

STATISTICS ON BARGRAPHS OF CATALAN WORDS

DAVID CALLAN, TOUFIK MANSOUR, AND JOSÉ L. RAMÍREZ

ABSTRACT. Catalan words are a particular case of growth-restricted words. Here we give the bivariate generating function for bargraphs associated to Catalan words according to the semiperimeter and area statistics. Exact formulas to count Catalan bargraphs according to the two statistics are also found. We show a connection to the Narayana numbers. Finally, we give similar results for the exterior corner statistic.

1. INTRODUCTION

A word $w = w_1 w_2 \cdots w_n$ over the set of positive integers is called a *Catalan word* if $w_1 = 1$ and $1 \leq w_i \leq w_{i-1} + 1$ for $i = 2, \dots, n$. Let $\mathcal{C}_n$ denote the set of the Catalan words of length n . For example, $\mathcal{C}_3 = \{111, 112, 121, 122, 123\}$. The cardinality of the set $\mathcal{C}_n$ is given by the Catalan number $C_n = \frac{1}{n+1} \binom{2n}{n}$, see [20, Exercise 80]. There is a natural way to associate a Dyck path of length $2n$ with a Catalan word in $\mathcal{C}_n$. Remember that a *Dyck path* of length $2n$ is a lattice path of $\mathbb{Z} \times \mathbb{Z}$ running from $(0, 0)$ to $(2n, 0)$ that never passes below the x -axis and whose permitted steps are the up diagonal step $(1, 1)$ and the down diagonal step $(1, -1)$, called *rise* and *fall*, respectively. For a Dyck path of length $2n$, we associate the word formed by the y -coordinate of each final point of the rises steps. This word is clearly a Catalan word of length n and the construction is a bijection. For example, in Figure 1 we show the Dyck path associated to the Catalan word $12211123433421 \in \mathcal{C}_{15}$. Catalan words have been studied in the context of exhaustive generation of Gray codes for growth-restricted words [16]. Recently, Barril et al. [1] study the distribution of descents on the sets of Catalan words avoiding a pattern of length at most three.

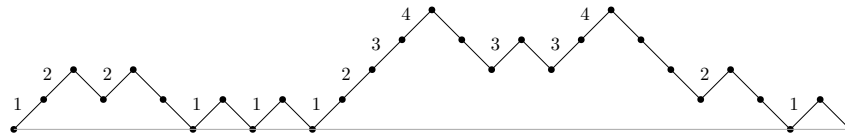


FIGURE 1. Dyck path of the Catalan word 12211123433421.

In this paper, we consider the area, semiperimeter and exterior corner statistics of the Catalan words over their associated bargraphs. Remember that a *bargraph* is a self-avoiding lattice path in the first quadrant with steps $U = (0, 1)$, $H = (1, 0)$, and $D = (0, -1)$ that starts at the origin and ends on the x -axis. A Catalan word $w = w_1 \cdots w_n$ can be represented as a bargraph, whose i -th column contains w_i cells for $1 \leq i \leq n$. For example, Figure 2 shows

Date: August 10, 2020.

2010 Mathematics Subject Classification. 05A15, 05A19.

Key words and phrases. Catalan word, generating function, area, semiperimeter, external corners.

the bargraph associated to the Catalan word 12211123433421. The distribution of several bargraph statistics has been given in [2, 3, 4, 10, 14, 15] and references contained therein.

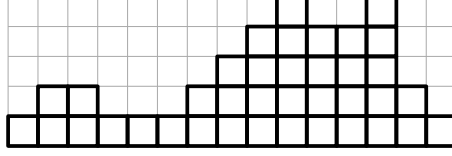


FIGURE 2. Bargraph of the Catalan word 12211123433421.

The outline of the paper is this: Section 2 contains the bivariate generating function of the joint distribution of the area and semiperimeter statistics over the set of Catalan words of length n . Recurrence relations for the joint distribution are given for the Catalan words that end with letter i . Section 3 contains the generating function for the number of nonempty Catalan words according to length and semiperimeter, as well as an explicit expression in terms of binomial coefficients. Finally, Sections 4 and 5 contain analogous results for the area and exterior corners statistics.

2. AREA AND SEMIPERIMETER STATISTICS

Let w be a Catalan word. We denote by $\mathbf{area}(w)$ and $\mathbf{sper}(w)$ the area and semiperimeter of the bargraph representation of the Catalan word w . For example, for the bargraph in Figure 2 we have $\mathbf{area}(w) = 38$ and $\mathbf{sper}(w) = 22$. Let $\mathcal{C}_{n,i}$ denote the set of Catalan words of length n whose last symbol is i . Define the generating function

$$A_i(x; p, q) := \sum_{n \geq 1} x^n \sum_{w \in \mathcal{C}_{n,i}} p^{\mathbf{sper}(w)} q^{\mathbf{area}(w)}.$$

By definitions, we have

$$(1) \quad A_1(x; p, q) = p^2 qx + pqx \sum_{j \geq 1} A_j(x; p, q),$$

$$(2) \quad A_i(x; p, q) = p^2 q^i x A_{i-1}(x; p, q) + pq^i x \sum_{j \geq i} A_i(x; p, q), \quad i \geq 2.$$

Define the bivariate generating function $A(x; p, q; v) = \sum_{i \geq 1} A_i(x; p, q) v^{i-1}$. Then by multiplying (2) by v^{i-1} and summing over $i \geq 2$ with using (1), we obtain

$$A_1(x; p, q) = p^2 qx + pqx A(x; p, q; 1),$$

$$A(x; p, q; v) - A_1(x; p, q) = p^2 q^2 xv A(x; p, q; qv) + \frac{pq^2 xv}{1 - qv} (A(x; p, q; 1) - A(x; p, q; qv)),$$

which leads to

$$(3) \quad A(x; p, q; v) = p^2 qx + pqx A(x; p, q; 1) + p^2 q^2 xv A(x; p, q; qv) + \frac{pq^2 xv}{1 - qv} (A(x; p, q; 1) - A(x; p, q; qv)).$$

Note that when $p = q = 1$, we have

$$\frac{1 - v + xv^2}{1 - v} A(x; 1, 1; v) = x + \frac{x}{1 - v} A(x; 1, 1; 1).$$

By taking $v = C(x) = \frac{1-\sqrt{1-4x}}{2x}$ the generating function for the Catalan numbers, we see that $A(x; 1, 1; 1) = C(x) - 1$, as expected. Hence,

$$(4) \quad A(x; 1, 1; v) = x \frac{C(x) - v}{1 - v + xv^2}.$$

Let $a_{n,i}(p, q)$ denote the joint distribution over $\mathcal{C}_{n,i}$ for the area and semiperimeter, that is,

$$a_{n,i}(p, q) := \sum_{w \in \mathcal{C}_{n,i}} p^{\text{sper}(w)} q^{\text{area}(w)}.$$

Theorem 2.1. *For $n \geq 3$ and $3 \leq i \leq n$*

$$a_{n,i}(p, q) = qa_{n,i-1}(p, q) + (p-1)pq^i a_{n-1,i-1}(p, q) - p^2 q^i a_{n-1,i-2}(p, q),$$

with the initial conditions $a_{n,1}(p, q) = pq \sum_{i=1}^{n-1} a_{n-1,i}(p, q)$ and $a_{n,2}(p, q) = qa_{n,1}(p, q) + pq^2 a_{n-1,1}(p, q)$ for $n \geq 2$, and $a_{1,1}(p, q) = p^2 q$.

Proof. For $n \geq 2$ and $1 \leq i \leq n$, we have

$$(5) \quad a_{n,i}(p, q) = pq^i \sum_{j=i}^{n-1} a_{n-1,j}(p, q) + p^2 q^i a_{n-1,i-1}(p, q).$$

Indeed, any Catalan word $w \in \mathcal{C}_{n,i}$ can be decomposed as $w = w'i$, where $w' \in \mathcal{C}_{n-1,j}$ for $i-1 \leq j \leq n-1$. If $j = i-1$ then the contribution is $p^2 q^i a_{n-1,i-1}(p, q)$, see Figure 3 (left). For $i \leq j \leq n-1$ the contribution is given by $pq^i a_{n-1,j}(p, q)$, see Figure 3 (right). Summing over j we obtain the sum on the right side of (5).

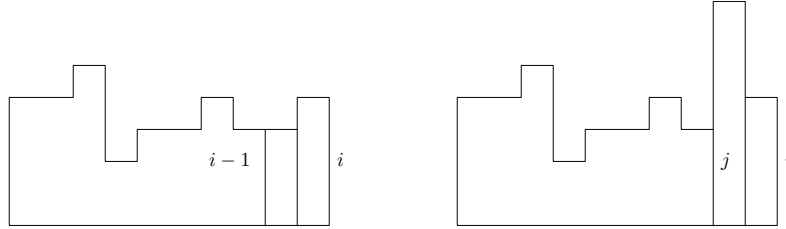


FIGURE 3. Decomposition of the Catalan word in $\mathcal{C}_{n,i}$.

From (5) we have that

$$a_{n,i}(p, q) - qa_{n,i-1}(p, q) = p^2 q^i a_{n-1,i-1}(p, q) - pq^i a_{n-1,i-1}(p, q) - p^2 q^i a_{n-1,i-2}(p, q).$$

The initial values can be obtained from the definition. $\square$

We recall that an infinite lower triangular matrix is called a (proper) Riordan array [18] if its k th column satisfies the generating function $g(z)(f(z))^k$ for $k \geq 0$, where $g(z)$ and $f(z)$ are formal power series with $g(0) \neq 0$, $f(0) = 0$ and $f'(0) \neq 0$. The matrix corresponding to the pair $f(z), g(z)$ is denoted by $(g(z), f(z))$. If we multiply (g, f) by a column vector $(c_0, c_1, \dots)^T$ with the generating function $h(z)$, then the resulting column vector has generating function $gh(f)$. This property is known as the Fundamental Theorem of Riordan arrays or summation property.

The product of two Riordan arrays $(g(z), f(z))$ and $(h(z), l(z))$ is defined by

$$(g(z), f(z)) * (h(z), l(z)) = (g(z)h(f(z)), l(f(z))).$$

We recall that the set of all Riordan matrices is a group under the operator “ $*$ ” [18]. The identity element is $I = (1, z)$, and the inverse of $(g(z), f(z))$ is

$$(6) \quad (g(z), f(z))^{-1} = (1/(g \circ \bar{f})(z), \bar{f}(z)),$$

where $\bar{f}(z)$ is the compositional inverse of $f(z)$. Rogers [17] introduced the concept of the A -sequence. Specifically, Rogers observed that every element $d_{n+1,k+1}$ of a Riordan matrix (not belonging to row 0 or column 0) can be expressed as a linear combination of the elements in the preceding row. In particular, an infinite lower triangular array $\mathcal{D} = (d_{n,k})_{n,k \geq 0}$ is a Riordan array if and only if there is a sequence $(a_0 \neq 0, a_1, a_2, \dots)$ such that for every non-negative integers n, k

$$d_{n+1,k+1} = a_0 d_{n,k} + a_1 d_{n,k+1} + a_2 d_{n,k+2} + \dots,$$

The sequence $(a_n)_{n \geq 0}$ is called A -sequence. Moreover, if $\mathcal{D} = (g(x), f(x))$, then the generating function $A(x)$ of the A -sequence of $\mathcal{D}$ satisfies $f(x) = xA(f(x))$, see [13].

Let $c(n, i)$ denote the number of Catalan words of length n that end with letter i . In Theorem 2.2 we show that the infinite matrix $[c(n, i)]_{n,i \geq 1}$ is a Riordan array.

Theorem 2.2. *The infinite triangular matrix $\mathcal{T} = [c(n+1, k+1)]_{n,k \geq 0}$ has a Riordan array expression given by*

$$\mathcal{T} = (C(x), xC(x)),$$

where $C(x) = \frac{1-\sqrt{1-4x}}{2x}$ is the generating function of the Catalan numbers. Moreover,

$$c(n, k) = \frac{k}{n} \binom{2n-1-k}{n-1}.$$

Proof. From the definition we have that

$$c(n, 1) = \sum_{i=1}^{n-1} c(n-i, 1)c(i, 1), \quad n \geq 2,$$

with $c(1, 1) = 1$. Therefore, $c(n+1, 1) = C_n$ and $\sum_{n \geq 0} c(n+1, 1)x^n = C(x)$. Given any word w in $\mathcal{C}_{n+1,k}$ for $k \geq 2$, it can be decomposed as $w = w'k$, where $w' \in \mathcal{C}_{n,i}$, for $k-1 \leq i$. Then we have the relation

$$c(n+1, k) = c(n, k-1) + c(n, k) + c(n, k+1) + \dots + c(n, n).$$

Then $\mathcal{T} = (g, f)$ is a Riordan array such that $g(x) = C(x)$, whose generating function of the A -sequence is $A(x) = 1/(1-x)$. Therefore $f(x) = xA(f(x)) = \frac{x}{1-f(x)}$, so $f(x)^2 - f(x) + x = 0$ and $f(x) = xC(x)$. Finally, from the definition of Riordan array and the combinatorial identity (see [12, Ec. (5.70)])

$$C(x)^r = \sum_{k \geq 0} \binom{2k+r}{k} \frac{r}{2k+r} x^k,$$

we have that

$$c(n+1, k+1) = [x^n] x^k C(x)^{k+1} = [x^{n-k}] \sum_{\ell \geq 0} \binom{2\ell+k+1}{\ell} \frac{k+1}{2\ell+k+1} x^\ell.$$

By comparing the coefficients we obtain the combinatorial expression for the sequence $c(n, k)$. □

From the summation property of the Riordan arrays we get again the bivariate generating function $A(x; 1, 1, v)$. Indeed,

$$(C(x), xC(x)) \left(\frac{x}{1-vx} \right) = \frac{xC(x)}{1-vxC(x)} = x \frac{C(x)-v}{1-v+ xv^2}.$$

3. THE SEMIPERIMETER STATISTIC

In this section we study the semiperimeter statistic. In particular we give a connection with the Narayana sequence. By (3) with $q = 1$, we obtain

$$(7) \quad \left(1 - p^2 xv + \frac{pxv}{1-v} \right) A(x; p, 1; v) = p^2 x + \left(px + \frac{pxv}{1-v} \right) A(x; p, 1; 1).$$

By taking $v = v_0$ such that $1 - p^2 xv_0 + \frac{pxv_0}{1-v_0} = 0$, namely

$$v_0 = \frac{1 + p^2 x + px - \sqrt{p^4 x^2 - 2p^3 x^2 + p^2 x^2 - 2p^2 x - 2px + 1}}{2p^2 x},$$

we obtain $A(x; p, 1; 1) = p(v_0 - 1)$. By definition of Narayana number (see sequence A001263 in [19]), we can state the following result.

Theorem 3.1. *The generating function for the number of nonempty Catalan words according to the length and the semiperimeter is given by*

$$\begin{aligned} N(px, p) &= \frac{1 - p(1+p)x - \sqrt{1 - 2p(1+p)x - p^2(1-p)^2 x^2}}{2px} \\ &= p^2 x + p^3(1+p)x + p^4(1+3p+p^2)x^2 + \dots, \end{aligned}$$

where

$$N(x, t) = \frac{1 - (1+t)x - \sqrt{1 - 2(1+t)x + ((t-1)x)^2}}{2x}$$

is the generating function for the Narayana numbers. Moreover, the number of Catalan words of length $n \geq 1$ with perimeter $k = n + 1, n + 2, \dots, 2n$ is given by

$$(8) \quad \frac{1}{k-n} \binom{n}{k-n-1} \binom{n-1}{k-n-1}.$$

Theorem 3.2. *There is a bijection between the Catalan words in $\mathcal{C}_n$ of semiperimeter $2n+1-k$ for $k = 1, 2, \dots, n$ and the Dyck paths of length $2n$ with exactly k peaks.*

Proof. Let $w = w_1 \cdots w_n$ be any Catalan word of semiperimeter $2n+1-k$ for $k = 1, 2, \dots, n$. Start reading the reverse word $w^R = w_n w_{n-1} \cdots w_1$. If you read the symbol w_i , then you make up steps until height $y = i$ and one down step. Notice that we obtain a Dyck path of length $2n$ with exactly k peaks. This algorithm is a bijection. $\square$

The distribution of the Dyck path of length $2n$ and k peaks is given by the Narayana numbers. So, from the above bijection we can also obtain (8). In Figure 4 we show the Catalan words in $\mathcal{C}_3$ and their corresponding Dyck paths under this bijection. Notice that Deutsch and Elizalde [10] give a bijection between the bargraphs and Dyck paths.

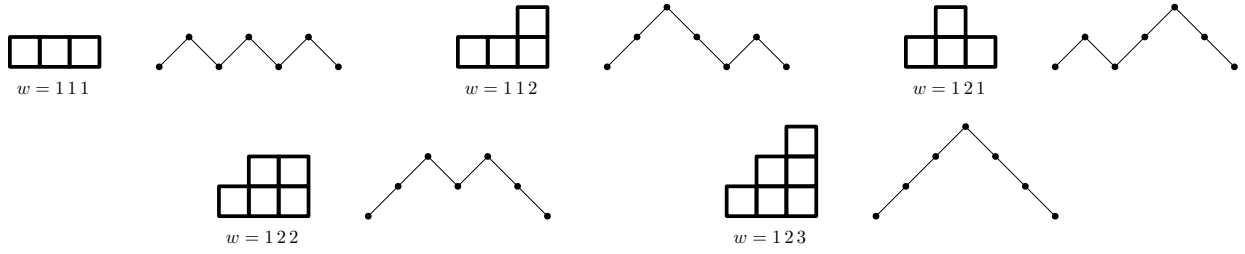


FIGURE 4. Dyck paths and their associated Catalan words and bargraphs.

From the same bijection given in Theorem 3.2 we obtain the following corollary.

Corollary 3.3. *There is a bijection between the Catalan words of length n that end with letter 1 of semiperimeter $2n - k$ for $k = 1, 2, \dots, n - 1$ and the Dyck paths of length $2(n - 1)$ with exactly k peaks.*

The following theorem uses two bijections on Dyck paths denoted DUtoDXD and DeutschInvol in [5]. We recall that DUtoDXD sends DU 's to DXD 's, where a DXD is a DUD or DDD , and DeutschInvol interchanges DU 's and UU 's (thereby explaining the symmetry $N(n, k) = N(n, n + 1 - k)$ of the Narayana numbers). See [9] for other results on corners in parallelogram polyominoes.

Theorem 3.4. *There is a bijection between the Catalan words in $\mathcal{C}_{n+1}$ that end with letter 2 and have semiperimeter $n + 3 + k$ and the parallelogram polyominoes of semiperimeter $n + 1$ having k corners for $n \geq 2$ and $k = 0, 1, \dots, n - 2$.*

For example, for $n = 4$ and $k = 1$ there are 8 parallelogram polyominoes of semiperimeter $n + 1 = 5$ with 1 corner, see Figure 5.

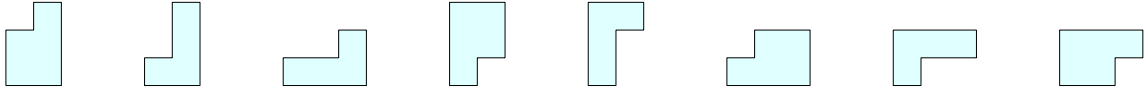


FIGURE 5. Parallelogram polyominoes of semiperimeter 5 with 1 corner.

That is the same as the number of Catalan words of length $n + 1 = 5$ that end with letter 2 and have semiperimeter $n + k + 3 = 8$, see Figure 6.

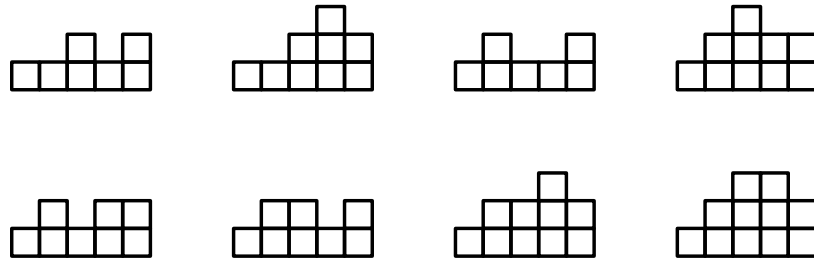


FIGURE 6. Catalan words of length 5 that end with letter 2 and have semiperimeter 8.

Proof of Theorem 3.4. The parallelogram polyominoes of semiperimeter $n + 1$ with k corners correspond to Dyck n -paths with k long interior inclines (see [5] for definitions of these terms). This is because corners on the upper boundary correspond to long ascents, and on the lower boundary to long descents, under the standard bijection from parallelogram polyominoes (or equivalently, Levine pairs) to Dyck paths. This standard bijection is used again below and is given, for example, in Sec. 5 of [5], where a Levine (r, s) -pair is the ordered pair of nonintersecting paths obtained from a parallelogram polyomino that ends at (r, s) by deleting the first and last steps from each of the upper and lower boundary.

For a Dyck n -path P , the semiperimeter of the corresponding Catalan bargraph is $n + 1 + \#UU$ s in P , and for the path P , $\#UU$'s + $\#UD$'s = n and of course, a UD is a peak. Hence the Catalan words in $\mathcal{C}_{n+1}$ that end with letter 2 and have semiperimeter $n + 3 + k$ correspond to Dyck $(n + 1)$ -paths that end UDD and have $n - k$ peaks. Partition these paths into two subsets: $\mathcal{A}_0$, those with a short (i.e. of length 1) last ascent, and $\mathcal{B}_0$, those with a long last ascent. Thus paths in $\mathcal{B}_0$ necessarily end $UUDD$. We have the sequence of bijections

$$\begin{aligned} \mathcal{A}_0 &\xrightarrow{\text{delete last peak}} \text{Dyck } n\text{-paths with } n - k - 1 \text{ peaks and long last descent} \\ &\xrightarrow{\text{ReversePath} \circ \text{DeutschInvol}} \text{Dyck } n\text{-paths with } k + 2 \text{ peaks and short last ascent,} \end{aligned}$$

and the latter Dyck paths correspond to the set $\mathcal{A}$ in Sec. 6 in [5], namely, Levine- $(k + 2, n - 1 - k)$ pairs in which the top path ends E .

Similarly, for $\mathcal{B}_0$, we have

$$\begin{aligned} \mathcal{B}_0 &\xrightarrow{\text{delete last 4 steps}} \text{Dyck } (n - 1)\text{-paths with } n - k - 1 \text{ peaks} \\ &\xrightarrow{\text{DeutschInvol}} \text{Dyck } (n - 1)\text{-paths with } k + 1 \text{ peaks} \\ &\xrightarrow[\text{at last peak}]{\text{insert } UD} \text{Dyck } n\text{-paths with } k + 1 \text{ peaks and long last ascent,} \end{aligned}$$

and the latter Dyck paths correspond to the set $\mathcal{B}$ in Sec. 6 in [5], namely, Levine- $(k + 1, n - k)$ pairs in which the top path ends N .

Now, Sec. 6 in [5] presents a bijection from $\mathcal{A}$ (resp. $\mathcal{B}$) to Dyck n -paths containing k long interior inclines and ending DD (resp. UD). In view of the opening remark above, this completes the proof. $\square$

Let $s(n, i)$ denote the semiperimeter of the Catalan words of length n that end with letter i . The first few values are

$$[s(n, i)]_{n, i \geq 1} = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 & 0 \\ 9 & 10 & 6 & 0 & 0 & 0 \\ 30 & 32 & 21 & 8 & 0 & 0 \\ 105 & 110 & 75 & 36 & 10 & 0 \\ 378 & 392 & 273 & 144 & 55 & 12 \\ \vdots & & & & & \end{pmatrix}$$

In Theorem 3.5 we give a recurrence relation for the sequence $s(n, i)$.

Theorem 3.5. For $n \geq 2$ and $2 \leq i \leq n$

$$s(n, i) = \frac{n(3i-1) - i(i+1)}{(2n-i)(2n-i-1)} \binom{2n-i}{n} + \sum_{j=i-1}^{n-1} s(n-1, j),$$

with $s(n, 1) = \sum_{j=1}^{n-1} s(n-1, j) + C_{n-1}$ for $n \geq 2$ and $s(1, 1) = 2$, where C_n is a Catalan number.

Proof. From decomposition given in Figure 3 we have for $n \geq 2$ and $2 \leq i \leq n$,

$$s(n, i) = s(n-1, i-1) + 2c(n-1, i-1) + \sum_{j=i}^{n-1} (s(n-1, j) + c(n-1, j)).$$

From the closed formula given in Theorem 2.2 we have

$$\begin{aligned} s(n, i) &= s(n-1, i-1) + 2 \frac{i-1}{n-1} \binom{2n-2-i}{n-2} + \sum_{j=i}^{n-1} \left(s(n-1, j) + \frac{j}{n-1} \binom{2n-3-j}{n-2} \right) \\ &= 2 \frac{i-1}{n-1} \binom{2n-2-i}{n-2} + \frac{i+1}{n} \binom{2n-2-i}{n-1-i} + \sum_{j=i-1}^{n-1} s(n-1, j) \\ &= \frac{n(3i-1) - i(i+1)}{(2n-i)(2n-i-1)} \binom{2n-i}{n} + \sum_{j=i-1}^{n-1} s(n-1, j). \end{aligned}$$

Note that we use the combinatorial identity

$$(9) \quad \sum_{j=i}^n c(n, j) = \sum_{j=i}^n \frac{j}{n} \binom{2n-1-j}{n-1} = \frac{i+1}{n+1} \binom{2n-i}{n-i}.$$

The initial values can be obtained in a similar way. □

If we consider the difference $s(n, i) - s(n, i-1)$, then for $n \geq 2$ and $3 \leq i \leq n$,

$$s(n, i) - s(n, i-1) = \frac{n(7-3i) + i^2 - 5}{(2n-i)(2n-i-1)} \binom{2n-i}{n-1} + s(n, i-1) - s(n-1, i-2).$$

Let $s(n)$ be the total semiperimeter over all Catalan words of length n . The first few values are

2, 7, 25, 91, 336, 1254, 4719, 17875, 68068, 260338, ... (sequence A097613).

Notice that our combinatorial interpretation is new. In Corollary 3.6 we give an explicit relation for the sequence $s(n)$.

Corollary 3.6. The total semiperimeter over all Catalan words of length n is given by

$$s(n) = \frac{1}{2} \left(\binom{2n+2}{n+1} - \binom{2n}{n} \right).$$

Proof. By differentiating $A(x; p, q; 1)$ at $p = 1$, Theorem 3.1 gives

$$A' = \frac{\partial}{\partial p} A(x; p, 1; 1) \big|_{p=1} = \frac{1-x}{2x\sqrt{1-4x}} - \frac{1}{2} - \frac{1}{x}.$$

By using the fact that $\sum_{n \geq 0} \binom{2n}{n} x^n = \frac{1}{\sqrt{1-4x}}$, we obtain that the coefficient of x^n in A' , namely the total semiperimeter over all Catalan words of length n , is given by

$$\frac{1}{2} \left(\binom{2n+2}{n+1} - \binom{2n}{n} \right),$$

as required. $\square$

4. THE AREA STATISTIC

The goal of this section is analyze the area statistic. In particular, we give a continued fraction expression for the generating function. By (3) with $p = 1$ we obtain

$$A(x; 1, q; v) = qx + \frac{qx}{1 - qv} A(x; 1, q; 1) - \frac{q^3 x v^2}{1 - qv} A(x; 1, q; qv).$$

By iterating this equation an infinite number of times (here we assume $|x| < 1$ or $|q| < 1$), we obtain

$$A(x; 1, q; v) = \sum_{j \geq 1} (-1)^{j-1} q^{j^2} x^j v^{2j-2} (1 - q^j v + A(x; 1, q; 1)) \prod_{i=1}^j \frac{1}{1 - q^i v}.$$

By setting $v = 1$, and solving for $A(x; 1, q; 1)$, we can state the following result.

Theorem 4.1. *The generating function for the number of nonempty Catalan words according to their length and area is given by*

$$\begin{aligned} A(x; 1, q; 1) &= \frac{\sum_{j \geq 1} (-1)^{j-1} \frac{q^{j^2} x^j}{\prod_{i=1}^{j-1} (1 - q^i)}}{1 - \sum_{j \geq 1} (-1)^{j-1} \frac{q^{j^2} x^j}{\prod_{i=1}^j (1 - q^i)}} \\ &= qx + (q+1)q^2 x^2 + q^3(q^3 + q^2 + 2q + 1)x^3 + (q+1)(q^5 + 2q^3 + q^2 + 2q + 1)q^4 x^4 + \dots \end{aligned}$$

We can also obtain the above result using automata theory. The problem of calculating the area of a given Catalan word is equivalent to calculating the sum of the labels over the rise steps in its associated Dyck path (see Figure 1). To this problem we can associate the transition diagram of a finite automaton as shown in Figure 7, whose states are represented by circles and final states by double circles, and the initial state is labelled by an arrow entering that state and each transition between states is represented by a labelled arrow. For our case, the label $q^\ell x$ from the node $q_{\ell-1}$ to the node q_ℓ represents one rise step in the Dyck path labelled by ℓ .

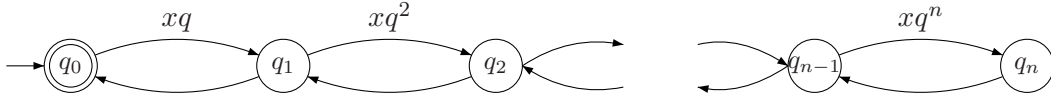


FIGURE 7. Counting automaton associated to the Dyck path.

The technique to find the ordinary generating function corresponding to an automaton (or in general for a context-free language) is known as the *Chomsky-Schützenberger methodology*, (see for example [8, 7]). The idea behind this procedure is to obtain a system of equations using a deterministic finite automaton, where the unknowns are the ordinary generating functions

associated with each state of the automaton. The finite automaton in Figure 7 gives rise to a system of equations for the associated multivariate generating functions:

$$\begin{cases} L_0(x, q) &= xqL_0(x, q) + 1, \\ L_1(x, q) &= xq^2L_2(x, q) + L_0(x, q), \\ &\vdots \\ L_{n-1}(x, q) &= xq^nL_n(x, q) + L_{n-2}(x, q), \\ L_n(x, q) &= L_{n-1}(x, q). \end{cases}$$

Solving this system, we obtain the generating function associated to the automaton. Note that the initial state of the automaton is q_0 , so

$$L_0(x, z) = \frac{1}{1 - \frac{qx}{1 - \frac{q^2x}{1 - \frac{\vdots}{1 - \frac{q^{n-1}}{1 - q^nx}}}}} \Bigg\} n \text{ times}$$

Since $A(x; 1, q; 1)$ is the limit as n approaches to infinity, we obtain the following theorem.

Theorem 4.2. *The generating function for the number of Catalan words, including the empty word, according to the length and area is given by the infinite continued fraction*

$$1 + A(x; 1, q; 1) = \frac{1}{1 - \frac{qx}{1 - \frac{q^2x}{\ddots}}}.$$

The above continued fraction expression was given by Carlitz [6] for the ordinary generating function $\sum_{n \geq 0} \sum_{D \in \text{Dyck}(n)} q^{\text{area}(D)+n}$, where $\text{Dyck}(n)$ is the set of the Dyck paths of length $2n$ (see also [11]).

Let $t(n, i)$ denote the area of the Catalan words of length n that end with letter i . The first few values are

$$[t(n, i)]_{n, i \geq 1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 & 0 & 0 \\ 7 & 9 & 6 & 0 & 0 & 0 \\ 27 & 32 & 24 & 10 & 0 & 0 \\ 107 & 121 & 93 & 50 & 15 & 0 \\ 428 & 470 & 363 & 214 & 90 & 21 \\ \vdots & & & & & \end{pmatrix}$$

From a similar argument as in Theorem 3.5 we obtain the following result.

Theorem 4.3. *For $n \geq 2$ and $2 \leq i \leq n$*

$$t(n, i) = \frac{i^2}{2n-i} \binom{2n-i}{n} + \sum_{j=i-1}^{n-1} t(n-1, j),$$

with $t(n, 1) = \sum_{j=1}^{n-1} t(n-1, j) + C_{n-1}$ for $n \geq 2$ and $c(1, 1) = 1$, where C_n is a Catalan number.

Let $t(n)$ be the total area of over all Catalan words of length n . The first few values are

1, 5, 22, 93, 386, 1586, 6476, 26333, 106762, 431910, ... (sequence A000346).

Corollary 4.4. *The total area over all Catalan words of length n is given by*

$$t(n) = \frac{1}{2} \left(4^n - \binom{2n}{n} \right).$$

Proof. Let $B(x; v) = \frac{\partial}{\partial q} A(x; 1, q; v) |_{q=1}$. Then by differentiating (3) respect to q we obtain,

$$\begin{aligned} B(x; v) &= x + \frac{x}{(1-v)^2} A(x; 1, 1; 1) + \frac{x}{1-v} B(x; 1) \\ &\quad - \frac{xv^2(3-2v)}{(1-v)^2} A(x; 1, 1; v) - \frac{xv^2}{1-v} (B(x; v) + \frac{\partial}{\partial v} A(x; 1, 1; v)). \end{aligned}$$

By using (4), namely $A(x; 1, 1; v) = x \frac{C(x)-v}{1-v+ xv^2}$, we obtain

$$\frac{(1-v+xv^2)^3}{1-v} B(x; v) = \frac{x(1-v+xv^2)^2}{1-v} B(x; 1) + x(1-xv^2)C(x) + xv(v-2).$$

By differentiating twice this equation respect to v , and then taking $v = C(x)$, we obtain that $B(x, 1) = \frac{1-\sqrt{1-4x}}{2(1-4x)}$, which, by the fact that $\sum_{n \geq 0} \binom{2n}{n} x^n = \frac{1}{\sqrt{1-4x}}$, leads to the desired result. $\square$

5. THE EXTERIOR CORNER STATISTIC

An *exterior corner* in a bargraph is a point of intersection of a horizontal (H) step with an up (U) step or the intersection of a down (D) step with a horizontal (H) step. Let $\text{cor}(w)$ denote the number of *exterior corners* associated to the bargraph of the Catalan word w . Let $\text{cor}_{hu}(w)$ and $\text{cor}_{dh}(w)$ denote the number of corners of type hu and dh , respectively. For example, for the bargraph in Figure 8 we denote by red (resp. blue) the corners of type hu (resp. dh). Moreover, $\text{cor}(w) = 10$, $\text{cor}_{hu}(w) = 6$, and $\text{cor}_{dh}(w) = 4$. This distribution has been studied for example for parallelogram polyominoes [9] and for the bargraphs associated to set partitions and compositions [15].

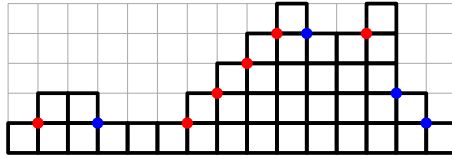


FIGURE 8. Exterior corners of a bargraph.

Define

$$B_i(x; p, q) := \sum_{n \geq 1} x^n \sum_{w \in \mathcal{C}_{n,i}} p^{\text{cor}_{hu}(w)} q^{\text{cor}_{dh}(w)}.$$

By definitions, we have

$$(10) \quad B_1(x; p, q) = x + xB_1(x; p, q) + qx \sum_{j \geq 2} B_j(x; p, q),$$

$$(11) \quad B_i(x; p, q) = px B_{i-1}(x; p, q) + x B_i(x; p, q) + xq \sum_{j \geq i+1} B_j(x, t), \quad i \geq 2.$$

Define $B(x; p, q; v) = \sum_{i \geq 1} B_i(x; p, q) v^{i-1}$. Then by multiplying (11) by v^{i-1} and summing over $i \geq 2$ with using (10), we obtain

$$\begin{aligned} B_1(x; p, q) &= x + xB_1(x; p, q) + qx (B(x; p, q; 1) - B_1(x; p, q)), \\ B(x; p, q; v) - B_1(x; p, q) &= pxv B(x; p, q; v) + x (B(x; p, q; v) - B_1(x; p, q)) \\ &\quad + \frac{xqv}{1-v} (B(x; p, q; 1) - B_1(x; p, q)) - \frac{xq}{1-v} (B(x; p, q; v) - B_1(x; p, q)), \end{aligned}$$

which leads to

$$(12) \quad B(x; p, q; v) = x + \left(pxv + x - \frac{xq}{1-v} \right) B(x; p, q; v) + \frac{qx}{1-v} B(x; p, q; 1).$$

If $p = q$ in (12) we obtain

$$(13) \quad B(x; q; v) := B(x; q, q; v) = x + \left(qxv + x - \frac{xq}{1-v} \right) B(x; q; v) + \frac{qx}{1-v} B(x; q; 1).$$

Then

$$\left(1 - x - qxv + \frac{xq}{1-v} \right) B(x; q; v) = x + \frac{qx}{1-v} B(x; q; 1).$$

By taking $v = v_0$ such that $1 - x - qxv_0 + \frac{xq}{1-v_0} = 0$, namely

$$v_0 = \frac{1 - x + qx - \sqrt{1 - 2x(1+q) + (1 + 2q - 3q^2)x^2}}{2qx},$$

we obtain $B(x; q; 1) = (v_0 - 1)/q$.

Theorem 5.1. *The generating function for the number of nonempty Catalan words according to the length and the exterior corners is given by*

$$\begin{aligned} B(x; q; 1) &= \frac{1 - (1+q)x - \sqrt{1 - 2(1+q)x + (1 + 2q - 3q^2)x^2}}{2q^2x} \\ &= x + (1+q)x^2 + (1 + 2q + 2q^2)x^3 + (1 + 3q + 6q^2 + 4q^3)x^4 + \dots, \end{aligned}$$

Let $w(n, i)$ denote the number of exterior corners of the Catalan words of length n that end with letter i . The first few values are

$$[w(n, i)]_{n, i \geq 1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 & 0 \\ 9 & 9 & 6 & 3 & 0 & 0 \\ 36 & 36 & 24 & 12 & 4 & 0 \\ 140 & 140 & 95 & 50 & 20 & 5 \\ & & \vdots & & & \end{pmatrix}$$

In Theorem 5.2 we give a recurrence relation for the sequence $w(n, i)$.

Theorem 5.2. For $n \geq 2$ and $2 \leq i \leq n$

$$w(n, i) = \frac{i-1}{n-1} \binom{2n-2-i}{n-2} + \frac{i+2}{n} \binom{2n-3-i}{n-1} + \sum_{j=i-1}^{n-1} w(n-1, j),$$

with $w(n, 1) = \sum_{j=1}^{n-1} w(n-1, j) + 3 \frac{(2n-4)!}{n!(n-3)!}$ for $n \geq 2$ and $w(1, 1) = 0$.

Proof. From decomposition given in Figure 3 we have for $n \geq 2$ and $2 \leq i \leq n$,

$$w(n, i) = w(n-1, i-1) + c(n-1, i-1) + w(n-1, i) + \sum_{j=i+1}^{n-1} (w(n-1, j) + c(n-1, j)).$$

From the closed formula given in Theorem 2.2 and the combinatorial identity (9) we obtain

$$\begin{aligned} w(n, i) &= w(n-1, i-1) + \frac{i-1}{n-1} \binom{2n-2-i}{n-2} + w(n-1, i) + \sum_{j=i+1}^{n-1} (w(n-1, j) + c(n-1, j)) \\ &= \frac{i-1}{n-1} \binom{2n-2-i}{n-2} + \frac{i+2}{n} \binom{2n-i-3}{n-i-2} + w(n-1, i-1) + \sum_{j=i-1}^{n-1} w(n-1, j). \end{aligned}$$

The initial values can be obtained in a similar way. $\square$

Let $w(n)$ be the total number of exterior corners over all Catalan words of length n . The first few values are

$$0, 1, 6, 27, 112, 450, 1782, 7007, 27456, 107406, \dots \quad (\text{sequence A220101}).$$

Corollary 5.3. The total number of exterior corners over all Catalan words of length n is given by

$$w(n) = \frac{3(n-1)}{2n-1} \binom{2n-1}{n-2}.$$

Let $\overline{w}(n)$ denote the total number of exterior corners, including the origin and termination point of the bargraph, over all Catalan words of length n . It is clear that $\overline{w}(n) = w(n) + C_n$, where C_n is the n -th Catalan numbers. From Corollary 5.3 we have

$$\overline{w}(n) = (3n-2)C_{n-1}.$$

The first few values of the sequence $\overline{w}(n)$ are

$$2, 5, 16, 55, 196, 714, 2640, 9867, 37180, 140998, \dots \quad (\text{sequence A051960}).$$

Finally, if $p = 1$ in (12) we obtain

$$(14) \quad \left(1 - x - xv + \frac{xq}{1-v}\right) B(x; 1, q; v) = x + \frac{qx}{1-v} B(x; 1, q; 1).$$

By taking $v = v_0$ such that $1 - x - xv_0 + \frac{xq}{1-v_0} = 0$, namely

$$v_0 = \frac{1 - \sqrt{1 - 4x - (4q-4)x^2}}{2x},$$

we have $B(x; 1, q; 1) = (v_0 - 1)/q$.

Theorem 5.4. *The generating function for the number of nonempty Catalan words according to the length and the exterior corners of type dh is given by*

$$\begin{aligned} B(x; 1, q; 1) &= \frac{1 - 2x - \sqrt{1 - 4(1 - (1 - q)x)x}}{2qx} \\ &= x + 2x^2 + (q + 4)x^3 + 2(3q + 4)x^4 + 2(q^2 + 12q + 8)x^5 + \cdots \end{aligned}$$

Let $w_{dh}(n)$ be the total number of exterior corners of type dh over all Catalan words of length n . The first few values are

$$0, 0, 1, 6, 28, 120, 495, 2002, 8008, 31824, \dots \quad (\text{sequence A002694}).$$

Corollary 5.5. *The total number of exterior corners of type dh over all Catalan words of length n is given by*

$$w_{dh}(n) = \binom{2n - 2}{n - 3}.$$

If $q = 1$ in (12) we can obtain similar results. Notice that we get again the Narayana distribution.

Theorem 5.6. *The generating function for the number of nonempty Catalan words according to the length and the exterior corners of type hu is given by*

$$\begin{aligned} B(x; p, 1; 1) &= -\frac{1 - (1 + p)x - \sqrt{(1 + (p - 1)x)^2 - 4px}}{2px} \\ &= x + (p + 1)x^2 + (p^2 + 3p + 1)x^3 + (p^3 + 6p^2 + 6p + 1)x^4 + \cdots \end{aligned}$$

Let $w_{hu}(n)$ be the total number of exterior corners of type hu over all Catalan words of length n . The first few values are

$$0, 1, 5, 21, 84, 330, 1287, 5005, 19448, 75582, \dots \quad (\text{sequence A002054}).$$

Corollary 5.7. *The total number of exterior corners of type hu over all Catalan words of length n is given by*

$$w_{hu}(n) = \binom{2n - 1}{n - 2}.$$

6. ACKNOWLEDGEMENT

The second author was partially supported by Universidad Nacional de Colombia, Project No. 46240.

REFERENCES

- [1] 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.
- [2] A. Blecher, C. Brennan, and A. Knopfmacher. Levels in bargraphs. *Ars Math. Contemp.* **9** (2015), 287–300.
- [3] A. Blecher, C. Brennan, and A. Knopfmacher. Peaks in bargraphs. *Trans. Royal Soc. S. Afr.* **71** (2016), 97–103.
- [4] A. Blecher, C. Brennan, and A. Knopfmacher. Combinatorial parameters in bargraphs. *Quaest. Math.* **39** (2016), 619–635.
- [5] D. Callan, Some identities for the Catalan and Fine numbers, preprint, 17 pages, (2005), available at <https://arxiv.org/abs/math/0502532>.
- [6] L. Carlitz. Problem: q -analog of the Lagrange expansion. In *Abstracts and Problems from the Conference on Eulerian Series and Applications*. Pennsylvania State University, 1974.

- [7] R. De Castro, A. Ramírez, and J. L. Ramírez. Applications in enumerative combinatorics of infinite weighted automata and graphs. *Sci. Ann. Comput. Sci.* **24**(1) (2014), 137–171.
- [8] M. Delest. Algebraic languages: a bridge between combinatorics and computer science. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **24** (1996), 71–88.
- [9] M. Delest, J. P. Dubernard, and I. Dutour. Parallelogram polyominoes and corners. *J. Symbolic Computation* **20** (1995), 503–515.
- [10] E. Deutsch and S. Elizalde. A bijection between bargraphs and Dyck paths. *Discrete Appl. Math.* **251** (2018), 340–344.
- [11] P. Flajolet. Combinatorial aspects of continued fractions. *Discrete Math.* **32** (1980), 125–161.
- [12] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, 1994.
- [13] T.-X. He and R. Sprugnoli. Sequence characterization of Riordan arrays. *Discrete Math.* **309** (2009), 3962–3974.
- [14] T. Mansour and A. Sh. Shabani. Enumerations on bargraphs. *Discrete Math. Lett.* **2** (2019), 65–94.
- [15] T. Mansour, A. Sh. Shabani, and M. Shattuck. Counting corners in compositions and set partitions presented as bargraphs. *J. Difference Equ. Appl.* **24** (2018), 992–1015.
- [16] T. Mansour and V. Vajnovszki. Efficient generation of restricted growth words. *Inform. Process. Lett.* **113** (2013), 613–616.
- [17] D. G. Rogers. Pascal triangles, Catalan numbers and renewal arrays. *Discrete Math.* **22** (1978), 301–310.
- [18] L. W. Shapiro, S. Getu, W. Woan, and L. Woodson. The Riordan group. *Discrete Appl. Math.* **34** (1991), 229–239.
- [19] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences, <http://oeis.org/>.
- [20] R. Stanley. *Catalan numbers*. Cambridge University Press, Cambridge, 2015.

DEPARTMENT OF STATISTICS, UNIVERSITY OF WISCONSIN-MADISON, MADISON, WI
E-mail address: callan@stat.wisc.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HAIFA, 3498838 HAIFA, ISRAEL
E-mail address: tmansour@univ.haifa.ac.il

DEPARTAMENTO DE MATEMÁTICAS, UNIVERSIDAD NACIONAL DE COLOMBIA, BOGOTÁ, COLOMBIA
E-mail address: jlramirezr@unal.edu.co