Числа Каталана

По статье А.В. Спивака [1]

Задача 1. Вычисления произведений. (Э.Каталан), 1838г. Сколькими способами можно вычислить произведение n чисел?

Произведение abc можно понимать двояко: (ab)c и a(bc). Конечно, по закону ассоциативности, результат не зависит от порядка умножений. Но промежуточные результаты — зависят!

Произведение abcd можно понимать пятью способами: ((ab)c)d, (a(bc))d, a((bc)d), a(b(cd)) и (ab)(cd).

Произведение abcde — четырнадцатью способами. Чтобы убедиться в этом, не обязательно их все выписывать. Достаточно заметить, что есть 5 способов вида a(bcde), 2 способа вида (ab)(cde) , 2 способа вида (abc)(de) и 5 способов вида (abcd)e.

Произведение abcdef можно понимать 42 способами: 14 способов вида a(bcdef), 5 способов вида (ab)(cdef) , $2\cdot 2$ способа вида (abc)(def), 5 способов вида (abcd)(ef) и 14 способов вида (abcde)f .

И т.д.

Определение. *Последовательность Каталана*[1] (обозначается $\{C_n\}_1^\infty$) — это числовая последовательность, n-й член которой выражает *число способов* расставить скобки в произведении n множителей.

При этом договоримся[2] считать, что $C_1 = 1, C_2 = 1$.

Выведем рекуррентную формулу для C_n (то есть формулу, выражающую очередной член последовательности через предыдущие). Пронумеруем знаки умножения

$$a_1 \underset{(1)}{\times} a_2 \underset{(2)}{\times} a_3 \underset{(3)}{\times} \ldots \underset{(n-2)}{\times} a_{n-1} \underset{(n-1)}{\times} a_n \ .$$

Все способы расставить скобки в произведении можно разделить на $n-1$ недублирующихся случаев, когда в последнюю очередь выполняется умножение (1), или (2), … , или $(n-1)$.

В каждом текущем r-м случае произведение $(a_1\times\ldots\times a_r)\times(a_{r+1}\times\ldots\times a_n)$ получается, как произведение как-то скомбинированных первых r символов на

---

[1] Эжен Шарль Каталан (1814—1894) — бельгийский математик.

[2] Если число одно, то формально даже нельзя написать произведение; если чисел два: ab, то скобки писать излишне. Но для дальнейших рассуждений удобно положить первые два числа Каталана равными 1, а не 0.

произведение каким-то образом скомбинированных остальных $n-r$ символов.

Первые r символов по определению могут быть скомбинированы C_r способами, последние $n-r$ символов — C_{n-r} способами. В силу правил произведения и суммы имеет место формула:

$$C_n = C_1C_{n-1} + C_2C_{n-2} + \cdots + C_{n-1}C_1, \text{ где } n \in \mathbb{N}\backslash\{1,2\} \qquad (*)$$

Другими словами, чтобы посчитать, например, C_7, достаточно выписать одно за одним, первые шесть чисел Каталана,

1	1	2	5	14	42
42	14	5	2	1	1

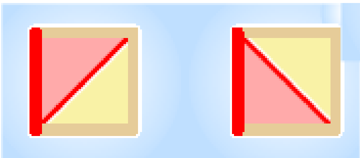
а под ними — те же числа в обратном порядке; умножив каждое верхнее число на соответствующее нижнее и сложив, получаем $C_7 = 132$.

Существует *несколько десятков* эквивалентных формулировок задач, с помощью которых можно ввести числа Каталана. Мы приведём ещё только одну (ну, может быть две, ☺).

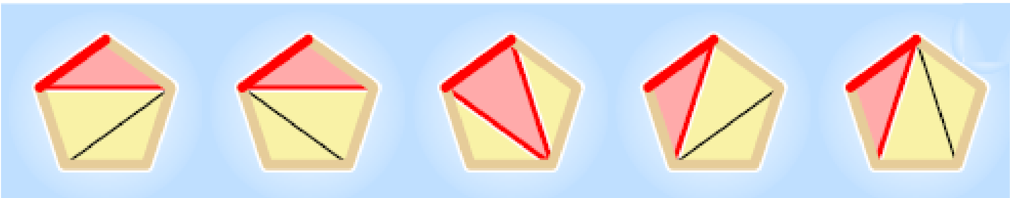
Задача 2. Разрезания на треугольники[3] (Л.Эйлер)

Сколькими способами можно выпуклый $(n+1)$*-угольник разрезать на треугольники диагоналями, не пересекающимися внутри этого* n*-угольника?* Ответ для треугольника $n=2$ тривиален: никаких диагоналей проводить не надо.

Для четырехугольника $(n=3)$ можно провести любую из двух диагоналей, так что способов два.

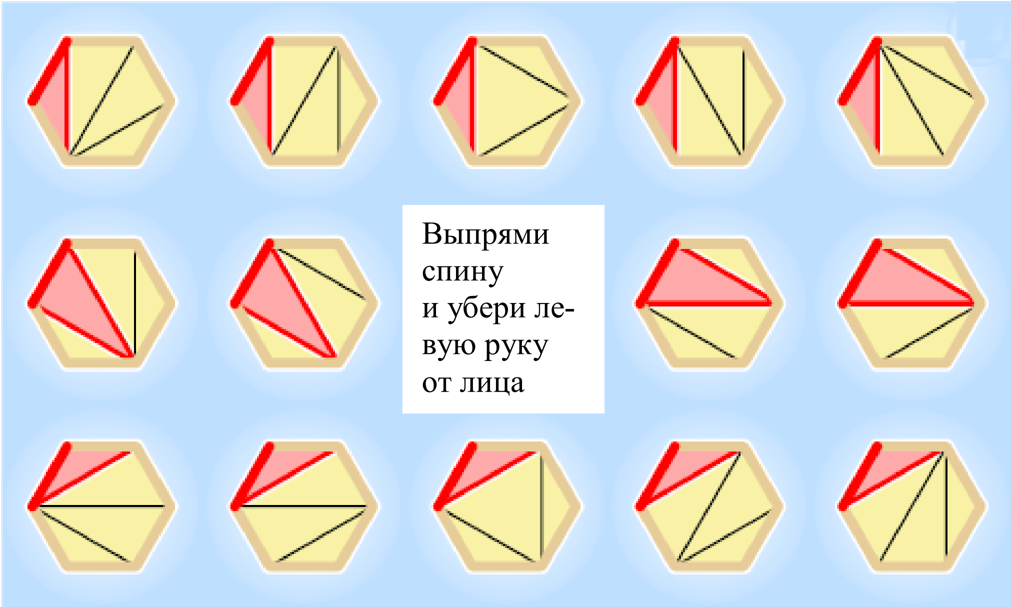


Для пятиугольника $(n=4)$ — из любой вершины две диагонали, 5 способов.



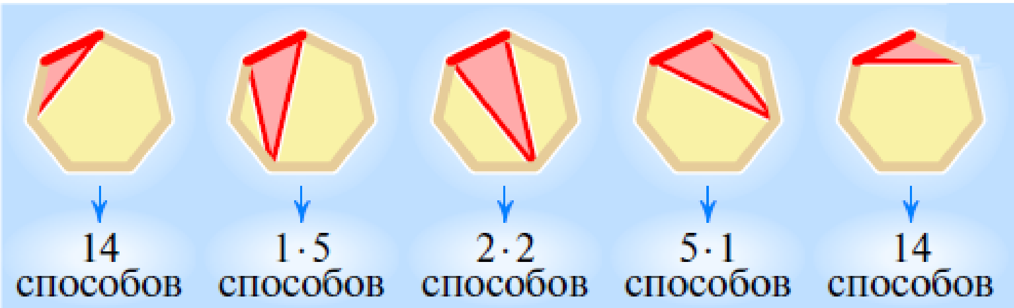
Для шестиугольника $(n=5)$ получаем первый нетривиальный ответ: 14 способов.

[3] При формулировке Леонардом Эйлером этой задачи последовательность 1,1, 2, 5, 14, 42, 139, … впервые и возникла. Эжен Каталан жил 100 лет спустя Леонарда Эйлера. Но, как часто бывает в математике, объект не получил имени первооткрывателя.



Далее нужно придумать систему, будем действовать так: фиксируем одну из сторон (выделена жирно красным) и «сажаем» на неё треугольники, рассматривая разрезания в зависимости от того, какая точка является третьей вершиной. Третья вершина как бы делает обход оставшихся $(n-1)$ вершин многоугольника (у нас — против часовой стрелки).

Для семиугольника (при $n=6$) имеем 5 разных случаев. В первом и последнем из них количество разбиений равно 14, ибо после отрезания треугольника остаётся шестиугольник. Во втором и четвёртом случаях при вырезании



треугольника семиугольник распадается на треугольник и пятиугольник. В третьем случае семиугольник распадается на два четырёхугольника. Поскольку каждый из них можно разбить двумя способами, получаем 2 ·2 = 4 варианта. Итак, семиугольник можно разбить всего 14+5+2· 2+5+14 = 42 способами.

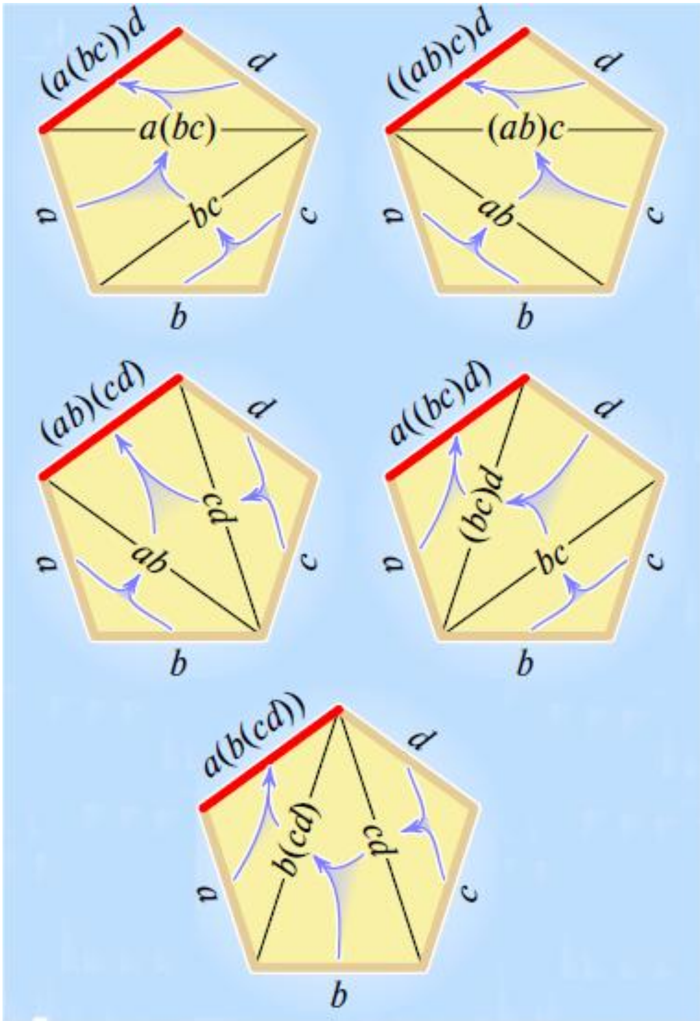
И так далее.

Если мы обозначим $\tilde{C}_n$ — число способов разрезать $(n+1)$-угольник непересекающимися диагоналями (последних $(n-2)$ штуки) на $(n-1)$ треугольник, то, договорившись[4] считать $\tilde{C}_1 = \tilde{C}_2 = 1$, мы получим для её об-

[4] Для отрезка и для треугольника нельзя провести ни одной диагонали. Но нам по прежнему удобно считать первые два числа последовательности равными 1, а не нулю. Удобство состоит в том, что тогда (*) работает для всех натуральных n, начиная с $n=3$.

щего члена формулу аналогичную (*). Сделать самостоятельно (почти всё уже сделано). Так как рекуррентные формулы и первые два члена последовательностей $\{C_n\}$ и $\{\tilde{C}_n\}$ — одинаковы, то речь идет об одной и той же последовательности чисел — последовательности Каталана.

Между задачами 1 и 2 есть более явная связь: взаимно-однозначное соответствие (**биекция**) между разбиениями на треугольники и способами подсчета произведений. Как заметил в 1961 г. Фордер[5], можно выделить одну сторону $(n+1)$-угольника и написать сомножители около других его сторон, по одной букве у каждой стороны, а затем «стягивать» треугольники, на двух сторонах которых уже что-то написано, записывая произведение на третью сторону. При этом порядок сомножителей определяет обход треугольников, в нашем примере, против часовой стрелки. См. рис. справа для пятиугольника.



Способ дает биекцию между задачами.

Упражнение. Порядковые номера всех нечётных чисел Каталана являются степенями двойки. Строже: число C_n нечётно тогда и только тогда, когда n является натуральной степенью двойки.

Указание. Поскольку рекуррентная формула (*) одинаково читается как слева направо, так и справа налево, то чётны все числа C_{2n+1}, где $n \in \mathbb{N}$. По той же причине число C_{2n} чётно тогда и только тогда, когда чётно число C_n.

Замечание. Выше мы нумеровали последовательность Каталана с 1 (ведь неудобно считать, когда на руках нет пальцев). Но часто математики нумеровать начинают с нуля (отсутствие пальцев их не смущает). Понимая, что натуральный ряд биективен любой своей бесконечной части, а, также, любой бесконечной части множества целых чисел, мы понимаем, что нумеровать можно вообще с любого целого отрицательного (можно развить мысль про счет на отрицательном числе пальцев). Все дело в том, какое удобство будет

[5] Фордер Генри Джордж (1889—1981) — англ. математик, большую часть жизни прожил в Новой Зеландии.

нести перенумерация. В он-лайн энциклопедии целочисленных последовательностей (On-Line Encyclopedia of Integer Sequences, OEIS) https://oeis.org/A000108 числа Каталана нумеруются с нуля. Ниже мы свяжем числа Каталана с биномиальными коэффициентами и нам будет удобно, что нумерация начнётся с нуля. Формулу (*) перепишем так:

$$C_n = C_0 C_{n-1} + C_2 C_{n-2} + \cdots + C_{n-1} C_0 = \sum_{i=0}^{n-1} C_i C_{n-1-i}\,, \qquad (0)$$

где $n \in \mathbb{N}$, $C_0 = 1$.

Треугольник Каталана

Вспомним треугольник Паскаля — по боковым сторонам треугольника стоят единицы, числа внутри треугольника расположены в шахматном порядке, и каждое внутреннее число равно сумме двух чисел, стоящих непосредственно над ним. Треугольник Паскаля можно получить также, если стартовать со строки из бесконечного числа нулей и одной единицы и ниже каждое очередное число писать в шахматном порядке по тому же правилу. Нули потом стереть. Напоминаем, что числа треугольника Паскаля суть биномиальные коэффициенты C_n^k, $n,k = 0,1,\ldots$, $k \le n$. Ведь правило его образования дублирует свойство $C_n^k = C_{n-1}^k + C_{n-1}^{k-1}$.

0	0		0		0		0		0		0		1		0		0		0		0		0
1		0		0		0		0		0		1		1		0		0		0		0	
2	0		0		0		0		0		1		2		1		0		0		0		0
3		0		0		0		0		1		3		3		1		0		0		0	
4			0		0		0		1		4		6		4		1		0		0		0
5		0		0		0		1		5		10		10		5		1		0		0	
6			0		0		1		6		15		20		15		6		1		0		0
7		0		0		1		7		21		35		35		21		7		1		0	
8			0		1		8		28		56		70		56		28		8		1		0
9		0		1		9		36		84		126		126		84		36		9		1	
10			1		10		45		120		210		252		210		120		45		10		1

Треугольник Паскаля (первые 11 строк)

Треугольник Каталана генерируется схожим образом: есть строка из бесконечного числа нулей и одной единицы, пишем ниже числа в шахматном порядке по тому же правилу, НО левее центральной вертикальной линии всегда пишем нули. Когда надоест рисовать треугольник — нули стираем.

0			1		0		0		0		0		0
1		0		1		0		0		0		0	
2			1		1		0		0		0		0
3		0		2		1		0		0		0	
4			2		3		1		0		0		0
5		0		5		4		1		0		0	
6			5		9		5		1		0		0
7		0		14		14		6		1		0	
8			14		28		20		7		1		0
9		0		42		48		27		8		1	
10			42		90		75		35		9		1

⇦ Треугольник Каталана (первые 11 строк)

На вертикальной линии (выделено темно-синим) — числа Каталана!

Причина в том, что любое число треугольника Каталана равно количеству путей, которыми можно прийти в соответствующую точку из вершины, двигаясь вниз-влево или вниз-вправо (задачу рассказать на семинаре, не обсуждали биекцию….)

Между треугольниками Паскаля и Каталана есть тесная связь.
Теорема. Каждое число треугольника Каталана есть разница соседних чисел в треугольнике Паскаля, то есть имеет место формула

$$K_n^k = C_n^k - C_n^{k-1}, \quad n,k \in \mathbf{N}, k \le n \quad K_0^0 = 1. \qquad (1)$$

где K_n^k — числа треугольника Каталана.
Доказательство (индукцией по n).
База индукции. Будем вычитать соседние числа треугольника Паскаля, записывая разность красным в свободные клеточки рядом, правее уменьшаемого. Из крайних единичек вычитаем нули — снова получаем единички. Видно, что для первых 10 строк формула (1) имеет место. Треугольник Каталана как бы «прорастает» сквозь треугольник Паскаля.

0	0		1	1	0		0		0		0		0	
1		1	0	1	1	0		0		0		0		
2	1		2	1	1	1	0		0		0		0	
3		3	0	3	2	1	1	0		0		0		
4	4		6	2	4	3	1	1	0		0		0	
5		10	0	10	5	5	4	1	1	0		0		
6	15		20	5	15	9	6	5	1	1	0		0	
7		35	0	35	14	21	14	7	6	1	1	0		
8	56		70	14	56	28	28	20	8	7	1	1	0	
9		126	0	126	42	84	48	36	27	9	8	1	1	
10	210		252	42	210	90	120	75	45	35	10	9	1	1

«Два в одном»

Пусть наше утверждение верно для какой-то строки с номером k.
Сделаем шаг индукции. Обозначим x, y, z произвольную тройку чисел треугольника Паскаля идущих в k-й строке подряд слева-направо. В силу пред-

положения между ними располагаются числа треугольника Каталана $x-y$ и $y-z$.
По правилу образования треугольников числа под $x, (x-y), y, (y-z), z$ в $(k+1)$-й строке будут такими:

x	x-y	y	y-z	z
	x+y	x-y+y-z	y+z	

Вопрос: будет ли число в красной рамочке разностью двух соседних, то есть будет ли оно подчиняться правилу (1)? Да, очевидно, будет. В силу принципа математической индукции формула (1) доказана.

Прямые формулы для C_n. Зная, что на вертикальной левой линии в треугольнике Каталана стоят числа Каталана, имеем из (1) прямую формулу для их подсчета:

$$C_n = K_n^1 = C_{2n}^n - C_{2n}^{n-1}, \ n = 1, \dots, \quad C_0 = 1 \ . \tag{2}$$

Другими словами, число Каталана C_n равно разности центрального биномиального коэффициента и соседнего с ним в той же строке треугольника Паскаля.
Комментарий. Обратите внимание, что числа Каталана в одноименном треугольнике идут через строчку (следствие шахматного порядка), поэтому сочетания в формуле (2) берутся из $2n$.

Применив формулу для сочетаний, получим из (2)

$$C_n = C_{2n}^n - C_{2n}^{n-1} = \frac{(2n)!}{n!n!} - \frac{(2n)!}{(n-1)!(n+1)!} = \frac{(2n)!}{(n-1)!n!}\left(\frac{1}{n} - \frac{1}{n+1}\right) = \frac{(2n)!}{n!(n+1)!}.$$

Т.е.

$$C_n = \frac{(2n)!}{n!(n+1)!}, \quad n \in \mathbf{N} \cup \{0\}. \tag{3}$$

Или, свернув обратно по формуле сочетаний, получим из (3)

$$C_n = \frac{1}{n+1} C_{2n}^n, \quad n \in \mathbf{N} \cup \{0\}. \tag{4}$$

Или, выделив в правой части (3) $C_{n-1} = \dfrac{(2n-2)!}{(n-1)!n!}$, получим ещё одну **рекуррентную** формулу для чисел Каталана

$$C_n = \frac{2(2n-1)}{(n+1)} C_{n-1}, \quad n \in \mathbf{N}, \quad C_0 = 1. \tag{5}$$

Производящие функции

Материал выходит за рамки школьной программы, вместе с тем, и без понимания «как это работает» позволяет решать задачи.

Цель: получить явную формулу для чисел Каталана (4) иным способом. Способ являет собой общий мощный метод (метод производящих функций), поэтому говорим о нём.

Определение. *Производящая функция (ПФ) последовательности* $\{a_n\}_0^\infty$, — это формальный степенной ряд

$$\sum_{n=0}^{\infty} a_n x^n .$$

Исходя из определения, производящая функция для последовательности Каталана 1, 1, 2, 5, 14, 42, 132, 429, . . . — это функция

$$f(x) = 1 + x + 2x^2 + 5x^3 + 14x^4 + \cdots . = \sum_{n=0}^{\infty} C_n x^n .$$

ПФ – это мощный инструмент, который позволяет получать формулы, которые иногда очень сложно, если вообще возможно, получить.

Давайте помножим $f(x)$ на себя

$$f(x)f(x) = (1 + x + 2x^2 + 5x^3 + 14x^4 + \cdots)(1 + x + 2x^2 + 5x^3 + 14x^4 + \cdots)=$$

$$= \sum_{n=0}^{\infty} C_n x^n . \sum_{k=0}^{\infty} C_k x^k \ = \ C_0C_0 + (C_1C_0 + C_1C_0)x + \cdots =$$

$$= \ C_0C_0 + (C_1C_0 + C_1C_0)x + \cdots =$$

перемножая ряд на ряд, нужно сортировать слагаемые по степеням x: при нулевой степени будет, очевидно, C_1, при первой степени C_2, и т.д., т.е. в качестве коэффициентов снова выступают числа Каталана!

$$= \ C_1 + C_2x + C_3x^2 + \cdots =$$

«подправим», чтобы выделить $f(x)$

$$= \frac{x(C_1 + C_2x + C_3x^2 + \cdots)}{x} = \frac{C_1x + C_2x^2 + C_3x^3 + \cdots}{x} = \frac{f(x) - C_0}{x}$$

Таким образом, мы имеем уравнение на $f(x)$:

$$xf^2 - f + 1 = 0,$$

решая его, как квадратное относительно f, получаем[6]

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

Далее, нужно знание в матанализе, чтобы разложить функцию справа в степенной ряд (это делается по стандартным формулам). Получив это разложение, увидим, что для коэффициентов разложения (то есть чисел Каталана) имеет место формула (4).

Список источников

1. Спивак А.В. Числа Каталана.

http://www.mccme.ru/circles/oim/materials/spivak-04-1.pdf (версия без рисунков)

2. Р. Грэхем, Д. Кнут, О. Паташник. Конкретная математика. Основания информатики.

3. М.Гарднер. Числа Каталана. Квант, 1978 №7

http://kvant.mccme.ru/1978/07/chisla_katalana.htm

На будущее:

Определение. *Сверткой* двух последовательностей $\{a_n\}_0^\infty$, $\{b_n\}_0^\infty$, называют последовательность $\{c_n\}_0^\infty$, где

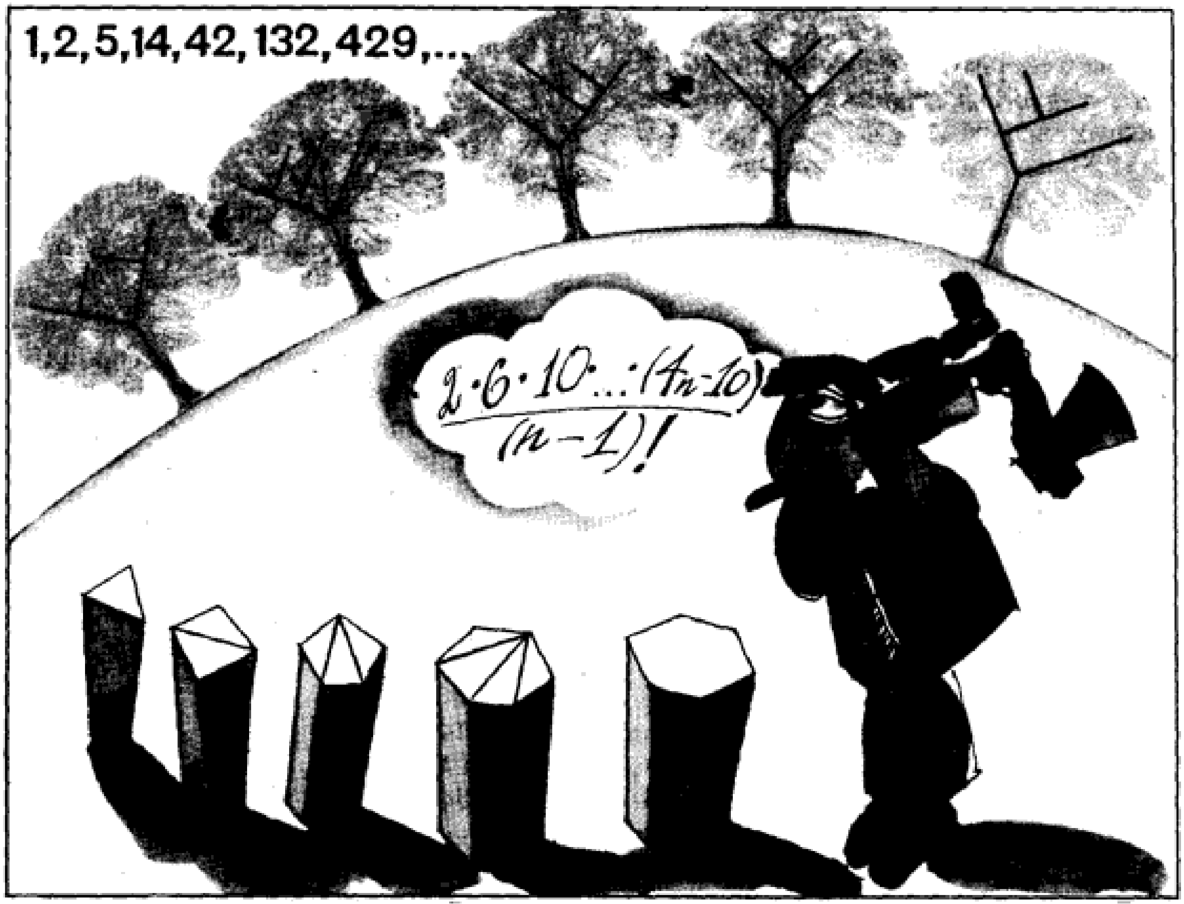
$$c_n = \sum_{i=0}^{n} a_i b_{n-i} .$$

Формулу (0) можно переписать в виде

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

Свёртке двух последовательностей соответствует умножение их производящих функций.

[6] В формуле взяли знак минус, т.к. плюс не отвечает условию $f(0) = 1$.


1,2,5,14,42,132,429,...
$\frac{2\cdot 6\cdot 10\cdot\ldots\cdot(4n-10)}{(n-1)!}$