INFORMATIQUE THÉORIQUE ET APPLICATIONS

D. GARDY
D. GOUYOU-BEAUCHAMPS

Enumerating Davenport-Schinzel sequences

Informatique théorique et applications, tome 26, nº 5 (1992), p. 387-402.
<http://www.numdam.org/item?id=ITA_1992__26_5_387_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 26, n° 5, 1992, p. 387 à 402)

ENUMERATING DAVENPORT-SCHINZEL SEQUENCES (*)

by D. GARDY [1] and D. GOUYOU-BEAUCHAMPS [1]

Communicated by J. BERSTEL

Abstract. – *Davenport-Schinzel sequences of order s are words with no subsequence* ababa... *of length* $s+2$. *We give enumerating results for the case* $s=2$. *In particular, we relate some of these sequences to Catalan and Schröder numbers.*

Résumé. – *Nous étudions les suites de Davenport-Schinzel d'ordre* 2; *ce sont des mots sans sous-suite* abab. *Nous obtenons la fonction génératrice de dénombrement de ces suites, suivant la longueur d'un mot et le nombre de lettres distinctes qu'il contient. En particulier, certaines de ces suites sont énumérées par les nombres de Catalan ou de Schröder.*

1. INTRODUCTION

Davenport-Schinzel sequences are words with forbidden subsequences, which were first defined by Davenport and Schinzel [6] in connection with the general solution of a (homogeneous) linear differential equation with constant coefficients, of order $s+1$:

$$F(D) f(x)=0. \tag{1}$$

Here D denotes the derivative operator, and F is a polynomial of degree $s+1$. If we suppose that $F(D)$ has real coefficients, and that the roots of $F(\lambda)=0$ are all real, not necessarily distinct, the form of any solution is:

$$f(x)=P_1(x)\, e^{\lambda_1 x}+\ldots+P_k(x)\, e^{\lambda_k x}. \tag{2}$$

(*) Received April 1991, revised September 1991.
This work was partially supported by PRC Mathématique-Informatique and by Esprit Basic Research Action No. 3075 (project ALCOM).

[1] Laboratoire de Recherche en Informatique, CNRS UA 410, Université de Paris-Sud, 91405 Orsay (France).

Informatique théorique et Applications/Theoretical Informatics and Applications
0988-3754/92/05 387 16/$3.60/© AFCET-Gauthier-Villars

The $\lambda_1, \ldots, \lambda_k$ are the distinct roots of $F(\lambda)=0$, of respective multiplicities $m_1, \ldots, m_k$, so that $m_1+\ldots+m_k=s+1$; $P_1(x), \ldots, P_k(x)$ are polynomials of degrees at most $m_1-1, \ldots, m_k-1$. Let

$$f_1(x), \ldots, f_n(x) \tag{3}$$

be n distinct (but not necessarily independent) solutions of (1). For each real number x, apart from a finite number of exceptions, there will be just one of the functions (3) which is greater than all the others. So we can dissect the real line into N intervals:

$$(-\infty, x_1), (x_1, x_2), \ldots, (x_{N-1}, \infty) \tag{4}$$

such that inside any one of the intervals (x_{j-1}, x_j) a particular one of the functions (3) is the largest, and that this largest function is not the same for two consecutive intervals.

The problem is to find how large N can be, for given s and given n. If we remark that any function of type (2) has at most s distinct zeros, the problem can be transformed into the following purely combinatorial problem: Find the maximal length of a word on the alphabet $A=\{a_1, \ldots, a_n\}$ with no immediate repetition and which contains no subword (1) of the form:

$$abab\ldots \quad \text{with} \quad s+2 \text{ letters and } a \neq b. \tag{5}$$

This number is usually denoted by $\lambda_s(n)$. Much work has been devoted to $\lambda_s(n)$; we recall the main results below:

- **s=1:** $\lambda_1(n)=n$ (Davenport and Schinzel [6]).
- **s=2:** $\lambda_2(n)=2n-1$ (Davenport and Schinzel [6]).
- **s=3:** $\lambda_s(n)$ is of order $\Theta(n\alpha(n))$, with $\alpha(n)$ the functional inverse of the Ackermann function (Davenport and Schinzel [6], Davenport [5], Szemerédi [20], Hart and Sharir [10], Komjáth [12], Wiernik [21]). This is more than linear, but $\alpha(n)$ is less than 4 for all purposes.
- **s=4:** $\lambda_4(n)$ is of order $\Theta(n\,2^{\alpha(n)})$ (Szemerédi [20], Sharir [16], Agarwal [1], Agarwal, Sharir and Shor [2]). As in the case $s=3$, it is theoretically superlinear, but almost linear for all realistic values of n.

(1) We recall that the letters of the subword can be intertwined with other letters. For example, the word $w=abcadb$ contains the subword $abab$, although $abab$ is not a factor of w.

• **s≧5:** The exact order of $\lambda_s(n)$ is not yet known; Agarwal, Sharir and Shor [2]) have proved that

$$\lambda_{2s+2}(n) = O(n\, 2^{\alpha(n)^{s}(1+o(1))}) \qquad \text{and} \qquad \lambda_{2s+3}(n) = O(n\, 2^{\alpha(n)^{s}(1+o(1))\log \alpha(n)}).$$

Again this is slightly more than linear.

It is of little surprise that these sequences have found a new field of application in combinatorial and computational geometry. The most classical application provides a combinatorial characterization of the lower envelope of continuous functions, in the following manner. Let $f_1, \ldots, f_n$ be n continuous real functions, such that any two functions intersect in at most s points; the sequence of indices of functions which form the lower bound of the graphs is a Davenport-Schinzel sequence of order (n, s).

The papers of Sharir [17] and Sharir *et al.* [18] review recent results in computer graphics, motion planning and computational geometry, that are based on Davenport-Schinzel sequences. For example, here are some problems where Davenport-Schinzel sequences appear:

• preprocessing of 2-D polyhedral terrain so as to support fast ray shooting queries from a fixed point [4];

• determining whether two disjoint interlocking simple polygons can be separated from one another by a sequence of translations [15];

• determining whether a given convex polygon can be translated and rotated so as to fit into another given polygonal region [13];

• motion planning for a convex polygon in the plane amidst polygonal barriers [11];

• analysis of the combinatorial complexity of the lower envelope of a collection of bivariate piecewise linear functions, whose graphs consist of n faces altogether [14];

• finding the boundary of any region in a subdivision of the plane induced by a set of n rays [3].

For all these algorithms, the worst case complexity requires constructing extremal configurations of a lower envelope, that force the algorithms to perform a large number of operations. This gives a worst-case complexity expressed in terms of $\lambda_s(n)$, which is typically only slightly more than linear. Our main objective is to find the average-case complexity of these algorithms. This leads us to study parameters of Davenport-Schinzel sequences such as average length of words or average number of distinct letters.

This paper presents some results for these parameters in the special case $s=2$. It is organized as follows: We give in Section 2 a formal definition of Davenport-Schinzel sequences, and we compute the generating function enumerating sequences according to their length and to the number of distinct letters they contain. Consequences are given in Section 3; in particular we show that sequences on a given number of letters are enumerated by Schröder numbers, and that sequences of maximal length $2n-1$ are enumerated by Catalan numbers. Finally, we indicate in Section 4 why our method fails when s is at least equal to 3.

2. DECOMPOSITION OF A SEQUENCE

2.1. Definitions

Let s be a positive integer, and A an alphabet of size $n: A=\{a_1, \ldots, a_n\}$. A Davenport-Schinzel sequence of order (n, s) is classically defined as a word $w=u_1 \ldots u_k$ of A^+, such that:

1. any two consecutive letters are distinct;

2. w has no subsequence $u_{i_1} \ldots u_{i_{s+2}}$ $(1 \leqq i_1 < i_2 < \ldots < i_{s+2} \leqq k)$ on any two distinct letters a and b, satisfying:

$$u_{i_1}=u_{i_3}=\ldots=a \qquad \text{and} \qquad u_{i_2}=u_{i_4}=\ldots=b.$$

We first precise the objects we are interested in. In the lower envelope problem we alluded to in the introduction, the names of the functions (letters) are arbitrary: Our interest is in the Davenport-Schinzel sequence *up to a renaming of the letters*. This seems to hold for all topics where Davenport-Schinzel sequences appear: The object of interest is not a word, but its equivalence class when subjected to any renaming of the letters. This leads us to propose the following notation, which we shall use throughout the rest of the paper:

> *A Davenport-Schinzel* word *on an alphabet A is a word of A^+ satisfying the properties* 1 *and* 2 *above ("word" has the sense usual in formal language theory); a* sequence *is an equivalence class of words.*

For example, *ab* and *ba* are two words corresponding to the same sequence. We shall denote by $N_{n,k}$ the number of words of Davenport-Schinzel of length k on n letters, and by $a_{n,k}$ the number of sequences up to a renaming of the letters.

We define $DS(s)$ as the set of sequences (equivalence classes of words), for a fixed s and for all alphabets. Our goal is to find a suitable decomposition of the sequences of $DS(2)$, from which we can deduce an equation on their generating function. We shall see that, for $s=2$, this is an algebraic equation of degree two which can easily be solved. We first simplify the problem by introducing complete words and obtain a simple result on the generating function of complete sequences of $DS(1)$, which we shall need in the sequel.

2.2. Complete words

It turns out that the most natural words or sequences, from an enumeration point of view, are those in which all the letters of the alphabet appear; we call them *complete* words. The words on a given alphabet, including those where some letters are missing, can be enumerated easily once the generating function of the complete words is known. This leads to the following definition:

> *A* complete *word w on an alphabet A of size n is such that all letters of A appear in w.*

The length k of a complete word for $s=2$ on an alphabet of size n belongs to the set $\{n, \ldots, 2n-1\}$. Let $M_{n,k}$ be the number of such words and let $b_{n,k}=M_{n,k}/n!$ be the number of related sequences. The following relations hold:

$$N_{n,k}=\sum_{q=1}^{n}\binom{n}{q}M_{q,k}; \tag{6}$$

$$a_{n,k}=\sum_{q=1}^{n} b_{q,k}. \tag{7}$$

We define in a similar way complete sequences, as sequences in which all the letters of A appear at least once.

2.3. Generating functions: definitions

Let $\Psi(x, y)$ be the exponential generating function of the $N_{n,k}$, where variable x marks the length of a word, and variable y marks the size of the alphabet:

$$\Psi(x, y)=\sum_{k,\, n \geqq 1} N_{n,k}\, x^k \frac{y^n}{n!}.$$

Similarly, denote by $\Xi(x, y)$ the ordinary generating function of the $a_{n,k}$:

$$\Xi(x, y)=\sum_{k, n \geqq 1} a_{n, k} x^k y^n,$$

and by $\Phi(x, y)$ the generating function counting complete sequences:

$$\Phi(x, y)=\sum_{n \geqq 1} \sum_{k=n}^{2n-1} b_{n, k} x^k y^n=\sum_{n \geqq 1} \sum_{k=n}^{2n-1} M_{n, k} x^k \frac{y^n}{n!}.$$

The relation (6) between the numbers $N_{n,k}$ and $M_{q,k}$ allows us to express Φ in terms of the function Ψ:

$$\Psi(x, y)=e^y \Phi(x, y).$$

This can also be obtained directly, as e^y is the function enumerating the *sets* of letters which do not appear in a word.

Now the functions Φ and Ξ are themselves related and the relation (7) gives:

$$\Xi(x, y)=\frac{\Phi(x, y)}{1-y}.$$

Again this can be obtained directly: $1/(1-y)$ is the function enumerating the *sequences* of letters that have to be added to the letters of a complete sequence, to get the whole alphabet.

We sum up these results in the following theorem:

THEOREM 1: *The generating functions $\Xi(x, y)$ and $\Psi(x, y)$, enumerating respectively words and sequences according to their length and to the size of the alphabet can be expressed using the generating function $\Phi(x, y)$ enumerating complete words or sequences:*

$$\Xi(x, y)=\Phi(x, y)/(1-y);$$
$$\Psi(x, y)=\Phi(x, y) e^y.$$

2.4. Enumeration of complete sequences of $DS(1)$

Let $g_{n,k}$ be the number of sequences for $s=1$, of length k, and such that all the letters of alphabet A, of size n, appear in the sequence. From the definition of $DS(1)$, its sequences have no repeated letter, and there is basically just one such sequence. This shows that:

$$\text{if} \quad n \neq k, \qquad g_{n, k}=0;$$

$$\text{if} \quad n=k, \qquad g_{n,n}=1.$$

Define $g(x,y)=\sum_{k,\, n \geqq 1} g_{n,k} x^k y^n$; we have: $g(x,y)=\sum_{n \geqq 1} x^n y^n = xy/(1-xy)$.

2.5. Decomposition of a sequence of $DS(2)$

Let w be a sequence of $DS(2)$; w may actually be a sequence of $DS(1)$, with no repeated letter. If at least one letter is repeated, we decompose w unambiguously, according to the occurrences of the *first* such letter, *i.e.* the leftmost letter in w which will be repeated later on. Let us denote this letter, which is repeated $p \geqq 2$ times, by a. We have:

$$w = w_1 a w_2 \ldots w_p a w_{p+1}.$$

Each $w_i (1 \leqq i \leqq p+1)$ defines a sub-alphabet $A_i \subset A \setminus \{a\}$. The sequence w_1 has no repeated letter, and belongs to $DS(1)$; for $i \geqq 2$, each w_i is a sequence of $DS(2)$. The w_i cannot be empty, except maybe w_1 or w_{p+1}. Moreover, the following condition holds:

$$\text{For} \quad 1 \leqq i < j \leqq p+1, \qquad A_i \cap A_j = \varnothing. \tag{8}$$

This gives the general decomposition of $DS(2)$, with the condition (8) relative to sub-alphabets:

$$DS(2) = DS(1) \oplus ((\varepsilon \oplus DS(1)) . a . (DS(2) . a)^+ . (\varepsilon \oplus DS(2))), \tag{9}$$

Some letters of the alphabet A may not occur in the sequence w; let B denote the set of such letters. A sequence w is a complete sequence if and only if $B = \varnothing$. We now have a way to partition A, according to the decomposition of w:

$$A = \{a\} \oplus A_1 \oplus A_2 \oplus \ldots \oplus A_p \oplus A_{p+1} \oplus B.$$

2.6. The generating function $\Phi(x, y)$

The decomposition (9) given above is valid both on complete and non-complete sequences. However, if we want to mark the different letters, and to translate this decomposition on generating functions, we must restrict ourselves to complete sequences. We recall that x marks the length of a complete sequence, and y the number of distinct letters in the sequence.

Equation (9) translates into an equation on the function $\Phi(x, y)$:

$$\Phi = g + (1+g)\,xy\,\frac{x\,\Phi}{1-x\,\Phi}(1+\Phi).$$

Injecting the value of $g(x, y)$, we obtain: $x\,\Phi^2 + (xy-1)\,\Phi + xy = 0$, which is easily solved:

THEOREM 2: *The generating function enumerating the number of distinct complete sequences according to their length (marked by x) and number of distinct letters (marked by y) is:*

$$\Phi(x, y) = (1 - xy - \sqrt{(1-xy)^2 - 4\,x^2\,y})/(2\,x).$$

3. CONSEQUENCES

3.1. Number of sequences of given length

The number of complete sequences of length k on an alphabet of size n is simply $b_{n,k} = [x^k\,y^n]\,\Phi(x, y)$. We can express it using Catalan numbers $C_n = (2\,n)\,!/(n)\,!\,(n+1)\,!$. First, $1-xy$ can be factored out of the square root; the choice between $1-xy$ and $xy-1$ is made so that the solution is continuous at the origin:

$$b_{n,k} = -\frac{1}{2}[x^{k+1}\,y^n]\sqrt{(1-xy)^2 - 4\,x^2\,y}$$
$$= -\frac{1}{2}[x^{k+1}\,y^n]\left\{(1-xy)\sqrt{1-\frac{4\,x^2\,y}{(1-xy)^2}}\right\}.$$

Define $\alpha_{i,j} = [x^i\,y^j]\sqrt{1-4\,x^2\,y/(1-xy)^2}$; we have:

$$b_{n,k} = -\frac{1}{2}\alpha_{k+1,n} + \frac{1}{2}\alpha_{k,n-1}.$$

Now we evaluate $\alpha_{i,j}$. It can be rewritten as:

$$\alpha_{i,j} = 4^{i-j}\,[x^{i-j}\,t^j]\sqrt{1-xt/(1-t)^2}.$$

Using the equality:

$$[z^n]\sqrt{1-z} = -(2\,n-2)\,!/(2^{2\,n-1}\,n\,!\,(n-1)\,!) = -C_{n-1}/2^{2\,n-1},$$

we obtain:

$$\alpha_{i,j} = -2\,C_{i-j-1}\,[t^{2j-i}]\frac{1}{(1-t)^{2i-2j}}.$$

Now $[t^k]\{1/(1-t)^p\} = \binom{k+p-1}{p-1}$, and we have:

$$\alpha_{i,j} = -2\,C_{i-j-1}\binom{i-1}{2j-i} \qquad \text{if} \quad j<i<2j;$$
$$=0 \qquad \text{otherwise.}$$

Plugging our evaluations of $\alpha_{k+1,n}$ and $\alpha_{k,n-1}$ into $b_{n,k}$, we get:

$$b_{n,k} = C_{k-n}\left(\binom{k}{2n-k-1} - \binom{k-1}{2n-k-2}\right) = C_{k-n}\binom{k-1}{2n-k-1}.$$

The number of non-complete sequences, or the number of words, either complete or non-complete, are easily expressed in terms of $b_{n,k}$. We sum up these results in the following theorem:

THEOREM 3: *The number of complete sequences of $DS(2)$, of length k on an alphabet of size n, is:*

$$b_{n,k} = C_{k-n}\binom{k-1}{2n-k-1}.$$

The number of sequences of $DS(2)$, of length k on n letters, is:

$$a_{n,k} = \sum_{q=1}^{n} C_{k-q}\binom{k-1}{2q-k-1}.$$

The number $M_{n,k}$ of complete words and the number $N_{n,k}$ of non complete words, of length k on n letters, satisfy:

$$M_{n,k} = n!\,C_{k-n}\binom{k-1}{2n-k-1};$$

$$N_{n,k} = n!\sum_{r=0}^{n-1} C_{k+r-n}\binom{k-1}{2n-k-1-2r}\Big/ r!.$$

3.2. Sequences of maximal length

The complete sequences of maximal length are obtained for $k=2n-1$. Their number is $b_{n,\,2n-1}=C_{n-1}$.

Let D be the classical *restricted Dyck language* on two letters (x and $\bar{x}$). The words w of D, or *Dyck words*, are characterized by the two following conditions:

1. for any left factor u of w, the number of occurrences of the letter x in u is greater or equal to the number of occurrences of the letter $\bar{x}$ in u ($|u|_x \geqq |u|_{\bar{x}}$);

2. the number of occurrences of the letter x in w is equal to the number of occurrences of the letter $\bar{x}$ in w ($|w|_x = |w|_{\bar{x}}$).

It is well-known that the Dyck words of length $2n$ are enumerated by the Catalan number C_n. Thus it is possible to find a bijection between complete sequences of maximal length and Dyck words. We describe below two maps B_1 and B_2 that realize this bijection.

Example: The $C_3=5$ complete sequences of maximal length on four letters are given by:

$$abacada, \quad abacdca, \quad abcbada, \quad abcbdba, \quad abcdcba.$$

and the $C_3=5$ Dyck words of length six are given by:

$$xxxxxx, \quad xxxxxx, \quad xxxxxx, \quad xxxxxx, \quad xxxxxx.$$

We first point out some properties useful for the sequel. *Every word (or sequence) of maximal length is complete.* If not, its length could be increased by adding a letter that does not occur in the word. *In a sequence of maximal length w, the first letter a is always a repeated letter.* If not, the letter a could be concatenated at the end of the sequence and we would obtain another, longer, sequence. The same argument shows that *the last letter of a maximal sequence is always equal to its first letter.*

Using these properties, we can split every maximal sequence w on $n>1$ letters as follows:

$$w=a.w_1.a.w_2,$$

where w_1 is not the empty word, is maximal on $q>0$ letters and does not contain the letter a, and where aw_2 is also a maximal word, on the $n-q$ letters that do not occur in w_1.

We define below the first map B_1 from maximal sequences to Dyck words:

DEFINITION OF B_1:

- *if the length of w is one, then $B_1(w)=\varepsilon$ (the empty word);*
- *if the length of w is greater than one, then $w=a.w_1.a.w_2$ and $B_1(w)=x\,B_1(w_1)\,\bar{x}\,B_1(aw_2)$.*

It is easy to verify that B_1 is a bijection (w_1 and w_2 are uniquely defined) and that, for each maximal sequence w of $DS(2)$, $|B_1(w)|=|w|-1$, where $|w|$ denotes the length of w. We denote by L the set of words $B_1(w)$ where w is a maximal sequence of $DS(2)$. The language L satisfies the equation:

$$L=\varepsilon \oplus x.L.\bar{x}.L$$

which is the classical equation of the restricted Dyck language D.

The second map B_2 is defined non recursively and is obtained in one pass from left to right over the maximal sequence. We do not code the first letter; every new letter is coded by x and every already encountered letter (including the first letter) is coded by $\bar{x}$. More formally, we define B_2 in terms of some intermediate operators τ_E, where E is a subset of A which contains the letters already encountered:

DEFINITION OF B_2: *The coding associated to a maximal sequence aw, beginning with a, is $B_2(aw)=\tau_{\{a\}}(w)$, with the operators τ_E defined as follows, for all $E \subset A$:*

- $\tau_E(\varepsilon)=\varepsilon$;
- *for $a\in A$,*
  - *If $a\in E$, $\tau_E(aw)=\bar{x}\,\tau_E(w)$.*
  - *If $a\notin E$, $\tau_E(aw)=x\,\tau_{E\cup\{a\}}(w)$.*

The two following properties are easy to verify and we leave the proof to the reader:

1. *For any subset $E\subset A$ and any sequence w, $\tau_E(w)=\tau_{E\cap \text{Alphabet}(w)}(w)$.*

2. *If we can write a sequence w as aw_1, with $a\in A$, $a\notin$ Alphabet(w_1), and $w_1\neq\varepsilon$, then $B_2(w)=x\,B_2(w_1)$.*

We can now state the following result:

PROPOSITION: *B_1 and B_2 define the same bijection between complete sequences of maximal length on $n>0$ letters and Dyck words of length $2n-2$.*

Proof: If $n=1$, the result is obvious. Now suppose it also holds for $1 \leq n < l$. Let w denote a complete sequence of maximal length on l letters. We have seen before that $w = aw_1aw_2$, with $a \in A$, $a \notin \text{Alphabet}(w_1)$ and $w_1 \neq \varepsilon$. Moreover w_1 is a complete sequence of maximal length on $q > 0$ letters, aw_2 is a complete sequence of maximal length on $l-q$ letters $(0 < l-q < l)$ and w_1 has no letters in common with w_2.

Hence we have $B_1(w) = x\, B_1(w_1)\, \bar{x}\, B_1(aw_2)$ and $B_2(w) = \tau_{\{a\}}(w_1\, aw_2)$. But as $a \notin \text{Alphabet}(w_1)$ and $\text{Alphabet}(w_1) \cap \text{Alphabet}(w_2) = \varnothing$, by Property 1 above $\tau_{\{a\}}(w_1aw_2) = \tau_{\{a\}}(w_1)\, \bar{x}\, \tau_{\{a\}}(w_2)$, *i.e.*

$$B_2(w) = \tau_{\{a\}}(w_1)\, \bar{x}\, \tau_{\{a\}}(w_2).$$

By definition $\tau_{\{a\}}(w_1)$ is equal to $B_2(aw_1)$ and $\tau_{\{a\}}(w_2)$ to $B_2(aw_2)$; by induction we obtain that $B_2(aw_1) = B_1(aw_1)$ and $B_2(aw_2) = B_1(aw_2)$. Moreover $a \notin \text{Alphabet}(w_1)$; hence by Property 2 $B_2(aw_1) = x\, B_2(w_1)$. Thus we get

$$B_2(w) = x\, B_2(w_1)\, \bar{x}\, B_2(aw_2) = x\, B_1(w_1)\, \bar{x}\, B_1(aw_2),$$

i.e. $B_2(w) = B_1(w)$ as desired. ■

3.3. Sequences on a given number of letters

The total number of complete sequences, for a given size n of alphabet A, is $[y^n]\,\Phi(1, y)$. We have $\Phi(1, y) = (1 - y - \sqrt{1 - 6y + y^2})/2$; this can be expressed using the generating function $r(t)$ of Schröder numbers [9]:

$$\Phi(1, y) = y r(y) \qquad \text{for} \quad r(t) = \sum_n R_n t^n = (1 - t - \sqrt{1 - 6t + t^2})/2t.$$

The Schröder words on the alphabet $\{x, \bar{x}, y\}$ are given by the language equation:

$$\mathcal{S} = 1 + yy\,\mathcal{S} + x\,\mathcal{S}\,\bar{x}\,\mathcal{S}.$$

We give below a bijection C between Schröder words and complete sequences, which extends the bijection B_1 given for sequences of maximal length:

- Let w be a word on a single letter: $C(w) = \varepsilon$.
- If $|w| > 1$, and if the first letter of w is not repeated: $C(aw') = yy\, C(w')$.
- If the first letter of w is repeated: let $w = aw_1aw_2$, for $a \notin w_1$. Then $C(w) = x\; C(w_1)\, \bar{x}\, C(aw_2)$.

Hence the total number of complete sequences on $n+1$ letters is the Schröder number $\mathcal{S}_n=\sum_{r=0}^{n} C_r\binom{n+r}{2r}$. The following property can be easily proved by recurrence on n:

PROPOSITION: *Every complete sequence w on n letters is coded by a word $C(w)$ of length $2n-2$.*

3.4. Average length and number of letters

The average length of a complete sequence on n letters is:

$$\frac{[y^n]\,\Phi'_x(1,y)}{[y^n]\,\Phi(1,y)}.$$

To compute it, we first evaluate the denominator:

$$\Phi(1,y)=(1/2)(1-y-\sqrt{1-6y+y^2}).$$

The function $y\mapsto\sqrt{1-6y+y^2}$ has for dominant singularity $y_0=3-2\sqrt{2}=0{,}17157287...$; the other singularity is $y_1=3+2\sqrt{2}$. We can easily get an asymptotic expression of $[y^n]\,\Phi(1,y)$ by a transfer lemma [8]: $\sqrt{1-6y+y^2}=\sqrt{(1-y/y_0)(1-y/y_1)}$ and thus:

$$\begin{aligned}[y^n]\,\Phi(1,y)&\approx-\frac{1}{2}\sqrt{1-\frac{y_0}{y_1}}[y^n]\sqrt{1-\frac{y}{y_0}}\\&\approx-\frac{1}{2}\sqrt{1-\frac{y_0}{y_1}}y_0^{-n}[z^n]\sqrt{1-z}.\end{aligned}$$

Using the formula $[z^n]\sqrt{1-z}\approx-1/2n\sqrt{\pi n}$, we obtain:

$$[y^n]\,\Phi(1,y)\approx\sqrt{1-\frac{y_0}{y_1}}\,\frac{1}{4y_0^n n\sqrt{\pi n}}.$$

We next study the coefficient of y^n in

$$\Phi'_x(1,y)=-\frac{1}{2}-\frac{y-1}{2\sqrt{1-6y+y^2}}.$$

We have:

$$[y^n]\,\Phi'_x(1,y)=-\frac{1}{2}[y^n]\frac{y-1}{\sqrt{1-6y+y^2}}.$$

And using again a transfer lemma, we state:

$$[y^n]\frac{y-1}{\sqrt{1-6y+y^2}} \approx \frac{y_0-1}{\sqrt{1-y_0/y_1}} y_0^{-n}[z^n]\frac{1}{\sqrt{1-z}}.$$

The formula $[z^n]\{1/\sqrt{1-z}\} \approx 1/\sqrt{\pi n}$ gives finally

$$[y^n]\Phi'_x(1, y) \approx -\frac{1-y_0}{2y_0^n\sqrt{1-y_0/y_1}\sqrt{\pi n}}.$$

Hence the theorem:

THEOREM 4: *The average length of a complete sequence on n letters is asymptotically equal to* $(1+1/\sqrt{2})n = 1.70710677...n$.

In the same way, we evaluate the average number of letters in a complete sequence of length k: $[x^k]\{\Phi'_y(x, 1)\}/[x^k]\{\Phi(x, 1)\}$. First the singularities of the function

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

are -1 and $+1/3$. As above, we show that:

$$[x^k]\Phi(x, 1) \approx -\frac{1}{\sqrt{3}}[x^k]\sqrt{1-3x} \approx \frac{\sqrt{3}}{2}\frac{3^k}{k\sqrt{\pi k}}.$$

Then we evaluate $\Phi'_y(x, 1) = (1/2)(-1+\sqrt{(1+x)/(1-3x)})$. We clearly have:

$$[x^k]\Phi'_y(x, 1) \approx \frac{1}{\sqrt{3}}[x^k]\frac{1}{\sqrt{1-3x}} \approx \frac{3^k}{\sqrt{3\pi k}}.$$

Finally we establish the following result:

THEOREM 5: *The average number of distinct letters in a complete sequence of length k is asymptotically equal to* $2k/3$.

These results also hold for non-complete sequences and words: Informally, the asymptotic equivalents are determined by the singularity of smallest modulus of the generating function, and multiplying Φ by e^y to get Ψ adds no singularity. As for Ξ, the factor $1/(1-y)$ adds the singularity 1, which is farther than the singularities $3-2\sqrt{2}$ or $1/3$.

3.5. Probability distribution of the length of a sequence or the number of letters

We have plotted the curve of the probability distribution of the length of a complete sequence, when the number n of letters is fixed, and the probability distribution of the number of letters in a complete sequence of fixed length k. Empirically, these distributions are found to follow a normal law. This can be proved either from the expression of the $b_{n,k}$, using the Stirling approximation for $n!$, or from results in combinatorial statistics which characterize classes of functions defining normal distributions [19].

4. THE CASE $s \geqq 3$

Our method fails in the case where s is at least equal to 3: We can still decompose w as in equation (9), but with different conditions on the words w_i and the sub-alphabets they define. In particular, these sub-alphabets are no longer disjoint: Condition (8) does not hold, and we cannot translate the decomposition into an equation on the bivariate generating function.

Moreover, suppose that the generating function $f(x) = \sum_n \alpha(n) y^n$ is algebraic (*i.e.* solution of an algebraic equation). Then the singularities of f will be algebraico-logarithmic and their growth will be infinitely larger than the growth of $\alpha(n)$ [7]. So we cannot hope to obtain an algebraic generating function for $s > 2$.

ACKNOWLEDGEMENTS

The authors thank an anonymous referee for pointing out an error in the original paper.

REFERENCES

1. P. Agarwal, Intersection and decomposition algorithms for arrangements of curves in the plane. Ph. D., New York University, Courant Institute of Mathematical Sciences, 1989.
2. P. Agarwal, M. Sharir and P. Shor, Sharp upper and lower bounds on the length of general Davenport-Schinzel sequences. *J. Combinat. Theory Ser. A*, 52, (2): 228-274, 1989.
3. P. Alevizos, J. D. Boissonnat and F. Preparata, An optimal algorithm for the boundary of a cell in a union of rays. *Algorithmica*, 5, (4): 573-590, 1990.

4. R. COLE and M. SHARIR, Visibility of a polyhedral surface from a point. *Technical Report* 266, Comp. Science Dept., Courant Institute of Mathematical Sciences, New York University, December 1986.
5. H. DAVENPORT, A combinatorial problem connected with differential equations II. *Acta Arithmetica*, XVII: 363-372, 1971.
6. H. DAVENPORT and A. SCHINZEL, A combinatorial problem connected with differential equations. *Amer. J. Math.*, 87: 684-694, 1965.
7. P. FLAJOLET, Analytic models and ambiguity of context-free languages. *Theoretical Computer Science*, 49, (2): 283-310, 1987.
8. P. FLAJOLET and A. ODLYZKO, Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3, (2): 216-240, 1990.
9. D. GOUYOU-BEAUCHAMPS and B. VAUQUELIN, Deux propriétés combinatoires des nombres of Schröder. *Informatique Théorique et Applications*, 22, (3): 361-388, 1988.
10. S. HART and M. SHARIR, Nonlinearity of Davenport-Schinzel sequences and of generalized path compression schemes. *Combinatorica*, 6: 151-177, 1986.
11. K. KEDEM and M. SHARIR, An efficient algorithm for planning collision-free translational motion of a convex polygonal object in 2-dimensional space amidst polygonal obstacles. In *ACM Symp. on Computational Geometry*, pp. 75-80, 1985.
12. P. KOMJÁTH, A simplified construction of nonlinear Davenport-Schinzel sequences. *J. Combin. Theory Ser. A*, 49, (2): 262-267, 1988.
13. D. LEVEN and M. SHARIR, On the number of critical free contacts of a convex polygonal object moving in 2-dimensional polygonal space. *Discrete Comp. Geom.*, 2: 255-270, 1987.
14. J. PACH and M. SHARIR, The upper envelope of a piecewise linear function and the boundary of a region enclosed by convex plates: Combinatorial Analysis. *Discrete Comp. Geom.*, 4, (4): 291-309, 1989.
15. R. POLLACK, M. SHARIR and S. SIFRONY, Separating two simple polygons by a sequence of translations. *Discrete Comp. Geom.*, 3: 123-136, 1988.
16. M. SHARIR, Almost linear upper bounds on the length of general Davenport-Schinzel sequences. *Combinatorica*, 7, (1): 131-143, 1987.
17. M. SHARIR, *Davenport-Schinzel sequences and their geometric applications*, chapter Theoretical Foundations of Computer Graphics and CAD, pp. 253-278. Springer-Verlag, NATO ASI Series, Vol. F-40, R. A. Earnshaw edition, 1988.
18. M. SHARIR, R. COLE, K. KEDEM, D. LEVEN, R. POLLACK and S. SIFRONY, Geometric applications of Davenport-Schinzel sequences. In *27th Symposium on Foundations of Computer Science*, pp. 77-86, Toronto (Canada), 1986.
19. M. SORIA, Méthodes d'analyse pour les constructions combinatoires et les algorithmes. *Thèse d'État*, L.R.I., Université Paris-Sud (Orsay), Juillet 1990.
20. E. SZEMERÉDI, On a problem by Davenport and Schinzel. *Acta Arithmetica*, XXV: 213-224, 1974.
21. A. WIERNIK, Planar realizations of nonlinear Davenport-Schinzel sequences by segments. In *27th Symposium on Foundations of Computer Science*, pp. 97-106, Toronto (Canada), 1986.