Probabilistic study of the recurrence function of Sturmian sequences

Pablo Rotondo

Ongoing work with

Valérie Berthé, Eda Cesaratto, Brigitte Vallée, and Alfredo Viola

Universidad de la República, Uruguay

Aléa, 16-20 March, 2015.

Study in combinatorics of words.

Main aim: description of the finite factors of an infinite word u

- How many factors of length n? $\longrightarrow$ Complexity
- What are the gaps between them? $\longrightarrow$ Recurrence

Very easy when the word is eventually periodic !

Study in combinatorics of words.

Main aim: description of the finite factors of an infinite word u

– How many factors of length n? $\longrightarrow$ Complexity

– What are the gaps between them? $\longrightarrow$ Recurrence

Very easy when the word is eventually periodic !

Sturmian words:

the "simplest" binary infinite words that are not eventually periodic

Study in combinatorics of words.

Main aim: description of the finite factors of an infinite word u

– How many factors of length n? $\longrightarrow$ Complexity

– What are the gaps between them? $\longrightarrow$ Recurrence

Very easy when the word is eventually periodic !

Sturmian words:

the "simplest" binary infinite words that are not eventually periodic

The recurrence function is widely studied for Sturmian words.

Classical study : for each fixed Sturmian word,

what are the extreme bounds for the recurrence function?

Study in combinatorics of words.

Main aim: description of the finite factors of an infinite word u

– How many factors of length n? $\longrightarrow$ Complexity

– What are the gaps between them? $\longrightarrow$ Recurrence

Very easy when the word is eventually periodic !

Sturmian words:

the "simplest" binary infinite words that are not eventually periodic

The recurrence function is widely studied for Sturmian words.

Classical study : for each fixed Sturmian word,

what are the extreme bounds for the recurrence function?

Here, in a convenient probabilistic model, we perform a probabilistic study:

For a "random" sturmian word, and for a given "position",

– what is the mean value of the recurrence?

– what is the limit distribution of the recurrence?

Table of Contents

Complexity

Note: $\mathcal{L}_u(n)$ denotes the set of factors of length n in u.
Complexity function of an infinite word $u \in \mathcal{A}^{\mathbb{N}}$

$$p_u \colon \mathbb{N} \to \mathbb{N}\,, \qquad p_u(n) = |\mathcal{L}_u(n)|\ .$$

Some simple facts

$$p_u(n) \le |\mathcal{A}|^n\,, \qquad p_u(n) \le p_u(n+1)\,.$$

Property

$u \in \mathcal{A}^{\mathbb{N}}$ is not eventually periodic

$$\iff p_u(n) < p_u(n+1)$$

$$\implies n+1 \le p_u(n)\,.$$

Recurrence

Definition (Uniformly recurrent)

A word $u \in \mathcal{A}^{\mathbb{N}}$ is uniformly recurrent iff each finite factor appears infinitely often and with bounded gaps.

Recurrence

Definition (Uniformly recurrent)

A word $u \in \mathcal{A}^{\mathbb{N}}$ is uniformly recurrent iff each finite factor appears infinitely often and with bounded gaps.

Definition (Recurrence function)

Let $u \in \mathcal{A}^{\mathbb{N}}$

$$R_{\langle u\rangle}(n) = \inf\{m \in \mathbb{N} : \text{for each } w \in \mathcal{L}_u(m)$$
$$\text{, every } U \in \mathcal{L}_u(n) \text{ is a factor of } w\}\,.$$

Recurrence

Definition (Uniformly recurrent)

A word $u \in \mathcal{A}^{\mathbb{N}}$ is uniformly recurrent iff each finite factor appears infinitely often and with bounded gaps.

Definition (Recurrence function)

Let $u \in \mathcal{A}^{\mathbb{N}}$

$$R_{\langle u \rangle}(n) = \inf\{m \in \mathbb{N} : \text{for each } w \in \mathcal{L}_u(m) \\ \text{, every } U \in \mathcal{L}_u(n) \text{ is a factor of } w\}\,.$$

In words, given any window of length $m = R_{\langle u \rangle}(n)$ in u, we can find every factor U of length $|U| = n$ within it.

Sturmian words

$\Longrightarrow$ "simplest" words that are not eventually periodic.

Sturmian words

$\Longrightarrow$ "simplest" words that are not eventually periodic.

Definition

A word $u \in \{\mathbf{0}, \mathbf{1}\}^{\mathbb{N}}$ is Sturmian iff $p_u(n) = n + 1$ for each $n \geq 0$.

- All Sturmian words are uniformly recurrent.

Sturmian words

$\Longrightarrow$ "simplest" words that are not eventually periodic.

Definition

A word $u \in \{\mathbf{0}, \mathbf{1}\}^{\mathbb{N}}$ is Sturmian iff $p_u(n) = n + 1$ for each $n \geq 0$.

- All Sturmian words are uniformly recurrent.

Explicit construction

A word u is Sturmian iff there are $\alpha, \beta \in [0, 1[$, with α irrational, such that

$$u_n = \lfloor \alpha\,(n+1) + \beta \rfloor - \lfloor \alpha\, n + \beta \rfloor \text{ for all } n \geq 0\,,$$

or

$$u_n = \lceil \alpha\,(n+1) + \beta \rceil - \lceil \alpha\, n + \beta \rceil \text{ for all } n \geq 0\,.$$

Denote by $\underline{S}(\alpha, \beta)$ and $\overline{S}(\alpha, \beta)$ respectively the words produced.

The recurrence function of Sturmian words

Let u be $\underline{S}(\alpha, \beta)$ or $u = \overline{S}(\alpha, \beta)$, then

- $R_{\langle u \rangle}(n)$ depends only on α. Thus we write $R_{\alpha}(n)$.

The recurrence function of Sturmian words

Let u be $\underline{S}(\alpha, \beta)$ or $u = \overline{S}(\alpha, \beta)$, then

- $R_{\langle u \rangle}(n)$ depends only on α. Thus we write $R_{\alpha}(n)$.
- $(R_{\alpha}(n))_{n \in \mathbb{N}}$ depends only on the continuants of α.

The continuant $q_k(\alpha)$ is the denominator of the k-th convergent of the continued fraction expansion of α. It is an increasing sequence.

Theorem (Morse, Hedlund, 1940)

We have

$$R_{\alpha}(n) = n - 1 + q_{k-1}(\alpha) + q_k(\alpha)\,,$$

when $q_{k-1}(\alpha) \leq n < q_k(\alpha)$.

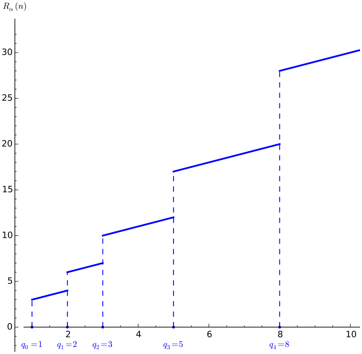


Figure: Recurrence function for $\alpha = \varphi^2$, with $\varphi = \frac{\sqrt{5}-1}{2}$.

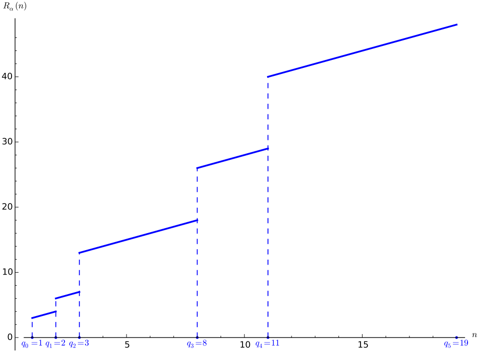


Figure: Recurrence function for $\alpha = 1/e$.

Classical results

Proposition

For all α

$$\liminf \frac{R_\alpha(n)}{n} \leq 3\,.$$

Theorem (Morse, Hedlund, 1940)

For α-a.e.

$$\limsup \frac{R_\alpha(n)}{n\,\log n} = \infty\,, \quad \textit{and} \quad \limsup \frac{R_\alpha(n)}{n\,(\log n)^c} = 0\,,$$

whenever $c > 1$.

Classical results

Proposition

For all α

$$\liminf \frac{R_\alpha(n)}{n} \leq 3\,.$$

Proof.

Take the sequence $n_k = \left\lfloor \frac{q_k+q_{k-1}}{2} \right\rfloor$. $\square$

Theorem (Morse, Hedlund, 1940)

For α-a.e.

$$\limsup \frac{R_\alpha(n)}{n\,\log n} = \infty\,, \quad \textit{and} \quad \limsup \frac{R_\alpha(n)}{n\,(\log n)^c} = 0\,,$$

whenever $c > 1$.

Usual studies of $R_\alpha(n)$

- consider all possible sequences of indices n.
- give information on extreme cases.
- give results for a.e. α.

Here we

- study particular sequences of indices n depending on α. Fixed relative position on the intervals $[q_{k-1}(\alpha), q_k(\alpha)[$.
- perform a probabilistic study (random α !).

We work with particular families of indices n

Relative position sequence for α

Given $\mu \in]0,1]$ the sequence

$$n_k^{\langle \mu \rangle}(\alpha) = q_{k-1}(\alpha) + \lfloor \mu \left(q_k(\alpha) - q_{k-1}(\alpha) \right) \rfloor$$

is called the subsequence of position μ of α.

We work with particular families of indices n

Relative position sequence for α

Given $\mu \in]0,1]$ the sequence

$$n_k^{\langle\mu\rangle}(\alpha) = q_{k-1}(\alpha) + \lfloor \mu\,(q_k(\alpha) - q_{k-1}(\alpha)) \rfloor$$

is called the subsequence of position μ of α.

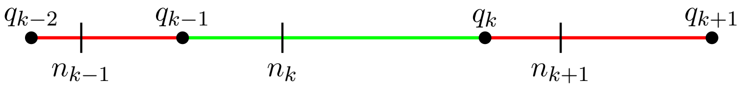


Figure: Sequence of indices n for $\mu = 1/3$.

We study

- the behaviour of

$$\frac{R_\alpha(n)}{n}, \quad n = n_k^{\langle\mu\rangle} = q_{k-1} + \lfloor \mu\,(q_k - q_{k-1}) \rfloor$$

  for a fixed relative position μ within $[q_{k-1}, q_k[$.

  Note. $n_k^{\langle\mu\rangle}$ is a variable depending on $\alpha \in \mathcal{I}$.

- what happens when α is drawn uniformly from $\mathcal{I} = [0, 1]$.

We study

- the behaviour of

$$\frac{R_\alpha(n)}{n}, \quad n = n_k^{\langle\mu\rangle} = q_{k-1} + \lfloor \mu\,(q_k - q_{k-1}) \rfloor$$

  for a fixed relative position μ within $[q_{k-1}, q_k[$.

  Note. $n_k^{\langle\mu\rangle}$ is a variable depending on $\alpha \in \mathcal{I}$.

- what happens when α is drawn uniformly from $\mathcal{I} = [0, 1]$.

We perform an asymptotic study of the sequence

$$S_k^{\langle\mu\rangle} = \frac{R_\alpha(n)+1}{n}, \quad n = n_k^{\langle\mu\rangle}$$

- Limit of its expected value with $\mu \in]0, 1]$ fixed.
- Limit of its distribution with $\mu \in [0, 1]$ fixed.

Expectation

Theorem

For each $\mu \in]0,1]$*, the sequence of random variables* $S_k^{\langle\mu\rangle}$ *satisfies*

$$\mathbb{E}[S_k^{\langle\mu\rangle}] = 1 + \frac{1}{\log 2}\,\frac{|\log\mu|}{1-\mu} + O\left(\frac{\varphi^{2k}}{\mu}\right) + O\left(\varphi^k\,\frac{|\log\mu|}{1-\mu}\right),$$

where $\varphi = \frac{\sqrt{5}-1}{2} \doteq 0.6180339\ldots$ *and the constants corresponding to the* O*-terms are uniform in* μ *and* k*.*

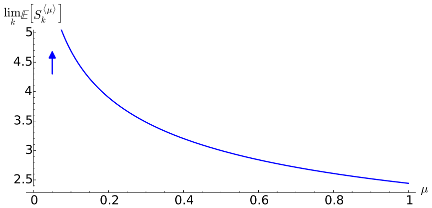


Figure: Limit of the expected value as a function of μ.

Distribution

Theorem

For each $\mu \in [0, 1]$ *with* $\mu \neq \frac{1}{2}$, *the sequence* $S_k^{\langle \mu \rangle}$ *has a limit density*

$$s_\mu(x) = \frac{1}{\log 2\,(x-1)\ |2-\mu-x\,(1-\mu)|}\,\mathbf{1}_{I_\mu}(x)\,,$$

with I_μ *being the interval with endpoints* 3 *and* $1 + 1/\mu$.
For all $b \geq \min\{3,\ 1 + \frac{1}{\mu}\}$

$$\Pr\left(S_k^{\langle \mu \rangle} \leq b\right) = \int_0^b s_\mu(x)dx + \frac{1}{b}\,O\left(\varphi^k\right)\,,$$

where the constant of the O*-term is uniform in* b *and* k.
When $|\mu - \frac{1}{2}| \geq \epsilon$ *for a fixed* $\epsilon > 0$, *it is also uniform in* μ.

Interval	Asymptotic Pr	Empirical Pr
$[3.0, 3.0]$	0.0	0.0
$[3.0, 3.5]$	0.4854	0.485237
$[3.0, 4.0]$	0.7369	0.737139
$[3.0, 4.5]$	0.8931	0.893511
$[3.0, 5.0]$	1.0	1.0

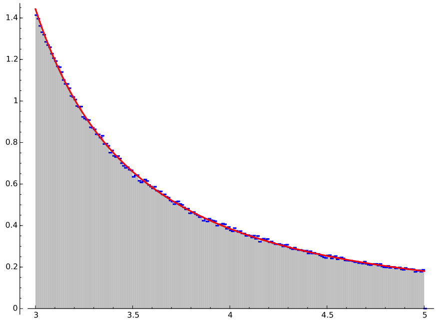


Figure: For $\mu = \frac{1}{4}$, the limiting distribution for $S_k^{\langle \mu \rangle}$ compared to the results of a simulation with $N = 10^6$ iterations and $k = 25$.

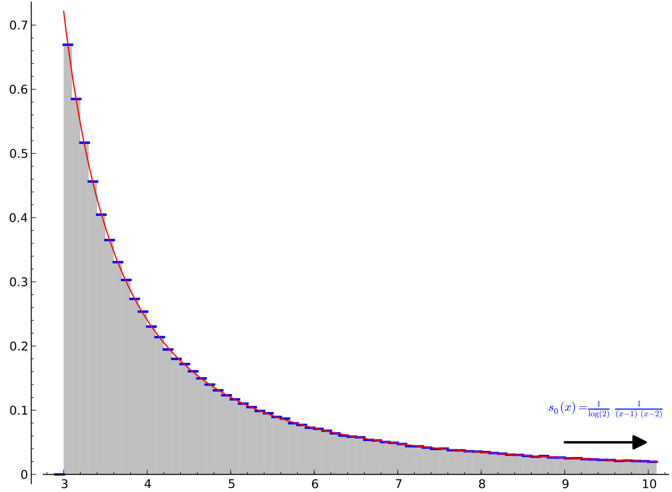


Figure: Scaled histogram for $\mu = 0$ compared to $s_0(x)$. Here $k = 25$, $N = 10^6$, and the bin width is $\delta = \frac{1}{10}$.

General overview of the proof

The proof is divided into four steps

i) Drop the integer part in $S_k^{\langle\mu\rangle}$ getting

$$\tilde{S}_k^{\langle\mu\rangle} = 1 + \frac{q_k + q_{k-1}}{q_{k-1} + \mu\,(q_k - q_{k-1})}\,,$$

which depends only on $\frac{q_{k-1}}{q_k}$. Indeed

$$\tilde{S}_k^{\langle\mu\rangle} = f_\mu\left(\frac{q_{k-1}}{q_k}\right)\,, \qquad f_\mu(x) = 1 + \frac{1+x}{x + \mu\,(1-x)}\,.$$

ii) The expected value and the distribution of $\tilde{S}_k^{\langle\mu\rangle}$ are expressed in terms of the iterates of the Perron-Frobenius operator **H**.

iii) Using spectral properties of **H**, when acting on the space of bounded variation $\Longrightarrow$ we obtain the asymptotics.

iv) Finally we return from $\tilde{S}_k^{\langle\mu\rangle}$ to $S_k^{\langle\mu\rangle}$.

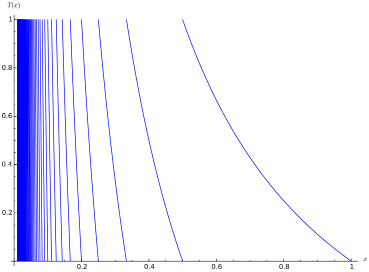
The Dynamic System

The Gauss map

$$T(x) = \left\{ \frac{1}{x} \right\} = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor .$$

The inverse branches of the Gauss map are given by

$$\mathcal{H} = \left\{ h_m \colon x \mapsto \frac{1}{m+x} \quad : \quad m \geq 1 \right\} .$$



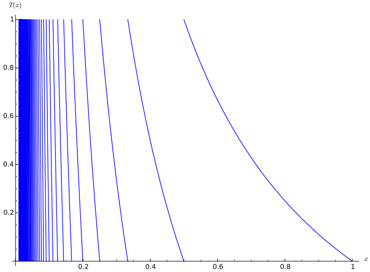
The Dynamic System

The Gauss map

$$T(x) = \left\{ \frac{1}{x} \right\} = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor .$$

The inverse branches of the Gauss map are given by

$$\mathcal{H} = \left\{ h_m \colon x \mapsto \frac{1}{m+x} \quad : \quad m \geq 1 \right\} .$$



Finally, the inverse branches for the composition T^k are

$$\mathcal{H}^k = \{ h_{m_1, m_2, \ldots m_k} = h_{m_1} \circ h_{m_2} \circ \ldots \circ h_{m_k} \quad : \quad m_1, \ldots, m_k \geq 1 \} .$$

The linear fractional transformation $h_{m_1,\ldots,m_k} \in \mathcal{H}^k$ can be written as

$$h_{m_1,\ldots,m_k}(x) = \cfrac{1}{m_1 + \cfrac{1}{\ddots + \cfrac{1}{m_k + x}}} = \frac{p_{k-1}\, x + p_k}{q_{k-1}\, x + q_k}\,,$$

and satisfies the mirror property

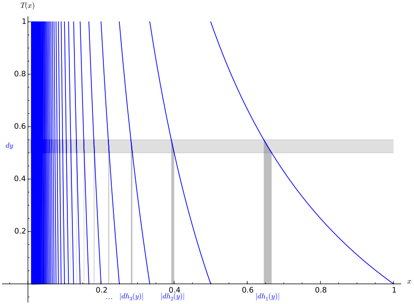
$$h_{m_k,\ldots,m_1}(x) = \cfrac{1}{m_k + \cfrac{1}{\ddots + \cfrac{1}{m_1 + x}}} = \frac{p_{k-1}\, x + {\color{blue}q_{k-1}}}{{\color{red}p_k}\, x + q_k}\,.$$

The Perron-Frobenius operator **H**

Question: If $g \in \mathcal{C}^0(\mathcal{I})$ were the density of $\alpha \Longrightarrow$ density of $T(\alpha)$?

The Perron-Frobenius operator $\mathbf{H}$

Question: If $g \in \mathcal{C}^0(\mathcal{I})$ were the density of $\alpha \Longrightarrow$ density of $T(\alpha)$?



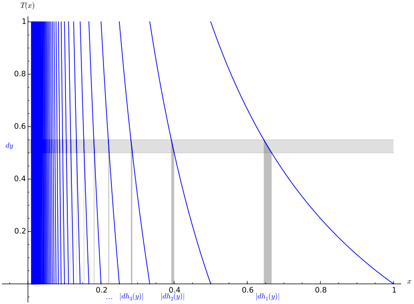
Answer: The density is

$$\mathbf{H}[g](x) = \sum_{h\in\mathcal{H}} \left|h'(x)\right| \, g\left(h(x)\right)$$

$$= \sum_{m=1}^{\infty} \frac{1}{(m+x)^2}\, g\left(\frac{1}{m+x}\right) .$$

The Perron-Frobenius operator $\mathbf{H}$

Question: If $g \in \mathcal{C}^0(\mathcal{I})$ were the density of $\alpha \Longrightarrow$ density of $T(\alpha)$?



Answer: The density is

$$\mathbf{H}[g](x) = \sum_{h \in \mathcal{H}} \left| h'(x) \right| \, g\left(h(x)\right)$$
$$= \sum_{m=1}^{\infty} \frac{1}{(m+x)^2} \, g\left(\frac{1}{m+x}\right) .$$

In general $T^k(\alpha)$ has density

$$\mathbf{H}^k[g](x) = \sum_{h \in \mathcal{H}^k} \left| h'(x) \right| \, g\left(h(x)\right) .$$

Evaluating at $x = 0$

$$\mathsf{H}^k[g](0) = \sum_{m_1,\ldots,m_k \geq 1} \frac{1}{q_k^2}\, g\left(\frac{p_k}{q_k}\right) .$$

Evaluating at $x = 0$

$$\mathbf{H}^k[g](0) = \sum_{m_1,\ldots,m_k \geq 1} \frac{1}{q_k^2}\, g\left(\frac{p_k}{q_k}\right) .$$

Evaluating at $x = 0$

$$\mathbf{H}^k[g](0) = \sum_{m_1,\ldots,m_k \geq 1} \frac{1}{q_k^2}\, g\left(\frac{p_k}{q_k}\right) .$$

Applying the mirror property of continued fractions implies, since we sum over all k-tuples $\Longrightarrow$

$$\mathbf{H}^k[g](0) = \sum_{m_1,\ldots,m_k \geq 1} \frac{1}{q_k^2}\, g\left(\frac{q_{k-1}}{q_k}\right) .$$

The expectation

- $\tilde{S}_k^{\langle\mu\rangle}$ is a step function, constant on $h_{m_1,\ldots,m_k}(\mathcal{I})$.
- The length of $h_{m_1,\ldots,m_k}(\mathcal{I})$ is $\frac{1}{q_k(q_k+q_{k-1})}$.
- The value of $\tilde{S}_k^{\langle\mu\rangle}$ on $h_{m_1,\ldots,m_k}(\mathcal{I})$ is $f_\mu\left(\frac{q_{k-1}}{q_k}\right)$.

The expectation

- $\tilde{S}_k^{\langle\mu\rangle}$ is a step function, constant on $h_{m_1,\ldots,m_k}(\mathcal{I})$.
- The length of $h_{m_1,\ldots,m_k}(\mathcal{I})$ is $\frac{1}{q_k\,(q_k+q_{k-1})}$.
- The value of $\tilde{S}_k^{\langle\mu\rangle}$ on $h_{m_1,\ldots,m_k}(\mathcal{I})$ is $f_\mu\left(\frac{q_{k-1}}{q_k}\right)$.

Since $\{h(\mathcal{I}) : h \in \mathcal{H}^k\}$ is a partition of $(0,1)$

$$\begin{aligned}
\mathbb{E}\left[\tilde{S}_k^{\langle\mu\rangle}\right] &= \sum_{m_1,\ldots,m_k\geq 1} \frac{1}{q_k\;(q_k+q_{k-1})}\, f_\mu\left(\frac{q_{k-1}}{q_k}\right) \\
&= \sum_{m_1,\ldots,m_k\geq 1} \frac{1}{q_k^2}\,\frac{f_\mu(x)}{1+x} \qquad \left(\text{where } x = \frac{q_{k-1}}{q_k}\right) \\
&= \mathbf{H}^k\left[\frac{f_\mu(x)}{1+x}\right](0)\,,
\end{aligned}$$

and

$$\Pr\left(\tilde{S}_k^{\langle\mu\rangle} \leq b\right) = \mathbb{E}\left[\mathbf{1}_{\leq b}\circ \tilde{S}_k^{\langle\mu\rangle}\right] = \mathbf{H}^k\left[\frac{\mathbf{1}_{\leq b}\circ f_\mu(x)}{1+x}\right](0)\,.$$

Analytic study of **H**

The operator **H** acts on the Banach space $\mathrm{BV}(\mathcal{I})$ of functions of bounded variation, with norm

$$\|f\|_{BV} = V_0^1(f) + \|f\|_1 \,.$$

Analytic study of **H**

The operator **H** acts on the Banach space BV($\mathcal{I}$) of functions of bounded variation, with norm

$$\|f\|_{BV} = V_0^1(f) + \|f\|_1 \,.$$

The following properties are classical

- Dominant eigenvalue: $\lambda = 1$, it is simple too.
- Dominant eigenfunction: $\psi(x) = \frac{1}{\log 2}\,\frac{1}{1+x}$.
- Spectral gap: subdominant spectral radius φ^2 for $\varphi = \frac{\sqrt{5}-1}{2}$.

For any $g \in \mathrm{BV}(\mathcal{I})$

$$\mathbf{H}^k[g](x) = \frac{1}{\log 2}\,\frac{1}{1+x}\,\int_0^1 g(x)dx + O\left(\varphi^{2k}\,\|g\|_{\mathsf{BV}}\right) \,.$$

Going back

$$\begin{aligned}\mathbb{E}\left[\tilde{S}_k^{\langle\mu\rangle}\right] &= \mathbf{H}^k\left[\frac{f_\mu(x)}{1+x}\right](0)\\ &= \frac{1}{\log 2}\int_0^1 \frac{f_\mu(x)}{1+x}dx + O\left(\varphi^{2k}\left\|\frac{f_\mu(x)}{1+x}\right\|_{\mathsf{BV}}\right)\\ &= 1+\frac{1}{\log 2}\frac{|\log\mu|}{1-\mu}+O\left(\frac{\varphi^{2k}}{\mu}\right),\end{aligned}$$

the other error term in our theorem comes from $\left|\tilde{S}_k^{\langle\mu\rangle} - S_k^{\langle\mu\rangle}\right|$.

Going back

$$\begin{aligned}\mathbb{E}\left[\tilde{S}_k^{\langle\mu\rangle}\right] &= \mathbf{H}^k\left[\frac{f_\mu(x)}{1+x}\right](0)\\ &= \frac{1}{\log 2}\int_0^1 \frac{f_\mu(x)}{1+x}dx + O\left(\varphi^{2k}\left\|\frac{f_\mu(x)}{1+x}\right\|_{\mathsf{BV}}\right)\\ &= 1+\frac{1}{\log 2}\frac{|\log\mu|}{1-\mu} + O\left(\frac{\varphi^{2k}}{\mu}\right),\end{aligned}$$

the other error term in our theorem comes from $\left|\tilde{S}_k^{\langle\mu\rangle} - S_k^{\langle\mu\rangle}\right|$.

$$\begin{aligned}\Pr\left(\tilde{S}_k^{\langle\mu\rangle}\le b\right) &= \mathbf{H}^k\left[\frac{\mathbf{1}_{\le b}\circ f_\mu(x)}{1+x}\right](0)\\ &= \frac{1}{\log 2}\int_0^1 \frac{\mathbf{1}_{\le b}\circ f_\mu(x)}{1+x}dx + O\left(\varphi^{2k}\left\|\frac{\mathbf{1}_{\le b}\circ f_\mu(x)}{1+x}\right\|_{\mathsf{BV}}\right)\\ &= \text{here consider the inverse of } f_\mu \text{ in } I_\mu\,.\end{aligned}$$

Possible extensions

It is possible to extend the study to

- Make $\mu_k \to 0$ as $k \to \infty$ as well.
- Reals with bounded $m_k \leq M \implies$ Hausdorff measure.
- Quadratic irrationals $\implies$ their CFE is eventually periodic.

References

P. Alessandri, and V. Berthé,
Three distance theorems and combinatorics on words,
L'Enseignement Mathématique, 44, pp. 103–132, 1998.

E. Cesaratto, and B. Vallée, Pseudo-randomness of a random Kronecker sequence. An instance of dynamical analysis,
To appear as a chapter in the book Combinatorics, Words and Symbolic Dynamics (ed. V. Berthé and M. Rigo).

J. Bourdon, B. Daireaux, and B. Vallée
Dynamical analysis of α-Euclidean Algorithms,
Journal of Algorithms, 44, pp. 246-285, 2002.

M. Iosifescu, and C. Kraaaikamp
Metrical Theory of Continued Fractions,
Collection Mathematics and Its Applications, Vol 547, Kluwer Academic Press, 2002.