Publ. I.R.M.A. Strasbourg, 1993, 460/S–23
Actes 23^e Séminaire Lotharingien, p. 77–83

ENUMERATION OF SOME DAVENPORT-SCHINZEL SEQUENCES

BY

DANIÈLE GARDY (*) and DOMINIQUE GOUYOU-BEAUCHAMPS (*)

ABSTRACT. — Davenport-Schinzel sequences of order s are words with no subsequence $ababa\ldots$ 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.

1 Introduction

Davenport-Schinzel sequences are words with forbidden subsequences, which were first defined by Davenport and Schinzel [1] in connection with the general solution of a differential equation, and which have found a new field of application in computational geometry. We indicate here only the most classical application, and refer the interested reader to the survey paper of Sharir *et al.* [4] for details. Let $f_1, \ldots, f_n$ be n countinuous 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).

We give below the formal definition of Davenport-Schinzel sequences and recall some previous results, which are all relevant to the maximal length of a sequence. We do not know of any work counting the number of distinct sequences. We next give a decomposition of sequences for $s = 2$ in Section 2, from which we obtain the function enumerating sequences according to their length and to the number of distinct letters they contain. Some 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 are enumerated by Catalan numbers. Finally, we indicate in Section 4 why our method fails in the case $s \geq 3$.

*This work was partially supported by PRC Mathématique-Informatique and by Esprit Basic Research Action No. 3075 (project ALCOM).

1.1 Definition

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

- any two consecutive letters are distinct;
- w has no subsequence $u_{i_1} \cdots u_{i_{s+2}}$ $(1 \leq i_1 < i_2 < \cdots < i_{s+2} \leq k)$ on any two distinct letters a and b, satisfying:

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

We shall denote by $DS(s)$ the set of Davenport-Schinzel sequences, for a fixed s and for all alphabets.

1.2 Maximal length

A major consequence of the definition is that the number of Davenport-Schinzel sequences of order (n, s) is finite; this number is usually denoted by $\lambda_s(n)$. Much work has been devoted to the estimation of $\lambda_s(n)$. We recall here the main results and refer to [4] for a detailed bibliography:

- **s=1:** $\lambda_1(n) = n$.
- **s=2:** $\lambda_2(n) = 2n - 1$.
- **s=3:** $\lambda_s(n)$ is of order $\Theta(n\alpha(n))$, with $\alpha(n)$ the functional inverse of the Ackermann function. This is more than linear, but $\alpha(n)$ is less than 4 for all purposes.
- **s $\geq$ 4:** it is possible to give a bound on $\lambda_s(n)$. As in the case $s = 3$, it is theoretically superlinear, but is almost linear for all realistic values of n.

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

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 be true in general: The object of interest is not a word, but its structure subjected to a permutation of the letters.

Let us note that, from a counting point of view, there is not much difference: If $N_{n,k}$ is the number of words of Davenport-Schinzel on n letters and of length k, the number of sequences up to a permutation is $a_{n,k} = \frac{N_{n,k}}{n!}$. Denote by $\Psi(x, y)$ the ordinary generating function of the $a_{n,k}$, where variable x marks the length of a word, and variable y marks the size of the alphabet. We have:

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

Function Ψ is also an exponential generating function of the $N_{n,k}$, according to the size of the alphabet.

Our goal is to find a suitable decomposition of the words of $DS(2)$, from which we can deduce an equation on their generating function. This equation is of degree two, and can easily be solved. We first simplify the problem by introducing primary words and obtain a simple result on the generating function of primary words of $DS(1)$, which we shall need in the sequel. In the sequel, a *word* has the usual sense, and a *sequence* is a word, modulo a permutation of the letters.

2.1 Primary words

We introduce the notion of a *primary* word:

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

The length k of a primary word of $DS(2)$, on an alphabet of size n, belongs to the interval $[n, 2n-1]$. Let $M_{n,k}$ be the number of primary words of $DS(2)$, of length k, and $b_{n,k} = \frac{M_{n,k}}{n!}$ the number of related sequences. The following relation holds:

$$N_{n,k} = \sum_{q=1}^{n} \binom{n}{q} M_{q,k}.$$

We define in a similar way primary sequences. Let $\Phi(x,y)$ be the generating function counting primary sequences:

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

We obtain easily:

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

2.2 Enumeration of primary words of $DS(1)$

Let $g_{k,n}$ be the number of words of Davenport-Schinzel for $s = 1$, of length k, and such that all letters of alphabet A, of size n, appear in the word. We deduce immediately from the definition of $DS(1)$ that such a word has no repeated letter. This shows that, if $n \neq k$: $g_{k,n} = 0$, and if $n = k$: $g_{n,n} = n!$. Define $g(x,y) = \sum_{k,n\geq 1} g_{n,k} x^k \frac{y^n}{n!}$; we have: $g(x,y) = \sum_{n\geq 1} x^n y^n = \frac{xy}{1-xy}$.

2.3 Decomposition of a word of $DS(2)$

Let w be a word of $DS(2)$; w may actually be a word 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. Let a be this letter, which is repeated $p \geq 2$ times. We have:

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

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

$$\text{For } 1 \leq i < j \leq p+1, \quad A_i \cap A_j = \emptyset. \tag{1}$$

Some letters of alphabet A may not occur in word w; Let B denote the set of such letters. We now have a way to partition A, according to the decomposition of w:

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

A word w is a primary word if and only if $B = \emptyset$. This gives the general decomposition of $DS(2)$, with condition (1) relative to sub-alphabets:

$$DS(2) = DS(1) \bigoplus \Big((\epsilon \oplus DS(1)) \cdot a \cdot (DS(2) \cdot a)^{+} \cdot (\epsilon \oplus DS(2)) \Big). \tag{2}$$

2.4 Generating functions

The decomposition given above is valid both on primary and non-primary words. However, if we want to mark the different letters that appear in a word, and to translate this decomposition on generating functions, we must restrict ourselves to primary words. We recall that x marks the length of a primary word, and y the number of distinct letters in the word. Equation (2) translates into an equation on 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:

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

3 Consequences

3.1 Number of words of given length

The number of primary sequences on an alphabet of size n, and of length k, is simply $b_{n,k} = [x^k y^n]\Phi(x,y)$. We can express it using Catalan numbers $C_n = \frac{(2n)!}{n!(n+1)!}$.

The number of primary 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:

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

The number $M_{n,k}$ of primary words and the number $N_{n,k}$ of non primary words of $DS(2)$ satisfy:

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

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

3.2 Sequences of maximal length

The primary sequences of maximal length are obtained for $k = 2n-1$. Their number is $b_{n,2n-1} = C_{n-1}$. It is possible (but we do not give it here) to find a bijection between these sequences and known combinatorial objects enumerated by Catalan numbers, such as Dyck words.

3.3 Sequences on a given number of letters

The total number of primary 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 [3]:

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

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

$$S = 1 + yyS + xS\bar{x}S.$$

We give below a bijection C between Schröder words and primary sequences.

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

The following property can be proved by recurrence on n: *Every primary 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 primary sequence on n letters is:

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

It is easy to obtain an asymptotic equivalent of this expression. We first determine the singularity of smallest modulus of the function $y \mapsto \sqrt{1-6y+y^2}$, which is $y_0 = 3-2\sqrt{2} \approx 0.17157287...$ A transfer lemma [2] gives:

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

with $y_1 = 3+2\sqrt{2}$ (y_1 is the other singularity of the previous function). The numerator can be studied in the same way, and we obtain:

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

Hence the result:

> *The average length of a primary sequence on n letters is asymptotically equal to $(1+\frac{1}{\sqrt{2}})n$.*

The average number of letters in a primary sequence of length k is obtained exactly in the same way:

> *The average number of distinct letters in a primary sequence of length k is asymptotically equal to $2k/3$.*

These results also hold for non-primary sequences (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).

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 primary sequence, when the number n of letters is fixed, and the probability distribution of the number of letters in a primary sequence of fixed length k. Empirically, these distributions are found to follow a normal law. This can be proved by mathematical technics, and is a direct consequence of results obtained by M. Soria [5].

4 The case $s \geq 3$

Our method fails in the case where s is at least equal to 3: We can still decompose w as in equation (2), 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 (1) does not hold, and we cannot translate the decomposition into an equation on the bivariate generating function.

References

[1] H. DAVENPORT, A. SCHINZEL: "A combinatorial problem connected with differential equations." Amer. J. Math. 87, pp. 684-694, 1965.

[2] P. FLAJOLET, A. ODLYZKO: "Singularity analysis of generating functions." Research Report No. 826, Inria, April 1988. To appear in Siam Journal of Discrete Mathematics, 1990.

[3] D. GOUYOU-BEAUCHAMPS, B. VAUQUELIN: "Deux propriétés combinatoires des nombres of Schröder." Informatique théorique et Applications, Vol. 22, No. 3, pp. 361-388, 1988.

[4] M. SHARIR, R. COLE, K. KEDEM, D. LEVEN, R. POLLACK, S. SIFRONY: "Geometric Applications of Davenport-Schinzel Sequences." 27^{th} Symposium on Foundations Of Computer Science, Toronto (Canada), 27-29 Oct. 1986, pp. 77-86.

[5] M. SORIA: *Méthodes d'analyse pour les constructions combinatoires et les algorithmes.* Thèse d'Etat, Université Paris-Sud, 1990.

Danièle GARDY,
Dominique GOUYOU-BEAUCHAMPS,
Laboratoire de Recherche en Informatique,
C.N.R.S. U.A. 410,
Université de Paris-Sud,
91405 Orsay, France.