Generalized Pascal triangles and binomial coefficients of words

Manon STIPULANTI FRIA grantee University of Liège m.stipulanti@ulg.ac.be
Joint work with Julien LEROY and Michel RIGO

Pascal triangle and Sierpiński gasket

Pascal triangle

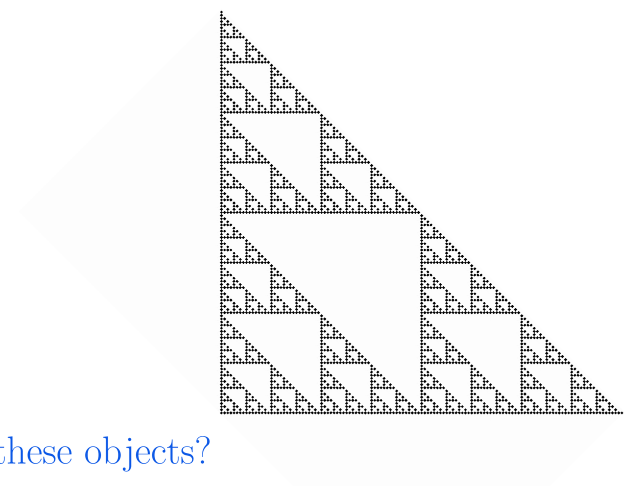
Classical binomial coefficient of integers

$\binom{m}{k}$ $m, k \in \mathbb{N}$

Pascal's rule: $\binom{m}{k} = \binom{m-1}{k-1} + \binom{m-1}{k}$

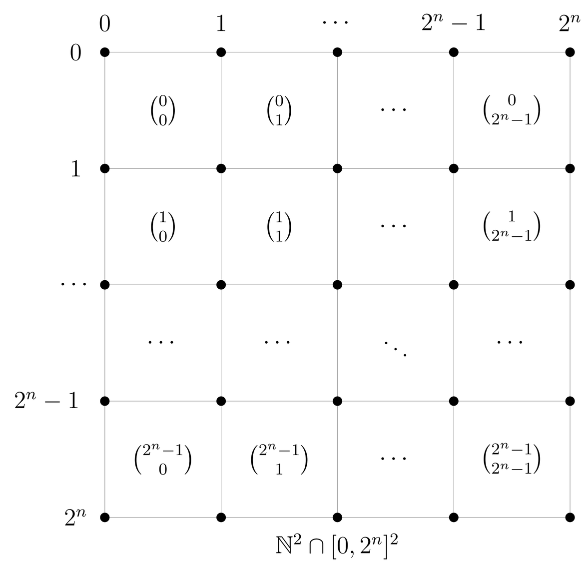
	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0
2	1	2	1	0	0	0	0	0
3	1	3	3	1	0	0	0	0
4	1	4	6	4	1	0	0	0
5	1	5	10	10	5	1	0	0
6	1	6	15	20	15	6	1	0
7	1	7	21	35	35	21	7	1

Sierpiński gasket



Link between these objects?

For each $n \in \mathbb{N}$, consider the intersection of the lattice $\mathbb{N}^2$ with the region $[0, 2^n] \times [0, 2^n]$:

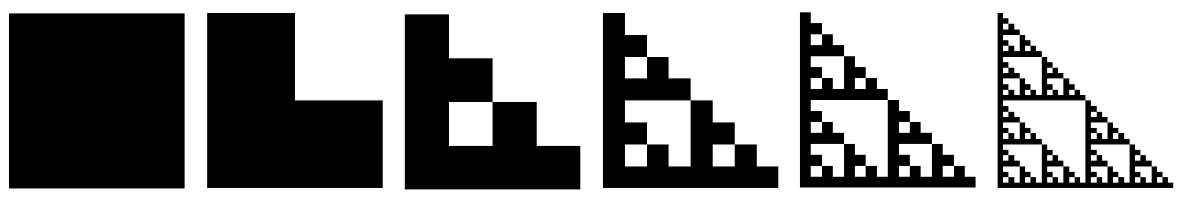


Color the unit square associated with the binomial coefficient $\binom{m}{k}$

in white if $\binom{m}{k} \equiv 0 \bmod 2$

in black if $\binom{m}{k} \equiv 1 \bmod 2$.

If we normalize this region by a homothety of ratio $1/2^n$, we get a sequence of compacts in $[0, 1] \times [0, 1]$.



The elements of the latter sequence corresponding to $n \in \{0, \ldots, 5\}$

Due to a folklore fact, this sequence converges, for the Hausdorff distance, to the Sierpiński gasket when n tends to infinity.

Binomial coefficients of words

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). For example, if $u = 101001$ and $v = 101$, then $\binom{u}{v} = 6$ since all the occurrences of v inside of u are

101001, 101001, 101001, 101001, 101001, 101001.

This concept is a natural generalization of the binomial coefficients of integers. For an alphabet containing only one letter a, we have

$$\binom{a^m}{a^k} = \binom{m}{k} \quad \forall m, k \in \mathbb{N}.$$

Moreover, due to the following result, we have the analogue of the Pascal's rule for binomial coefficients of words.

Lemma (Chapter 6, [6]): Let Σ be a finite alphabet. For all words $u, v \in \Sigma^*$ and all letters $a, b \in \Sigma$, we have

$$\binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}.$$

Generalized Pascal triangles

To define a new triangular array, we consider all the words over a finite alphabet $\{a_1, \ldots, a_\ell\}$ 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 $a_1 < a_2 < \cdots < a_\ell$).

If we take the case of a 2-letter alphabet $\{0, 1\}$, we consider the language of the base-2 expansions of integers, assuming without loss of generality that non-empty words start with 1:

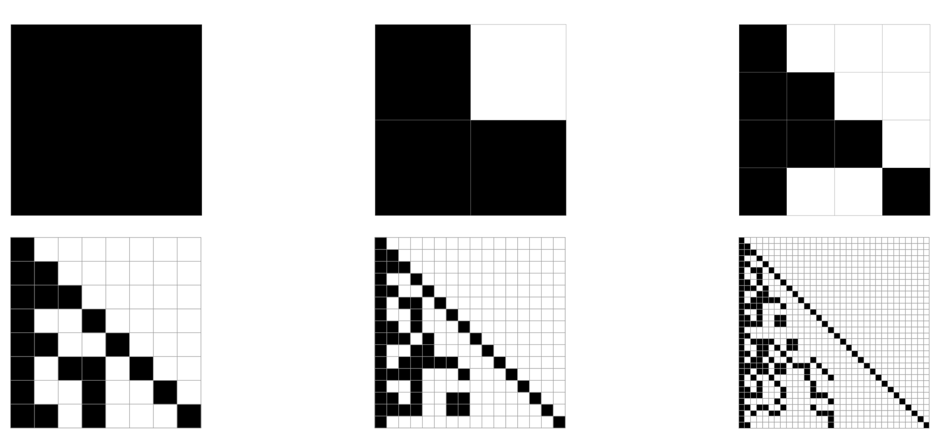
$$\mathrm{L}_2 = \mathrm{rep}_2(\mathbb{N}) = \{\varepsilon\} \cup 1\{0, 1\}^*.$$

The first few values of the generalized Pascal triangle P_2 are given in the following table.

	ε	1	10	11	100	101	110	111	
ε	**1**	0	0	0	0	0	0	0	1
1	**1**	**1**	0	0	0	0	0	0	2
10	1	1	1	0	0	0	0	0	3
11	**1**	**2**	0	**1**	0	0	0	0	3
100	1	1	2	0	1	0	0	0	4
101	1	2	1	1	0	1	0	0	5
110	1	2	2	1	0	0	1	0	5
111	**1**	**3**	0	**3**	0	0	0	**1**	4

If we consider the words of L_2 that only contain 1's, we obtain the elements of the usual Pascal triangle (in **bold**).

Using the same construction as before (namely coloring in black and white a grid containing binomial coefficients of words and then normalizing each region by a homothety), we get a sequence of compacts in $[0, 1] \times [0, 1]$.



The first six elements of the latter sequence

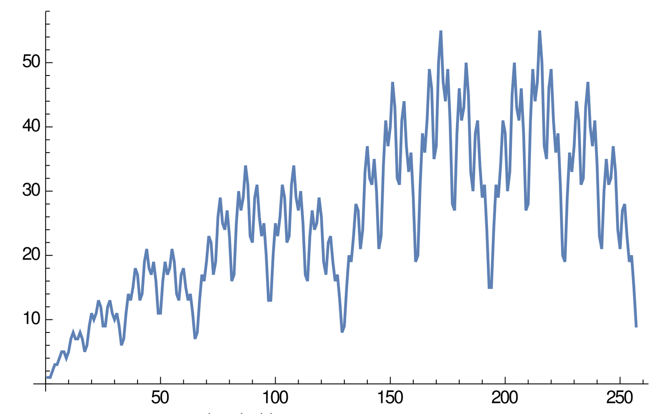
Theorem [3]: The sequence of compact sets defined previously converges, for the Hausdorff distance, to a limit object $\mathcal{L}$ that can be characterized using simple combinatorial properties.

It is straightforward to adapt our reasonings, constructions and results to a more general setting, namely we fix a prime number p and a rest $r \in \{1, \ldots, p-1\}$ and we color the squares in the grid in black if the corresponding binomial coefficient is congruent to r modulo p or white otherwise.

Counting positive binomial coefficients

For each $n \in \mathbb{N}_0$, we let $s_2(n)$ denote the number of positive binomial coefficients on the nth row of the generalized Pascal triangle P_2. We also set $s_2(0) := 1$. The first few terms of s_2 are

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



The sequence $(s_2(n))_{n \in \mathbb{N}}$ between 0 and a power of 2

Definition: Let $s = (s(n))_{n \in \mathbb{N}}$ be a sequence of integers and let $k \geq 2$. The k-kernel of s is the set of subsequence

$$\mathcal{K}_k(s) = \{(s(k^i \cdot n + j))_{n \in \mathbb{N}} \mid i \geq 0 \text{ and } 0 \leq j < k^i\}.$$

A sequence s is k-automatic if its k-kernel is finite. A sequence s is k-regular if there exist a finite number of sequences $(t_1(n))_{n \in \mathbb{N}}, \ldots, (t_\ell(n))_{n \in \mathbb{N}}$ such that each sequence $(t(n))_{n \in \mathbb{N}} \in \mathcal{K}_k(s)$ is a $\mathbb{Z}$-linear combination of the t_j's. A sequence s is k-synchronized if the language $\{\mathrm{rep}_k(n, s(n)) \mid n \in \mathbb{N}\}$ is accepted by some finite automaton.

Remark [2]: k-automatic $\subseteq$ k-synchronized $\subseteq$ k-regular.

Proposition [4]: The sequence s_2 is 2-regular but not 2-synchronized.

Extension to other numeration systems

	ε	1	10	100	101	1000	1001	1010	
ε	1	0	0	0	0	0	0	0	1
1	1	1	0	0	0	0	0	0	2
10	1	1	1	0	0	0	0	0	3
100	1	1	2	1	0	0	0	0	4
101	1	2	1	0	1	0	0	0	4
1000	1	1	3	3	0	1	0	0	5
1001	1	2	2	1	2	0	1	0	6
1010	1	2	3	1	1	0	0	1	6

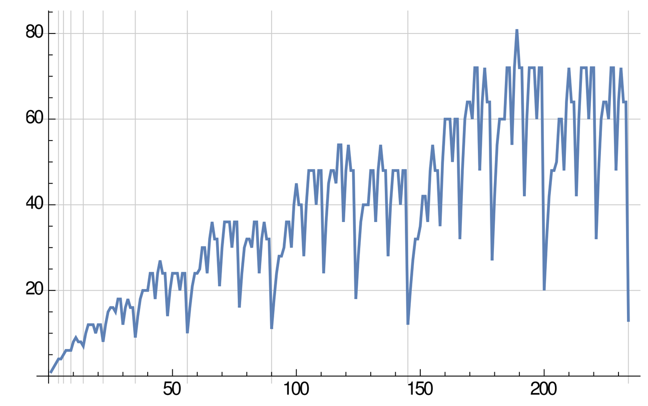
Instead of considering the language L_2, we restrict ourselves to words that contain no factor of the form 11. We are thus left with the language

$$L_F = \varepsilon \cup 1\{0, 01\}^*$$

which is the language of the Zeckendorff numeration system based on the Fibonacci numbers defined by $F(0) = 1$, $F(1) = 2$ and $F(n+2) = F(n+1) + F(n)$ for all $n \in \mathbb{N}$. We define a new generalized Pascal triangle P_F using those words.

For each $n \in \mathbb{N}$, we let $s_F(n)$ denote the number of positive binomial coefficients on the $(n+1)$th row of the generalized Pascal triangle P_F. The first few terms of s_F are

$$1, 2, 3, 4, 4, 5, 6, 6, 6, 8, 9, 8, 8, 7, 10, 12,$$
$$12, 12, 10, 12, 12, 8, 12, 15, 16, 16, 15, \ldots.$$



The sequence $(s_F(n))_{n \in \mathbb{N}}$ between 0 and a Fibonacci number

Extension of k-regularity [1, 7]: Let $s = (s(n))_{n \in \mathbb{N}}$ be a sequence of integers and let $k \geq 2$. The k-kernel $\mathcal{K}_k(s)$ of s can be obtained under the following process. First, fix a word $w \in \{0, 1, \ldots, k-1\}^*$ and select all the nonnegative integers whose base-k expansions with leading 0's end with this word w. Then, evaluate s at those integers to create a specific subsequence of the k-kernel. Let w vary in $w \in \{0, 1, \ldots, k-1\}^*$ to obtain the entire k-kernel.

The F-kernel $\mathcal{K}_F(s)$ of s can be obtained under the same technique. First, fix a word $w \in \{0, 1\}^*$ and select all the nonnegative integers whose Fibonacci representations with leading 0's end with this word w. Then, evaluate s at those integers to create a specific subsequence of the F-kernel. Let w vary in $w \in \{0, 1\}^*$ to obtain the entire F-kernel.

$w = 0$

n	$\mathrm{rep}_2(n)$	$s(n)$
0	ε	$s(0)$
1	1	$s(1)$
2	10	$s(2)$
3	11	$s(3)$
4	100	$s(4)$
5	101	$s(5)$

n	$\mathrm{rep}_F(n)$	$s(n)$
0	ε	$s(0)$
1	1	$s(1)$
2	10	$s(2)$
3	100	$s(3)$
4	101	$s(4)$
5	1000	$s(5)$

Definition: A sequence $s = (s(n))_{n \in \mathbb{N}}$ is F-automatic if its F-kernel is finite. A sequence s is F-regular if there exist a finite number of sequences $(t_1(n))_{n \in \mathbb{N}}, \ldots, (t_\ell(n))_{n \in \mathbb{N}}$ such that each sequence $(t(n))_{n \in \mathbb{N}} \in \mathcal{K}_F(s)$ is a $\mathbb{Z}$-linear combination of the t_j's.

Proposition [4]: The sequence s_F is F-regular.

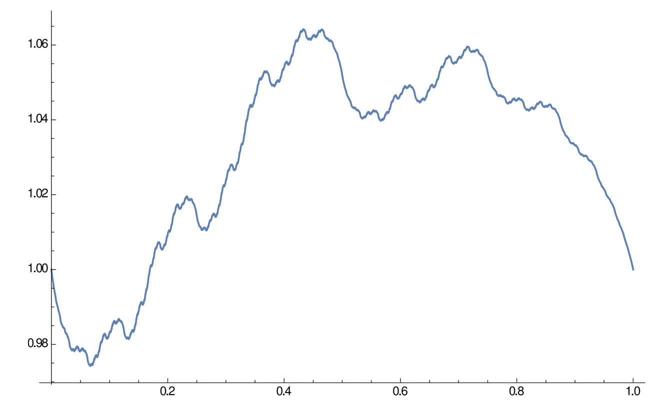
Asymptotic behavior of the summatory function of $(s_2(n))_{n \in \mathbb{N}_0}$

Definition: For each $n \in \mathbb{N}_0$, we define $A(n) = \sum_{1}^{n} s_2(n)$ and we set $A(0) := 0$. The sequence $(A(n))_{n \in \mathbb{N}_0}$ is the summatory function of the sequence $(s_2(n))_{n \in \mathbb{N}_0}$. The first few terms of $(A(n))_{\in \mathbb{N}_0}$ are

$$1, 3, 6, 9, 13, 18, 23, 27, 32, 39, 47, 54, 61, 69, 76, 81, 87, 96, 107, 117, \ldots.$$

Theorem [5]: There exists a continuous function Φ over $[0, 1)$ such that $\Phi(0) = 1$, $\lim_{\alpha \to 1^-} \Phi(\alpha) = 1$ and the sequence $(A(n))_{n \in \mathbb{N}_0}$ satisfies, for all $n \geq 1$,

$$A(n) = 3^{\log_2(n)}\, \Phi(\mathsf{relp}_2(n)) = N^{\log_2 3}\, \Phi(\mathsf{relp}_2(n)).$$



The function Φ in $[0, 1)$

[1] J.-P. Allouche, K. Scheicher, R. F. Tichy, Regular maps in generalized number systems, *Math. Slovaca* **50** (2000), 41–58.
[2] A. Carpi, C. Maggi, On synchronized sequences and their separators, *Theor. Inform. Appl.* **35** (2001), 513–524.
[3] J. Leroy, M. Rigo, M. Stipulanti, Generalized Pascal triangle for binomial coefficients of words, *Adv. Appl. Math.* **80** (2016), 24–47.
[4] J. Leroy, M. Rigo, M. Stipulanti, Counting the number of non-zero coefficients in rows of generalized Pascal triangles, submitted.
[5] J. Leroy, M. Rigo, M. Stipulanti, Summatory function of sequences counting subwords occurrences, submitted.
[6] M. Lothaire, *Combinatorics on Words*, Cambridge Mathematical Library, Cambridge University Press, (1997).
[7] J. Shallit, A generalization of automatic sequences, *Theoret. Comp. Sci.* **61** (1988), no. 1, 1–16.

LA LIBERTÉ DE CHERCHER

School of Combinatorics, Automata and Number Theory (CANT 2016)

Marseille, November 28 – December 2, 2016

Université de Liège