Advances in Applied Mathematics 80 (2016) 24–47

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

Generalized Pascal triangle for binomial coefficients of words

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

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

A R T I C L E I N F O

Article history:
Received 1 March 2016
Received in revised form 23 April 2016
Accepted 25 April 2016
Available online 4 May 2016

MSC:
primary 28A80
secondary 28A78, 11B85, 68R15

A B S T R A C T

We introduce a generalization of Pascal triangle based on binomial coefficients of finite words. These coefficients count the number of times a word appears as a subsequence of another finite word. Similarly to the Sierpiński gasket that can be built as the limit set, for the Hausdorff distance, of a convergent sequence of normalized compact blocks extracted from Pascal triangle modulo 2, we describe and study the first properties of the subset of $[0, 1] \times [0, 1]$ associated with this extended Pascal triangle modulo a prime p.

1. Introduction

Pascal triangle and the corresponding Sierpiński gasket are well studied objects (see, for instance, [15] for a survey). They exhibit self-similarity features and have connections with dynamical systems, cellular automata, number theoretic questions and automatic

* 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 a FNRS post-doctoral fellow.
[2] M. Stipulanti is supported by a FRIA grant R.FNRS.3791.

http://dx.doi.org/10.1016/j.aam.2016.04.006

sequences [2,1,5,10,13]. In this paper, we will consider a variation of these two objects by extending binomial coefficients to the free monoid A^* where A is a finite alphabet.

Let us start with basic combinatorial definitions. A *finite word* is simply a finite sequence of letters belonging to a finite set called *alphabet*. In combinatorics on words, one can introduce the binomial coefficient $\binom{u}{v}$ of two finite words u and v which is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword). As an example, we consider two particular words over $\{0, 1\}$ and

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

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 single 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 letters a. For more on these binomial coefficients, see for instance [11, Chap. 6].

In this paper, we are interested in Pascal triangle obtained when considering binomial coefficients of words. To define such a triangular array, we will consider all the words over a finite alphabet and we order them by genealogical ordering (i.e., first by length, then by the classical lexicographic ordering for words of the same length assuming $0 < 1$). For the sake of simplicity, we will mostly discuss the case of a 2-letter alphabet $\{0, 1\}$ and to relate these words to base-2 expansions of integers, we will assume without loss of generality that the non-empty words start with 1. (If leading zeroes were allowed, then different words could represent the same integer.)

Definition 1. We let $\text{rep}_2(n)$ denote the greedy base-2 expansion of $n \in \mathbb{N}_{>0}$ starting with 1 where the notation $\mathbb{N}_{>0}$ stands for the set of all positive integers. We set $\text{rep}_2(0)$ to be the empty word denoted by ε. Let $L = \{\varepsilon\} \cup 1\{0, 1\}^*$ be the set of base-2 expansions of the integers. We let L_n denote the set of words of length at most n belonging to L, i.e.,

$$L_n = (\{\varepsilon\} \cup 1\{0, 1\}^*) \cap \{0, 1\}^{\leq n}.$$

Note that $\#L_n = 2^n$ for all $n \geq 0$.

The first few values in the generalized Pascal triangle T that we will deal with are given in Table 1. These values correspond to the words $\varepsilon, 1, 10, 11, 100, 101, 110, 111$.