УДК 519.688

Многомерное расширение классических чисел Каталана для решения прикладных задач анализа случайных точечных изображений*

А.Л. Резник, А.А. Соловьев

Институт автоматики и электрометрии СО РАН (Новосибирск, Россия)

Solving the Applied Problem of Random-Dot Images Analysis Using a Multidimensional Extension of Classic Catalan Numbers*

A.L. Reznik, A.A. Soloviev

Institute of Automation and Electrometry SB RAS (Novosibirsk, Russia)

Введенное в статье понятие обобщенной последовательности Каталана является полезным инструментом при решении многих теоретических и прикладных вероятностно-комбинаторных задач. В сочетании с алгоритмами, осуществляющими программные аналитические преобразования, обобщенные числа Каталана упрощают решение многих задач информатики и прикладной математики. В частности, они оказываются эффективным средством решения задач, относящихся к регистрации случайных точечных изображений, при построении преобразований сигналов различной степени гладкости, при разработке оптимальных по быстродействию алгоритмов поиска импульсно-точечных объектов со случайным временем генерации мгновенных импульсов. Предложение авторов формулировать задачи перечислительной комбинаторики в словарно-символьной форме естественным образом приводит к многомерным расширениям классических чисел Каталана и обладает несколькими преимуществами. Совместное применение многомерных чисел Каталана и высокопроизводительных систем компьютерной алгебры позволило авторам решить ряд сложных прикладных задач, связанных с надежностью регистрации случайных точечных изображений.

***Ключевые слова**:* обобщенные числа Каталана, перечислительная комбинаторика, системы компьютерной алгебры, случайные точечные изображения.

The paper introduces the concept of generalized Catalan numbers as a useful tool for solving many theoretical and applied combinatorial-probabilistic problems. Combined with analytical transformation software algorithms, the generalized Catalan numbers simplify the solution of many problems in computer science and applied mathematics. In particular, they turn out to be an effective tool for solving problems related to the registration of random-dot images, in signal transformations of various degrees of smoothness, and when developing speed-optimal algorithms for searching for pulsed-point objects with a random generation time of super short pulses. The proposal of the authors to formulate the problems of enumerative combinatorics in a word-symbolic form naturally leads to multidimensional extensions of the classical Catalan numbers and has several advantages. The combined use of multidimensional Catalan numbers and high-performance computer algebra systems enables solving a number of complex applied problems related to the reliability of the registration of random-dot images.

***Keywords**:* generalized Catalan numbers, enumerative combinatorics, computer algebra systems, random bitmaps.

DOI: 10.14258/izvasu(2023)4-13

Введение

Числа Каталана встречаются в огромном количестве приложений теории вероятностей и математической статистики [1–3]. Классические числа Каталана могут быть заданы многими эквивалентными способами, например прямыми формулами

$$C_n = \frac{(2n)!}{n!(n+1)!} \tag{1}$$

* Настоящая работа поддержана Министерством науки и высшего образования Российской Федерации (проект № AAA-A17-117052410034-6).

или

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

либо одним из рекуррентных соотношений

$$\begin{aligned} &C_0 = 1, \\ &C_\text{n} = \sum_{k=0}^{n-1} C_\text{k} C_{n-k-1}, \quad n > 0 \end{aligned} \qquad (3)$$

или

$$\begin{aligned} &C_0 = 1, \\ &C_n = \frac{4n-2}{n+1} C_{n-1}, \quad n > 0 \end{aligned} . \qquad (4)$$

Цель настоящего сообщения состоит в представлении нового расширения классических чисел Каталана, базирующегося на словарно-символьной статистике. Описаны способы нахождения их явного аналитического вида и применение в задачах, относящихся к анализу случайных точечных изображений.

1. Метод отражения, примененный французским математиком Д. Андре для решения баллотировочной проблемы. Связь баллотировочной задачи с числами Каталана

Начнем с часто используемого в перечислительной комбинаторике метода отражения, автором которого считается французский математик Дезире Андре [4]. Один из таких примеров приводится в монографии В. Феллера [5, гл. III] и относится к баллотировочной задаче, которую в 1887 году сформулировал (и представил первое решение) Жозеф Бертран [6]: «Предположим, что кандидаты А и В участвуют в выборах. Если a голосов были отданы за кандидата А и b голосов были отданы за кандидата В, где $a>b$, то вероятность того, что А оставался впереди В в течение всего времени подсчета избирательных бюллетеней, равна $(a-b)/(a+b)$».

Баллотировочная задача и решеточные пути. В терминах решеточных путей предварительная (по отношению к баллотировочной задаче) подзадача формулируется следующим образом: «Даны целые положительные числа a и b, причем $a>b$. Требуется найти количество „хороших“ решеточных путей, т.е. путей, начинающихся в начале координат и состоящих из a „шагов вверх“ (1,1) и b „шагов вниз“ (1,-1), при условии, что никакой шаг не заканчивается на оси X».

Решение предварительной подзадачи с помощью метода отражения. Пусть T означает конечную точку пути $(a+b, a-b)$. Так как каждый «хороший» путь должен начинаться с «шага вверх», то количество «хороших» путей совпадает с количеством путей из точки (1,1) в точку T, которые никогда не касаются оси X. Множество «плохих» путей из точки (1,1) в точку T, которые где-либо касаются оси X, находится во взаимно-однозначном соответствии с множеством всех путей из точки (1,-1) в точку T. Это легко усмотреть, если начальные сегменты каждого из таких «плохих» путей, начинающихся в точке (1,1) и заканчивающихся его первым прикосновением к оси X, зеркально отразить относительно оси X. Вычитая общее число таких «отраженных» путей, начинающихся в точке (1,-1) и заканчивающихся в точке $(a+b, a-b)$, из общего числа всех путей, соединяющих точку (1, 1) с точкой $(a+b, a-b)$, получим общее количество «хороших путей»:

$$\binom{a+b-1}{a-1} - \binom{a+b-1}{b-1} = \frac{a-b}{a+b}\binom{a+b}{a}. \qquad (5)$$

Предварительная подзадача решена. Поскольку вероятность того, что кандидат А всегда опережал кандидата В течение всего времени подсчета избирательных бюллетеней, есть отношение количества «хороших» путей (5) к общему количеству всех путей из начала координат в точку $(a+b, a-b)$, которое равно $\binom{a+b}{a}$, искомая вероятность равна $(a-b)/(a+b)$. Это и есть ответ на баллотировочную задачу Бертрана.

Теперь о том, как связана баллотировочная задача с числами Каталана. Эта связь самая прямая: если в формуле (5) положить $a=n+1$ и $b=n$, то получим классические числа Каталана, причем правая часть равенства (5) перейдет в формулу (1), а левая часть — в формулу (2).

2. Обобщение классических чисел Каталана с применением словарно-символьной статистики

В перечислительной комбинаторике известно немало расширений чисел Каталана, для которых классическая последовательность Каталана является лишь частным случаем. В настоящем разделе предлагается новое расширение классических чисел Каталана, базирующееся на словарно-символьном подходе к постановке комбинаторных задач. Основным достоинством этого расширения, как уже отмечалось, является универсальность формулировок и простота обобщения словарно-символьных задач на многомерный случай.

2.1. Обобщение классической последовательности Каталана на двумерный случай в терминах словарно-символьной статистики. Решение двумерной задачи «методом отражения»

ЗАДАЧА № 1

В образовании слова участвуют N_a символов «a» и N_b символов «b». Сколько из этих слов таких, что при просмотре слова слева направо количество встреченных символов «b» никогда не превышает количества встреченных символов «a»? (Естественно, предполагается, что $N_a \geq N_b$.)

Решение:

$$WORD1(N_a, N_b) = \frac{(N_a + N_b)!}{N_a! N_b!} - \frac{(N_a + N_b)!}{(N_a + 1)!(N_b - 1)!}. \quad (6)$$

Двумерная последовательность $WORD1(N_a, N_b)$ является простейшим расширением классических чисел Каталана, которые получаются из нее при $N_a = N_b$. Формулу (6) нетрудно получить, если задачу № 1 заменить на эквивалентную ей задачу с монотонными путями в прямоугольнике (см. рис. 1). В получающейся задаче с решеточными путями требуется найти количество «хороших» монотонных путей, которые выходят из начала координат (0,0) и ведут в точку (N_a, N_b), не пересекая диагонали левого квадрата со стороной N_b.

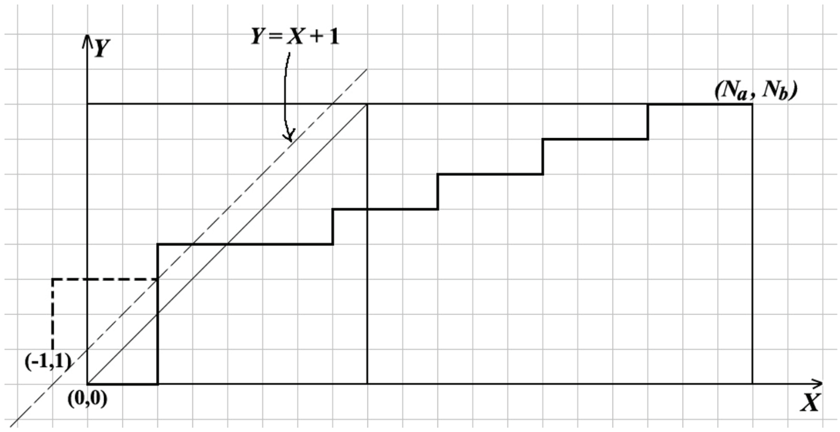

Рис. 1. Иллюстрация к задаче № 1 о простейшем расширении чисел Каталана

Как и в разделе 1, на предварительном этапе будет необходимо определить общее количество «плохих» путей, пересекающих диагональ левого квадрата. С применением уже использованного «метода отражения» легко доказывается, что их число совпадает с общим числом монотонных путей, выходящих из точки (-1,+1) и ведущих в точку (N_a, N_b). Количество же «хороших» путей есть разница между всеми монотонными путями и всеми «плохими» путями, которая и описывается формулой (6).

В свою очередь, задача № 1 допускает следующее обобщение:

ЗАДАЧА № 2

В образовании слова участвуют N_a символов «a» и N_b символов «b». Сколько из этих слов таких, что при просмотре слова слева направо количество встреченных символов «b» никогда не превышает количества встреченных символов «a» более чем на k? (Предполагается, что $k \geq 0$ и $N_a + k \geq N_b$.)

Решение:

$$WORD2(N_a, N_b, k) = \frac{(N_a + N_b)!}{N_a! N_b!} - \frac{(N_a + N_b)!}{(N_a + k + 1)!(N_b - k - 1)!}. \quad (7)$$

Это дальнейшее расширение чисел Каталана, а именно: при k=0 задача № 2 эквивалентна задаче № 1, а при $N_a = N_b$ и k=0 получаем классические числа Каталана. Иллюстрацией к задаче № 2 служит рисунок 2. Формула (7) соответствует количеству монотонных путей из точки (0,0) в точку (N_a, N_b), лежащих в заштрихованной зоне.

2.2. Трехмерное обобщение классической последовательности Каталана

ЗАДАЧА № 3

В образовании слова участвуют N_a символов «a», N_b символов «b» и N_c символов «c». Требуется указать общее число слов $WORD3(N_a, N_b, N_c)$, для которых одновременно выполняются два условия:

1) при просмотре слова слева направо количество встреченных символов «b» никогда не превышает количества встреченных символов «a»;

2) при просмотре слова справа налево количество встреченных символов «c» никогда не превышает количества встреченных символов «a».

Задача № 3 может быть решена чисто геометрическими средствами с помощью последовательного трехкратного применения уже упоминавшегося нами метода «зеркального отражения». Для этого она сначала должна быть переведена в эквивалентную ей задачу отыска-

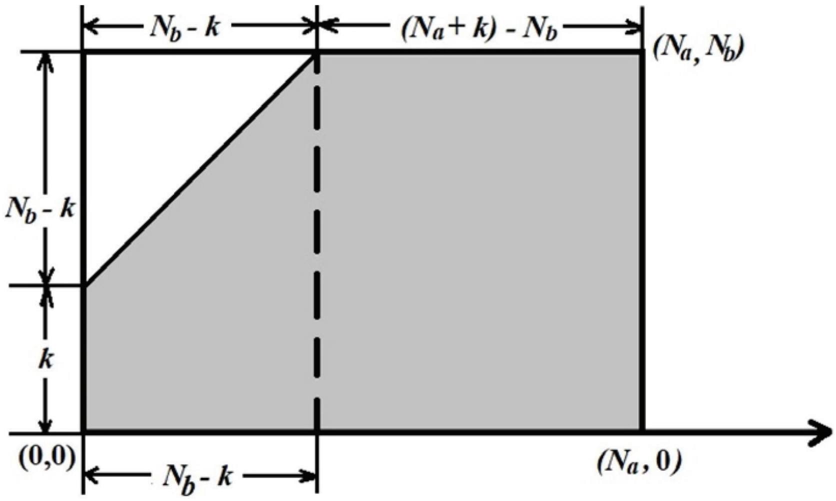

Рис. 2. Иллюстрация к задаче № 2

ния общего числа монотонных путей, каждый из которых характеризуется тем, что соединяет вершины (0,0,0) и (N_a,N_b,N_c) трехмерного ограничивающего параллелепипеда, не пересекая при этом ни одной из двух отсекающих плоскостей P_1: $X=Y$ и P_2: $Z=X+N_c-N_a$.

Решение этой переформулированной задачи проводится следующим образом: из общего количества монотонных путей S_0, ведущих из точки (0,0,0) в точку (N_a,N_b,N_c), которое равно $S_0=\dfrac{(N_a+N_b+N_c)!}{N_a!N_b!N_c!}$, необходимо вычесть количество путей, которые выходят за пределы хотя бы одного из полупространств $X\geq Y$ или $Z\geq X+N_c-N_a$. Иными словами, требуется сложить количество путей S_1, выходящих за отсекающую плоскость P_1, с количеством S_2 путей, выходящих за отсекающую плоскость P_2, а затем вычесть из полученной суммы количество S_{12} путей, выходящих как за плоскость P_1, так и за плоскость P_2 (поскольку в сумме они учтены дважды).

В результате получим следующее общее решение и для задачи № 3 о количестве трехсимвольных слов, и для эквивалентной ей задачи о количестве монотонных путей, ведущих из точки (0,0,0) в точку (N_a,N_b,N_c) и не пересекающих при этом ни одну из плоскостей P_1: $X=Y$ и P_2: $Z=X+N_c-N_a$:

$$WORD3(N_a,N_b,N_c)=S_0-S_1-S_2+S_{12}=\frac{(N_a+N_b+N_c)!}{N_a!N_b!N_c!}-\frac{(N_a+N_b+N_c)!}{(N_a+1)!(N_b-1)!N_c!}-$$
$$-\frac{(N_a+\mathrm{N_b}+N_c)!}{(N_a+1)!N_b!(N_c-1)!}+\frac{(N_a+\mathrm{N_b}+N_c)!}{(N_a+2)!(N_b-1)!(N_c+1)!},\quad N_a>N_b+N_c-2. \tag{8}$$

Важно отметить, что рассчитанное соотношение для S_{12} и, как следствие этого, полученная формула (8) справедливы только при выполнении условия $N_a>N_b+N_c-2$. В этом случае удается показать, что любой путь, удовлетворяющий поставленным в задаче ограничениям, не может пересечься с плоскостью P_2 раньше пересечения с плоскостью P_1. При отсутствии же такого условия решение задачи № 3 усложняется, и для его отыскания необходимы более тонкие методы, выходящие за рамки нашего рассмотрения [7].

Заметим, что при $N_c=0$ задача № 3 переходит в задачу № 2, а это означает, что решение задачи № 3 при $N_c>0$ привело к еще одному расширению чисел Каталана — на этот раз на трехмерный случай.

2.3. Прикладные научные задачи, решаемые с использованием обобщенных чисел Каталана

Числа Каталана (включая описанные в настоящей статье обобщенные числа «словарно-символьного происхождения») по самой своей природе являются инструментом дискретного анализа (а точнее — одним из инструментов перечислительной комбинаторики), с их помощью могут быть решены многие прикладные вероятностные задачи непрерывного характера. Так, при проведении исследований, связанных с оценкой надежности цифровой регистрации случайных точечных полей, возникает следующая простая по постановке вероятностная задача: «Требуется найти вероятность $P_{n,k}(\varepsilon)$ того, что при случайном бросании n точек на интервал (0,1) не будет образовано ни одного подынтервала размером ε, содержащего более k точек».

Несмотря на кажущуюся простоту постановки, замкнутое аналитическое решение приведенной задачи известно лишь для частного случая k=1 [8–9]. Но с применением разработанных нами высокопроизводительных программных систем компьютерной алгебры [10–11] и расширенных чисел Каталана удалось установить и затем строго доказать ряд новых ранее неизвестных аналитических соотношений. В частности, получены точные формулы:

$$P_{2m,2}(\varepsilon)=\begin{cases} C_{2m}^{m}(1-(m-1)\varepsilon)^{2m}-C_{2m}^{m-1}(1-(m-1)\varepsilon)^{2m}-C_{2m}^{m-2}(1-m\varepsilon)^{m+2}(1-(m-2)\varepsilon)^{m-2}+ \\ +2C_{2m}^{m-3}(1-m\varepsilon)^{m+3}(1-(m-2)\varepsilon)^{m-3}-C_{2m}^{m-4}(1-m\varepsilon)^{m+4}(1-(m-2)\varepsilon)^{m-4}, \quad \dfrac{1}{m+1}<\varepsilon<\dfrac{1}{m}; \\ \dfrac{1}{m+1}C_{2m}^{m}(1-(m-1)\varepsilon)^{2m}, \quad \dfrac{1}{m}<\varepsilon<\dfrac{1}{m-1}; \end{cases}$$

$$P_{2m+1,2}(\varepsilon)=\begin{cases} C_{2m+1}^{m+1}(1-m\varepsilon)^{m+1}(1-(m-1)\varepsilon)^{m}-2C_{2m+1}^{m+2}(1-m\varepsilon)^{m+2}(1-(m-1)\varepsilon)^{m-1}+ \\ +C_{2m+1}^{m+3}(1-m\varepsilon)^{m+3}(1-(m-1)\varepsilon)^{m-2}, \quad \dfrac{1}{m+1}<\varepsilon<\dfrac{1}{m}; \end{cases}$$

и рассчитан ряд замкнутых вероятностных зависимостей и асимптотических соотношений.

Библиографический список

1. Gardner M. Mathematical Games, Catalan numbers: an integer sequence that materializes in unexpected places // Scientific American. 1976.

2. Hayes B. A Question of Numbers // American Scientist. 1996. № 1.

3. Saracevic M., Sharma S., Ahmad K. A novel block encryption method based on Catalan random walks // Multimedia Tools and Applications. 2021. Vol. 8. № 7. Doi: 10.1007/s11042-021-11497-5

4. André, D. Solution directe du problème résolu par M. Bertrand // Comptes Rendus Acad. Sci. Paris. 1887. Vol. 105.

5. Feller W. An Introduction to Probability Theory and its Applications, 2nd ed. John Wiley: New York, 1957.

6. Bertrand J. Solution d'un probleme // Comptes Rendus de l'Academie des Sciences, Paris. 1887. Vol. 105.

7. Gessel I., Zeilberger D. Random walk in a Weyl chamber // Proc. Amer. Math. Soc. 1992. Vol. 115.

8. Wilks S. Mathematical Statistics. Princeton: Princeton Univ. Press. 1944.

9. Parzen E. Modern Probability Theory and Its Applications. John Wiley and Sons: New York. 1960.

10. Reznik A., Tuzikov A., Soloviev A., Torgov A. Analysis of random point images with the use of symbolic computation codes and generalized Catalan numbers // Optoelectronics Instrumentation and Data Processing. 2016. Vol. 52. № 6. Doi: 10.3103/S8756699016060017.

11. Резник А.Л., Тузиков А.В., Соловьев А.А., Торгов А.В. Интеллектуальная программная поддержка в задачах анализа случайных цифровых изображений // Вычислительные технологии. 2018. Т. 23. № 5. Doi: 10.25743/ICT.2018.23.5.007