Discrete Mathematics 340 (2017) 862–881

Contents lists available at ScienceDirect

Discrete Mathematics

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

Counting the number of non-zero coefficients in rows of generalized Pascal triangles

Julien Leroy [1], Michel Rigo [*], Manon Stipulanti [2]

Université de Liège, Institut de mathématique, Allée de la découverte 12 (B37), 4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 17 May 2016
Received in revised form 27 December 2016
Accepted 2 January 2017

Keywords:
Binomial coefficients
Subword
Farey tree
Regular sequences
Fibonacci numeration system

ABSTRACT

This paper is about counting the number of distinct (scattered) subwords occurring in a given word. More precisely, we consider the generalization of the Pascal triangle to binomial coefficients of words and the sequence $(S(n))_{n\geq 0}$ counting the number of positive entries on each row. By introducing a convenient tree structure, we provide a recurrence relation for $(S(n))_{n\geq 0}$. This leads to a connection with the 2-regular Stern–Brocot sequence and the sequence of denominators occurring in the Farey tree. Then we extend our construction to the Zeckendorf numeration system based on the Fibonacci sequence. Again our tree structure permits us to obtain recurrence relations for and the F-regularity of the corresponding sequence.

1. Introduction

A *finite word* is simply a finite sequence of letters belonging to a finite set called the *alphabet*. We recall from combinatorics on words that the binomial coefficient $\binom{u}{v}$ of two finite words u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword). As an example, consider the following binomial coefficient for two words over $\{0, 1\}$

$$\binom{101001}{101} = 6.$$

Indeed, if we index the letters of the first word $u_1u_2\cdots u_6 = 101001$, we have

$$u_1u_2u_3 = u_1u_2u_6 = u_1u_4u_6 = u_1u_5u_6 = u_3u_4u_6 = u_3u_5u_6 = 101.$$

Observe that this concept is a natural generalization of the binomial coefficients of integers. For a one-letter alphabet $\{a\}$, we have

$$\binom{a^m}{a^n} = \binom{m}{n}, \quad \forall\, m, n \in \mathbb{N} \tag{1}$$

where a^m denotes the concatenation of m copies of the letter a. For more on these binomial coefficients, see, for instance, [26, Chap. 6]. There is a vast literature on the subject with applications in formal language theory (e.g., Parikh matrices,

* Corresponding author.
E-mail addresses: J.Leroy@ulg.ac.be (J. Leroy), M.Rigo@ulg.ac.be (M. Rigo), M.Stipulanti@ulg.ac.be (M. Stipulanti).

1 J. Leroy is FNRS post-doctoral fellow.
2 M. Stipulanti is supported by a FRIA grant.

http://dx.doi.org/10.1016/j.disc.2017.01.003
0012-365X/

Table 1
The first few values in the generalized Pascal triangle P_2.

		ε	1	10	11	100	101	110	111	S
$\text{rep}_2(0)$	ε	**1**	0	0	0	0	0	0	0	1
$\text{rep}_2(1)$	1	**1**	**1**	0	0	0	0	0	0	2
$\text{rep}_2(2)$	10	1	1	1	0	0	0	0	0	3
$\text{rep}_2(3)$	11	**1**	**2**	0	**1**	0	0	0	0	3
$\text{rep}_2(4)$	100	1	1	2	0	1	0	0	0	4
$\text{rep}_2(5)$	101	1	2	1	1	0	1	0	0	5
$\text{rep}_2(6)$	110	1	2	2	1	0	0	1	0	5
$\text{rep}_2(7)$	111	**1**	**3**	0	**3**	0	0	0	**1**	4

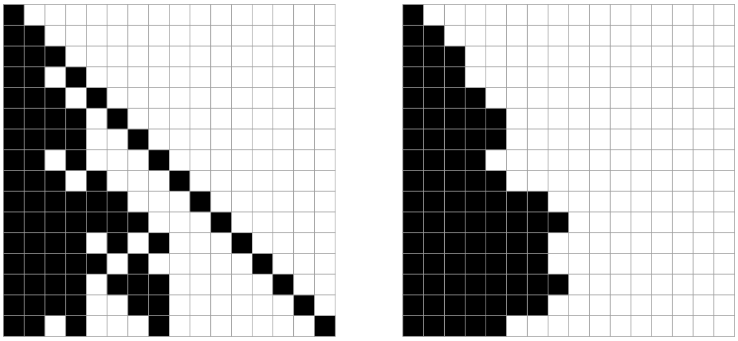

Fig. 1. Positive values in the generalized Pascal triangle P_2 (on the left) and the compressed version (on the right).

p-group languages or piecewise testable languages [21,22]) and combinatorics on words (e.g., avoiding binomial repetitions [28]). One of the first combinatorial questions was to determine when it is possible to uniquely reconstruct a word from some of its binomial coefficients; for instance, see [12].

The connection between the Pascal triangle and the Sierpiński gasket is well understood. Considering the intersection of the lattice $\mathbb{N}^2$ with the region $[0, 2^j] \times [0, 2^j]$, the first 2^j rows and columns of the usual Pascal triangle $(\binom{m}{n} \bmod 2)_{m,n<2^j}$ provide a coloring of this lattice. If we normalize this region by a homothety of ratio $1/2^j$, we get a sequence in $[0, 1] \times [0, 1]$ converging, for the Hausdorff distance, to the Sierpiński gasket when j tends to infinity. In [24] we extend this result for a generalized Pascal triangle defined below.

Definition 1. We let $\text{rep}_2(n)$ denote the (greedy) base-2 expansion of $n \in \mathbb{N}\setminus\{0\}$ starting with 1 (these expansions correspond to expansions with most significant digit first). We set $\text{rep}_2(0)$ to be the empty word denoted by ε. Let $L_2 = \{\varepsilon\} \cup 1\{0, 1\}^*$ be the set of base-2 expansions of the integers. For all $i \geq 0$, we denote by w_i the ith word in L_2. Observe that $w_i = \text{rep}_2(i)$ for $i \geq 0$.

To define an array P_2, we will consider all the words over $\{0, 1\}$ (starting with 1) and we order them by the radix order (i.e., first by length, then by the classical lexicographic ordering for words of the same length assuming $0 < 1$). For all $i, j \geq 0$, the element in the ith row and jth column of the array P_2 is defined as the binomial coefficient $\binom{w_i}{w_j}$ of the words w_i, w_j from Definition 1. This array is a generalized Pascal triangle that was introduced in [24] and its first few values are given in Table 1. These values correspond to the words ε, 1, 10, 11, 100, 101, 110, 111. In Table 1, the elements of the classical Pascal triangle are written in bold. A visual representation is given in Fig. 1 where we have represented the positive values and a compressed version of the same figure. In Table 1, we also add an extra column in which we count, on each row of P_2, the number of positive binomial coefficients as explained in Definition 2.

Definition 2. We are interested in the sequence[3] $(S(n))_{n\geq 0}$ whose nth term, $n \geq 0$, is the number of non-zero elements in the nth row of P_2. Hence, the first few terms of this sequence are

1, 2, 3, 3, 4, 5, 5, 4, 5, 7, 8, 7, 7, 8, 7, 5, 6, 9, 11, 10, 11, 13, 12, 9, 9, 12, 13, 11, 10, ...

Otherwise stated, for $n \geq 0$, we define

$$S(n) := \#\left\{v \in L_2 \mid \binom{\text{rep}_2(n)}{v} > 0\right\}. \tag{2}$$

[3] The sequence obtained by adding an extra 1 as a prefix of our sequence of interest matches exactly the sequence A007306 found in Sloane's On-Line Encyclopedia of Integer Sequences [33].