Thomas Koshy

The Ubiquitous Catalan Numbers

Just as Fibonacci and Lucas numbers are a great source of fun and excitement for both amateurs and professionals alike (Askey 2005, Koshy 2002), so are the less well-known Catalan numbers. They too are excellent candidates for mathematical activities such as experimentation, exploration, and conjecture. I was surprised to see that, like Fibonacci and Lucas numbers, Catalan numbers seemed to show up in several problems I had assigned to students over the years.

Example 1. Find the number of mountain ranges that can be drawn with n upstrokes and n downstrokes. In other words, find the number of different paths we can choose from the origin to the point $(2n, 0)$ on the xy-plane subject to the following conditions:

- We can touch the x-axis, but we cannot cross it.
- From the point (x, y), we can climb up to the point $(x + 1, y + 1)$ or climb down to the point $(x + 1, y - 1)$. See **figure 1**.

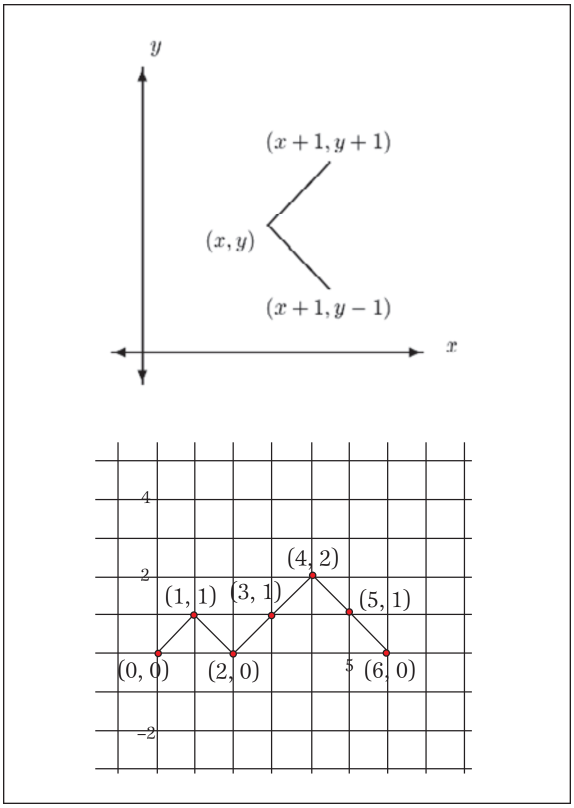

Fig. 1 The two possible paths from the point (x, y)

Solution. When we have no clue about the solution, it is often a good technique to solve the problem in a few simple cases and then look for a pattern, as illustrated in the following steps.

Step 1. Collect enough data by performing simple experiments. **Figure 2** shows the various possibilities for $n = 0, 1, 2, 3,$ and 4. Notice that some mountain ranges are reflections of others.
Step 2. Arrange the results of the five experiments in a table, as in **table 1**. The numbers in the bottom row that count the number of mountain ranges are called the Catalan numbers, C_n.
Step 3. Look for a pattern. (The pattern does not seem to be obvious.)
Step 4. If a pattern exists, can you conjecture a formula for the nth Catalan number C_n?

This department focuses on mathematics content that appeals to secondary school teachers. It provides a forum that allows classroom teachers to share their mathematics from their work with students, their classroom investigations and projects, and their other experiences. We encourage submissions that pose and solve a novel or interesting mathematics problem, expand on connections among different mathematical topics, present a general method for describing a mathematical notion or solving a class of problems, elaborate on new insights into familiar secondary school mathematics, or leave the reader with a mathematical idea to expand. Send submissions to "Delving Deeper" by accessing **mt.msubmit.net**.

"Delving Deeper" can accept manuscripts in ASCII or Word formats **only**.

*Edited by **Al Cuoco**, acuoco@edc.org*
Center for Mathematics Education, Education Development Center
Newton, MA 02458

***E. Paul Goldenberg**, pgoldenberg@edc.org*
Center for Mathematics Education, Education Development Center
Newton, MA 02458

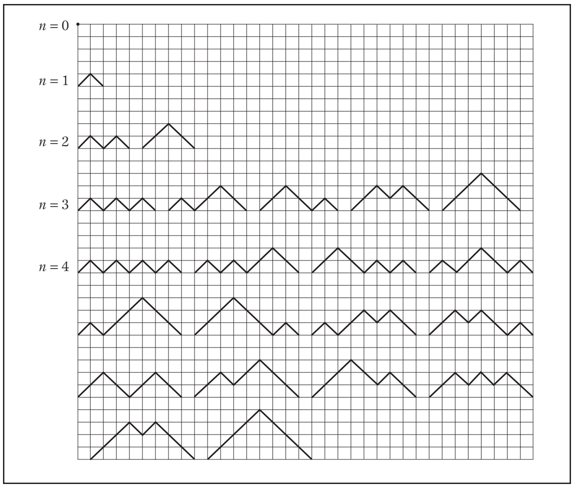

Fig. 2 Mountain ranges for n = 0, 1, 2, 3, and 4

Step 5. If you conjectured a formula, can you establish it?

The next example deals with partial sums. If $S = a_1 + a_2 + a_3 + \cdots + a_n$ is a sum of numbers, each of the sums a_1, $a_1 + a_2$, $a_1 + a_2 + a_3$, $\ldots$ $a_1 + a_2 + a_3 + \cdots + a_n$ is a partial sum of S.

Example 2. Arrange n 1s and n –1s in a row in such a way that every partial sum is a nonnegative integer. Find the number of such possible sequences.
Solution. Once again, we have no clue about the solution or the answer. So we proceed as in example 1.

Step 1. Collect enough data by performing simple experiments. Notice that *every* partial sum must be nonnegative, so every sequence must begin with a 1. **Figure 3** shows the various possible arrangements corresponding to $0 \le n \le 3$.
Step 2. Arrange the data in a table, as in **table 2**.
Step 3. Look for a pattern. (The pattern, as before, is not obvious.)
Step 4. If a pattern exists, can you conjecture a formula for the nth Catalan number C_n?
Step 5. If you conjectured a formula, can you establish it?

Example 3. Bertrand's ballot problem (Feller 1968) is another version of the first two examples:

> Two candidates A and B are running for president. Each person gets n votes. In how many ways can the $2n$ votes be counted in such a way that at each count the number of votes received by A is greater than or equal to the number of votes received by B?

Example 4. This example, originally studied by Catalan, deals with correctly parenthesized algebraic expressions. For example, for numbers $a, b, c,$ and $d,$ all of these expressions are correctly parenthesized:

$$((ab)c)d,\ (a(bc)d,\ (ab)(cd),\ (a(bc)d), \text{ and } a(b(cd)),$$

but this expression is not:

$$a)(bc)(d$$

Suppose we are given n pairs of left and right parentheses. Find the number of correctly parenthesized sequences that can be formed.
Solution. For convenience, we drop the letters. So "( )" and "(( ))( )" are correctly parenthesized but ")( )(" is not. The results corresponding to $0 \le n \le 3$ are in **table 3**. It would be a good exercise to list the possible sequences corresponding to $n = 4$.

Readers might try showing that examples

Table 1

Data from Example 1

n	1	2	3	4	5	…	n	
Number of mountain ranges	1	2	5	14	?	…	?	← Catalan numbers

$n = 1$	1	–1				
$n = 2$	1	1	–1	–1		
	1	–1	1	–1		
$n = 3$	1	1	1	–1	–1	–1
	1	1	–1	1	–1	–1
	1	–1	1	–1	1	–1
	1	–1	1	1	–1	–1
	1	1	–1	–1	1	–1

Fig. 3 Various arrangements corresponding to $0 \le n \le 3$

Table 2

Data from Example 2

n	1	2	3	4	5	…	n	
Number of partial sums	1	2	5	14	?	…	?	← Catalan numbers again

Table 3

Data from Example 4

n	Correctly parenthesized sequences				
1	( )				
2	( )( )	(( ))			
3	( )( )( )	(( ))( )	( )(( ))	(( )( ))	((( )))

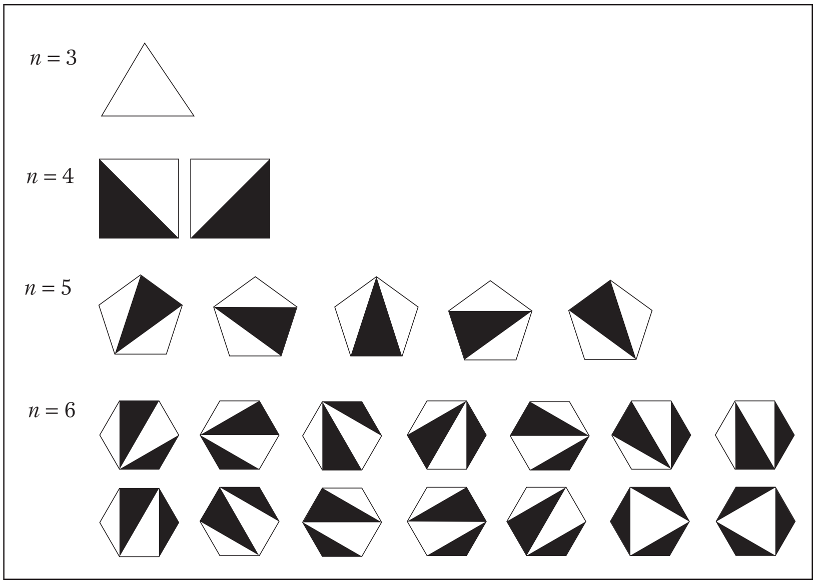

Fig. 4 Triangulations of n-gons, where $3 \le n \le 6$

1–4 are *isomorphic*—they are the same problems wrapped in different contexts.

We now turn to a triangulation problem studied by Euler. An n-gon (a polygon with n sides) is *convex* if every diagonal of the polygon lies inside it.

Example 5. Find the number of ways (the interior of) a convex n-gon can be partitioned into triangles by drawing non-intersecting diagonals, where $n \ge 3$.
Solution. The various possibilities for $3 \le n \le 6$ are shown in **figure 4**.

Using an induction argument, which he characterized as "quite laborious," Euler showed that the number of triangulations T_n of a convex n-gon is given by

$$(1) \qquad T_n = \frac{2 \cdot 6 \cdot 10 \cdot \cdots \cdot (4n-10)}{(n-1)!}, \quad n \ge 3.$$

For example,

$$T_3 = \frac{2}{2!} = 1, T_4 = \frac{2 \cdot 6}{3!} = 2, \text{ and } T_5 = \frac{2 \cdot 6 \cdot 10}{4!} = 5.$$

After a few more calculations, it looks as if these are the same Catalan numbers C_n shifted by two spaces to the right. That is, it seems as if $C_n = T_{k+2}$. If this were true, we would have

$$(2) \qquad C_n = \frac{2 \cdot 6 \cdot \cdots \cdot (4n-2)}{(n+1)!}, \quad n \ge 1.$$

Using formula (2), the value of C_n could then be rewritten as follows:

$$\begin{aligned} C_n &= \frac{4n-2}{n+1} \cdot \frac{2 \cdot 6 \cdot 10 \cdot \cdots \cdot (4n-6)}{n!} \\ &= \frac{4n-2}{n+1} \cdot C_{n-1} \end{aligned}$$

When $n = 1$, this yields $C_1 = C_0$; but $C_1 = 1$. So we could define C_0 to be 1. In other words, we have conjectured a recursive formula for C_n.

Conjecture. C_n satisfies the following recurrence:

$$\begin{aligned} C_0 &= 1 \\ C_n &= \frac{4n-2}{n+1} \cdot C_{n-1}, \quad n \ge 1 \end{aligned}$$

For example, here is a verification for $n = 4$:

$$\begin{aligned} C_4 &= \frac{14}{5} \cdot C_3 \\ &= \frac{14}{5} \cdot \frac{10}{4} \cdot C_2 \\ &= \frac{14}{5} \cdot \frac{10}{4} \cdot \frac{6}{3} \cdot C_1 \\ &= \frac{14}{5} \cdot \frac{10}{4} \cdot \frac{6}{3} \cdot \frac{2}{2} \cdot C_0 \\ &= 14 \end{aligned}$$

This agrees with our data.

The recurrence above is only a conjecture at this point, and readers might try proving it by showing that it correctly models any of the examples 1 through 4. There are two other curious phenomena, each worthy of a closer look:

- Why should the numbers produced by our recurrence be integers?
- How might one prove Euler's theorem about triangulations?

ANOTHER RECURSIVE FORMULA

In 1759, Johann Andreas von Segner developed another recursive formula for C_n:

$$\begin{aligned} (3) \quad C_n &= C_0C_{n-1} + C_1C_{n-2} + \cdots + C_{n-2}C_1 + C_{n-1}C_0 \\ &= (C_0, C_1, \ldots, C_{n-1}) \cdot (C_{n-1}, C_{n-2}, \ldots, C_0) \end{aligned}$$

where the dot indicates the dot product of the two vectors.

For example,

$$\begin{aligned} C_5 &= (C_0, C_1, C_2, C_3, C_4) \cdot (C_4, C_3, C_2, C_1, C_0) \\ &= (1, 1, 2, 5, 14) \cdot (14, 5, 2, 1, 1) \\ &= 1(14) + 1(5) + 2(2) + 5(1) + 14(1) \\ &= 42. \end{aligned}$$

Formulas (2) and (3) can be used to obtain an explicit formula for C_n (Guy 1990):

$$(4) \qquad C_n = \frac{1}{n+1}\binom{2n}{n}$$

For instance,

$$
\begin{aligned}
C_5 &= \frac{1}{6}\binom{10}{5} \\
&= \frac{252}{6} = 42.
\end{aligned}
$$

CATALAN MEETS PASCAL

Formula (4) contains a fascinating treasure: Every Catalan number C_n can be obtained by dividing the *central binomial coefficient* in Pascal's triangle by $n + 1$. For example, consider the Pascal's triangle in **figure 5**, where the various central binomial coefficients are boxed. To find C_4, read the central binomial coefficient 70 in row 8 (that is, $2n$); divide it by 5 (that is, $n + 1$). This yields $C_4 = 14$.

Catalan numbers can be obtained from Pascal's triangle in three other ways.

1. The difference of a central binomial coefficient and its adjacent entry in the same row is a Catalan number:

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

For example,

$$
\begin{aligned}
C_3 &= \binom{6}{3} - \binom{6}{2} \\
&= 20 - \boxed{15} \\
&= 5.
\end{aligned}
$$

See **figure 6**.

2. The difference of twice a central binomial coefficient, and the binomial coefficient that lies along the same diagonal and in the row below, is a Catalan number:

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

For example,

$$
\begin{aligned}
C_3 &= 2\binom{6}{3} - \binom{7}{3} \\
&= 2 \cdot 20 - \boxed{35} \\
&= 5.
\end{aligned}
$$

See **figure 7**.

3. Every Catalan number, except C_0, is the difference of the central binomial coefficient and the binomial coefficient two spaces away in the same row:

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

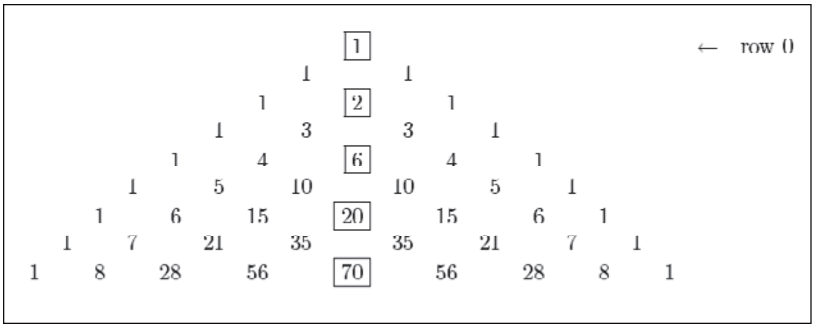

Fig. 5 Pascal's triangle

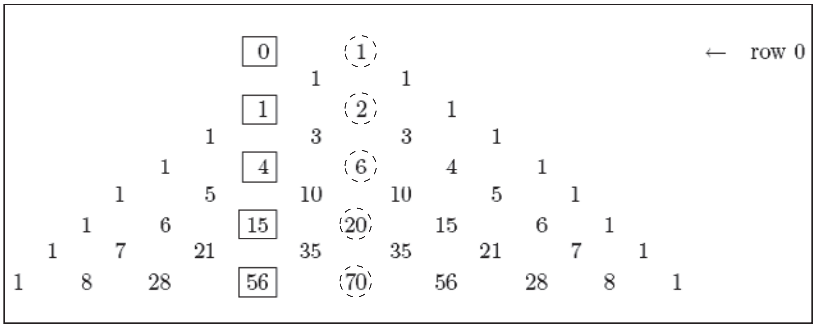

Fig. 6 Subtract the boxed numbers from the central cofficient.

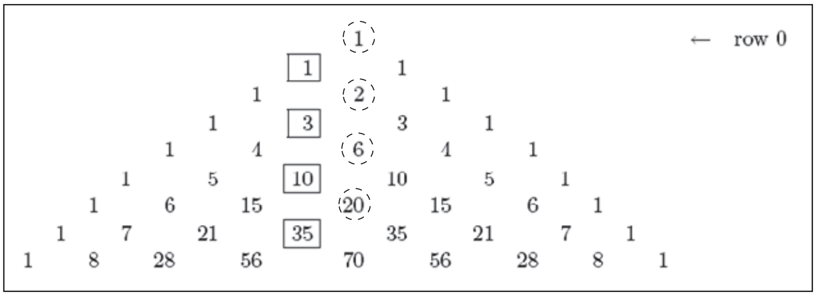

Fig. 7 Double a central coefficient and subtract its northwest neighbor.

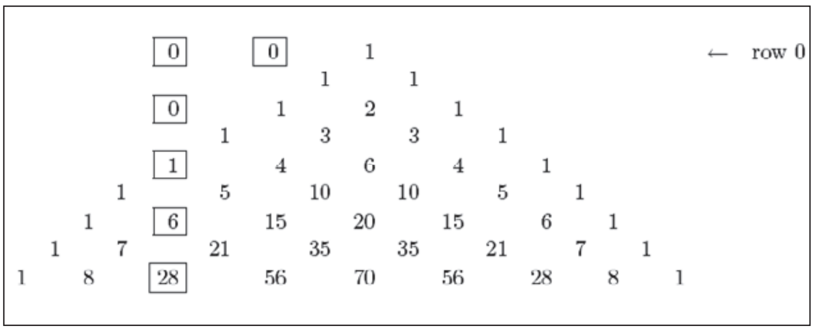

Fig. 8 Catalan numbers as a difference of near neighbors

For example,

$$
\begin{aligned}
C_5 &= \binom{8}{4} - \binom{8}{2} \\
&= 70 - 28 \\
&= 42.
\end{aligned}
$$

See **figure 8**.

The study of Catalan numbers raises some interesting questions for discussion.

1. A quick look at **figure 5** shows that every central binomial coefficient is an even integer. Is it always true? In other words, is

$$\binom{2n}{n}$$

always an even integer?
2. We have found that

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

is an integer for $0 \le n \le 4$. Is it always true? In other words, is $n + 1$ always a factor of

$$\binom{2n}{n}?$$

3. We have found four explicit formulas for C_n, namely, formulas (2) through (5). Are there other such formulas?
4. How can one establish formulas (3) through (6)?
5. Are there other ways, in addition to formulas (5) through (7), that Catalan numbers can be extracted from Pascal's triangle?
6. Can we obtain formula (4) from formula (2)?

Editors' notes: Thomas Koshy gives us much food for thought. Indeed, this article provides glimpses of many surprising connections, but it gives us very few proofs. We would be interested in submissions from readers that provide proofs for some of the questions that arise in the paper, like questions 1–6 above. Here are some others.

There are identities and results embedded in the article that make no mention of Catalan numbers. For example, if C_n satisfies both of these properties,

$$C_n = \begin{cases} 1 & \text{if } n = 0 \\ \dfrac{4n-2}{n+1}C_{n-1} & \text{if } n > 0 \end{cases} \quad \text{and} \quad C_n = \frac{1}{n+1}\binom{2n}{n},$$

then it would be true that

$$\frac{1}{n+1}\binom{2n}{n} = \left(\frac{4n-2}{n+1}\right)\frac{1}{n}\binom{2n-2}{n-1}.$$

Multiplying both sides by $n + 1$, we would have

$$\binom{2n}{n} = \frac{4n-2}{n}\binom{2n-2}{n-1}.$$

One can prove this by algebraic calculation, but is there a *combinatorial* proof? Speaking of combinatorial proofs, can readers find such proofs that formulas (5)–(7) all produce the same numbers?

In another direction, we have seen several articles in this department (e.g., Kobayashi 2006) that explain why the sum of the entries in the nth row of Pascal's triangle is 2^n. What about the sum of the *squares* of the entries in the nth row?

BIBLIOGRAPHY

Askey, Richard. "Fibonacci and Lucas Numbers." *Mathematics Teacher* 98 (May 2005): 610–15.

Feller, W. *An Introduction to Probability Theory and Its Applications.* Vol. 1, 3rd ed. New York: Wiley, 1968.

Gardner, M. "Mathematical Games." *Scientific American* 234 (June 1976): 120–25.

Guy, R. K. Letters to the Editor. *Mathematics Magazine* 61 (Oct. 1988): 269.

———. "The Second Law of Small Numbers." *Mathematics Magazine* 63 (Feb. 1990): 3–20.

Jarvis, F. "Catalan Numbers." *Mathematical Spectrum* 36 (2003–2004): 9–12.

Kobayashi, Yukio. "Relations among Powers of 2, Combinations, and Symbolic Algebra." *Mathematics Teacher* 99 (April 2006): 577–78.

Koshy, T. *Discrete Mathematics with Applications*. Burlington, MA: Elsevier/Academic Press, 2004.

———. *Fibonacci and Lucas Numbers with Applications.* New York: Wiley, 2002.

Larcombe, P. "The 18th Century Chinese Discovery of the Catalan Numbers." *Mathematical Spectrum* 32 (1999–2000): 5–6. ∞

THOMAS KOSHY, tkoshy@frc.mass.edu, is a professor of mathematics at Framingham State College, Framingham, MA 01701. An author of six mathematics books, he loves to share his passion for mathematics and its beauty and power with others.