitter; Elemente der Mathematik 48 (1993) pp. 45–60

[2] Macmahon, Combinatory Analysis, Chelsea, New York (1960)

[3] L. Moser and W. Zayachkowski, Lattice paths with diagonal steps; Scripta Mathematica 26 (1961–63) pp. 223–229

Judita Cofman
Mathematisches Institut
Universität Erlangen-Nürnberg
Bismarckstr. 1 1/2
D-91054 Erlangen
Germany

Judita Cofman hat über die ihre "Maths-Workshops" auch die folgenden zwei Bücher veröffentlicht:

Judita Cofman: *What to solve? Problems for young mathematicians*, Oxford University Press, 1990,

Judita Cofman: *Numbers and shapes revisited. More problems for young mathematicians*, Oxford University Press, 1995.