Discrete Mathematics 341 (2018) 2608–2615

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Descent distribution on Catalan words avoiding a pattern of length at most three

Jean-Luc Baril, Sergey Kirgizov, Vincent Vajnovszki

LE2I, Université de Bourgogne Franche-Comté, B.P. 47 870, 21078 DIJON-Cedex, France

ARTICLE INFO

Article history:
Received 18 March 2018
Received in revised form 24 May 2018
Accepted 1 June 2018

Keywords:
Enumeration
Catalan word
Pattern avoidance
Descent
Popularity

ABSTRACT

Catalan words are particular growth-restricted words over the set of non-negative integers, and they represent still another combinatorial class counted by the Catalan numbers. We study the distribution of descents on the sets of Catalan words avoiding a pattern of length at most three: for each such a pattern p we provide a bivariate generating function where the coefficient of x^ny^k in its series expansion is the number of length n p-avoiding Catalan words with k descents. As a byproduct, we enumerate the set of Catalan words avoiding p, and we provide the popularity of descents on this set.

1. Introduction and notation

A length n *Catalan word* is a word $w_1w_2\ldots w_n$ over the set of non-negative integers with $w_1 = 0$, and

$$0 \le w_i \le w_{i-1} + 1,$$

for $i = 2, 3, \ldots, n$. We denote by $\mathcal{C}_n$ the set of length n Catalan words, and $\mathcal{C} = \cup_{n\ge 0}\mathcal{C}_n$. For example, $\mathcal{C}_2 = \{00, 01\}$ and $\mathcal{C}_3 = \{000, 001, 010, 011, 012\}$. It is well known that the cardinality of $\mathcal{C}_n$ is given by the nth Catalan number $\frac{1}{n+1}\binom{2n}{n}$, see for instance [13, exercise 6.19.*u*, p. 222], which is the general term of the sequence A000108 in the On-line Encyclopedia of Integer Sequences (OEIS) [12]. See also [10] where Catalan words are considered in the context of the exhaustive generation of Gray codes for growth-restricted words.

A *pattern* p is a word satisfying the property that if x appears in p, then all integers in the interval $[0, x-1]$ also appear in p. We say that a word $w_1w_2\ldots w_n$ contains the pattern $p = p_1\ldots p_k$ if there is a subsequence $w_{i_1}w_{i_2}\ldots w_{i_k}$ of w, $i_1 < i_2 < \cdots < i_k$, which is order-isomorphic to p. For example, the Catalan word 01012312301 contains seven occurrences of the pattern 110 and four occurrences of the pattern 210. A word *avoids* the pattern p whenever it does not contain any occurrence of p. We denote by $\mathcal{C}_n(p)$ the set of length n Catalan words avoiding the pattern p, and $\mathcal{C}(p) = \cup_{n\ge 0}\mathcal{C}_n(p)$. For instance, $\mathcal{C}_4(012) = \{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111\}$, and $\mathcal{C}_4(101) = \{0000, 0001, 0010, 0011, 0012, 0100, 0110, 0111, 0112, 0120, 0121, 0122, 0123\}$.

A *descent* in a word $w = w_1w_2\ldots w_n$ is an occurrence w_iw_{i+1} such that $w_i > w_{i+1}$, and we denote by $d(w)$ the number of descents of w. The distribution of the number of descents has been widely studied on several classes of combinatorial objects such as permutations and words, since descents have some particular interpretations in fields as Coxeter groups or theory of lattice paths [4,6]. More specifically, there are natural one-to-one correspondences between descents in Catalan words and some patterns in other classical Catalan structures, and below we give two such examples.

E-mail addresses: barjl@u-bourgogne.fr (J. Baril), sergey.kirgizov@u-bourgogne.fr (S. Kirgizov), vvajnov@u-bourgogne.fr (V. Vajnovszki).

https://doi.org/10.1016/j.disc.2018.06.001

Let $\delta \mapsto w$ be the bijection which maps a semilength n Dyck word over $\{u, d\}$ into a length n Catalan word defined as: w is the sequence of the lowest ordinate of the up steps u in the Dyck word δ, in lattice path representation. Under this bijection, occurrences of consecutive patterns ddu in Dyck words correspond to descents in Catalan words. Similarly, Mäkinen's bijection [9] gives a one-to-one correspondence between descents in Catalan words (called *left-distance sequences* by the author of [9]) and particular nodes (left-child nodes having a right child) in binary trees.

A *statistic st* on a finite set S is an association of an integer to each element of S, and the *popularity* of st is $\sum_{x\in S} st(x)$, which is the cardinality of S times the expectation of st. The number of occurrences of a pattern or the number of descents are examples of statistics on words. See [5] where the notion of popularity was introduced in the context of pattern based statistics, and [1,7,11,2] for some related results.

The main goal of this paper is to study the descent distribution on Catalan words (see Table 1 for some numerical values). More specifically, for each pattern p of length at most three, we give the distribution of descents on the sets $\mathcal{C}_n(p)$ of length n Catalan words avoiding p. We denote by $C_p(x, y) = \sum_{n,k\geq 0} c_{n,k} x^n y^k$ the bivariate generating function for the cardinality of words in $\mathcal{C}_n(p)$ with k descents. Plugging $y = 1$

– into $C_p(x, y)$, we deduce the generating function $C_p(x)$ for the set $\mathcal{C}_n(p)$, and
– into $\frac{\partial C_p(x,y)}{\partial y}$, we deduce the generating function for the popularity of descents in $\mathcal{C}_n(p)$.

The proofs in this paper are based mainly on functional equations, and alternative bijective proofs would be of interest.

From the definition at the beginning of this section it follows that a Catalan word is either the empty word, or it can uniquely be written as $0(w' + 1)w''$, where w' and w'' are in turn Catalan words, and $w' + 1$ is obtained from w' by adding one to each of its entries. We call this recursive decomposition *first return decomposition* of a Catalan word, and it will be crucial in our further study. It follows that $C(x)$, the generating function for the cardinality of $\mathcal{C}_n$, satisfies:

$$C(x) = 1 + x \cdot C^2(x),$$

which corresponds precisely to the sequence of Catalan numbers.

The remainder of the paper is organized as follows. In Section 2, we study the distribution of descents on the set $\mathcal{C}$ of Catalan words. As a byproduct, we deduce the popularity of descents in $\mathcal{C}$. We consider also similar results for the obvious cases of Catalan words avoiding a pattern of length two. In Section 3, we study the distribution and the popularity of descents on Catalan words avoiding each pattern of length three.

2. The sets $\mathcal{C}$ and $\mathcal{C}(p)$ for $p \in \{00, 01, 10\}$

Here we consider both unrestricted Catalan words and those avoiding a length two pattern.

We denote by $C(x, y)$ the bivariate generating function where the coefficient of $x^n y^k$ of its series expansion is the number of length n Catalan words with k descents. When we restrict to Catalan words avoiding a pattern p, the corresponding generating function is denoted by $C_p(x, y)$.

Theorem 1. *We have*

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

Proof. Let $w = 0(w' + 1)w''$ be the first return decomposition of a non-empty Catalan word w with $w', w'' \in \mathcal{C}$. If w' (resp. w'') is empty then the number $d(w)$ of descents in w is the same as that of w'' (resp. w'); otherwise, we have $d(w) = d(w') + d(w'') + 1$ since there is a descent between $w' + 1$ and w''. So, we obtain the functional equation $C(x, y) = 1 + xC(x, y) + x(C(x, y) - 1) + xy(C(x, y) - 1)^2$ which gives the desired result. □

By the considerations in the introductory section, it turns out that the coefficient of $x^n y^k$ in the series expansion of $C(x, y)$ is also the number of semilength n Dyck words with k occurrences of the consecutive pattern ddu, which is given by the statistic St000386 in [3], see also A091894 in [12].

As expected, $C(x) = C(x, 1) = \frac{1-\sqrt{1-4x}}{2x}$ is the generating function for the Catalan numbers, and $\frac{\partial C(x,y)}{\partial y}|_{y=1}$ is the generating function for the descent popularity on $\mathcal{C}$, and we have the next corollary.

Corollary 1. *The popularity of descents on the set* $\mathcal{C}_n$ *is* $\binom{2n-2}{n-3}$, *and its generating function is* $\frac{1-4x+2x^2-(1-2x)\sqrt{1-4x}}{2x\sqrt{1-4x}}$ *(sequence* A002694 *in* [12]*).*

Catalan words of odd lengths encompass a smaller size Catalan structure. This result is stated in the next theorem, see the bold entries in Table 1.

Theorem 2. *Catalan words of length* $2n + 1$ *with* n *descents are enumerated by the nth Catalan number* $\frac{1}{n+1}\binom{2n}{n}$.