Generalized Pascal triangle for binomial coefficients of words: an overview

Joint work with Julien Leroy and Michel Rigo

Manon Stipulanti FRIA grantee

Mons Theoretical Computer Science Days
September 7, 2016

Pascal triangle

		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
m	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

Usual binomial coefficients of integers:

$$\binom{m}{k} = \frac{m!}{(m-k)!\,k!}$$

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Example: $u = 101001$ $\quad v = 101$

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Example: $u = \mathbf{101}001$ $v = 101$ 1 occurrence

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Example: $u = \mathbf{10}100\mathbf{1}$ $\qquad v = 101 \qquad$ 2 occurrences

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Example: $u = 101001$ $\quad v = 101 \quad$ 3 occurrences

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a “scattered” subword).

Example: $u = \mathbf{101001}$ $v = 101$ 4 occurrences

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Example: $u = 10\mathbf{1}\mathbf{0}0\mathbf{1}$ $\qquad v = 101 \qquad$ 5 occurrences

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Example: $u = 10\mathbf{1}0\mathbf{01}$ $\qquad v = 101 \qquad$ 6 occurrences

Binomial coefficient of finite words

Definition: A *finite word* is a finite sequence of letters belonging to a finite set called alphabet.

Binomial coefficient of words

Let u, v be two finite words.
The *binomial coefficient* $\binom{u}{v}$ of u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword).

Example: $u = 101001$ $\qquad v = 101$

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

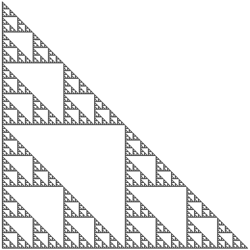
Remark:
Natural generalization of binomial coefficients of integers

With a one-letter alphabet $\{a\}$

$$\binom{a^m}{a^k} = \binom{\overbrace{a \cdots a}^{m \text{ times}}}{\underbrace{a \cdots a}_{k \text{ times}}} = \binom{m}{k} \quad \forall\, m, k \in \mathbb{N}$$

Link between the Pascal triangle and the Sierpiński gasket

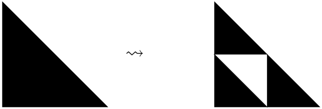
		k							
		0	1	2	3	4	5	6	7
	1	1	0	0	0	0	0	0	0
	1	1	1	0	0	0	0	0	0
	2	1	2	1	0	0	0	0	0
m	3	1	3	3	1	0	0	0	0
	4	1	4	6	4	1	0	0	0
	5	1	5	10	10	5	1	0	0
	6	1	6	15	20	15	6	1	0
	7	1	7	21	35	35	21	7	1



Usual binomial coefficients of integers:

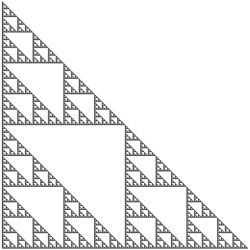
$$\binom{m}{k} = \frac{m!}{(m-k)!\,k!}$$

A way to build the Sierpiński gasket:



Link between the Pascal triangle and the Sierpiński gasket

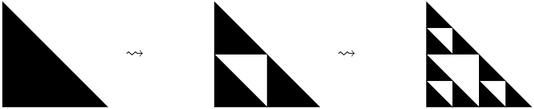
					k				
		0	1	2	3	4	5	6	7
	1	1	0	0	0	0	0	0	0
	1	1	1	0	0	0	0	0	0
	2	1	2	1	0	0	0	0	0
m	3	1	3	3	1	0	0	0	0
	4	1	4	6	4	1	0	0	0
	5	1	5	10	10	5	1	0	0
	6	1	6	15	20	15	6	1	0
	7	1	7	21	35	35	21	7	1



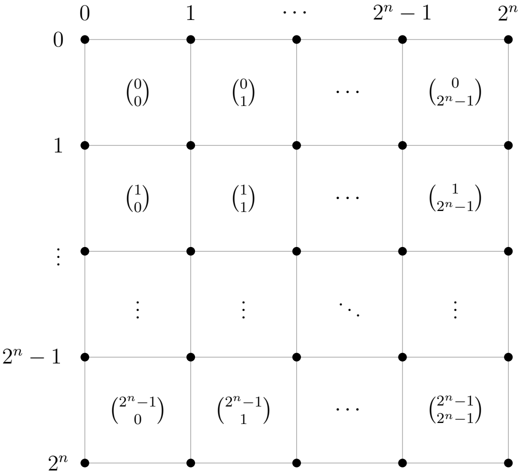
Usual binomial coefficients of integers:

$$\binom{m}{k} = \frac{m!}{(m-k)!\,k!}$$

A way to build the Sierpiński gasket:



- Grid: intersection between $\mathbb{N}^2$ and $[0, 2^n] \times [0, 2^n]$



$\mathbb{N}^2 \cap [0, 2^n]^2$

- Color the grid:
  Color the first 2^n rows and columns of the Pascal triangle

$$\left(\binom{m}{k} \bmod 2\right)_{0\le m,k<2^n}$$

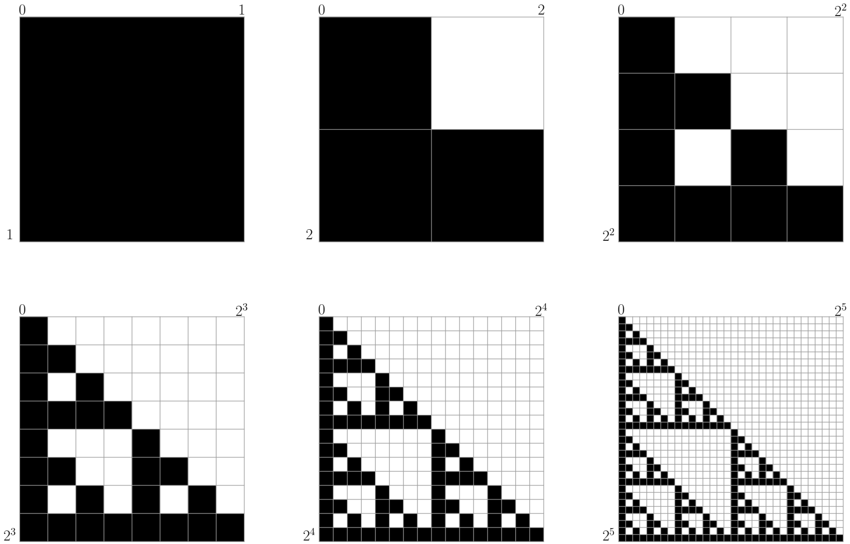
  in
  - white if $\binom{m}{k} \equiv 0 \bmod 2$
  - black if $\binom{m}{k} \equiv 1 \bmod 2$

- Color the grid:
  Color the first 2^n rows and columns of the Pascal triangle

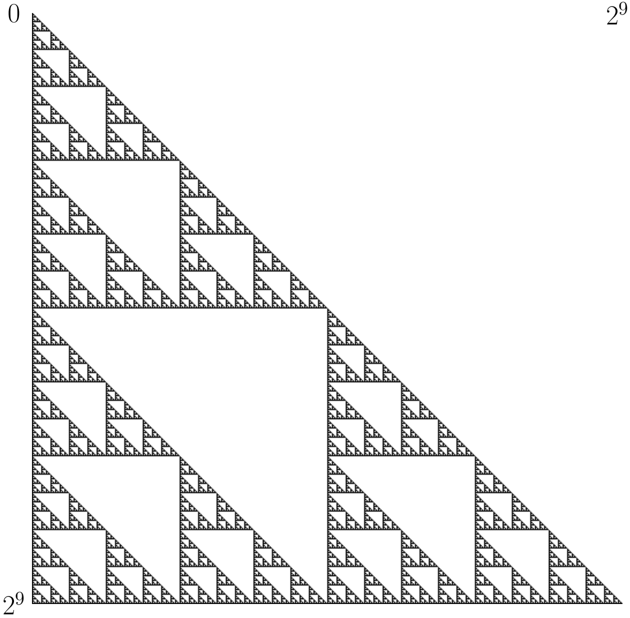
$$\left(\binom{m}{k} \bmod 2\right)_{0\leq m,k<2^n}$$

  in
  - white if $\binom{m}{k} \equiv 0 \bmod 2$
  - black if $\binom{m}{k} \equiv 1 \bmod 2$
- Normalize by a homothety of ratio $1/2^n$
  $\rightsquigarrow$ sequence belonging to $[0,1] \times [0,1]$

The first six elements of the sequence



The tenth element of the sequence



Theorem [von Haeseler, Peitgen, Skordev, 1992]

This sequence converges, for the Hausdorff distance, to the Sierpiński gasket (when n tends to infinity).

Theorem [von Haeseler, Peitgen, Skordev, 1992]

This sequence converges, for the Hausdorff distance, to the Sierpiński gasket (when n tends to infinity).

Definitions:

- *ϵ-fattening* of a subset $S \subset \mathbb{R}^2$

$$[S]_\epsilon = \bigcup_{x \in S} B(x, \epsilon)$$

- $(\mathcal{H}(\mathbb{R}^2), d_h)$ complete space of the non-empty compact subsets of $\mathbb{R}^2$ equipped with the *Hausdorff distance* d_h

$$d_h(S, S') = \min\{\epsilon \in \mathbb{R}_{\geq 0} \mid S \subset [S']_\epsilon \quad \text{and} \quad S' \subset [S]_\epsilon\}$$

Idea: binomial coefficients of integers
$\rightsquigarrow$ binomial coefficients of words

Idea: binomial coefficients of integers
$\rightsquigarrow$ binomial coefficients of words

Definitions:

- $\mathrm{rep}_2(n)$ greedy base-2 expansion of $n \in \mathbb{N}_{>0}$ beginning by 1
- $\mathrm{rep}_2(0) = \varepsilon$ where ε is the empty word

First values of the generalized Pascal triangle

$\rightsquigarrow$ base-2 expansions ordered genealogically (first by length, then lexicographically)

		v							
		ε	1	10	11	100	101	110	111
	ε	1	0	0	0	0	0	0	0
	1	1	1	0	0	0	0	0	0
	10	1	1	1	0	0	0	0	0
u	11	1	2	0	1	0	0	0	0
	100	1	1	2	0	1	0	0	0
	101	1	2	1	1	0	1	0	0
	110	1	2	2	1	0	0	1	0
	111	1	3	0	3	0	0	0	1

First values of the generalized Pascal triangle

$\rightsquigarrow$ base-2 expansions ordered genealogically (first by length, then lexicographically)

		v							
		ε	**1**	10	**11**	100	101	110	**111**
	ε	**1**	0	0	0	0	0	0	0
	1	**1**	**1**	0	0	0	0	0	0
	10	1	1	1	0	0	0	0	0
u	**11**	**1**	**2**	0	**1**	0	0	0	0
	100	1	1	2	0	1	0	0	0
	101	1	2	1	1	0	1	0	0
	110	1	2	2	1	0	0	1	0
	111	**1**	**3**	0	**3**	0	0	0	**1**

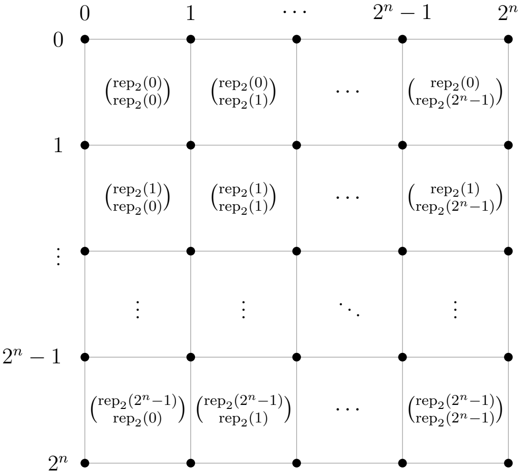
The classical Pascal triangle

Questions:

- After coloring and normalization can we expect the convergence to an analogue of the Sierpiński gasket?
- Could we describe this limit object ?

Same construction

- Grid: intersection between $\mathbb{N}^2$ and $[0, 2^n] \times [0, 2^n]$



$\mathbb{N}^2 \cap [0, 2^n]^2$

- Color the grid:
  Color the first 2^n rows and columns of the generalized Pascal triangle

$$\left(\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \bmod 2\right)_{0\le m,k<2^n}$$

  in

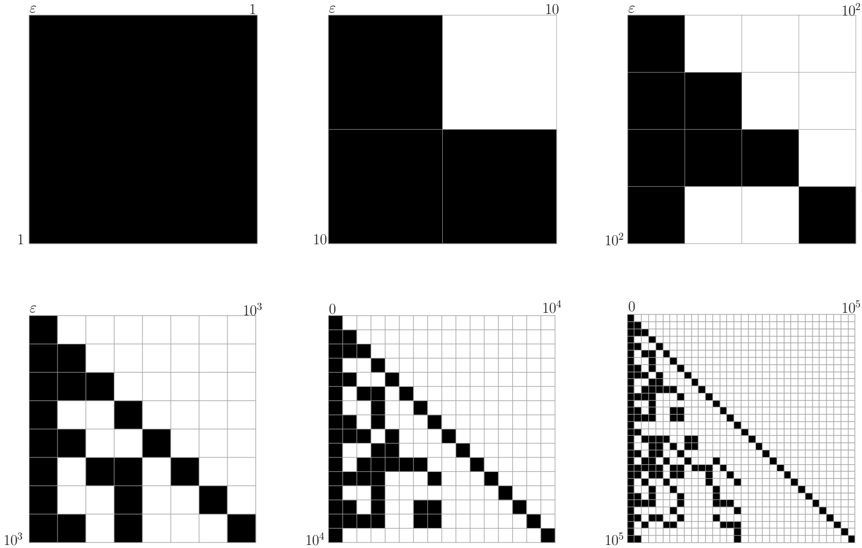
  - white if $\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \equiv 0 \bmod 2$
  - black if $\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \equiv 1 \bmod 2$

- Color the grid:
  Color the first 2^n rows and columns of the generalized Pascal triangle

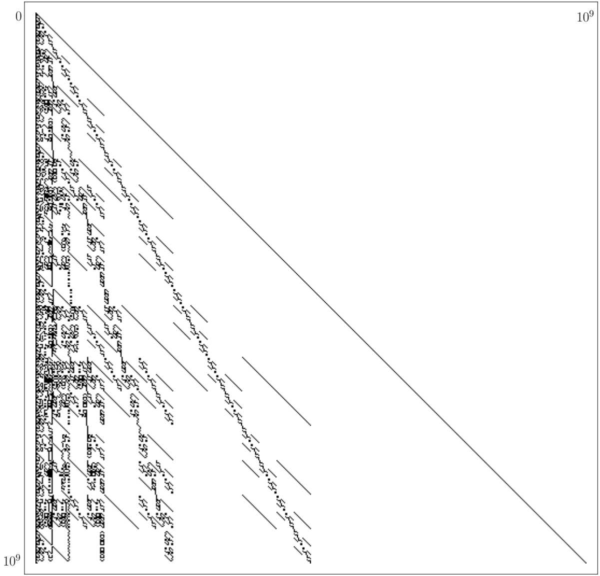
$$\left(\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \bmod 2 \right)_{0 \le m,k < 2^n}$$

  in
  - white if $\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \equiv 0 \bmod 2$
  - black if $\binom{\mathrm{rep}_2(m)}{\mathrm{rep}_2(k)} \equiv 1 \bmod 2$
- Normalize by a homothety of ratio $1/2^n$
  $\rightsquigarrow$ sequence belonging to $[0,1] \times [0,1]$

The first six elements of the sequence

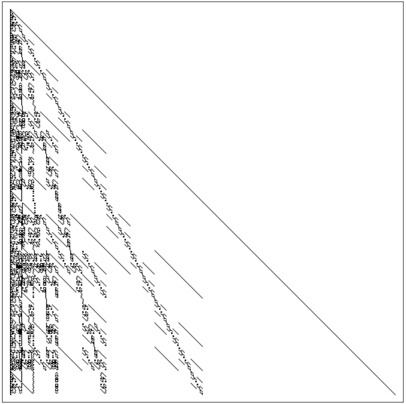


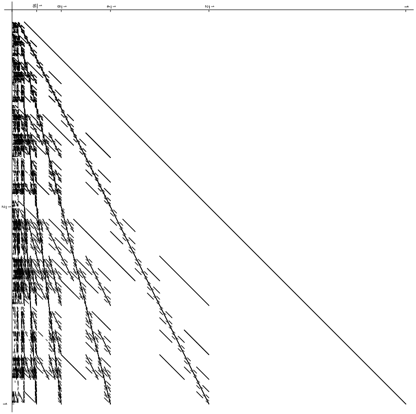
The tenth element of the sequence



Theorem [Leroy, Rigo, S., 2016]

The sequence converges to a limit object $\mathcal{L}$.





Topological closure of a union of segments described through a simple combinatorial property

Extension modulo p

Simplicity: coloring regarding the parity of binomial coefficients

Extension

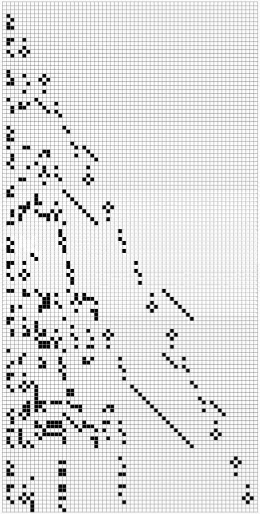
Everything still holds for binomial coefficients $\equiv r \bmod p$ with

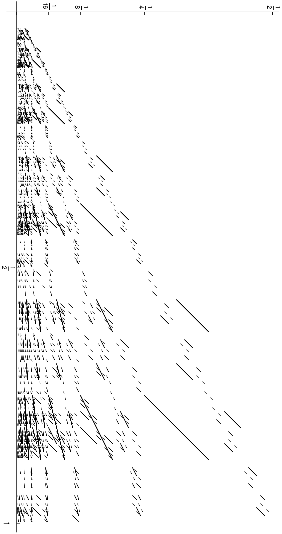
- base-2 expansions of integers
- p a prime
- $r \in \{1, \ldots, p-1\}$

Example with $p = 3$, $r = 2$

Left: binomial coefficients $\equiv 2 \bmod 3$

Right: estimate of the corresponding limit object





Counting the positive binomial coefficients

u \ v	ε	1	10	11	100	101	110	111	S
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**

Definition: $\forall n \geq 1$

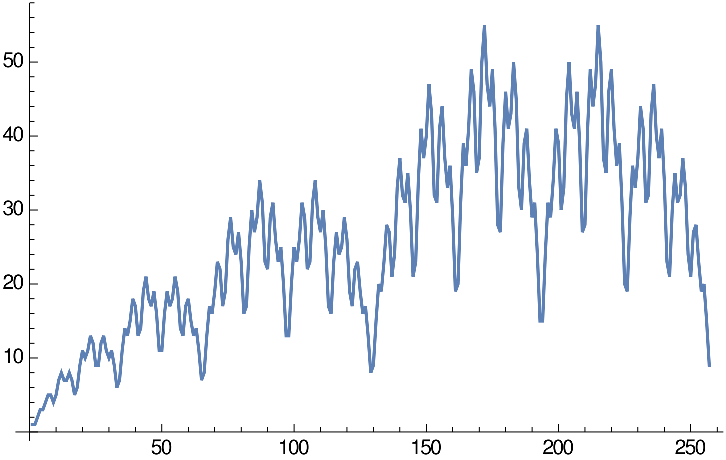
$$S(n) = \text{number of non-zero elements in the } n\text{th row of the generalized Pascal triangle} = \# \left\{ \binom{\text{rep}_2(n-1)}{\text{rep}_2(m)} > 0 \mid m \in \mathbb{N} \right\}$$

$$S(0) = 1$$

First few terms of $(S(n))_{n\geq 0}$:

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

The sequence $(S(n))_{n\geq 0}$ in the interval $[0, 256]$



Palindromic structure $\rightsquigarrow$ regularity

Some properties of the sequence $(S(n))_{n\geq 0}$

- 2-*kernel* of h

$$\begin{aligned}\mathcal{K}_2(h) &= \{h(n), h(2n), h(2n+1), h(4n), h(4n+1), h(4n+2), \ldots\} \\ &= \{(h(2^i n + j))_{n\geq 0} \mid i \geq 0 \text{ and } 0 \leq j < 2^i\}\end{aligned}$$

Some properties of the sequence $(S(n))_{n\geq 0}$

- 2-*kernel* of h

$$\begin{aligned}\mathcal{K}_2(h) &= \{h(n), h(2n), h(2n+1), h(4n), h(4n+1), h(4n+2), \ldots\} \\ &= \{(h(2^i n + j))_{n\geq 0} \mid i \geq 0 \text{ and } 0 \leq j < 2^i\}\end{aligned}$$

- 2-*regular* if there exist

$$(t_1(n))_{n\geq 0}, \ldots, (t_\ell(n))_{n\geq 0}$$

s.t. each $(t(n))_{n\geq 0} \in \mathcal{K}_2(h)$ is a $\mathbb{Z}$-linear combination of the t_j's

Some properties of the sequence $(S(n))_{n\geq 0}$

- 2-*kernel* of h

$$\begin{aligned}\mathcal{K}_2(h) &= \{h(n), h(2n), h(2n+1), h(4n), h(4n+1), h(4n+2), \ldots\} \\ &= \{(h(2^i n + j))_{n\geq 0} \mid i \geq 0 \text{ and } 0 \leq j < 2^i\}\end{aligned}$$

- 2-*regular* if there exist

$$(t_1(n))_{n\geq 0}, \ldots, (t_\ell(n))_{n\geq 0}$$

s.t. each $(t(n))_{n\geq 0} \in \mathcal{K}_2(h)$ is a $\mathbb{Z}$-linear combination of the t_j's

Proposition [Leroy, Rigo, S., 2016]

$(S(n))_{n\geq 0}$ is 2-regular.

- 2-*automatic* if the 2-kernel is finite

- 2-*automatic* if the 2-kernel is finite

Proposition [Leroy, Rigo, S., 2016]

$(S(n))_{n\geq 0}$ is not 2-automatic.

- 2-*automatic* if the 2-kernel is finite

Proposition [Leroy, Rigo, S., 2016]

$(S(n))_{n\geq 0}$ is not 2-automatic.

- 2-*synchronized* if the language

$$\{\mathrm{rep}_2(n, h(n)) \mid n \in \mathbb{N}\}$$

is accepted by some finite automaton

- 2-*automatic* if the 2-kernel is finite

Proposition [Leroy, Rigo, S., 2016]

$(S(n))_{n\geq 0}$ is not 2-automatic.

- 2-*synchronized* if the language

$$\{\text{rep}_2(n, h(n)) \mid n \in \mathbb{N}\}$$

is accepted by some finite automaton

Proposition [Leroy, Rigo, S., 2016]

$(S(n))_{n\geq 0}$ is not 2-synchronized.

Remark: 2-automatic $\subsetneq$ 2-synchronized $\subsetneq$ 2-regular.

The Fibonacci case

Definitions:

- $\text{rep}_F(n)$ greedy Fibonacci representation of $n \in \mathbb{N}_{>0}$ beginning by 1
- $\text{rep}_F(0) = \varepsilon$ where ε is the empty word

		v								
		ε	1	10	100	101	1000	1001	1010	S_F
	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**
u	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**

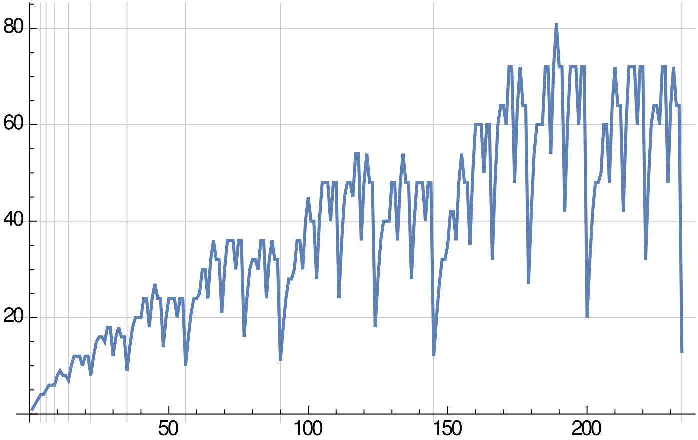
Definition: $\forall n \geq 0$

$$S_F(n) = \#\left\{ \binom{\text{rep}_F(n)}{\text{rep}_F(m)} > 0 \mid m \in \mathbb{N} \right\}$$

First few terms of $(S_F(n))_{n \geq 0}$:

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

The sequence $(S_F(n))_{n\geq 0}$ in the interval $[0, 233]$



F-kernel and F-regularity

2-kernel $\mathcal{K}_2(h)$ of a sequence h

- Select all the nonnegative integers whose base-2 expansion (with leading zeroes) ends with $w \in \{0,1\}^*$
- Evaluate h at those integers
- Let w vary in $\{0,1\}^*$

$\mathbf{w = 0}$

n	$\text{rep}_2(n)$	$h(n)$
0	ε	$\mathbf{h(0)}$
1	1	$h(1)$
2	**10**	$\mathbf{h(2)}$
3	11	$h(3)$
4	**100**	$\mathbf{h(4)}$
5	101	$h(5)$

F-kernel $\mathcal{K}_F(h)$ of a sequence h

- Select all the nonnegative integers whose Fibonacci representation (with leading zeroes) ends with $w \in \{0,1\}^*$
- Evaluate h at those integers
- Let w vary in $\{0,1\}^*$

n	$\text{rep}_F(n)$	$h(n)$
0	ε	$\mathbf{h(0)}$
1	1	$h(1)$
2	**10**	$\mathbf{h(2)}$
3	**100**	$\mathbf{h(3)}$
4	101	$h(4)$
5	**1000**	$\mathbf{h(5)}$

F-regular if there exist

$$(t_1(n))_{n\geq 0}, \dots, (t_\ell(n))_{n\geq 0}$$

s.t. each $(t(n))_{n\geq 0} \in \mathcal{K}_F(h)$ is a $\mathbb{Z}$-linear combination of the t_j's

F-regular if there exist

$$(t_1(n))_{n\geq 0}, \ldots, (t_\ell(n))_{n\geq 0}$$

s.t. each $(t(n))_{n\geq 0} \in \mathcal{K}_F(h)$ is a $\mathbb{Z}$-linear combination of the t_j's

Proposition [Leroy, Rigo, S., 2016]

$(S_F(n))_{n\geq 0}$ is F-regular.

In the literature, not so many sequences that have this kind of property

Work in progress

Study the

- behavior of the summatory function of $(S(n))_{n\geq 0}$
- behavior of the summatory function of $(S_F(n))_{n\geq 0}$

Work in progress

Study the

- behavior of the summatory function of $(S(n))_{n\geq 0}$
- behavior of the summatory function of $(S_F(n))_{n\geq 0}$

Example: $s_2(n)$ number of 1's in $\text{rep}_2(n)$
s_2 is 2-regular
summatory function $N \mapsto \sum_{j=0}^{N-1} s_2(j)$

Theorem [Delange, 1975]

$$\frac{1}{N}\sum_{j=0}^{N-1} s_2(j) = \frac{1}{2}\log_2 N + \mathcal{G}(\log_2 N) \qquad (1)$$

where $\mathcal{G}$ continuous, nowhere differentiable, periodic of period 1.

Theorem [Delange, 1975]

$$\frac{1}{N}\sum_{j=0}^{N-1} s_2(j) = \frac{1}{2}\log_2 N + \mathcal{G}(\log_2 N) \qquad (1)$$

where $\mathcal{G}$ continuous, nowhere differentiable, periodic of period 1.

Theorem [Allouche, Shallit, 2003]

Under some hypotheses, the summatory function of every k-regular sequence has a behavior analogous to (1).

Theorem [Delange, 1975]

$$\frac{1}{N}\sum_{j=0}^{N-1} s_2(j) = \frac{1}{2}\log_2 N + \mathcal{G}(\log_2 N) \qquad (1)$$

where $\mathcal{G}$ continuous, nowhere differentiable, periodic of period 1.

Theorem [Allouche, Shallit, 2003]

Under some hypotheses, the summatory function of every k-regular sequence has a behavior analogous to (1).

$\rightsquigarrow$ Replacing s_2 by S and S_F: same behavior as (1) but do not satisfy the previous theorem