RAIRO-Theor. Inf. Appl. 58 (2024) 1
https://doi.org/10.1051/ita/2024001

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

CONSECUTIVE PATTERN-AVOIDANCE IN CATALAN WORDS ACCORDING TO THE LAST SYMBOL

Jean-Luc Baril[1], Daniela Colmenares[2], José L. Ramírez[2,*], Emmanuel D. Silva[2], Lina M. Simbaqueba[2] and Diana A. Toquica[2]

Abstract. We study the distribution of the last symbol statistics on the sets of Catalan words avoiding a consecutive pattern of length at most three. For each pattern p, we provide a bivariate generating function, where the coefficient $\boldsymbol{c}_p(n,k)$ of $x^n y^k$ in its series expansion is the number of length n Catalan words avoiding p and ending with the symbol k. We deduce recurrence relations or closed forms for $\boldsymbol{c}_p(n,k)$ and we provide asymptotic approximations for the expectation of the last symbol on all Catalan words avoiding p. Finally, we characterize the sequence $\boldsymbol{c}_p(n,k)$ using Riordan arrays.

Mathematics Subject Classification. 05A15, 05A19.

Received February 3, 2023. Accepted January 11, 2024.

1. Introduction

Restricted growth words $w = w_1w_2\cdots w_n$ are words defined over the set of non-negative integers by $w_1 = \mathtt{0}$ and $0 \le w_i \le \mathtt{st}(w_1\cdots w_{i-1}) + 1$, where $\mathtt{st}$ is an integer statistic. Whenever the statistic $\mathtt{st}$ returns the last symbol of a word, *i.e.*, $\mathtt{st}(w_1\cdots w_{i-1}) = w_{i-1}$, the restricted growth words are called *Catalan words*, that is, the word $w = w_1w_2\cdots w_n$ is a Catalan word if $w_1 = \mathtt{0}$ and $0 \le w_i \le w_{i-1} + 1$ for $i = 2, \dots, n$. For $n \ge 0$, let $\mathcal{C}_n$ denote the set of Catalan words of length n. For instance, $\mathcal{C}_3 = \{\mathtt{000}, \mathtt{001}, \mathtt{010}, \mathtt{011}, \mathtt{012}\}$. The cardinality of $\mathcal{C}_n$ is given by the Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$, see [1], Exercise 80. Catalan words have already been studied in the context of exhaustive generation of Gray codes for growth-restricted words [2]. More recently, Baril *et al.* [3, 4] study the distribution of descents on restricted Catalan words avoiding a pattern or a pair of patterns of length at most three. Ramírez and Rojas [5] also study the distribution of descents for Catalan words avoiding consecutive patterns of length at most three. Baril, González, and Ramírez [6] enumerate Catalan words avoiding a classical pattern of length at most three according to the length and the value of the last symbol. They also give the exact value or an asymptotic for the expectation of the last symbol. Also, we refer to [7, 8], where the authors study several combinatorial statistics on the polyominoes associated with words in $\mathcal{C}_n$.

Keywords and phrases: Catalan word, consecutive pattern, generating function.

[1] LIB, Université de Bourgogne, BP 47 870, 21078 Dijon-Cedex France.
[2] Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia.

[*] Corresponding author: jlramirezr@unal.edu.co

The goal of this work is to complement all these studies by providing enumerative results for Catalan words avoiding a consecutive pattern of length at most three with respect to the length and the value of the last symbol. Using classical methods presented in [9, 10], we also give the exact value or an asymptotic approximation for the expectation of the last symbol for these words. The study of consecutive patterns was introduced by Elizalde and Noy [11] in the context of permutations. Since then, several results have appeared in the literature, see for example [12–14] and references therein.

The remainder of this paper is structured as follows. In Section 2, we introduce the notation that will be used in this work. In Section 3, we study the distribution of the last symbol in the set of Catalan words avoiding consecutive patterns of length 2 providing the bivariate generating function that counts Catalan words avoiding a given pattern with respect to the length and the last symbol and a matrix with the total number of words for each length and last symbol. Also, we present a recurrence and an asymptotic approximation for the expectation of the last symbol for these patterns. In Section 4, we present the same for the consecutive patterns of length 3 including bijections between some of them. Finally, in Section 5 we show how the matrices presented above are Riordan arrays.

2. Notations

For an integer $r \geq 2$, a *consecutive pattern* $p = \underline{p_1 p_2 \cdots p_r}$ is a word (of length r) over the set $\{0, 1, \ldots, r-1\}$ satisfying the condition: if $j > 0$ appears in p, then $j-1$ also appears in p. A Catalan word $w = w_1 w_2 \cdots w_n$ *contains* the consecutive pattern $p = \underline{p_1 p_2 \cdots p_r}$ if there exists a subsequence $w_i w_{i+1} \cdots w_{i+r-1}$ (for some $i \geq 1$) of w which is order-isomorphic to $p_1 p_2 \ldots p_r$. We say that w *avoids* the consecutive pattern p whenever w does not contain the consecutive pattern p. For example, the Catalan word $\texttt{0123455543}$ avoids the consecutive pattern $\underline{001}$ and contains one subsequence isomorphic to the pattern $\underline{210}$.

For $n \geq 0$, let $\mathcal{C}_n(p)$ denote the set of Catalan words of length n avoiding the consecutive pattern p. We denote by $\boldsymbol{c}_p(n)$ the cardinality of $\mathcal{C}_n(p)$, $\mathcal{C}(p) := \bigcup_{n\geq 0} \mathcal{C}_n(p)$, and $\mathcal{C}(p)^+ := \mathcal{C}(p) \setminus \{\epsilon\}$. We denote by $\mathsf{last}(w)$ the last symbol of w. Let $\mathcal{C}_{n,k}(p)$ denote the set of Catalan words $w \in \mathcal{C}_n(p)$ such that $\mathsf{last}(w) = k$, and let $\boldsymbol{c}_p(n,k) := |\mathcal{C}_{n,k}(p)|$. Obviously, we have $\boldsymbol{c}_p(n) = \sum_{k=0}^{n-1} \boldsymbol{c}_p(n,k)$.

We introduce the bivariate generating function

$$\boldsymbol{H}_p(x,y) := \sum_{w \in \mathcal{C}(p)^+} x^{|w|} y^{\mathsf{last}(w)} = \sum_{n\geq 1,\ k \geq 0} \boldsymbol{c}_p(n,k) x^n y^k,$$

and we set

$$\boldsymbol{H}_p(x) := \sum_{w \in \mathcal{C}(p)^+} x^{|w|} = \boldsymbol{H}_p(x,1).$$

Notice that these generating functions do not consider the empty word. Let $\mathcal{T}_p$ be the infinite matrix $\mathcal{T}_p := (\boldsymbol{c}_p(n,k))_{n\geq 1, k\geq 0}$.

The expectation of the last symbol on all Catalan words in $\mathcal{C}_n(p)$ is given by (see [9])

$$\boldsymbol{a}_p(n) := \frac{[x^n]\partial_y \boldsymbol{H}_p(x,y)|_{y=1}}{\boldsymbol{c}_p(n)} = \frac{[x^n]\partial_y \boldsymbol{H}_p(x,y)|_{y=1}}{[x^n]\boldsymbol{H}_p(x,1)}.$$

In order to obtain an asymptotic approximation for $\boldsymbol{a}_p(n)$, we will use classical methods presented in [9, 10] on the two generating functions $[x^n]\partial_y \boldsymbol{H}_p(x,y)|_{y=1}$ and $[x^n]\boldsymbol{H}_p(x,1)$.

Throughout this work, we will often use the first return decomposition of a Catalan word w, which is $w = \texttt{0}(w'+1)w''$, where w' and w'' are Catalan words, and where $(w'+1)$ is the word obtained from w' by adding 1 at all these symbols (for instance if $w' = \texttt{012012}$ then $(w'+1) = \texttt{123123}$). As an example, the first return decomposition of $w = \texttt{0122123011201}$ is given by setting $w' = \texttt{011012}$ and $w'' = \texttt{011201}$.

3. Consecutive patterns of length 2

The goal of this section is to study the distribution of the last symbol in the set of Catalan words avoiding a consecutive pattern of length 2. For each pattern $p \in \{\underline{00}, \underline{01}, \underline{10}\}$, we provide the bivariate generating function $\boldsymbol{H}_p(x, y)$ that counts Catalan words with respect to the length and the last symbol.

Theorem 3.1. *We have*

$$\boldsymbol{H}_{\underline{00}}(x, y) = \frac{2x}{1 + x - 2xy + \sqrt{1 - 2x - 3x^2}}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{00})$, and let $w = \mathtt{0}(w' + 1)w''$ be the first return decomposition, where $w', w'' \in \mathcal{C}(\underline{00})$. If $w'' = \epsilon$, then $w = \mathtt{0}(w' + 1)$, where w' is possibly empty. The generating function for this case is $x + xy\boldsymbol{H}_{\underline{00}}(x, y)$. If w'' is non-empty, then w' is necessarily non-empty and this produces the generating function $x\boldsymbol{H}_{\underline{00}}(x)\boldsymbol{H}_{\underline{00}}(x, y)$. Therefore, we have the functional equation

$$\boldsymbol{H}_{\underline{00}}(x, y) = x + xy\boldsymbol{H}_{\underline{00}}(x, y) + x\boldsymbol{H}_{\underline{00}}(x)\boldsymbol{H}_{\underline{00}}(x, y),$$

where $\boldsymbol{H}_{\underline{00}}(x) = \boldsymbol{H}_{\underline{00}}(x, 1)$, satisfies the functional equation $\boldsymbol{H}_{\underline{00}}(x) = x + x\boldsymbol{H}_{\underline{00}}(x) + x\boldsymbol{H}_{\underline{00}}(x)^2$. Solving these equations, we obtain the desired result. $\square$

Theorem 3.2. *For $n \geq 2$ and $0 \leq k < n$, we have*

$$\boldsymbol{c}_{\underline{00}}(n, k) = \sum_{i=k-1}^{n-1} \boldsymbol{c}_{\underline{00}}(n - 1, i) - \boldsymbol{c}_{\underline{00}}(n - 1, k),$$

anchored with $\boldsymbol{c}_{\underline{00}}(1, 0) = 1$ and $\boldsymbol{c}_{\underline{00}}(n, k) = 0$ for $k < 0$ or $1 \leq n \leq k$. Moreover, for $n \geq 2$, $2 \leq k < n$, we have

$$\boldsymbol{c}_{\underline{00}}(n, k) = \boldsymbol{c}_{\underline{00}}(n, k - 1) - \boldsymbol{c}_{\underline{00}}(n - 1, k) + \boldsymbol{c}_{\underline{00}}(n - 1, k - 1) - \boldsymbol{c}_{\underline{00}}(n - 1, k - 2).$$

Proof. Let $w = w_1 \cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{00})$, with $n \geq 2$. So, the subword $w_1 \cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{00})$ for some $i \geq k - 1$, $i \neq k$ (otherwise the word w would end with kk). Consequently, we obtain the first equality. In addition, if we consider the difference $\boldsymbol{c}_{\underline{00}}(n, k) - \boldsymbol{c}_{\underline{00}}(n, k - 1)$, then we deduce the second equality. $\square$

The first few rows of the matrix $\mathcal{T}_{\underline{00}}$ are

$$\mathcal{T}_{\underline{00}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{0} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 3 & 0 & 1 & 0 & 0 & 0 & 0 \\ 6 & 7 & 3 & 4 & 0 & 1 & 0 & 0 & 0 \\ 15 & 14 & 12 & 4 & 5 & 0 & 1 & 0 & 0 \\ 36 & 37 & 24 & 18 & 5 & 6 & 0 & 1 & 0 \\ 91 & 90 & 67 & 36 & 25 & 6 & 7 & 0 & 1 \end{pmatrix}$$

The Catalan words corresponding to the bold coefficients in the above array are

$$\mathcal{C}_{4,0}(\underline{00}) = \{\mathtt{0120}\}, \quad \mathcal{C}_{4,1}(\underline{00}) = \{\mathtt{0101}, \mathtt{0121}\}, \quad \mathcal{C}_{4,2}(\underline{00}) = \{\}, \quad \text{and} \quad \mathcal{C}_{4,3}(\underline{00}) = \{\mathtt{0123}\}.$$

$\phi(w) = UFDFFUUDUDFDF$

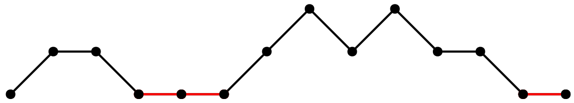

FIGURE 1. Motzkin path corresponding to the word $w = \texttt{01201234343423} \in \mathcal{C}_{14,3}(\underline{00})$.

The matrix $\mathcal{T}_{\underline{00}}$ coincides with the array A097609, in the OEIS [15], that counts the number of Motzkin paths with respect to the length and the number of horizontal steps at level 0. Recall that a length n Motzkin path is a lattice path in $\mathbb{N}^2$, starting at the origin, ending at $(n, 0)$, consisting of steps $U = (1, 1)$, $D = (1, -1)$, and $F = (1, 0)$. A constructive bijection ϕ between $\mathcal{C}(\underline{00})$ and the set of Motzkin paths can be defined recursively as follows: $\phi(\texttt{0}) = \epsilon$, $\phi(\texttt{0}(w' + 1)) = F\phi(w')$, and $\phi(\texttt{0}(w' + 1)w'') = U\phi(w')D\phi(w'')$. Clearly, if $w \in \mathcal{C}_{n,k}(\underline{00})$ then $\phi(w)$ is a Motzkin path of length $n - 1$ having k horizontal steps at level 0. For example, if $w = \texttt{01201234343423} \in \mathcal{C}_{14,3}(\underline{00})$, then we have

$$\begin{aligned}\phi(w) &= U\phi(\texttt{01})D\phi(\texttt{01234343423}) = UF\phi(\texttt{0})DF\phi(\texttt{0123232312}) = UFDFF\phi(\texttt{012121201})\\ &= UFDFFU\phi(\texttt{010101})D\phi(\texttt{01}) = UFDFFUU\phi(\texttt{0})D\phi(\texttt{0101})DF\phi(\texttt{0})\\ &= UFDFFUUDU\phi(\texttt{0})D\phi(\texttt{01})DF = UFDFFUUDUDFDF.\end{aligned}$$

Figure 1 shows the corresponding Motzkin path with 3 horizontal steps at level 0.

Corollary 3.3. *An asymptotic approximation for the expectation $\boldsymbol{a}_{\underline{00}}(n)$ of the last symbol over $\mathcal{C}_n(\underline{00})$ is* 2.

The last two theorems of this section have been already presented in [6], since Catalan words avoiding a consecutive pattern $p \in \{\underline{01}, \underline{10}\}$ are those avoiding the corresponding classical pattern.

Theorem 3.4. *We have*

$$\boldsymbol{H}_{\underline{01}}(x, y) = \frac{x}{1 - x}, \qquad \boldsymbol{c}_{\underline{01}}(n, k) = \begin{cases} 1, & \text{if } k = 0;\\ 0, & \text{otherwise;}\end{cases} \qquad \text{and} \qquad \boldsymbol{a}_{\underline{01}}(n) = 0.$$

Theorem 3.5. *We have*

$$\boldsymbol{H}_{\underline{10}}(x, y) = \frac{x}{1 - x(1 + y)}, \qquad \boldsymbol{c}_{\underline{10}}(n, k) = \binom{n-1}{k} \qquad \text{and} \qquad \boldsymbol{a}_{\underline{10}}(n) = \frac{n-1}{2}.$$

4. Consecutive patterns of length 3

In this section we investigate the distribution of the last symbol in the set of Catalan words avoiding a consecutive pattern of length 3.

4.1. The consecutive pattern $\underline{012}$

Theorem 4.1. *We have*

$$\boldsymbol{H}_{\underline{012}}(x, y) = \frac{1 - x + xy - 3x^2y - 2x^3y^2 - (1 + xy)\sqrt{1 - 2x - 3x^2}}{2x(1 - y + xy + x^2y^2)}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{012})$, and let $w = \texttt{0}(w' + 1)w''$ be the first return decomposition, where $w', w'' \in \mathcal{C}(\underline{012})$. If $w'' = \epsilon$, then $w = \texttt{0}$, or $w = \texttt{01}(w' + 1)$ with $w' \in \mathcal{C}(\underline{012})$. The corresponding generating function for these words is $x + x^2y(1 + \boldsymbol{H}_{\underline{012}}(x, y))$. If w'' is non-empty, then $w = \texttt{0}w''$, or

$w = \texttt{01}(w'+1)w''$, where w' is possibly empty. This case produces the generating function

$$x\boldsymbol{H}_{\underline{012}}(x,y) + x^2(1+\boldsymbol{H}_{\underline{012}}(x))\boldsymbol{H}_{\underline{012}}(x,y),$$

where (see Thm. 2.1 of [5])

$$\boldsymbol{H}_{\underline{012}}(x) = \frac{1-x-\sqrt{1-2x-3x^2}}{2x^2} - 1$$

is the generating function for the number of non-empty Catalan words avoiding $\underline{012}$. Therefore, we have the functional equation

$$\boldsymbol{H}_{\underline{012}}(x,y) = x + x^2y(1+\boldsymbol{H}_{\underline{012}}(x,y)) + x\boldsymbol{H}_{\underline{012}}(x,y) + x^2(1+\boldsymbol{H}_{\underline{012}}(x))\boldsymbol{H}_{\underline{012}}(x,y).$$

Solving this equation, we obtain the desired result. $\square$

Theorem 4.2. *For $n \geq 2$, $3 \leq k \leq n-1$, we have*

$$\boldsymbol{c}_{\underline{012}}(n,k) = \sum_{i=k-1}^{n-1} \boldsymbol{c}_{\underline{012}}(n-1,i) - \sum_{i=k-2}^{n-1} \boldsymbol{c}_{\underline{012}}(n-3,i),$$

and for $n \geq 2$, $k = 0, 1$, we have

$$\boldsymbol{c}_{\underline{012}}(n,k) = \sum_{i=k-1}^{n-1} \boldsymbol{c}_{\underline{012}}(n-1,i),$$

anchored with $\boldsymbol{c}_{\underline{012}}(1,0) = 1$ and $\boldsymbol{c}_{\underline{012}}(n,k) = 0$ for $k < 0$ or $1 \leq n \leq k$. Moreover, for $n \geq 4$ and $3 \leq k \leq n-1$, we have

$$\boldsymbol{c}_{\underline{012}}(n,k) = \boldsymbol{c}_{\underline{012}}(n,k-1) - \boldsymbol{c}_{\underline{012}}(n-1,k-2) + \boldsymbol{c}_{\underline{012}}(n-3,k-3).$$

Proof. Let $w = w_1 \cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{012})$, with $n \geq 2$. So, the subword $w_1 \cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{012})$ for $i \geq k-1$, but we have not to consider the cases where $w_{n-3} \geq w_{n-2} = k-2$ with $w_{n-1} = k-1$ whenever they can occur, *i.e.* when $k \geq 2$. Therefore, by summing over all possible values of i, we obtain the first equality; whenever $k = 0, 1$, all $i \geq k-1$ have to be considered which induces the second equality. Finally, by considering the difference $\boldsymbol{c}_{\underline{012}}(n,k) - \boldsymbol{c}_{\underline{012}}(n,k-1)$ whenever $n \geq 4$ and $k \geq 3$, the third equality follows. $\square$

The first few rows of the matrix $\mathcal{T}_{\underline{012}}$ are

$$\mathcal{T}_{\underline{012}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{4} & \mathbf{4} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 9 & 9 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 21 & 21 & 8 & 1 & 0 & 0 & 0 & 0 & 0 \\ 51 & 51 & 21 & 4 & 0 & 0 & 0 & 0 & 0 \\ 127 & 127 & 55 & 13 & 1 & 0 & 0 & 0 & 0 \\ 323 & 323 & 145 & 39 & 5 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\psi(w) = FUUFUDDUDD$

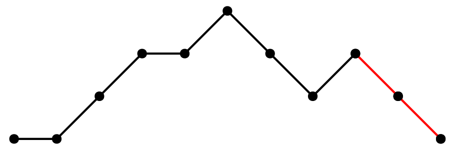

FIGURE 2. Motzkin path corresponding to the word $w = \texttt{0011222312} \in \mathcal{C}_{10,2}(\underline{012})$.

The Catalan words corresponding to the bold coefficients in the above array are

$$\mathcal{C}_{4,0}(\underline{012}) = \{\texttt{0000}, \texttt{0010}, \texttt{0100}, \texttt{0110}\}, \quad \mathcal{C}_{4,1}(\underline{012}) = \{\texttt{0001}, \texttt{0011}, \texttt{0101}, \texttt{0111}\}, \quad \text{and}$$
$$\mathcal{C}_{4,2}(\underline{012}) = \{\texttt{0112}\}.$$

The matrix $\mathcal{T}_{\underline{012}}$ coincides with the array A098979 that counts the number of Motzkin paths with respect to the length and the length of the final run of down steps. Moreover, the row sums of $\mathcal{T}_{\underline{012}}$ corresponds to the sequence of Motzkin numbers A001006. A constructive bijection ψ can be defined recursively as follows: $\psi(\epsilon) = \epsilon$, $\psi(\texttt{0}w') = F\psi(w')$, and $\psi(\texttt{01}(w'+1)w'') = U\psi(w')D\psi(w'')$. Clearly, if $w \in \mathcal{C}_{n,k}(\underline{012})$, then $\psi(w)$ is a Motzkin path of length n where the length of its final descent is k. For example, if $w = \texttt{0011222312} \in \mathcal{C}_{10,2}(\underline{012})$, then we have

$$\begin{aligned}\psi(w) &= F\psi(\texttt{011222312})F = FU\psi(\texttt{0111201})D\psi(\epsilon) = FUU\psi(\texttt{001})D\psi(\texttt{01})D\\ &= FUUFUDDUDD.\end{aligned}$$

Figure 2 shows the corresponding Motzkin path with a final descent of length 2.

Corollary 4.3. *An asymptotic approximation for the expectation $\boldsymbol{a}_{\underline{012}}(n)$ of the last symbol over $\mathcal{C}_n(\underline{012})$ is $5/4$.*

4.2. The consecutive pattern $\underline{001}$ and $\underline{011}$

Theorem 4.4. *We have*

$$\boldsymbol{H}_{\underline{001}}(x,y) = \frac{2x}{1 - x^2 - (2x - 2x^2)y + \sqrt{1 - 4x + 2x^2 + x^4}}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{001})$, and let $w = \texttt{0}(w'+1)w''$ be the first return decomposition, where $w', w'' \in \mathcal{C}(\underline{001})$. If $w'' = \epsilon$, then $w = \texttt{0}(w'+1)$ with w' possibly empty. The generating function for this case is $x + xy\boldsymbol{H}_{\underline{001}}(x,y)$. If w'' is non-empty, then $w = \texttt{00}\cdots\texttt{0}$ (with $w' = \epsilon$) or $w = \texttt{0}(w'+1)w''$ (with $w' \neq \epsilon$). This case produces the generating function

$$\frac{x^2}{1-x} + x\boldsymbol{H}_{\underline{001}}(x)\boldsymbol{H}_{\underline{001}}(x,y),$$

where (see Thm. 2.3 of [5])

$$\boldsymbol{H}_{\underline{001}}(x) = \frac{1 - x^2 - \sqrt{(1-x)(1-3x-x^2-x^3)}}{2(1-x)x} - 1$$

is the generating function for the number of non-empty Catalan words avoiding the consecutive pattern $\underline{001}$. Therefore, we have the functional equation

$$\boldsymbol{H}_{\underline{001}}(x,y) = x + xy\boldsymbol{H}_{\underline{001}}(x,y) + \frac{x^2}{1-x} + x\boldsymbol{H}_{\underline{001}}(x)\boldsymbol{H}_{\underline{001}}(x,y).$$

Solving this equation, we obtain the desired result. □

Theorem 4.5. *We have* $\boldsymbol{H}_{\underline{011}}(x,y) = \boldsymbol{H}_{\underline{001}}(x,y)$.

Proof. There exists a bijection between the Catalan words avoiding $\underline{011}$ and those avoiding $\underline{001}$ preserving the last symbol (see Thm. 2.4 of [5]). Indeed, the bijection comes from an algorithm explained in [5], which consists in replacing, from left to right, each factor $k^j(k+1)$ with the factor $k(k+1)^j$ $(j \geq 2)$. For instance, the bijection transforms $w = \texttt{000123222321112001012} \in \mathcal{C}(\underline{011})$ into $\texttt{012333233321222011012} \in \mathcal{C}(\underline{001})$. So, the two bivariate generating functions are equal. □

Theorem 4.6. *Let* $p \in \{\underline{011}, \underline{001}\}$. *Then, for* $n \geq 2$, $0 \leq k \leq n-1$, *we have*

$$\boldsymbol{c}_p(n,k) = \sum_{i=k-1}^{n-2} \boldsymbol{c}_p(n-1,i) - \boldsymbol{c}_p(n-2,k-1),$$

anchored with $\boldsymbol{c}_p(1,0) = 1$ *and* $\boldsymbol{c}_p(n,k) = 0$, *otherwise. Moreover, for* $n \geq 2$ *and* $1 \leq k \leq n$, *we have*

$$\boldsymbol{c}_p(n,k) = \boldsymbol{c}_p(n,k-1) - \boldsymbol{c}_p(n-1,k-2) - \boldsymbol{c}_p(n-2,k-1) + \boldsymbol{c}_p(n-2,k-2).$$

Proof. Let $w = w_1 \cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{001})$, with $n \geq 2$. So, the subword $w_1 \cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{001})$ for $i \geq k-1$, but we have not to consider the words where $w_{n-2} = w_{n-1} = k-1$. Therefore, by summing over all possible values of i, we obtain the first equality. The second equality follows from the difference $\boldsymbol{c}_{\underline{001}}(n,k) - \boldsymbol{c}_{\underline{001}}(n,k-1)$. □

The first few rows of the matrix $\mathcal{T}_p$, for $p \in \{\underline{011}, \underline{001}\}$ are

$$\mathcal{T}_p = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{4} & \mathbf{3} & \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 9 & 7 & 4 & 1 & 1 & 0 & 0 & 0 & 0 \\ 22 & 18 & 10 & 5 & 1 & 1 & 0 & 0 & 0 \\ 57 & 48 & 28 & 13 & 6 & 1 & 1 & 0 & 0 \\ 154 & 132 & 79 & 39 & 16 & 7 & 1 & 1 & 0 \\ 429 & 372 & 227 & 115 & 51 & 19 & 8 & 1 & 1 \end{pmatrix}.$$

This array does not appear in the OEIS [15], however the first column (and the row sums) corresponds to the sequence A105633 (this sequence also counts the number of Dyck paths of semilength $n+1$ avoiding $UUDU$). The Catalan words corresponding to the bold coefficients in the above array are

$$\mathcal{C}_{4,0}(\underline{001}) = \{\texttt{0000}, \texttt{0100}, \texttt{0110}, \texttt{0120}\}, \quad \mathcal{C}_{4,1}(\underline{001}) = \{\texttt{0101}, \texttt{0121}, \texttt{0111}\},$$
$$\mathcal{C}_{4,2}(\underline{001}) = \{\texttt{0122}\}, \quad \text{and} \quad \mathcal{C}_{4,3}(\underline{001}) = \{\texttt{0123}\}.$$

An asymptotic analysis of the generating functions $[x^n]\partial_y \boldsymbol{H}_p(x,y)|_{y=1}$ and $[x^n]\boldsymbol{H}_p(x,1)$, $p \in \{\underline{011}, \underline{001}\}$, gives the following result. The constant a is the solution closest to the origin of the polynomial z^3+z^2+3z-1. Maple provides an explicit version if one asks for a *simplification*.

Corollary 4.7. *For $p \in \{\underline{011}, \underline{001}\}$, an asymptotic approximation for the expectation $\boldsymbol{a}_p(n)$ of the last symbol over $\mathcal{C}_n(p)$ is*

$$\frac{a\,(1-a)^3\left(17a^2+22a+57\right)^2}{4\left(2a^2+3a+7\right)^2} \sim 1.6785735141,$$

where $a = \frac{1}{3}\left(26+6\sqrt{33}\right)^{\frac{1}{3}} - \frac{8}{3}\left(26+6\sqrt{33}\right)^{-\frac{1}{3}} - \frac{1}{3}$.

4.3. The consecutive pattern $\underline{010}$

Theorem 4.8. *We have*

$$\boldsymbol{H}_{\underline{010}}(x,y) = \frac{1-2xy+x^2-\sqrt{1-4x+2x^2-4x^3+x^4}}{2(1+x^2-y-x^2y+xy^2)}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{010})$, and let $w = \mathtt{0}(w'+1)w''$ be the first return decomposition, where $w', w'' \in \mathcal{C}(\underline{010})$. If $w'' = \epsilon$, then $w = \mathtt{0}(w'+1)$ with w' possibly empty. The generating function for this case is $x + xy\boldsymbol{H}_{\underline{010}}(x,y)$. If w'' is non-empty, then $w' \neq \mathtt{0}$. This case produces the generating function

$$x(\boldsymbol{H}_{\underline{010}}(x) - x + 1)\boldsymbol{H}_{\underline{010}}(x,y),$$

where (see Thm. 2.3 of [5])

$$\boldsymbol{H}_{\underline{010}}(x) = \frac{1+x^2-\sqrt{(1+x^2)(1-4x+x^2)}}{2x} - 1$$

is the generating function of the non-empty Catalan words avoiding the consecutive pattern $\underline{010}$. Therefore, we have the functional equation

$$\boldsymbol{H}_{\underline{010}}(x,y) = x + xy\boldsymbol{H}_{\underline{010}}(x,y) + x(\boldsymbol{H}_{\underline{010}}(x) - x + 1)\boldsymbol{H}_{\underline{010}}(x,y).$$

Solving this equation, we obtain the desired result. □

Theorem 4.9. *For $n \geq 2$, $0 \leq k \leq n-1$, we have*

$$\boldsymbol{c}_{\underline{010}}(n,k) = \sum_{i=k-1}^{n-2} \boldsymbol{c}_{\underline{010}}(n-1,i) - \boldsymbol{c}_{\underline{010}}(n-2,k),$$

anchored with $\boldsymbol{c}_{\underline{010}}(1,0) = 1$ and $\boldsymbol{c}_{\underline{010}}(n,k) = 0$, otherwise. Moreover, for $n \geq 2$ and $2 \leq k \leq n-1$, we have

$$\boldsymbol{c}_{\underline{010}}(n,k) = \boldsymbol{c}_{\underline{010}}(n,k-1) - \boldsymbol{c}_{\underline{010}}(n-1,k-2) - \boldsymbol{c}_{\underline{010}}(n-2,k) + \boldsymbol{c}_{\underline{010}}(n-2,k-1).$$

Proof. Let $w = w_1 \cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{010})$, with $n \geq 2$. So, the subword $w_1 \cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{010})$ for $i \geq k-1$, but we do not have to consider the words where $w_{n-2} = k$ and $w_{n-1} = k+1$. Therefore, by summing over all possible values of i we obtain the first equality. Moreover, if we consider the difference $\boldsymbol{c}_{\underline{010}}(n,k) - \boldsymbol{c}_{\underline{010}}(n,k-1)$, then we have the second equality. □

The first few rows of the matrix $\mathcal{T}_{\underline{010}}$ are

$$\mathcal{T}_{\underline{010}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 9 & 8 & 6 & 4 & 1 & 0 & 0 & 0 & 0 \\ 25 & 25 & 16 & 10 & 5 & 1 & 0 & 0 & 0 \\ 73 & 74 & 51 & 28 & 15 & 6 & 1 & 0 & 0 \\ 223 & 223 & 159 & 91 & 45 & 21 & 7 & 1 & 0 \\ 697 & 696 & 496 & 296 & 150 & 68 & 28 & 8 & 1 \end{pmatrix}$$

This array does not appear in the OEIS [15], however, the first column of this matrix is the sequence A101499 (this sequence also counts the number of peakless Motzkin paths of length n in which the horizontal steps at level greater than or equal to 1 come in 2 colors). The row sums of $\mathcal{T}_{\underline{010}}$ corresponds to the sequence A187256. The Catalan words corresponding to the bold coefficients in the above array are

$$\mathcal{C}_{4,0}(\underline{010}) = \{\texttt{0000}, \texttt{0110}, \texttt{0120}\}, \quad \mathcal{C}_{4,1}(\underline{010}) = \{\texttt{0001}, \texttt{0011}, \texttt{0111}\},$$
$$\mathcal{C}_{4,2}(\underline{010}) = \{\texttt{0012}, \texttt{0112}, \texttt{0122}\}, \quad \text{and} \quad \mathcal{C}_{4,3}(\underline{010}) = \{\texttt{0123}\}.$$

Corollary 4.10. *An asymptotic approximation for the expectation $\boldsymbol{a}_{\underline{010}}(n)$ of the last symbol over $\mathcal{C}_n(\underline{010})$ is* 2.

4.4. The consecutive pattern $\underline{201}$

Theorem 4.11. *We have*

$$\boldsymbol{H}_{\underline{201}}(x,y) = \frac{(1-x)\left(1+2x-\sqrt{1-4x+4x^3}\right)}{(2-x)\left(1-2xy+\sqrt{1-4x+4x^3}\right)}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{201})$ and let $w = \texttt{0}(w'+1)w''$ be first return decomposition, where $w', w'' \in \mathcal{C}(\underline{201})$. If $w'' = \epsilon$, then $w = \texttt{0}(w'+1)$ with w' possibly empty. The generating function for this case is $x + xy\boldsymbol{H}_{\underline{201}}(x,y)$. If $w'' \neq \epsilon$ and $w' = \epsilon$, then $w = \texttt{0}w''$ and the generating function for this case is $x\boldsymbol{H}_{\underline{201}}(x,y)$. If $w'' \neq \epsilon$ and $w' \neq \epsilon$, then $w = \texttt{0}(w'+1)w''$ and we consider two subcases:

(i) If w' ends with 0, then $w' = w'''\texttt{0}$ with $w''', w'' \in \mathcal{C}(\underline{201})$. The generating function for this case is $x^2(1+\boldsymbol{H}_{\underline{201}}(x))\boldsymbol{H}_{\underline{201}}(x,y)$, where

$$\boldsymbol{H}_{\underline{201}}(x) = \frac{1-2x+2x^2-\sqrt{1-4x+4x^3}}{2(2-x)x^2} - 1,$$

is the generating function for the total number of non-empty Catalan words avoiding $\underline{201}$ (see Thm. 2.5 of [5]).

(ii) If w' does not end with 0, then the last symbol of $(w'+1)$ is at least 2, which implies that w'' necessarily is 0 or $\texttt{0}w'''$ with $w''' \in \mathcal{C}(\underline{201})$. So, in this case the generating function is

$$x(\boldsymbol{H}_{\underline{201}}(x) - x\boldsymbol{H}_{\underline{201}}(x) - x)x(\boldsymbol{H}_{\underline{201}}(x,y)+1).$$

Therefore, bringing together all the above subcases, we have the functional equation

$$\begin{aligned}\boldsymbol{H}_{\underline{201}}(x,y) = x + xy\boldsymbol{H}_{\underline{201}}(x,y) + x\boldsymbol{H}_{\underline{201}}(x,y) + x^2(1+\boldsymbol{H}_{\underline{201}}(x))\boldsymbol{H}_{\underline{201}}(x,y) + \\ x(\boldsymbol{H}_{\underline{201}}(x) - x\boldsymbol{H}_{\underline{201}}(x) - x)x(\boldsymbol{H}_{\underline{201}}(x,y)+1).\end{aligned}$$

Solving this equation, we obtain the desired result. □

Theorem 4.12. *Let $n \geq 2$, $1 \leq k \leq n-1$, we have*

$$\boldsymbol{c}_{\underline{201}}(n,k) = \sum_{i=k-1}^{n-2} \boldsymbol{c}_{\underline{201}}(n-1,i) - \sum_{i=k+1}^{n-3} \boldsymbol{c}_{\underline{201}}(n-2,i),$$

and for $n \geq 2$, $k=0$,

$$\boldsymbol{c}_{\underline{201}}(n,0) = \sum_{i=0}^{n-2} \boldsymbol{c}_{\underline{201}}(n-1,i),$$

anchored with $\boldsymbol{c}_{\underline{201}}(1,0)=1$, and $\boldsymbol{c}_{\underline{201}}(n,k)=0$ otherwise. Moreover, for $n \geq 3$ and $3 \leq k < n$, we have

$$\boldsymbol{c}_{\underline{201}}(n,k) = \boldsymbol{c}_{\underline{201}}(n,k-1) - \boldsymbol{c}_{\underline{201}}(n-1,k-2) + \boldsymbol{c}_{\underline{210}}(n-2,k).$$

Proof. Let $w = w_1 \cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{201})$, with $n \geq 2$. So, the subword $w_1 \cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{201})$ for $i \geq k-1$, but we do not have to consider the words where $w_{n-2} > k$ and $w_{n-1} = k-1$. Therefore, by summing over all possible values of i we obtain the first equality whenever $k \geq 1$, and the second equality for $k=0$. Finally, by considering the difference $\boldsymbol{c}_{\underline{201}}(n,k) - \boldsymbol{c}_{\underline{201}}(n,k-1)$, we obtain the third equality. □

The first few rows of the matrix $\mathcal{T}_{\underline{201}}$ are

$$\mathcal{T}_{\underline{201}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5 & 5 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 14 & \mathbf{13} & 9 & 4 & 1 & 0 & 0 & 0 & 0 \\ 41 & 37 & 26 & 14 & 5 & 1 & 0 & 0 & 0 \\ 124 & 110 & 78 & 45 & 20 & 6 & 1 & 0 & 0 \\ 384 & 338 & 240 & 144 & 71 & 27 & 7 & 1 & 0 \\ 1212 & 1062 & 756 & 463 & 243 & 105 & 35 & 8 & 1 \end{pmatrix}$$

This array does not appear in OEIS [15], however, the first column of this matrix is the sequence A159769 (this sequence also counts the number of Dyck paths of semilength $n-1$ avoiding $DDUUU$).The row sums of $\mathcal{T}_{\underline{201}}$ corresponds to the sequence A159773. The Catalan words corresponding to the bold coefficient in the above array are

$$\mathcal{C}_{5,1}(\underline{201}) = \{\texttt{00001}, \texttt{00011}, \texttt{00101}, \texttt{00111}, \texttt{00121}, \\ \texttt{01001}, \texttt{01011}, \texttt{01101}, \texttt{01111}, \texttt{01121}, \texttt{01211}, \texttt{01221}, \texttt{01231}\}$$

Corollary 4.13. *An asymptotic approximation for the expectation $\boldsymbol{a}_{\underline{201}}(n)$ of the last symbol over $\mathcal{C}_n(\underline{201})$ is*

$$\frac{-128a^7 + 352a^5 - 284a^3 + 55a}{16a^5 + 24a^4 - 12a^3 - 24a^2 + 2a + 6} \sim 2.070578537,$$

where $a = -\frac{\sqrt{3}}{3}\cos\left(\frac{1}{6}\arctan\left(\frac{3\sqrt{111}}{5}\right) + \frac{\pi}{6}\right) + \sin\left(\frac{1}{6}\arctan\left(\frac{3\sqrt{111}}{5}\right) + \frac{\pi}{6}\right)$ is the solution closest to the origin of $4z^3 - 4z + 1$.

4.5. The consecutive pattern $\underline{101}$

Theorem 4.14. *We have*

$$\boldsymbol{H}_{\underline{101}}(x,y) = \frac{x\left(2-x+2x^2-x^3-x\sqrt{1-4x+2x^2-4x^3+x^4}\right)}{(1-x+x^2)(1-2xy+x^2+\sqrt{1-4x+2x^2-4x^3+x^4})}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{101})$, and let $w = \mathtt{0}(w'+1)w''$ be the first return decomposition, where $w', w'' \in \mathcal{C}(\underline{101})$. If $w'' = \epsilon$, then $w = \mathtt{0}(w'+1)$ with w' possibly empty. The generating function for this case is $x + xy\boldsymbol{H}_{\underline{101}}(x,y)$. If w'' is non-empty and $w' = \epsilon$, then the corresponding generating functions is $x\boldsymbol{H}_{\underline{101}}(x,y)$. If w' and w'' are non-empty, we distinguish two cases.

(i) If w' ends with $\mathtt{0}$ then $w' = v\mathtt{0}$ with $v \in \mathcal{C}(\underline{101})$. This implies that w'' cannot start with 01, which means that $w'' = 0w'''$ with $w''' \in \mathcal{C}(\underline{101})$. The generating function for this case is $x^2(1+\boldsymbol{H}_{\underline{101}}(x))x(1+\boldsymbol{H}_{\underline{101}}(x,y))$, where

$$\boldsymbol{H}_{\underline{101}}(x) = \frac{1-x^2-\sqrt{(1+x^2)(1-4x+x^2)}}{2x(1-x+x^2)} - 1$$

is the generating function for the number of all non-empty Catalan words avoiding $\underline{101}$ (see Thm. 2.3 of [5]).

(ii) If w' does not end with $\mathtt{0}$ then w'' has no more restriction, and the generating function for this case is

$$x(\boldsymbol{H}_{\underline{101}}(x) - x\boldsymbol{H}_{\underline{101}}(x) - x)\boldsymbol{H}_{\underline{101}}(x,y).$$

Therefore we have the functional equation

$$\begin{aligned}\boldsymbol{H}_{\underline{101}}(x,y) = x + xy\boldsymbol{H}_{\underline{101}}(x,y) + x\boldsymbol{H}_{\underline{101}}(x,y) + x^2(1+\boldsymbol{H}_{\underline{101}}(x))x(1+\boldsymbol{H}_{\underline{101}}(x,y)) +&\\ x(\boldsymbol{H}_{\underline{101}}(x) - x\boldsymbol{H}_{\underline{101}}(x) - x)\boldsymbol{H}_{\underline{101}}(x,y).&\end{aligned}$$

Solving this equation, we obtain the desired result. □

Theorem 4.15. *For $n \geq 2$, $1 \leq k \leq n$, we have*

$$\boldsymbol{c}_{\underline{101}}(n,k) = \sum_{i=k-1}^{n-2} \boldsymbol{c}_{\underline{101}}(n-1,i) - \boldsymbol{c}_{\underline{101}}(n-2,k),$$

and for $n \geq 2$,

$$\boldsymbol{c}_{\underline{101}}(n,0) = \sum_{i=0}^{n-2} \boldsymbol{c}_{\underline{101}}(n-1,i),$$

anchored with $\boldsymbol{c}_{\underline{101}}(1,0) = 1$, and $\boldsymbol{c}_{\underline{101}}(n,k) = 0$ otherwise. Moreover, for $n \geq 2$ and $2 \leq k \leq n$, we have

$$\boldsymbol{c}_{\underline{101}}(n,k) = \boldsymbol{c}_{\underline{101}}(n,k-1) - \boldsymbol{c}_{\underline{101}}(n-1,k-2) - \boldsymbol{c}_{\underline{101}}(n-2,k) + \boldsymbol{c}_{\underline{101}}(n-2,k-1).$$

Proof. Let $w = w_1 \cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{101})$, with $n \geq 2$. So, the subword $w_1 \cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{101})$ for $i \geq k-1$, but we do not have to consider the words where where $w_{n-2} = k$ and $w_{n-1} = k-1$. By summing over all possible values of i we obtain the first equality for $k \geq 1$, and the second for $k = 0$. Finally, by considering the difference $\boldsymbol{c}_{\underline{101}}(n,k) - \boldsymbol{c}_{\underline{101}}(n,k-1)$, we have the third equality. □

The first few rows of the matrix $\mathcal{T}_{\underline{101}}$ are

$$\mathcal{T}_{\underline{101}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{5} & \mathbf{4} & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 13 & 11 & 7 & 4 & 1 & 0 & 0 & 0 & 0 \\ 36 & 32 & 20 & 11 & 5 & 1 & 0 & 0 & 0 \\ 105 & 94 & 62 & 33 & 16 & 6 & 1 & 0 & 0 \\ 317 & 285 & 192 & 107 & 51 & 22 & 7 & 1 & 0 \\ 982 & 888 & 603 & 347 & 172 & 75 & 29 & 8 & 1 \end{pmatrix}$$

This array does not appear in the OEIS [15], however, the first column (and the row sums) of this matrix corresponds to the sequence A114465 (this sequence also counts the number of Dyck paths semilength n having no ascents of length 2 that start at an odd level). The Catalan words corresponding to the bold coefficients in the above array are

$$\mathcal{C}_{4,0}(\underline{101}) = \{\texttt{0000}, \texttt{0010}, \texttt{0100}, \texttt{0110}, \texttt{0120}\}, \quad \mathcal{C}_{4,1}(\underline{101}) = \{\texttt{0001}, \texttt{0011}, \texttt{0111}, \texttt{0121}\},$$
$$\mathcal{C}_{4,2}(\underline{101}) = \{\texttt{0012}, \texttt{0112}, \texttt{0122}\}, \quad \text{and} \quad \mathcal{C}_{4,3}(\underline{101}) = \{\texttt{0123}\}.$$

Corollary 4.16. *An asymptotic approximation for the expectation $\boldsymbol{a}_{\underline{101}}(n)$ of the last symbol over $\mathcal{C}_n(\underline{101})$ is*

$$\frac{13 + 8\sqrt{3}}{7 + 4\sqrt{3}} \sim 1.928203230.$$

4.6. The consecutive pattern $\underline{000}$

Theorem 4.17. *We have*

$$\boldsymbol{H}_{\underline{000}}(x, y) = \frac{2x(1 + x)}{1 + x + x^2(1 - 2y) - 2xy + \sqrt{1 - 2x - 5x^2 - 6x^3 - 3x^4}}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{000})$, and let $w = \texttt{0}(w' + 1)w''$ be the first return decomposition, where $w', w'' \in \mathcal{C}(\underline{000})$. If $w'' = \epsilon$, then $w = \texttt{0}(w' + 1)$ with w' possibly empty. The generating function for this case is $x + xy\boldsymbol{H}_{\underline{000}}(x, y)$. If w'' is non-empty and $w' = \epsilon$, then w'' cannot start with $\texttt{00}$. The corresponding generating function is $xH_{\underline{000},\texttt{00}}(x, y)$, where $H_{\underline{000},\texttt{00}}(x, y)$ is the bivariate generating function for Catalan words in $\mathcal{C}(\underline{000})$ that do not start with $\texttt{00}$. By counting by the complement we have

$$H_{\underline{000},\texttt{00}}(x, y) = x + xy\boldsymbol{H}_{\underline{000}}(x, y) + x\boldsymbol{H}_{\underline{000}}(x)\boldsymbol{H}_{\underline{000}}(x, y),$$

where (see Thm. 2.8 of [5])

$$\boldsymbol{H}_{\underline{000}}(x) = \frac{1 + x + x^2 - \sqrt{(1 - x - x^2)^2 - 4x^2(1 + x)^2}}{2x(1 + x)} - 1.$$

If w' and w'' are non-empty, then the generating function is $x\boldsymbol{H}_{\underline{000}}(x)\boldsymbol{H}_{\underline{000}}(x, y)$. Therefore, we have the functional equation

$$\boldsymbol{H}_{\underline{000}}(x, y) = x + xy\boldsymbol{H}_{\underline{000}}(x, y) + x(x + xy\boldsymbol{H}_{\underline{000}}(x, y) + x\boldsymbol{H}_{\underline{000}}(x)\boldsymbol{H}_{\underline{000}}(x, y)) + x\boldsymbol{H}_{\underline{000}}(x)\boldsymbol{H}_{\underline{000}}(x, y).$$

Solving this equation, we obtain the desired result. □

Theorem 4.18. *For $n \geq 2$, $1 \leq k \leq n$, we have*

$$\boldsymbol{c}_{\underline{000}}(n,k) = \sum_{i \geq k-1} \boldsymbol{c}_{\underline{000}}(n-1,i) - \boldsymbol{c}_{\underline{000}}(n-3,k-1) - \sum_{i \geq k+1} \boldsymbol{c}_{\underline{000}}(n-3,i),$$

and for $n \geq 2$,

$$\boldsymbol{c}_{\underline{000}}(n,0) = \sum_{m=0}^{n-1} \frac{1}{m+1} \sum_{i=\lceil n/2 \rceil - m-1}^{n-m-1} (-1)^i \binom{2m}{m} \binom{m+i}{i} \binom{m+i+1}{n-m-i-1},$$

anchored with $\boldsymbol{c}_{\underline{000}}(1,0) = 1$, and $\boldsymbol{c}_{\underline{000}}(n,k) = 0$, otherwise.

Proof. Let $w = w_1 \cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{000})$, with $n \geq 2$. So, the subword $w_1 \cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{000})$ for $i \geq k-1$, but we do not have to consider the words where $w_{n-2} = k$ and $w_{n-1} = k$. Since a word w' ending with kk in $\mathcal{C}_{n-1,k}(\underline{000})$ is of the form $w' = w''k-1kk$, or $w' = w''ikk, i \geq k+1$, the number of such words is given by $\boldsymbol{c}_{\underline{000}}(n-3,k-1) + \sum_{i \geq k+1} \boldsymbol{c}_{\underline{000}}(n-3,i)$. Therefore, we have

$$\boldsymbol{c}_{\underline{000}}(n,k) = \sum_{i \geq k-1} \boldsymbol{c}_{\underline{000}}(n-1,i) - \boldsymbol{c}_{\underline{000}}(n-3,k-1) - \sum_{i \geq k+1} \boldsymbol{c}_{\underline{000}}(n-3,i).$$

The second relation is given in the OEIS [15].

□

Notice that the sequence $\boldsymbol{c}_{\underline{000}}(n,0)$ corresponds to A061639. This sequence counts the number of planar planted trees with n non-root nodes and every 2-valent node isolated.

The first few rows of the matrix $\mathcal{T}_{\underline{000}}$ are

$$\mathcal{T}_{\underline{000}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{4} & \mathbf{3} & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 10 & 10 & 6 & 4 & 1 & 0 & 0 & 0 & 0 \\ 28 & 29 & 19 & 10 & 5 & 1 & 0 & 0 & 0 \\ 85 & 84 & 60 & 32 & 15 & 6 & 1 & 0 & 0 \\ 262 & 262 & 183 & 107 & 50 & 21 & 7 & 1 & 0 \\ 829 & 830 & 586 & 344 & 175 & 74 & 28 & 8 & 1 \end{pmatrix}$$

This array does not appear in the OEIS [15], however the row sums of $\mathcal{T}_{\underline{000}}$ corresponds to the sequence A247333. The Catalan words corresponding to the bold coefficients in the above array are

$$\mathcal{C}_{4,0}(\underline{000}) = \{\texttt{0010}, \texttt{0100}, \texttt{0110}, \texttt{0120}\}, \quad \mathcal{C}_{4,1}(\underline{000}) = \{\texttt{0011}, \texttt{0101}, \texttt{0121}\},$$
$$\mathcal{C}_{4,2}(\underline{000}) = \{\texttt{0012}, \texttt{0112}, \texttt{0122}\}, \quad \text{and} \quad \mathcal{C}_{4,3}(\underline{000}) = \{\texttt{0123}\}.$$

Corollary 4.19. *An asymptotic approximation for the expectation $\boldsymbol{a}_{\underline{000}}(n)$ of the last symbol over $\mathcal{C}_n(\underline{000})$ is 2.*

4.7. The consecutive pattern $\underline{210}$

Theorem 4.20. *We have*

$$\boldsymbol{H}_{\underline{210}}(x,y)=\frac{(2-x)x}{1-x-x^3-2xy+x^2(1+y)+\sqrt{1-4x+4x^3}}.$$

Proof. Let w denote a non-empty Catalan word in $\mathcal{C}(\underline{210})$, and let $w=\mathtt{0}(w'+1)w''$ be the first return decomposition, where $w',w''\in\mathcal{C}(\underline{210})$. If $w''=\epsilon$, then $w=\mathtt{0}(w'+1)$ with w' possibly empty. The generating function for this case is $x+xy\boldsymbol{H}_{\underline{210}}(x,y)$. If w'' is non-empty and $w'=\epsilon$, then the corresponding generating function is $x\boldsymbol{H}_{\underline{210}}(x,y)$. If w' and w'' are not empty, then w' cannot end with a descent, ab ($a>b$), that is, $w'=\mathtt{0}$ or $w'=vb$ where $v=v_1\dots v_r\in\mathcal{C}(\underline{210})$ with $r\geq 1$ and $b=v_r$ or $b=v_r+1$. Therefore, the generating function for this case is

$$x^2\boldsymbol{H}_{\underline{210}}(x,y)+2x^2\boldsymbol{H}_{\underline{210}}(x)\boldsymbol{H}_{\underline{210}}(x,y),$$

where (see Thm. 2.9 in [5])

$$\boldsymbol{H}_{\underline{210}}(x)=\frac{1-2x+2x^2-\sqrt{1-4x+4x^3}}{2(2-x)x^2}-1.$$

Therefore, we have the functional equation

$$\boldsymbol{H}_{\underline{210}}(x,y)=x+xy\boldsymbol{H}_{\underline{210}}(x,y)+x\boldsymbol{H}_{\underline{210}}(x,y)+x^2\boldsymbol{H}_{\underline{210}}(x,y)+2x^2\boldsymbol{H}_{\underline{210}}(x)\boldsymbol{H}_{\underline{210}}(x,y)$$

Solving this equation, we obtain the desired result. □

Theorem 4.21. *For $n\geq 2$, $1\leq k\leq n$, we have*

$$\boldsymbol{c}_{\underline{210}}(n,k)=\sum_{i\geq k-1}\boldsymbol{c}_{\underline{210}}(n-1,i)-\boldsymbol{c}_{\underline{210}}(n-3,k+1)-\sum_{i=1}^{n-2-k}(2i+1)\cdot\boldsymbol{c}_{\underline{210}}(n-3,k+i+1),$$

and for $n\geq 2$,

$$\boldsymbol{c}_{\underline{210}}(n,0)=\sum_{i\geq 0}\boldsymbol{c}_{\underline{210}}(n-1,i)-\boldsymbol{c}_{\underline{210}}(n-3,1)-\sum_{i=1}^{n-2}(2i+1)\cdot\boldsymbol{c}_{\underline{210}}(n-3,i+1),$$

anchored with $\boldsymbol{c}_{\underline{210}}(1,0)=1$, and $\boldsymbol{c}_{\underline{210}}(n,k)=0$ otherwise.

Proof. Let $w=w_1\cdots w_n$ denote a non-empty Catalan word in $\mathcal{C}_{n,k}(\underline{210})$, with $n\geq 2$. So, the subword $w_1\cdots w_{n-1}$ necessarily belongs to $\mathcal{C}_{n-1,i}(\underline{210})$ for $i\geq k-1$, but we do not have to consider the words where $w_{n-2}>w_{n-1}>k$. Since a word w' ending with ji, $j>i>k$ in $\mathcal{C}_{n-1,i}(\underline{210})$ is of the form $w'=w''ji$, where $w''\in\mathcal{C}_{n-3,j-1}(\underline{210})$ or $w''\in\mathcal{C}_{n-3,j}(\underline{210})$ the number of such words is given by

$$\begin{aligned}\sum_{i\geq k+1}\sum_{j\geq i+1}&\boldsymbol{c}_{\underline{210}}(n-3,j-1)+\sum_{i\geq k+1}\sum_{j\geq i+1}\boldsymbol{c}_{\underline{210}}(n-3,j)\\ &=\sum_{i=1}^{n-2-k}i\cdot\boldsymbol{c}_{\underline{210}}(n-3,k+i)+\sum_{i=0}^{n-2-k}(i+1)\cdot\boldsymbol{c}_{\underline{210}}(n-3,k+i+1)\\ &=\boldsymbol{c}_{\underline{210}}(n-3,k+1)+\sum_{i=1}^{n-2-k}(2i+1)\cdot\boldsymbol{c}_{\underline{210}}(n-3,k+i+1).\end{aligned}$$

Finally, we obtain the recursion

$$\boldsymbol{c}_{\underline{210}}(n,k)=\sum_{i\geq k-1}\boldsymbol{c}_{\underline{210}}(n-1,i)-\boldsymbol{c}_{\underline{210}}(n-3,k+1)-\sum_{i=1}^{n-2-k}(2i+1)\cdot\boldsymbol{c}_{\underline{210}}(n-3,k+i+1).$$

The case where $k=0$ can be obtained *mutatis mutandis.* □

Notice that the sequence $\boldsymbol{c}_{\underline{210}}(n,0)$ corresponds to A114465. This sequence counts the number of Dyck paths of length $2n$ having no ascents of length 2 that start at an odd level.

The first few rows of the matrix $\mathcal{T}_{\underline{210}}$ are

$$\mathcal{T}_{\underline{210}}=\begin{pmatrix}1&0&0&0&0&0&0&0&0\\1&1&0&0&0&0&0&0&0\\2&2&1&0&0&0&0&0&0\\\mathbf{5}&\mathbf{5}&\mathbf{3}&\mathbf{1}&0&0&0&0&0\\13&14&9&4&1&0&0&0&0\\36&40&28&14&5&1&0&0&0\\105&118&87&48&20&6&1&0&0\\317&359&273&161&75&27&7&1&0\\982&1118&869&536&270&110&35&8&1\end{pmatrix}$$

This array does not appear in the OEIS [15], however the row sums of $\mathcal{T}_{\underline{210}}$ corresponds to the sequence A159771. The Catalan words corresponding to the bold coefficients in the above array are

$$\mathcal{C}_{4,0}(\underline{210})=\{\texttt{0000},\texttt{0010},\texttt{0100},\texttt{0110},\texttt{0120}\},\quad \mathcal{C}_{4,1}(\underline{210})=\{\texttt{0001},\texttt{0011},\texttt{0101},\texttt{0111},\texttt{0121}\},$$
$$\mathcal{C}_{4,2}(\underline{210})=\{\texttt{0012},\texttt{0112},\texttt{0122}\},\quad\text{and}\quad \mathcal{C}_{4,3}(\underline{210})=\{\texttt{0123}\}.$$

Corollary 4.22. *An asymptotic approximation for the expectation $\boldsymbol{a}_{\underline{210}}(n)$ of the last symbol over $\mathcal{C}_n(\underline{210})$ is*

$$\frac{32a\left(a^3-2a^2+3a-1\right)\left(8976a^2+1552a-9297\right)^2(-2+a)}{13225\left(152a^2+64a-119\right)^2}\sim 2.943409552,$$

where $a=-\frac{\sqrt{3}}{3}\cos\left(\frac{1}{6}\arctan\left(\frac{3\sqrt{111}}{5}\right)+\frac{\pi}{6}\right)+\sin\left(\frac{1}{6}\arctan\left(\frac{3\sqrt{111}}{5}\right)+\frac{\pi}{6}\right)$ is the solution closest to the origin of $4z^3-4z+1$.

4.8. The consecutive patterns $\underline{120}$, $\underline{100}$, and $\underline{110}$

Theorem 4.23. *We have $\boldsymbol{H}_{\underline{100}}(x,y)=\boldsymbol{H}_{\underline{110}}(x,y)=\boldsymbol{H}_{\underline{120}}(x,y)$.*

Proof. There exists a bijection between Catalan words avoiding $\underline{100}$ and those avoiding $\underline{110}$ preserving the last symbol: from left to right, we replace each maximal factor $k^j(k-\ell)$, $j\geq 2$, $\ell\geq 1$, with the factor $k(k-\ell)^j$. For instance, the bijection transforms $w=\texttt{00012220122331}\in\mathcal{C}(100)$ into $\texttt{00012000122311}\in\mathcal{C}(110)$. Since this bijection preserves the last symbol, the two bivariate generating functions for the these sets are equal.

Also, there is a bijection between Catalan words avoiding $\underline{120}$ and those avoiding $\underline{110}$ preserving the last symbol: from left to right, we replace each factor $kk(k-\ell)$, $\ell\geq 1$, with the factor $k(k+1)(k-\ell)$. Then, the two bivariate generating functions for the these sets are equal. □

Theorem 4.24. *For $p \in \{\underline{120}, \underline{100}, \underline{110}\}$ we have*

$$\boldsymbol{H}_p(x,y) = \frac{1-2xy-\sqrt{1-4x+4x^3}}{2-2x^2-2y+2xy^2}.$$

Proof. Let w a non-empty Catalan word in $\mathcal{C}(\underline{120})$ whose first return decomposition is $w = \mathtt{0}(w'+1)w''$ such that $w', w'' \in \mathcal{C}(\underline{120})$. If $w'' = \epsilon$ then $w = \mathtt{0}(w'+1)$ with w' possibly empty. The generating function for this case is $x + xy\boldsymbol{H}_{\underline{120}}(x,y)$. If $w'' \neq \epsilon$ and $w' = \epsilon$ the corresponding generating function is $x\boldsymbol{H}_{\underline{120}}(x,y)$. If w' and w'' are non-empty, then $w'+1$ does not finish with an ascent $a(a+1)$, where $a \geq 1$. The corresponding generating function is

$$x(\boldsymbol{H}_{\underline{120}}(x) - \boldsymbol{H}'_{\underline{120}}(x))\boldsymbol{H}_{\underline{120}}(x,y),$$

where (see Thm. 2.6 in [5])

$$\boldsymbol{H}_{\underline{120}}(x) = \frac{1-2x^2-\sqrt{1-4x+4x^3}}{2x(1-x)} - 1$$

and $\boldsymbol{H}'_{\underline{120}}(x)$ is the generating function for non-empty Catalan words avoiding $\underline{120}$ and ending with an ascent, which induces that $\boldsymbol{H}'_{\underline{120}}(x) = x\boldsymbol{H}_{\underline{120}}(x)$.

Consequently, we have the functional equation

$$\boldsymbol{H}_{\underline{120}}(x,y) = x + xy\boldsymbol{H}_{\underline{120}}(x,y) + x\boldsymbol{H}_{\underline{120}}(x,y) + x(\boldsymbol{H}_{\underline{120}}(x) - x\boldsymbol{H}_{\underline{120}}(x))\boldsymbol{H}_{\underline{120}}(x,y)$$

and solving this, we obtain the desired result. $\square$

Theorem 4.25. *Let $p \in \{\underline{120}, \underline{100}, \underline{110}\}$. For $n \geq 2$ and $1 \leq k \leq n-1$, we have*

$$\boldsymbol{c}_p(n,k) = \boldsymbol{c}_p(n-2,k) - \boldsymbol{c}_p(n-1,k-2) + \boldsymbol{c}_p(n,k-1)$$

and for $n \geq 2$,

$$\boldsymbol{c}_p(n,0) = \frac{(-1)^{\frac{n-1-k}{2}}}{2}(1+(-1)^{n-1-k})\binom{k}{\lfloor\frac{n-1-k}{2}\rfloor}C_k,$$

where C_k is the k-th Catalan number.

Proof. For the first recurrence relation, it suffices to check that $\boldsymbol{H}_{\underline{120}}(x,y) - (x^2 - xy^2 + y)\boldsymbol{H}_{\underline{120}}(x,y) + xy$ does not depend on y. For the second relation, the sequence $\boldsymbol{c}_p(n,0)$ ($p \in \{\underline{120}, \underline{100}, \underline{110}\}$) is already known in the OEIS [15] as the sequence A157003. $\square$

Notice that the sequence A157003 also counts the number of Dyck paths of semilength n avoiding any one of the words $UDUDD$, $UUDDD$, $UUDUD$, or $UUUDD$. The first few rows of the matrix $\mathcal{T}_p$ for $p \in \{\underline{120}, \underline{100}, \underline{110}\}$

are

$$\mathcal{T}_p = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{4} & \mathbf{5} & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 10 & 12 & 9 & 4 & 1 & 0 & 0 & 0 & 0 \\ 27 & 32 & 25 & 14 & 5 & 1 & 0 & 0 & 0 \\ 78 & 90 & 72 & 44 & 20 & 6 & 1 & 0 & 0 \\ 234 & 266 & 213 & 137 & 70 & 27 & 7 & 1 & 0 \\ 722 & 812 & 650 & 428 & 235 & 104 & 35 & 8 & 1 \end{pmatrix}$$

This array does not appear in the OEIS [15], however the row sums of $\mathcal{T}_p$ corresponds to the sequence A087626. The Catalan words corresponding to the bold coefficients for $p = \underline{120}$ in the above array are

$$\mathcal{C}_{4,0}(\underline{120}) = \{\texttt{0000}, \texttt{0010}, \texttt{0100}, \texttt{0110}\}, \quad \mathcal{C}_{4,1}(\underline{120}) = \{\texttt{0001}, \texttt{0011}, \texttt{0101}, \texttt{0111}, \texttt{0121}\}$$
$$\mathcal{C}_{4,2}(\underline{120}) = \{\texttt{0012}, \texttt{0122}, \texttt{0112}\}, \quad \text{and} \quad \mathcal{C}_{4,3}(\underline{120}) = \{\texttt{0123}\},$$

Corollary 4.26. *For $p \in \{\underline{120}, \underline{100}, \underline{110}\}$, an asymptotic approximation for the expectation $\boldsymbol{a}_p(n)$ of the last symbol over $\mathcal{C}_n(p)$ is*

$$-\frac{4a\left(4a^2-5\right)^2(2a-1)}{\left(4a^2+2a-3\right)^2} \sim 2.340172972,$$

with $a = -\frac{\sqrt{3}}{3}\cos\left(\frac{1}{6}\arctan\left(\frac{3\sqrt{111}}{5}\right)+\frac{\pi}{6}\right)+\sin\left(\frac{1}{6}\arctan\left(\frac{3\sqrt{111}}{5}\right)+\frac{\pi}{6}\right)$.

5. Catalan matrices as Riordan arrays

In this section, we use Riordan arrays to describe the matrices introduced in the previous sections. We start by giving some background on Riordan arrays [16]. An infinite column vector $(a_0, a_1, \dots)^T$ has generating function $f(x)$ if $f(x) = \sum_{n\geq 0} a_n x^n$, and we index rows and columns starting at 0. A *Riordan array* is an infinite lower triangular matrix whose k-th column has generating function $g(x)f(x)^k$ for all $k \geq 0$, for some formal power series $g(x)$ and $f(x)$ with $g(0) \neq 0$, $f(0) = 0$, and $f'(0) \neq 0$. Such a Riordan array is denoted by $(g(x), f(x))$. If we multiply this matrix by a column vector $(c_0, c_1, \dots)^T$ having generating function $h(x)$, then the resulting column vector has generating function $g(x)h(f(x))$. The product of two Riordan arrays $(g(x), f(x))$ and $(h(x), l(x))$ is defined by

$$(g(x), f(x)) * (h(x), l(x)) = (g(x)h(f(x)), l(f(x))). \tag{5.1}$$

Under this operation, the set of all Riordan arrays is a group [16]. The identity element is $I = (1, x)$, and the inverse of $(g(x), f(x))$ is

$$(g(x), f(x))^{-1} = \left(1/\left(g \circ f^{<-1>}\right)(x), f^{<-1>}(x)\right), \tag{5.2}$$

where $f^{<-1>}(x)$ denotes the compositional inverse of $f(x)$.

For a consecutive pattern p and a formal power series $f(x) = \sum_{i\geq 0} f_i x^i$, such that $f_0 = 1$, we introduce the infinite matrix

$$\mathcal{M}_p^{f(x)} := \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{f} & \mathcal{T}_p \end{pmatrix},$$

where $\mathcal{T}_p = (\boldsymbol{c}_p(n,k))_{n\geq 1, k\geq 0}$, $\boldsymbol{f} = (f_1, f_2, \dots)^T$, and $\boldsymbol{0} = (0,0,\dots)$. In other words, let $\mathcal{T}_p[i]$ denote the i-th column of $\mathcal{T}_p$, then $\mathcal{M}_p^{f(x)} = (f(x), \mathcal{T}_p[0], \mathcal{T}_p[1], \dots)$.

Theorem 5.1. *For* $p \in \{\underline{010}, \underline{000}, \underline{210}, \underline{120}, \underline{100}, \underline{110}\}$ *and* $f(x) = 1$*, the matrix* $\mathcal{M}_p^{f(x)}$ *is a Riordan array given by* $(1, \boldsymbol{H}_p(x,0))$.

Proof. Let us deal with the case $p = \underline{010}$ and the remaining cases can be obtained *mutatis mutandis*. A Catalan word $w \in \mathcal{C}(\underline{010})$, whose last symbol is k, can be decomposed as $w_0(w_1+1)\cdots(w_k+k)$, where $w_i \in \cup_{n\geq 1}\mathcal{C}_{n,0}(\underline{010})$ for $0 \leq i \leq k$. Therefore, the generating function of $\mathcal{T}_p[k]$ is given by $\sum_{n\geq 0} \boldsymbol{c}_{\underline{010}}(n,k)x^n = \boldsymbol{H}_{\underline{010}}(x,0)^k$. Consequently, $\mathcal{M}_{\underline{010}}^{f(x)=1} = (1, \boldsymbol{H}_{\underline{010}}(x,0), \boldsymbol{H}_{\underline{010}}(x,0)^2, \dots)$ is a Riordan array. $\square$

For example,

$$\mathcal{M}_{\underline{010}}^{f(x)=1} = \left(1, \frac{1+x^2-\sqrt{1-4x+2x^2-4x^3+x^4}}{2(1+x^2)}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 3 & 3 & 1 & 0 & 0 & 0 \\ 0 & 9 & 8 & 6 & 4 & 1 & 0 & 0 \\ 0 & 25 & 25 & 16 & 10 & 5 & 1 & 0 \\ 0 & 73 & 74 & 51 & 28 & 15 & 6 & 1 \end{pmatrix}.$$

Theorem 5.2. *For* $p \in \{\underline{001}, \underline{011}\}$ *and* $f(x) = 1/(1-x)$*, the matrix* $\mathcal{M}_p^{f(x)}$ *is a Riordan array given by*

$$\left(f(x), \frac{\boldsymbol{H}_p(x,0)}{f(x)}\right) = \left(\frac{1}{1-x}, (1-x)\boldsymbol{H}_p(x,0)\right).$$

Proof. Let us deal with case $p = \underline{001}$, for the other case we can follow a similar argument. A Catalan word $w \in \mathcal{C}(\underline{001})$, whose last symbol is k, can be decomposed as $w_0(w_1+1)\cdots(w_k+k)$, where $w_i \in \cup_{n\geq 1}\mathcal{C}_{n,0}(\underline{001})$ for $0 \leq i \leq k$. The pattern $\underline{001}$ can be observed in between the blocks $(w_i+i)(w_{i+1}+i+1)$ if it has the form $(w_i'+i)\mathtt{0}^j(w_{i+1}'+i+1)$, for some $j \geq 1$. Therefore, the generating function of the column $\mathcal{T}_{\underline{001}}[k]$ is given by $\boldsymbol{H}_{\underline{001}}(x,0)^k/f(x)^{k-1}$. From the definition of Riordan array we obtain the desired result. $\square$

For example,

$$\mathcal{M}_{\underline{001}}^{f(x)=1/(1-x)} = \left(\frac{1}{1-x}, \frac{2(1-x)x}{1-x^2+\sqrt{1-4x+2x^2+x^4}}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 4 & 3 & 1 & 1 & 0 & 0 & 0 \\ 1 & 9 & 7 & 4 & 1 & 1 & 0 & 0 \\ 1 & 22 & 18 & 10 & 5 & 1 & 1 & 0 \\ 1 & 57 & 48 & 28 & 13 & 6 & 1 & 1 \end{pmatrix}.$$

Theorem 5.3. *The matrix* $\mathcal{T}_{\underline{101}}$ *is a Riordan array given by* $\big(\boldsymbol{H}_{\underline{101}}(x,0)/x, \boldsymbol{H}_{\underline{010}}(x,0)\big)$.

Proof. Multiplying the right-hand side of the equality by the vector $(1, y, y^2, \dots)^T$, which has generating function $1/(1-xy)$, the resulting vector has bivariate generating function

$$\begin{aligned}\left(\boldsymbol{H}_{\underline{101}}(x,0)/x, \boldsymbol{H}_{\underline{010}}(x,0)\right)\frac{1}{1-xy} &= \frac{\boldsymbol{H}_{\underline{101}}(x,0)}{x}\frac{1}{1-\boldsymbol{H}_{\underline{010}}(x,0)y}\\ &= \frac{2-x+2x^2-x^3-x\sqrt{1-4x+2x^2-4x^3+x^4}}{(1-x+x^2)(1-2xy+x^2+\sqrt{1-4x+2x^2-4x^3+x^4})}\\ &= \frac{\boldsymbol{H}_{\underline{101}}(x,y)}{x},\end{aligned}$$

by Theorem 4.14. □

From a similar argument as in the previous theorem we obtain the following theorem.

Theorem 5.4. *The matrix $\mathcal{T}_{\underline{201}}$ is a Riordan array given by $\left(\boldsymbol{H}_{\underline{201}}(x,0)/x, \boldsymbol{H}_p(x,0)\right)$, where $p \in \{\underline{120}, \underline{100}, \underline{110}\}$.*

Notice that the matrix related to the pattern $\underline{012}$ can not be a Riordan array.

From Theorems 5.3 and 5.4 we obtain the following combinatorial identities:

$$\boldsymbol{c}_{\underline{101}}(n,k) = \sum_{\ell+\ell_1+\cdots+\ell_k=n} \boldsymbol{c}_{\underline{101}}(\ell,0)\boldsymbol{c}_{\underline{010}}(\ell_1,0)\cdots\boldsymbol{c}_{\underline{010}}(\ell_k,0)$$

and

$$\boldsymbol{c}_{\underline{201}}(n,k) = \sum_{\ell+\ell_1+\cdots+\ell_k=n} \boldsymbol{c}_{\underline{201}}(\ell,0)\boldsymbol{c}_p(\ell_1,0)\cdots\boldsymbol{c}_p(\ell_k,0),$$

where $p \in \{\underline{120}, \underline{100}, \underline{110}\}$.

Finally, we will use a characterization of the Riordan arrays given by Rogers in [17]. That is, every element not belonging to row 0 or column 0 in a Riordan array can be expressed as a fixed linear combination of the elements in the preceding row. The A-sequence is defined to be the sequence coefficients of this linear combination. Similarly, Merlini *et al.* [18] introduced the Z-sequence, that characterizes the elements in column 0, except for the top one.

Theorem 5.5 ([18])**.** *An infinite lower triangular array $\mathcal{F} = (d_{n,k})_{n,k\geq 0}$ is a Riordan array if and only if $d_{0,0} \neq 0$ and there exist two sequences $(a_0, a_1, a_2, \dots)$, with $a_0 \neq 0$, and $(z_0, z_1, z_2, \dots)$ (called the A-sequence and the Z-sequence, respectively), such that*

$$\begin{aligned} d_{n+1,k+1} &= a_0 d_{n,k} + a_1 d_{n,k+1} + a_2 d_{n,k+2} + \cdots && \text{for } n,k \geq 0,\\ d_{n+1,0} &= z_0 d_{n,0} + z_1 d_{n,1} + z_2 d_{n,2} + \cdots && \text{for } n \geq 0.\end{aligned}$$

Note that the A-sequence, the Z-sequence, and the upper-left element completely characterize a Riordan array.

Theorem 5.6 ([18, 19])**.** *Let $\mathcal{F} = (g(x), f(x))$ be a Riordan array with inverse $\mathcal{F}^{-1} = (d(x), h(x))$. Then the A-sequence and the Z-sequence of $\mathcal{F}$ have generating functions*

$$A(x) = \frac{x}{h(x)}, \qquad Z(x) = \frac{1}{h(x)}\left(1 - d_{0,0}d(x)\right),$$

respectively.

From the definition of the A-sequence and Z-sequence for the Riordan arrays we can give additional recurrence relations for the sequences $\boldsymbol{c}_p(n,k)$.

Theorem 5.7. *If C_n denotes the n-th Catalan number, then for $n \geq 2$ and $k \geq 0$,*

$$\boldsymbol{c}_{\underline{010}}(n,k) = \sum_{\ell=0}^{n-1} \boldsymbol{c}_{\underline{010}}(n-1, k-1-\ell)\, a_\ell,$$

where

$$a_n := 1 + \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^{i+1} \binom{n-i-1}{i} \overline{C}_{n-i-1} \quad \textit{and} \quad \overline{C}_n := \begin{cases} C_{\frac{n-1}{2}}, & \textit{if } n \textit{ is odd;} \\ 0, & \textit{otherwise.} \end{cases}$$

Proof. By equation (5.2), the inverse of the matrix $\mathcal{M}_{\underline{010}}^{f(x)=1}$ is given by

$$(\mathcal{M}_{\underline{010}}^{f(x)=1})^{-1} = \left(1, \frac{1-\sqrt{1-4x^2+8x^3-4x^4}}{2(1-x)x}\right).$$

Therefore, by Theorem 5.6, the A-sequence and Z-sequence of the Riordan array $\mathcal{M}_{\underline{010}}^{f(x)=1}$ have generating functions given by

$$A(x) = \sum_{n\geq 0} a_n x^n = \frac{1+\sqrt{1-4((1-x)x)^2}}{2(1-x)} \quad \text{and} \quad Z(x) = 0.$$

Notice that $A(x) = xF(x(1-x))$, where $F(x) := \sum_{n\geq 0} f_n x^n = (1+\sqrt{1-4x^2})/(2x)$. It is possible to prove that $f_n = \overline{C}_n$, where $\overline{C}_n$ is as in the statement of the theorem. By comparing coefficients in $A(x)$ and the recurrences from Theorem 5.5 we now obtain the desired result. $\square$

The first few values of the sequence a_n for $n \geq 0$ are

$$1, \quad 1, \quad 0, \quad 2, \quad 0, \quad 4, \quad -4, \quad 12, \quad -24, \quad 56, \quad -128, \ldots.$$

We can obtain similar identities for the remaining sequences.

6. Final remarks

We use generating functions to enumerate Catalan words avoiding any consecutive pattern of length two and three with respect to the length and the value of the last symbol. We note that much of the combinatorial interpretations given by us are (probably) new and they do not correspond to those that appear in the OEIS. In this context, it can be of interest to obtain bijections that show the links between the different combinatorial objects.

Finally, we now give a potential open problem. What is the combinatorial interpretation of the inverse Riordan arrays introduced in the previous theorem? For example, the absolute value of the second and third columns of the inverse matrix $(\mathcal{M}_{\underline{010}}^{f(x)=1})^{-1}$ are the sequences A104545 (number of Motzkin paths of length n having no consecutive $(1,0)$ steps) and A256169, respectively.

$$(\mathcal{M}_{\underline{010}}^{f(x)=1})^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & -3 & 3 & -3 & 1 & 0 & 0 & 0 \\ 0 & 5 & -8 & 6 & -4 & 1 & 0 & 0 \\ 0 & -11 & 17 & -16 & 10 & -5 & 1 & 0 \\ 0 & 25 & -38 & 39 & -28 & 15 & -6 & 1 \end{pmatrix}.$$

Acknowledgements. We thanks the referees for the detailed comments and corrections that help us to improve the paper.

References

[1] R. Stanley, Catalan Numbers. Cambridge University Press, Cambridge (2015).
[2] T. Mansour and V. Vajnovszki, Efficient generation of restricted growth words. *Inform. Process. Lett.* **113** (2013) 613–616.
[3] J.-L. Baril, S. Kirgizov and V. Vajnovszki, Descent distribution on Catalan words avoiding a pattern of length at most three. *Discrete Math.* **341** (2018) 2608–2615.
[4] J.-L. Baril, C. Khalil and V. Vajnovszki, Catalan words avoiding pairs of length three patterns. *Discret. Math. Theor. Comput. Sci.* **22** (2021) 5.
[5] J.L. Ramírez and A. Rojas-Osorio, Consecutive patterns in Catalan words and the descent distribution. *Bol. Soc. Mat. Mex.* **29** (2023) Article 60.
[6] J.-L. Baril, J.F. González and J.L. Ramírez, Last symbol distribution in pattern avoiding Catalan words. Submitted (2023).
[7] D. Callan, T. Mansour and J.L. Ramírez, Statistics on bargraphs of Catalan words. *J. Autom. Lang. Comb.* **26** (2021) 177–196.
[8] T. Mansour, J.L. Ramírez and D.A. Toquica, Counting lattice points on bargraphs of Catalan words. *Math. Comput. Sci.* **15** (2021) 701–713.
[9] P. Flajolet and R. Sedgewick, Analytic Combinatorics. Cambridge University Press, Cambridge (2009).
[10] A.G. Orlov, On asymptotic behavior of the Taylor coefficients of algebraic functions. *Sib. Math. J.* **25** (1994) 1002–1013.
[11] S. Elizalde and M. Noy, Consecutive patterns in permutations. *Adv. Appl. Math.* **30** (2003) 110–125.
[12] J.S. Auli and S. Elizalde, Consecutive patterns in inversion sequences. *Discret. Math. Theor. Comput. Sci.* **21** (2019) 6.
[13] A. Mendes and J. Remmel, Permutations and words counted by consecutive patterns. *Adv. Appl. Math.* **37** (2006) 443–480.
[14] A. Panholze, Consecutive permutation patterns in trees and mappings. *J. Comb.* **12** (2021) 17–54.
[15] N.J.A. Sloane, The On-Line Encyclopedia of Integer Sequences. Available at https://oeis.org/.
[16] L.W. Shapiro, S. Getu, W. Woan and L. Woodson, The Riordan group. *Discrete Appl. Math.* **34** (1991) 229–239.
[17] D.G. Rogers, Pascal triangles, Catalan numbers and renewal arrays. *Discrete Math.* **22** (1978) 301–310.
[18] D. Merlini, D.G. Rogers, R. Sprugnoli and M.C. Verri, On some alternative characterizations of Riordan arrays. *Can. J. Math.* **49** (1997) 301–320.
[19] T.-X. He and R. Sprugnoli, Sequence characterization of Riordan arrays. *Discrete Math.* **309** (2009) 3962–3974.