Discrete Mathematics Letters
www.dmlett.com

Discrete Math. Lett. **13** (2024) 117–121
DOI: 10.47443/dml.2024.033

Research Article

Letters of a given size in Catalan words

Aubrey Blecher, Arnold Knopfmacher*

School of Mathematics, University of Witwatersrand, Johannesburg, South Africa

(Received: 8 February 2024. Received in revised form: 27 May 2024. Accepted: 29 May 2024. Published online: 15 July 2024.)

Abstract

A Catalan word with n letters, $w = w_1w_2\dots w_n$ over the set of positive integers is a word in which $w_1 = 1$ and for each $i \in \{2,3\dots,n\}$, the inequality $w_i \le w_{i-1}+1$ holds. We find generating functions $A(x,q_i)$ for the number of letters equal to i in Catalan words with n letters. We also develop the generating function $A(x,q_i|j)$ that tracks the number of letters i in Catalan words ending in a letter j. The relationship between $A(x,q_i)$ and $A(x,q_i|j)$ is fundamental to the kernel method used to obtain these generating functions.

Keywords: Catalan words; letters; generating functions; kernel method.

2020 Mathematics Subject Classification: 05A05, 05A15.

1. Introduction

A Catalan word with n letters, $w = w_1w_2\dots w_n$ over the set of positive integers is a word in which $w_1 = 1$ and for each $i \in \{2,3\dots,n\}$, $w_i \le w_{i-1}+1$. We denote the set of Catalan words of length n by $\mathcal{C}_n$, and let $\mathcal{C} = \cup_{n\ge1}\mathcal{C}_n$. Catalan words have received considerable attention from researchers in recent years; for example, see [1, 2, 7, 9]. The number of these words is given by the nth Catalan number $c(n) = \frac{1}{n+1}\binom{2n}{n}$. See Stanley's book [13] for the ways in which $c(n)$ tracks a surprisingly large number of different combinatorial objects. Also, see the introduction to [7] for an explanation of why $c(n)$ counts the number of Catalan words with n letters.

Catalan words may be represented as bargraphs, where a letter i is represented as a column of height i and the letter order of the Catalan word is preserved in the bargraph. Catalan words thus form a subset of objects that can be represented as bargraphs. For a comprehensive survey on bargraphs, see the paper [10]. For other recent papers on bargraphs, see [3–6, 8].

What we aim to do in this paper is to provide a bivariate generating function $A(x,q_i)$ that counts the number of occurrences of letters i, tracked by q_i in Catalan words with n letters, tracked by variable x. In other words,

$$A(x,q_i) := \sum_{n\ge1}\sum_{w\in\mathcal{C}_n} x^n q_i^{|w|_i},$$

where $|w|_i$ is the number of letters i in w.

The method we use is to introduce Catalan words ending with a letter j with generating function $A(x,q_i|j)$. Note that if a Catalan word ends with j, then it necessarily contains all values $i \le j$. Thereafter we introduce a third variable v which tracks the size of the last letter in any Catalan word, in addition to trackers for the number of letters and the number of letters i. This is set up in Equation (6) and solved using the kernel method in Equation (7) and what follows.

2. Counting letters i

Let x track the number of letters and q_i track the number of letters i. We denote the set of Catalan words of length n ending with a letter j by $\mathcal{C}_{n,j}$, and for $j < i$ let $A(x,q_i|j) = \sum_{n\ge1} x^n \sum_{w\in\mathcal{C}_{n,j}} q_i^{|w|_i}$. In other words, for $j<i$, $A(x,q_i|j)$ is the generating function for the number of Catalan words ending in j, according to the number of letters i. Likewise $A(x,q_i|bj)$ is the generating function for the number of such Catalan words ending in bj.

Assume first $i = 1$. Then

$$A(x,q_1|1) = xq_1 + \sum_{b\ge1} A(x,q_1|bj) = xq_1 + xq_1\sum_{b\ge1} A(x,q_1|b).$$

*Corresponding author (arnold.knopfmacher@wits.ac.za).

For $j > 1$,

$$A(x, q_1|j) = \sum_{b \geq j-1} A(x, q_1|bj) = x \sum_{b \geq j-1} A(x, q_1|b).$$

Now, define $A(x, q_1; v) := \sum_{j \geq 1} A(x, q_1|j) v^{j-1}$. We multiply the previous equations by v^{j-1} and sum on $j \geq 1$ to obtain

$$\begin{aligned} A(x, q_1; v) &= xq_1 + xq_1 \sum_{b \geq 1} A(x, q_1|b) + \sum_{j>1} x \sum_{b \geq j-1} A(x, q_1|b) v^{j-1} \\ &= xq_1 + xq_1 \sum_{b \geq 1} A(x, q_1|b) + x \sum_{b \geq 1} \sum_{j=2}^{b+1} A(x, q_1|b) v^{j-1} \\ &= xq_1 + xq_1 \sum_{b \geq 1} A(x, q_1|b) + xv \sum_{b \geq 1} A(x, q_1|b) \frac{1 - v^b}{1 - v} \\ &= xq_1 + xq_1 A(x, q_1; 1) + \frac{xv}{1 - v} (A(x, q_1; 1) - vA(x, q_1; v)). \end{aligned} \tag{1}$$

From this we obtain

$$(1 - v + xv^2) A(x, q_1; v) = (1 - v)(xq_1 + xq_1 A(x, q_1; 1)) + xvA(x, q_1; 1). \tag{2}$$

By the kernel method (see [11, 12]), we choose v by setting $1 - v + xv^2 = 0$. The relevant root for the kernel method is $v = \frac{1 - \sqrt{1-4x}}{2x} = C(x)$, where $C(x)$ is the generating function for the Catalan numbers. Using this value, we obtain

$$(1 - C(x))(xq_1 + xq_1 A(x, q_1; 1)) + xC(x)A(x, q_1; 1) = 0,$$

with solution

$$A(x, q_1) := A(x, q_1; 1) = \frac{q_1 \left(2x + \sqrt{1-4x} - 1\right)}{q_1 \left(2x + \sqrt{1-4x} - 1\right) - \sqrt{1-4x} + 1} = \frac{q_1(1 - C(x))}{q_1 C(x) - C(x) - q_1}. \tag{3}$$

Using variable q_1 to track the number of 1's in Catalan words, we deduce from the above equation that

$$A(x, q_1|1) = xq_1 A(x, q_1; 1) = \frac{xq_1^2(1 - C(x))}{q_1 C(x) - C(x) - q_1}. \tag{4}$$

Next, let q_i track the number of letters i. Assume $i \neq 1$.
For $i \neq j = 1$,

$$A(x, q_i|j) = x + \sum_{b \geq j-1} A(x, q_i|bj) = x + x \sum_{b \geq j-1} A(x, q_i|b).$$

For $i \neq j > 1$,

$$A(x, q_i|j) = \sum_{b \geq j-1} A(x, q_i|bj) = x \sum_{b \geq j-1} A(x, q_i|b),$$

and

$$A(x, q_i|i) = xq_i \sum_{b \geq i-1} A(x, q_i|b).$$

Now, as before, define $A(x, q_i; v) := \sum_{j \geq 1} A(x, q_i|j) v^{j-1}$. We multiply the previous equations by v^{j-1} and sum on $j \geq 1$ to obtain

$$\begin{aligned} A(x, q_i; v) &= x + x \sum_{b \geq 1} A(x, q_i|b) + \sum_{j>1} x \sum_{b \geq j-1} A(x, q_i|b) v^{j-1} - x(1 - q_i) \sum_{b \geq i-1} A(x, q_i|b) v^{i-1} \\ &= x + x \sum_{b \geq 1} A(x, q_i|b) + x \sum_{b \geq 1} \sum_{j=2}^{b+1} A(x, q_i|b) v^{j-1} - x(1 - q_i) \sum_{b \geq i-1} A(x, q_i|b) v^{i-1} \\ &= x + x \sum_{b \geq 1} A(x, q_i|b) + xv \sum_{b \geq 1} A(x, q_i|b) \frac{1 - v^b}{1 - v} - x(1 - q_i) \sum_{b \geq i-1} A(x, q_i|b) v^{i-1} \\ &= x + xA(x, q_i; 1) + \frac{xv}{1 - v} (A(x, q_i; 1) - vA(x, q_i; v)) - \frac{1 - q_i}{q_i} v^{i-1} A(x, q_i|i). \end{aligned} \tag{5}$$

Solving for $A(x, q_i; v)$, we obtain

$$(1 - v + xv^2)A(x, q_i; v) = (1 - v)(x + xA(x, q_i; 1)) + xvA(x, q_i; 1) - \frac{1 - q_i}{q_i}(1 - v)v^{i-1}A(x, q_i|i). \tag{6}$$

We apply the kernel method, hence putting $v = C(x)$ in the above equation, leads to

$$(1 - C(x))(x + xA(x, q_i; 1)) + xC(x)A(x, q_i; 1) - \frac{1 - q_i}{q_i}(1 - C(x))C(x)^{i-1}A(x, q_i|i) = 0. \tag{7}$$

Now, let q_2 track the number of 2's in Catalan words. Substituting $i = 2$ in Equation (7), we obtain

$$(1 - C(x))(x + xA(x, q_2; 1)) + xC(x)A(x, q_2; 1) - \frac{1 - q_2}{q_2}(1 - C(x))C(x)A(x, q_2|2) = 0. \tag{8}$$

Since $A(x, q_2|2) = q_2 xA(x, q_2; 1)$, we have

$$(1 - C(x))(x + xA(x, q_2; 1)) + xC(x)A(x, q_2; 1) - \frac{1 - q_2}{q_2}(1 - C(x))C(x)q_2 xA(x, q_2; 1) = 0. \tag{9}$$

Letting $A(x, q_2) := A(x, q_2; 1)$, we obtain the solution

$$\begin{aligned} A(x, q_2) &= -\frac{x\left(2x + \sqrt{1-4x} - 1\right)}{q_2\left(-\left(\left(\sqrt{1-4x} - 3\right)x\right) + \sqrt{1-4x} - 1\right) + (x-1)\left(2x + \sqrt{1-4x} - 1\right)} \\ &= \frac{1 - C(x)}{(q_2 - 1)(C(x)^2 - C(x)) - 1}. \end{aligned} \tag{10}$$

In what follows, we introduce the following convention: When we use the generating function $\hat{A}(x, q_i|i)$, the q_i tracks all i's other than the last one and when we use $\hat{A}(x, q_i|j)$ in the case where $j < i$, the q_i tracks the total number of i's.

We note that for any integer $i \geq 3$,

$$\hat{A}(x, q_i|i) = x(\hat{A}(x, q_i; 1) - \sum_{j=1}^{i-2} \hat{A}(x, q_i|j)),$$

since a Catalan word ending in such i must have at least 2 letters, but we also need to ensure by the Catalan word property that the previous letter does not end with a letter in $\{1, 2, \ldots, i-2\}$. Thus, it follows that

$$\hat{A}(x, q_i|i - 1) = x(\hat{A}(x, q_i; 1) - \sum_{j=1}^{i-3} \hat{A}(x, q_i|j)).$$

From this we deduce that for $i \geq 3$,

$$\hat{A}(x, q_i|i) = \hat{A}(x, q_i|i - 1) - x\hat{A}(x, q_i|i - 2),$$

with initial values given by $\hat{A}(x, q_i|1) = xA(x, q_i; 1) + x$ and $\hat{A}(x, q_i|2) = A(x, q_i; 1)$. Solving this recursion, and noting that $A(x, q_i|i) = q_i\hat{A}(x, q_i|i)$ we obtain the next result.

Proposition 2.1. *For $i \geq 1$ the generating function for the number of letters i in Catalan words ending in i (tracked by q_i), with letters tracked by x, is given by*

$$A(x, q_i|i) = q_i \frac{(A(x, q_i; 1) + 1)x\left((xC(x))^i - (1 - xC(x))^i\right) - (xC(x))^i + xC(x)\left((xC(x))^i + (1 - xC(x))^i\right)}{2xC(x) - 1}. \tag{11}$$

Substituting $A(x, q_i|i)$ into Equation (7), we obtain the next theorem.

Theorem 2.1. *The generating function for the number of letters i, tracked by q_i in Catalan words with n letters tracked by x is given by*

$$\begin{aligned} A(x, q_i) := A(x, q_i; 1) = &\frac{(C(x) - 1)\left(-xC(x) + 2x^2C(x)^2 + (q_i - 1)xC(x)^{1+i}B_2(x)\right)}{x\left(-C(x) + 2xC(x)^2 + (q_i - 1)C(x)^i B_1(x)(1 - C(x))\right)} \\ &+ \frac{(C(x) - 1)(q_i - 1)C(x)^i\left(xB_1(x) - (xC(x))^i\right)}{x\left(-C(x) + 2xC(x)^2 + (q_i - 1)C(x)^i B_1(x)(1 - C(x))\right)}, \end{aligned} \tag{12}$$

where we use the abbreviations $B_1(x) := (xC(x))^i - (1 - xC(x))^i$ and $B_2(x) := (xC(x))^i + (1 - xC(x))^i$.

3. Generating functions for the number of letters i in the cases where $i \leq 5$

By using Theorem 2.1, we obtain the following generating functions and series expansions:

For $i = 1$, we obtain the generating function

$$A(x, q_1; 1) = \frac{q_1(1 - C(x))}{q_1 C(x) - C(x) - q_1} \tag{13}$$

with series expansion that begins

$$\begin{aligned} &q_1 x + \left(q_1 + q_1^2\right) x^2 + \left(2q_1 + 2q_1^2 + q_1^3\right) x^3 + \left(5q_1 + 5q_1^2 + 3q_1^3 + q_1^4\right) x^4 \\ &+ \left(14q_1 + 14q_1^2 + 9q_1^3 + 4q_1^4 + q_1^5\right) x^5 + \left(42q_1 + 42q_1^2 + 28q_1^3 + 14q_1^4 + 5q_1^5 + q_1^6\right) x^6. \end{aligned} \tag{14}$$

When $i = 2$, we obtain the generating function

$$A(x, q_2; 1) = \frac{1 - C(x)}{(q_2 - 1)(C(x)^2 - C(x)) - 1} \tag{15}$$

with series expansion that begins

$$\begin{aligned} &x + (1 + q_2)x^2 + \left(1 + 3q_2 + q_2^2\right) x^3 + \left(1 + 7q_2 + 5q_2^2 + q_2^3\right) x^4 \\ &+ \left(1 + 16q_2 + 17q_2^2 + 7q_2^3 + q_2^4\right) x^5 + \left(1 + 39q_2 + 51q_2^2 + 31q_2^3 + 9q_2^4 + q_2^5\right) x^6. \end{aligned} \tag{16}$$

When $i = 3$, we obtain the generating function

$$A(x, q_3; 1) = \frac{x^2(C(x) - 1)((q_3 - 1)C(x) - q_3)}{(q_3 - 1)\left(2x^2 - 3x + 1\right) C(x) - q_3(x - 1)^2 - 2x + 1} \tag{17}$$

with series expansion that begins

$$\begin{aligned} &x + 2x^2 + (4 + q_3)x^3 + \left(8 + 5q_3 + q_3^2\right) x^4 + \left(16 + 18q_3 + 7q_3^2 + q_3^3\right) x^5 \\ &+ \left(32 + 57q_3 + 33q_3^2 + 9q_3^3 + q_3^4\right) x^6 + \left(64 + 170q_3 + 131q_3^2 + 52q_3^3 + 11q_3^4 + q_3^5\right) x^7. \end{aligned} \tag{18}$$

When $i = 4$, we obtain the generating function

$$A(x, q_4; 1) = \frac{x^2(C(x) - 1)((q_4 - 1)(x - 1)C(x) + q_4 + x - 1)}{(q_4 - 1)\left(2x^3 - 7x^2 + 5x - 1\right) C(x) + q_4(1 - 2x)^2 + x^3 - 4x^2 + 4x - 1} \tag{19}$$

with series expansion that begins

$$\begin{aligned} &x + 2x^2 + 5x^3 + (13 + q_4)x^4 + \left(34 + 7q_4 + q_4^2\right) x^5 \\ &+ \left(89 + 33q_4 + 9q_4^2 + q_4^3\right) x^6 + \left(233 + 132q_4 + 52q_4^2 + 11q_4^3 + q_4^4\right) x^7. \end{aligned} \tag{20}$$

When $i = 5$, we obtain the generating function

$$A(x, q_5; 1) = \frac{x^2(C(x) - 1)\left((q_5 - 1)\left(2x^2 - 3x + 1\right) C(x) - q_5(x - 1)^2 - 2x + 1\right)}{(q_5 - 1)\left(3x^4 - 13x^3 + 16x^2 - 7x + 1\right) C(x) - q_5\left(x^2 - 3x + 1\right)^2 - 6x^3 + 11x^2 - 6x + 1} \tag{21}$$

with series expansion that begins

$$\mathbf{x} + \mathbf{2x^2} + \mathbf{5x^3} + \mathbf{14x^4} + (\mathbf{41} + \mathbf{q_5})x^5 + \left(122 + \mathbf{9q_5} + q_5^2\right) x^6 + \left(365 + 52q_5 + 11q_5^2 + q_5^3\right) x^7. \tag{22}$$

We illustrate the bold terms in the above series expansion. It is clear that words of lengths 1 to 4 have no letters of size 5. Exactly one of the 42 Catalan words of length 5 has a letter 5; namely the word 12345. For the bold coefficient of x^6, the nine words of this type are $112345, 122345, 123345, 123445, 123451, 123452, 123453, 123454, 123456$.

We note that the case $i = 1$ is in the OEIS [14], where it has reference number A033184 and it refers to the Catalan triangle transposed. We leave as an open question for the reader whether there are some direct relations between Catalan words counted with respect to the number of letters and the number of $1s$, and the Catalan triangle.

For cases $i = 3, 4, 5$, the coefficients of $x^n q_i^0$ are found in [14] as the number of plane trees with n edges and height at most $i - 1$. These results follow from the bijections between Catalan words, Dyck paths and plane trees. The bijection between Catalan words and Dyck paths is found in [7]. A standard bijection between Dyck paths and plane trees is found in [13]. As a result of these bijections $A(x, q_i)$ defined in Theorem 2.1 also enumerates plane trees with $n + 1$ nodes and k nodes at level i.

References

[1] J.-L. Baril, C. Khalil, V. Vajnovszki, Catalan words avoiding pairs of length three patterns, *Discrete Math. Theor. Comput. Sci.* **22(2)** (2021) #5.

[2] J.-L. Baril, S. Kirgizov, V. Vajnovszki, Descent distribution on Catalan words avoiding a pattern of length at most three, *Discrete Math.* **341** (2018) 2608–2615.

[3] A. Blecher, C. Brennan, A. Knopfmacher, Combinatorial parameters in bargraphs, *Quaest. Math.* **39** (2016) 619–635.

[4] A. Blecher, C. Brennan, A. Knopfmacher, The inner site-perimeter of compositions, *Quaest. Math.* **43** (2020) 55–66.

[5] A. Blecher, C. Brennan, A. Knopfmacher, The inner site-perimeter of bargraphs, *Online J. Anal. Comb.* **16** (2021) #2.

[6] A. Blecher, C. Brennan, A. Knopfmacher, The site-perimeter of compositions, *Discrete Math. Appl.* **32** (2022) 75–89.

[7] D. Callan, T. Mansour, J. L. Ramírez, Statistics on bargraphs of Catalan words, *J. Autom. Lang. Comb.* **26** (2021) 177–196.

[8] T. Mansour, Semi-perimeter and inner site-perimeter of k-ary words and bargraphs, *Art Discrete Appl. Math.* **4** (2021) #P1.06.

[9] T. Mansour, J. L. Ramírez, D. A. Toquica, Counting lattice points on bargraphs of Catalan words, *Math. Comput. Sci.* **15** (2021) 701–713.

[10] T. Mansour, A. Shabani, Enumerations on bargraphs, *Discrete Math. Lett.* **2** (2019) 65–94.

[11] H. Prodinger, The kernel method: A collection of examples, *Sém. Lothar. Combin.* **50** (2004) #B50f.

[12] H. Prodinger, Partial skew Dyck paths – a kernel method approach, *Graphs Combin.* **38** (2022) #135.

[13] R. Stanley, *Catalan Numbers*, Cambridge University, Cambridge, 2015.

[14] The OEIS Foundation, *The On-line Encyclopedia of Integer Sequences*, Available at https://oeis.org/.