Two Arithmetical Sources and Their Associated Tries

Valérie Berthé

Université de Paris, CNRS, IRIF, F-75006, Paris, France
https://www.irif.fr/~berthe/
berthe@irif.fr

Eda Cesaratto

CONICET and Instituto del Desarrollo Humano, Universidad Nacional de General Sarmiento, Buenos Aires, Argentina
ecesaratto@campus.ungs.edu.ar

Frédéric Paccaut

LAMFA CNRS UMR 7352, Université de Picardie Jules Verne, Amiens, France
frederic.paccaut@u-picardie.fr

Pablo Rotondo

LITIS, Université de Rouen, France
pablo.rotondo@univ-rouen.fr

Martín D. Safe

Departamento de Matemática, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina
INMABB, Universidad Nacional del Sur (UNS)-CONICET, Bahía Blanca, Argentina
msafe@uns.edu.ar

Brigitte Vallée

GREYC, CNRS and Université de Caen, France
brigitte.vallee@unicaen.fr

Abstract

This article is devoted to the study of two arithmetical sources associated with classical partitions, that are both defined through the mediant of two fractions. The Stern-Brocot source is associated with the sequence of all the mediants, while the Sturm source only keeps mediants whose denominator is "not too large". Even though these sources are both of zero Shannon entropy, with very similar Renyi entropies, their probabilistic features yet appear to be quite different. We then study how they influence the behaviour of tries built on words they emit, and we notably focus on the trie depth.

The paper deals with Analytic Combinatorics methods, and Dirichlet generating functions, that are usually used and studied in the case of good sources with positive entropy. To the best of our knowledge, the present study is the first one where these powerful methods are applied to a zero-entropy context. In our context, the generating function associated with each source is explicit and related to classical functions in Number Theory, as the ζ function, the double ζ function or the transfer operator associated with the Gauss map. We obtain precise asymptotic estimates for the mean value of the trie depth that prove moreover to be quite different for each source. Then, these sources provide explicit and natural instances which lead to two unusual and different trie behaviours.

2012 ACM Subject Classification Theory of computation → Pattern matching; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases Combinatorics of words, Information Theory, Probabilistic analysis, Analytic combinatorics, Dirichlet generating functions, Sources, Partitions, Trie structure, Continued fraction expansion, Farey map, Sturm words, Transfer operator

Digital Object Identifier 10.4230/LIPIcs.AofA.2020.4

31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020).
Editors: Michael Drmota and Clemens Heuberger; Article No. 4; pp. 4:1–4:19
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Funding Laboratory IRP SINFIN, Projects DynA3S and CoDys (French ANR), Project AleaEnAmSud (STICAmSud), Projet RIN Alenor (Regional Project from French Normandy)
Eda Cesaratto: Partially supported by Universidad Nacional de General Sarmiento Grant 30/3307.
Martín D. Safe: Partially supported by Universidad Nacional del Sur Grant PGI 24/L115.

1 Introduction

A source is a way of producing random words from a given alphabet (see Definition 1). We study two arithmetical sources, each of them being associated with a classical family of partitions. These two families are defined through the mediant $(a + b)/(c + d)$ of the two fractions a/b and c/d. The first one, the Stern-Brocot source, is defined with the sequence of all the mediants, whereas the second one, the Sturm source, only keeps mediants whose denominator is "not too large". Even though the probabilistic features of the two sources appear to be quite different, they are both of zero Shannon entropy and their Renyi entropies appear to be quite similar. In Information Theory contexts, the trie structure built on words emitted by the source is a powerful tool for comparing words emitted by a source, and the shape of the trie – notably the average length of a branch, called the depth – can be viewed as a "measure" of the quality of the source. We then may expect that the trie depth behaves in a different way for each source, and provides a tool which strongly differentiates the two sources.

The probabilistic behaviour of the depth D_n of a trie built on n independent infinite words emitted by simple sources (memoryless and Markov sources) has been largely studied (see the book by Szpankowski [20] for a complete review in this case). In the context of "good" dynamical sources (with an entropy $\mathcal{E} > 0$) introduced in [23], the average-case analysis was first developed in [5], and [11], then a distributional analysis that exhibits a limit Gaussian law for D_n was performed in [4]. In all cases, for $n \to \infty$, the asymptotic mean value $\mathbb{E}[D_n]$ is of logarithmic order and involves the entropy, with the estimate $\mathbb{E}[D_n] \sim (1/\mathcal{E}) \log n$. The moments $\mathbb{E}[D_n^k]$ are proven to be of order $\Theta(\log^k n)$.

For sources of zero entropy, and to the best of our knowledge, the analysis of trie depth has not yet been performed in a general context. We study here the two arithmetical sources, associated with classical partitions, that have been previously presented. We obtain two results. First, for the two sources, all the moments of order $k \geq 2$ of the trie depth D_n behave in a similar way, as they are all infinite. However, the mean value $\mathbb{E}[D_n]$ exhibits a strong difference between the two sources, as $\mathbb{E}[D_n]$ is of order $\Theta(\log^2 n)$ for the Stern-Brocot source, and of order $\Theta(\sqrt{n})$ for the Sturm source.

Plan of the paper. Section 2 recalls the general context of sources, and focuses on the Analytic Combinatorics point of view. It introduces the two sources, and provides expressions for their Dirichlet generating functions (DGF's). Section 3 is devoted to tries, and focuses on the trie depth. It also presents the main analytical tool, the Rice formula, whose application is based on the tameness of the DGF's. Section 4 proves the tameness of DGF's in the two cases, and states the main result.

2 Sources, partitions, Dirichlet generating functions

We first recall two definitions of sources, and introduce the generating functions. We explain their roles in the analysis, notably for good sources. We then present the two sources of interest, with their DGF, and explain why the various notions of entropies are essentially the same for the two sources.

2.1 Sources and partitions

We first give a definition of a source, as it appears in Information Theory contexts.

▶ **Definition 1.** *A* probabilistic source $\mathcal{S}$ *over the finite (ordered) alphabet* $\Sigma := [0..r-1]$ *is a sequence* $Y := (Y_0, Y_1, \ldots, Y_i, \ldots)$ *of random variables* Y_i *with values in* Σ.

The value of the random variable Y_i is the symbol emitted by the source at the discrete time $t = i$, and the value of the sequence $Y \in \Sigma^{\mathbb{N}}$ is the (infinite) word emitted by the source.

Consider a finite word $w \in \Sigma^\star$, and denote by p_w the probability that Y begins with the prefix w. The set $(p_w)_{w\in\Sigma^\star}$ is called the set of *fundamental probabilities*, and the set $(p_w)_{w\in\Sigma^k}$ is the set of fundamental probabilities of *depth* k. We moreover assume

$$\pi_k := \max\{p_w \mid w \in \Sigma^k\} \text{ tends to 0 as } k \to \infty\,.$$

With Kolmogorov's extension theorem, the probabilistic source defines a probability $\mathbb{P}$ on the space $\Sigma^{\mathbb{N}}$ which is completely specified by the set $(p_w)_{w\in\Sigma^k}$ of fundamental probabilities.

There are particular instances of sources that are defined via a family of (labelled) partitions. We first recall this notion.

▶ **Definition 2.** *A family* $(\mathcal{P}_k)_{k\geq 0}$ *of labelled partitions associated with the alphabet* $\Sigma := [0..r-1]$ *is built in a recursive way as follows.*

(i) One begins with $\mathcal{P}_0 = \{[0,1]\}$ *and we let* $\mathcal{I}_\varepsilon := [0,1]$

(ii) For each $k \geq 0$, *the partition* $\mathcal{P}_{k+1}$ *is a refinement of the partition* $\mathcal{P}_k$.

(a) $\mathcal{P}_{k+1}$ *arises from* $\mathcal{P}_k$ *by dividing each (closed) interval of* $\mathcal{P}_k$ *into* r *(closed) intervals using* $r-1$ *points of the interval.*

(b) *Each interval of* $\mathcal{P}_k$ *is a closed interval labelled as* $\mathcal{I}_w$ *with* $w \in \Sigma^k$ *and gives rise to* r *closed intervals that are labelled from the left to the right as* $\mathcal{I}_{w\cdot a}$ *with* $a \in \Sigma$.

(c) *The diameter* $\delta(\mathcal{P}_k) := \max\{|\mathcal{I}_w| \mid w \in \Sigma^k\}$ *tends to 0 for* $k \to \infty$.

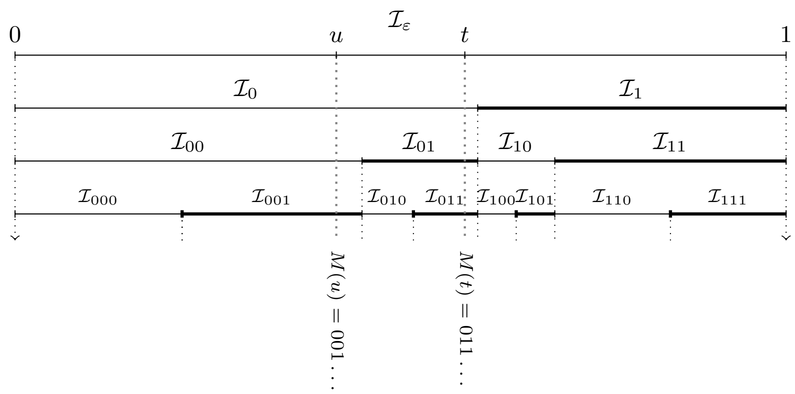


Figure 1 Example of a source defined via a family of partitions.

One now associates with this family $(\mathcal{P}_k)_{k\geq 0}$ of partitions a mapping $M : [0,1] \to \Sigma^{\mathbb{N}}$ that is defined outside the set $\mathcal{D}$ that gathers all the end-points of the intervals, as follows.

If $u \notin \mathcal{D}$, there exists, indeed, for each $k \geq 0$, a unique interval of $\mathcal{P}_k$ which contains (in its interior) the real u. Such an interval is labelled with a prefix $w \in \Sigma^k$. This prefix depends on the depth k and the real u, and is denoted as $w_k(u)$. As the partition $\mathcal{P}_{k+1}$ is a refinement of $\mathcal{P}_k$, and the diameter $\delta(\mathcal{P}_k) \to 0$, this sequence $w_k(u)$ of finite prefixes converges to a unique infinite word over the alphabet Σ that defines the value at u of the mapping M. (See Fig. 1).

In this way, the interval $\mathcal{I}_w$ gathers (up to a denumerable set) the reals u for which the word $M(u)$ begins with the prefix w. This is the fundamental interval of the prefix w, and its length is exactly the probability that $M(u)$ begins with w. Then, the mapping M defines a probabilistic source, whose fundamental probabilities are $p_w := |\mathcal{I}_w|$.

The paper deals with two instances of such a framework, over the binary alphabet $\Sigma := \{0, 1\}$. In both cases, the end-points of the partition are rational numbers, and the point which is added in the interval $[a/c, b/d]$ is the mediant $(a + b)/(c + d)$. It is always added in the first case, and gives rise to the Stern-Brocot partition. It is only added in the second case when its denominator is not too large: this partition is then defined via the Farey sequence, and, labelled in a convenient way, gives rise to what we call the Sturm partition.

2.2 Generating functions

Following the analytic combinatorics principles described in [10], and the main ideas introduced in [23], we associate, with a complex variable s, and a source, various Dirichlet generating functions (DGF's); first, $\Lambda_k(s)$ is relative to a given depth $k \geq 0$; second, $\Lambda(s)$ is associated with all possible depths; third, $\Lambda(s, v)$ is the bivariate generating function where the variable v "marks" the depth:

$$\Lambda_k(s) := \sum_{w\in\Sigma^k} p_w^s\,, \qquad \Lambda(s) := \sum_{w\in\Sigma^\star} p_w^s = \sum_{k\geq 0}\Lambda_k(s)\,, \qquad \Lambda(s,v) := \sum_{k\geq 0} v^k \Lambda_k(s)\,. \tag{1}$$

The bivariate DGF $\Lambda(s, v)$ proves very useful, due to the identity[1] $\Lambda_k(s) = [v^k]\Lambda(s, v)$ which possibly leads to use of singularity analysis.

All the main objects of a source that appear in a general Information Theory context – entropies, coincidence, trie parameters – are expressed with these series, as it is shown in [23] and [5], and now recalled: *Entropies* are the first classical parameters that describe the probabilistic properties of a source. They are defined with the DGF $\Lambda_k(s)$.

$$\text{[Shannon entropy]} \qquad \mathcal{E} = \lim_{k\to\infty} (1/k)\,\mathcal{E}_k, \qquad \mathcal{E}_k := \sum_{w\in\Sigma^k} p_w|\log p_w| = -\Lambda'_k(1)\,. \tag{2}$$

$$\text{[Renyi entropies of depth } k \text{ and exponent } \sigma > 1] \qquad \frac{1}{1-\sigma}\,\log\Lambda_k(\sigma)\,. \tag{3}$$

The *coincidence* between n words emitted by the source is defined as the length of their largest common prefix. Then, when the n words are independently drawn from the source, the coincidence becomes a random variable C_n defined on $[\Sigma^{\mathbb{N}}]^n$ whose distribution $\Pr[C_n \geq k + 1]$ exactly coincides with $\Lambda_k(n)$ and expectation $\mathbb{E}[C_n]$ coincides with $\Lambda(n)$.

The present paper mainly deals with another characteristic of the source, the *trie-depth*, denoted by D_n, which is expressed with the DGF $\Lambda(s)$, as recalled in (8).

2.3 A detour: review on the results for good sources

As it will be proven in Section 2.6, the two sources of interest are of zero entropy. They thus provide instances of sources that do not have the same behaviour as "good" sources, for instance memoryless sources or (ergodic) Markov Chains. Dynamical systems associated with expanding surjective maps of the interval provide other instances of "good" sources.

---

[1] As usual, the notation $[v^k]A(v)$ denotes the coefficient of v^k in $A(v)$.

We then recall (in a quite informal way) the main properties of the DGF's in the case of "good" sources. The DGF $\Lambda(s)$ is tame at $s = 1$ of order 1 (in the sense of Definition 9). Furthermore, there exists a basic function $s \mapsto \lambda(s)$, attached to each source, that is analytic on a neighborhood of the real axis $\Re s > d$ (for some $d < 1$, possibly equal to $-\infty$), and satisfies $\lambda(1) = 1$. This mysterious function is just equal to the sum $\sum_{i=0}^{r-1} p_i^s$ in the memoryless case. In the Markov chain case, this is the dominant eigenvalue of the matrix $P_s := (p_{i|j}^s)$ that extends the transition matrix P_1. Finally, in the case of ergodic dynamical sources, this is the dominant eigenvalue of the secant transfer operator $\mathbf{G}_s$ of the system.
The DGF's $\Lambda(s, v)$ and $\Lambda(s)$ essentially behave as quasi-inverses, for s close to the real axis,

$$\Lambda(s, v) \approx (1 - v\lambda(s))^{-1}, \qquad \Lambda(s) \sim_{s\to 1} (1 - \lambda(s))^{-1} \sim_{s\to 1} \frac{-1}{\lambda'(1)} \frac{1}{s-1}\,.$$

With properties of the bivariate DGF $\Lambda(s, v)$ and singularity analysis, the function $\Lambda_k(s)$ is a k-th quasi-power for s close to the real axis, and

$$\Lambda_k(s) = [v^k]\Lambda(s, v) \sim a(s)\lambda(s)^k \qquad (k \to \infty), \qquad \text{in particular} \quad \mathcal{E} = -\lambda'(1)\,.$$

These results, valid for "good" sources, will be used as comparison references in the rest of the paper. We now focus on the analysis of the two sources presented in the end of the Section 2.1 and obtain nice expressions for the DGF's described in Propositions 3 and 4.

2.4 The Stern-Brocot source

The *Stern-Brocot partition* of depth k, denoted as $\mathcal{B}_k$, is defined recursively as follows:
(i) One begins with $\mathcal{B}_0 = \{[0/1, 1/1]\}$;
(ii) For each $k \geq 1$, $\mathcal{B}_k$ arises from $\mathcal{B}_{k-1}$
by dividing each interval $[a/c, b/d]$ of $\mathcal{B}_{k-1}$ by its mediant $(a + b)/(c + d)$.

We now recall the relation between this family of partitions and the Farey map

$$T : [0, 1] \to [0, 1], \quad T(x) = x/(1-x) \ \text{ for } x \in [0, 1/2]\,, \quad T(x) = (1-x)/x \ \text{ for } x \in [1/2, 1]\,.$$

The set of the inverse branches of T is $\mathcal{H} := \{a, b\}$ with

$$a : [0, 1] \to [0, 1/2], \quad a(x) = x/(1 + x)\,; \qquad b : [0, 1] \to [1/2, 1], \quad b(x) = 1/(1 + x)\,;$$

Then, the set $\mathcal{H}^k := \{a, b\}^k$ of inverse branches of the iterate T^k generates the partition $\mathcal{B}_k$, and the set $\mathcal{B}_k$ gathers the fundamental intervals of depth k of the Farey map,

$$\mathcal{B}_k = \{[h(0), h(1)] \mid h \in \mathcal{H}^k\}.$$

The secant transfer operator $\mathbf{H}_s$ of the Farey map, defined as the sum $\mathbf{H}_s = \mathbf{A}_s + \mathbf{B}_s$, with

$$\mathbf{A}_s[G](x, y) := \left|\frac{a(x) - a(y)}{x - y}\right|^s G(a(x), a(y)),$$
$$\mathbf{B}_s[G](x, y) := \left|\frac{b(x) - b(y)}{x - y}\right|^s G(b(x), b(y))\,,$$

provides, via its iterates, the following expressions for the DGF's of the Stern-Brocot source,

$$\Lambda_k(s) = \mathbf{H}_s^k[1](0, 1) \quad \text{for } k \geq 0, \qquad \Lambda(s, v) = (I - v\mathbf{H}_s)^{-1}[1](0, 1)\,.$$

The decomposition $\mathcal{H}^\star = \{a^\star b\}^\star \cdot a^\star$ leads to the analogous decomposition for the quasi-inverse $(I - v\mathbf{H}_s)^{-1}$, namely $(I - v\mathbf{H}_s)^{-1} = (I - v\mathbf{A}_s)^{-1}(I - v\mathbf{B}_s(I - v\mathbf{A}_s)^{-1})^{-1}\,.$

For any $m \geq 1$, the LFT [2] $a^{m-1} \circ b$ coincides with the LFT $g_m : x \mapsto 1/(m+x)$ which is an inverse branch of the Gauss map. Then, the operator $v\mathbf{B}_s \circ (I - v\mathbf{A}_s)^{-1}$ coincides with a weighted version $\mathbf{G}_{s,v}$ of the secant transfer operator $\mathbf{G}_s$ of the Euclid DS, namely,

$$\mathbf{G}_{s,v}[F](x,y) := \sum_{m \geq 1} v^m \left| \left(\frac{1}{m+x}\right)\left(\frac{1}{m+y}\right)\right|^s F\left(\frac{1}{m+x}, \frac{1}{m+y}\right),$$

and coincides at $v = 1$ with the secant transfer operator $\mathbf{G}_s$. We have proven:

▶ **Proposition 3.** *The DGF's of the Stern-Brocot source satisfy*

$$\Lambda(s,v) = (I - v\mathbf{H}_s)^{-1}[1](0,1) = (I - v\mathbf{A}_s)^{-1}(I - \mathbf{G}_{s,v})^{-1}[1](0,1)\,;$$
$$\Lambda(s) = (I - \mathbf{H}_s)^{-1}[1](0,1) = (I - \mathbf{A}_s)^{-1}(I - \mathbf{G}_s)^{-1}[1](0,1)\,.$$

2.5 The Sturm source

We consider the source, called the Sturm source, which emits the Sturm characteristic words. More precisely, for each $\alpha \in [0,1]$, it emits the characteristic Sturm word $S(\alpha)$, whose definition is now recalled. Consider α of the interval $[0,1]$, the two intervals $I_0(\alpha) = [1-\alpha, 1[$ and $I_1(\alpha) = [0, 1-\alpha[$ it defines, together with the Kronecker-Weyl sequence $n \mapsto \{n\alpha\}$ (where $\{x\}$ denotes the fractional part of x). By definition, the n-th symbol of the word $S(\alpha)$ equals $j \in \{0,1\}$ if and only if $\{(n+1)\alpha\}$ belongs to $I_j(\alpha)$.

The partition $\mathcal{S}_k$ associated with the Sturm source of order k is defined by its end-points, that are the elements of the Farey sequence of depth $k+1$,

$$\mathcal{F}_{k+1} := \{\tfrac{a}{c} \mid a, c \geq 1,\ \gcd(a,c) = 1,\ c \leq k+1\}.$$

The partition $\mathcal{S}_k$ is built from $\mathcal{S}_{k-1}$ in a similar recursive way as the Stern-Brocot partition $\mathcal{B}_k$ is built from $\mathcal{B}_{k-1}$:

(i) One begins with $\mathcal{S}_0 = \mathcal{B}_0 = \{[0/1, 1/1]\}$

(ii) For each $k \geq 1$, $\mathcal{S}_k$ arises from $\mathcal{S}_{k-1}$

by dividing each interval $[a/c, b/d]$ of $\mathcal{S}_{k-1}$ by its mediant $(a+b)/(c+d)$
provided the denominator $c+d$ *be at most* $k+1$.

Due to the previous condition, the partition $\mathcal{S}_k$ is thus a pruning of the partition $\mathcal{B}_k$, and satisfies two classical properties

(P1) For any k, for any interval $[a/c, b/d]$ of $\mathcal{S}_k$, one has $ad - bc = -1$.

(P2) The set $\mathcal{C}_k$ which gathers the pairs (c,d) which appear as denominators of the intervals $[a/c, b/d] \in \mathcal{S}_{k-1}$ is equal to

$$\mathcal{C}_k := \{(c,d) \mid \max(c,d) \leq k < c+d, \quad \gcd(c,d) = 1\}\,.$$

Moreover, each pair (c,d) appears at most once. Then, the partition $\mathcal{S}_k$ has a polynomial number of intervals [of order $O(k^2)$] whereas the partition $\mathcal{B}_k$ has exactly 2^k intervals.

We now describe how to encode the partition $\mathcal{S}_k$ with the prefixes of length k of characteristic Sturmian words. The prefix of length k of the word $S(\alpha)$, denoted as $[S(\alpha)]_k$, satisfies two properties, described in [2, Proposition 3]:

(i) The two words $0 \cdot [S(\alpha)]_k$ and $1 \cdot [S(\alpha)]_k$ are both factors of the infinite word $S(\alpha)$;

(ii) For an interval $[a/c, b/d] \in \mathcal{S}_k$ with $a/c \neq 0$ and $b/d \neq 1$, one has the characterization

$$\forall \alpha \in]a/c, b/d[, \qquad [S(\alpha)]_k \text{ is a palindrome} \Longleftrightarrow c+d = k+2\,.$$

$\mathcal{S}_0 :$	(0)						ε						(1)
$\mathcal{S}_1 :$	(0)			0			(1/2)			1			(1)
$\mathcal{S}_2 :$	(0)		00		(1/3)	01	(1/2)	10	(2/3)		11		(1)
$\mathcal{S}_3 :$	(0)	000	(1/4)	001	(1/3)	010	(1/2)	101	(2/3)	110	(3/4)	111	(1)

Figure 2 The coding of the partition $\mathcal{S}_k$ for $k \leq 3$.

This property is used as follows for the coding of the partition $\mathcal{S}_k$: We begin with the coding of $\mathcal{S}_0$ as the empty word ϵ. Then, there are two cases: when the mediant can be added, the interval is subdivided and the coding of the two sub-intervals is made in the usual way, in the lexicographic order. When the mediant cannot be added, the interval is not subdivided, and its new coding is obtained in a unique way with a palindromic completion. (See Fig. 2)

We now study the generating functions.

▶ **Proposition 4.** *The generating series of the Sturm source can be expressed with double zeta functions $\zeta(a, b)$ together with the usual zeta function $\zeta(a)$ in the form*

$$\Lambda(s) = 1 + 2\frac{\zeta(s, s-1)}{\zeta(2s-1)} \qquad \textit{with} \quad \zeta(a) := \sum_{c\geq 1} \frac{1}{c^a}, \qquad \zeta(a, b) := \sum_{c\geq 1} \frac{1}{c^b} \sum_{d|d>c} \frac{1}{d^a}. \tag{4}$$

Proof. With Property $(P1)$, the length of an interval $[a/c, b/d] \in \mathcal{S}_k$ encoded by w, is equal to $p_w = 1/(cd)$. With $(P2)$, and the set $\mathcal{C}_k$ defined there, one obtains

$$\Lambda_{k-1}(s) = \sum_{(c,d)\in\mathcal{C}_k} \left(\frac{1}{cd}\right)^s, \qquad \Lambda(s) = \sum_{\substack{(c,d)\\ \gcd(c,d)=1}} \left(\frac{1}{cd}\right)^s \sum_{k\geq 0} [\![(c, d) \in \mathcal{C}_k]\!] ,$$

where $[\![\cdot]\!]$ is Iverson's bracket. Then,

$$\Lambda(s) = \sum_{\substack{(c,d)\\ \gcd(c,d)=1}} \left(\frac{1}{cd}\right)^s [(c + d) - \max(c, d)] = \sum_{\substack{(c,d)\\ \gcd(c,d)=1}} \frac{\min(c, d)}{(cd)^s} .$$

As the general term of the previous sum is homogeneous of weight $1 - 2s$, the two sums – the sum relative to general pairs (c, d) and the sum relative to coprime pairs – are related. The denominator $\zeta(2s - 1)$ appears, and, with the equality

$$\sum_{(c,d)} \frac{\min(c, d)}{c^s d^s} = \zeta(2s - 1) + 2\zeta(s, s - 1) ,$$

this yields the expression given in (4). ◀

The long version of the paper will provide an expression of the bivariate Sturm DGF $\Lambda(s, v)$ in terms of polylogarithms.

[2] LFT = linear fractional transformation

2.6 Entropies

The Shannon entropy defined in (2) is the first parameter associated with a source.

▶ **Proposition 5.** *The two sources are of zero Shannon entropy.*

Proof. (a) This is clear for the Sturm source. Jensen's inequality applied to the concave map $\phi(t) = t|\log t|$ relates $\mathcal{E}_k$ and the number A_k of (non empty) fundamental intervals of depth k, via the inequality $\mathcal{E}_k \le \log A_k$. One has indeed:

$$\frac{\mathcal{E}_k}{A_k} = \frac{1}{A_k}\sum_{\boldsymbol{w}\in\Sigma^k}\phi(p_w) \le \phi\left(\frac{1}{A_k}\sum_{w\in\Sigma^k} p_w\right) = \phi\left(\frac{1}{A_k}\right) = \frac{\log A_k}{A_k}, \quad \text{and thus} \quad \mathcal{E}_k \le \log A_k\,.$$

Then a source for which A_k is polynomial has a zero entropy. This applies to the Sturm source for which $A_k = O(k^2)$.

(b) This is also the case for the Stern-Brocot source. However, we do not find any direct proof of this fact in the literature. This is why we provide two proofs in the annex. ◀

The Renyi entropies of depth k and exponent $\sigma > 1$, defined via $\Lambda_k(\sigma)$ in (3), are already studied in at least two papers: The first result, due to Moshchevitin and Zhigljavsky in [16] describes the Stern-Brocot case. The second one, due to Hall [13] and Kanemitsu et al, describes in [14] the Sturm case.

▶ **Proposition 6.** [16, 13, 14] *As $k \to \infty$, the following asymptotic estimates hold for the DGF's $\Lambda_k(\sigma)$,*

$$[\text{Stern-Brocot case}, \sigma > 1] \qquad \Lambda_k(\sigma) = \frac{2}{k^\sigma}\frac{\zeta(2\sigma-1)}{\zeta(2\sigma)}\left[1 + O\left(\frac{\log k}{k^{(\sigma-1)/(2\sigma)}}\right)\right]; \tag{5}$$

$$[\text{Sturm case}, \sigma \ge 4] \qquad \Lambda_k(\sigma) = \frac{2}{k^\sigma}\frac{\zeta(\sigma-1)}{\zeta(\sigma)}\left[1 + O\left(\frac{1}{k}\right)\right]. \tag{6}$$

The previous estimates for the two sources are then quite similar, with the same polynomial behaviour in $O(1/k^\sigma)$ for $k \to \infty$. Even though the dominant constants are not the same, they are however of the same spirit.

We then study another parameter of the source, the trie-depth, with the hope that it may "differentiate" the two sources in a stronger way.

3 Tries built from words emitted by a source

The trie structure is an important data structure in algorithmics [12] that also plays a central role in Theoretical Information Theory contexts. This is why it has already been deeply analyzed, at least in the context of good sources. See [20] for analyses in the context of simple sources and [5, 11, 4] for analyses in the context of dynamical sources.

3.1 Trie and its depth

A trie is a tree structure, used as a dictionary, which compares words emitted by a source $\mathcal{S}$ via their prefixes. The trie $\mathcal{T}(\boldsymbol{x})$ is built on a finite sequence $\boldsymbol{x}$ of (infinite) words emitted by the source $\mathcal{S}$ and is defined recursively by the following three rules which involve the cardinality $N(\boldsymbol{x})$ of the sequence $\boldsymbol{x}$:

(a) If $N(\boldsymbol{x}) = 0$, then $\mathcal{T}(\boldsymbol{x}) = \emptyset$

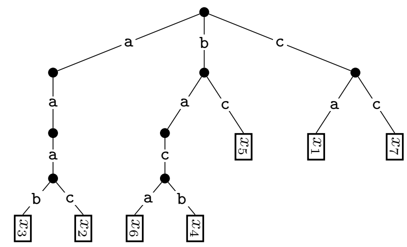
(b) If $N(\boldsymbol{x}) = 1$, with $\boldsymbol{x} = (x)$, then $\mathcal{T}(\boldsymbol{x})$ is a leaf labeled by x.

(c) If $N(\boldsymbol{x}) \geq 2$, then $\mathcal{T}(\boldsymbol{x})$ is formed with an internal node and r subtries equal to

$$\mathcal{T}(\boldsymbol{x}_{\langle 0\rangle}), \ldots, \mathcal{T}(\boldsymbol{x}_{\langle r-1\rangle}),$$

where $\boldsymbol{x}_{\langle\sigma\rangle}$ denotes the sequence consisting of words of $\boldsymbol{x}$ which begin with symbol σ, stripped of their initial symbol σ. If the set $\boldsymbol{x}_{\langle\sigma\rangle}$ is non-empty, the edge which links the subtrie $\mathcal{T}(\boldsymbol{x}_{\langle\sigma\rangle})$ to the internal node is labelled with the symbol σ.

x_1	cacccaaac...
x_2	aaacacbac...
x_3	aaabcbbca...
x_4	bacbaabcc...
x_5	bcabbbbbc...
x_6	bacaabbba ...
x_7	ccbacbcbb...



In this paper, we perform a probabilistic analysis of the shape of $\mathcal{T}(\boldsymbol{x})$. In our probabilistic model, the sequence $\boldsymbol{x}$ is formed with words that are independently drawn from the source $\mathcal{S}$, and we are interested in the asymptotics when the cardinality n of $\boldsymbol{x}$ tends to ∞. Here, we focus on a particular parameter of the trie, called the depth, and denoted by D_n: it is defined as the depth D_n of a *random* branch, and it is now described.

Given a sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$, the trie $\mathcal{T}(\boldsymbol{x})$ has exactly n branches, and the length (or the depth) of a branch is the number of the nodes it contains. For $i \in [1..n]$, the length of the i–th branch of the trie (corresponding to the word x_i) is denoted by $D_n^{(i)}$. Inside our model, the depth D_n of a random branch satisfies

$$\Pr[D_n \geq k+1] = \frac{1}{n}\sum_{i=1}^{n} \Pr[D_n^{(i)} \geq k+1]\,. \tag{7}$$

The parameter D_n will be simply called the trie depth. For a fixed source, this is a random variable which depends on the set $\boldsymbol{x}$ of words emitted by the source, and the (well-known) next proposition studies its moments. We remark that Assertion (iii) is less classical.

▶ **Proposition 7.** *Consider a probabilistic source and a set of n infinite words independently emitted by the source. Then, the depth D_n of the trie built on this set satisfies the following:*

(i) *The distribution of D_n involves the fundamental probabilities of the source,*

$$\Pr[D_n \geq k+1] = \sum_{|w|=k} p_w[1-(1-p_w)^{n-1}] = \frac{1}{n}\sum_{|w|=k}\sum_{\ell=2}^{n}(-1)^{\ell}\binom{n}{\ell}\,\ell\, p_w^{\ell} \qquad \text{for } k \geq 0.$$

(ii) *If the generating series $\Lambda(s)$ is well-defined for $s \geq 2$, then the expectation $\mathbb{E}[D_n]$ is expressed as an alternating sum which involves the values $\Lambda(\ell)$, for $\ell \geq 2$*

$$\mathbb{E}[D_n] = \frac{1}{n}\sum_{\ell=2}^{n}(-1)^{\ell}\binom{n}{\ell}\,\ell\,\Lambda(\ell)\,. \tag{8}$$

(iii) *As soon as the source contains, for each $k \geq 1$, a prefix of length $k-1$ whose probability is at least Ak^{-a} with $a \leq 1$ for some $A > 0$, all the moments of D_n of order k for any $n \geq 2$ and $k \geq 2$ are infinite.*

Proof.

(i) The event $D_n^{(i)} \geq k+1$ means that there exists a prefix w of length k that is common to the word x_i and at least to another word x_j of the sequence. Inside our model, this entails the first expression in (i), and, with a binomial expansion, the second one.

(ii) Now, if $\Lambda(2)$ is finite, any $\Lambda(\ell)$ is also finite for $\ell \geq 2$. Taking the sum over k of the previous expression entails (ii), after the exchange of two summations (over k and over n).

(iii) The sequence $n \mapsto \Pr[D_n \geq k]$ is increasing, and it is thus sufficient to deal with the case $n = 2$. Under the assumption, one has

$$\Pr[D_2 \geq k] = \sum_{|w|=k-1} p_w^2 \geq A^2\, k^{-2a}\,,$$

$$\begin{aligned}\mathbb{E}[D_2^2] &= \sum_{k\geq 1} k^2 \Big(\Pr[D_2 \geq k] - \Pr[D_2 \geq k+1] \Big) \\ &= \sum_{k\geq 1} (2k-1) \Pr[D_2 \geq k] \geq A^2 \sum_{k\geq 1} k^{1-2a}\,.\end{aligned}$$

All the moments $\mathbb{E}[D_n^k]$ are thus infinite for $n \geq 2, k \geq 2$ as soon as $a \leq 1$. ◀

▶ **Proposition 8.** *For the two sources, the moments $\mathbb{E}[D_n^k]$ of order $k \geq 2$ are infinite.*

Proof. For each source, the prefix 0^k has a probability $1/(k+1)$. Then (iii) applies. ◀

3.2 Survey for the Rice method

The rest of the paper then studies the mean value $\mathbb{E}[D_n]$, starting with its expression given in (8). We will use here the Rice method, that is dedicated to the study of sequences $n \mapsto f(n)$ that are expressed as a binomial sum which involves another sequence $n \mapsto p(n)$,

$$f(n) = \sum_{\ell=a}^{n} (-1)^\ell \binom{n}{\ell} p(\ell)\,, \qquad (a \text{ integer, } a \geq 0;\text{ here } a = 2). \tag{9}$$

The method was introduced by Nörlund [17, 18] and widely used in analytic combinatorics since the seminal papers of Flajolet and Sedgewick [9, 8]. See also the survey in [24]. The main role is then played by the analytical extension ψ of the sequence $n \mapsto p(n)$, provided it be *tame at $s = c$* for $c < a$. We now recall this notion:

▶ **Definition 9.** *A function $\psi(s)$ is tame at c with order d if there exists $\delta_0 > 0$, called the tameness width, for which*

(i) *$\psi(s)$ is analytic on $\Re s > c - \delta_0$ except at $s = c$ where it admits a pole of order $d > 0$, with a singular expression at $s = c$ of the form $\psi(s) \asymp a_d(s-c)^{-d} + \ldots + a_1(s-c)^{-1} + a_0$.*

(ii) *For any $\delta < \delta_0$, $\psi(s)$ is of polynomial growth on $\Re s \geq c - \delta$ as $|\Im s| \to \infty$.*

There are three main steps in the method.

Step 1. The inequality $c < a$ entails that ψ is analytic on $\Re s > a$, and the binomial formula is transfered into a Rice integral formula.

Consider a real c with $c \in]a-1, a[$. Then, for any $b \in]c, a[$ and $n \geq n_0$, the sequence $f(n)$ admits an integral representation

$$f(n) = \sum_{\ell=a}^{n} (-1)^\ell \binom{n}{\ell} p(\ell) = \frac{1}{2i\pi} \int_{b-i\infty}^{b+i\infty} L_n(s) \cdot \psi(s)\, ds\,, \qquad L_n(s) := \frac{\Gamma(n+1)\Gamma(-s)}{\Gamma(n+1-s)}\,.$$

Step 2. This integral representation is valid for any abscissa $b \in]c, a[$. The vertical line $\Re s = b$ is shifted to the left, using the tameness properties of ψ on the left,

For any $\delta < \delta_0$, the following asymptotic formula involving the tameness width δ holds:

$$f(n) = \text{Res}\,[L_n(s) \cdot \psi(s); s = c] + O(n^{c-\delta}), \quad (n \to \infty)\,.$$

Step 3. The residue $\text{Res}\,[L_n(s) \cdot \psi(s); s = c]$ admits the estimate

$$A_n[\psi] := \text{Res}\,[L_n(s) \cdot \psi(s); s = c] = n^c \cdot P(\log n)\,[1 + O(1/n)]\,, \tag{10}$$

and involves a polynomial P that is computed from the singular expansion of $\psi(s)$ and $\Gamma(-s)$ at $s = c$, with two cases:

(i) If c is integer, $\Gamma(-s)$ has also a pole at $s = c$, and the product $\Gamma(-s) \cdot \psi(s)$ has a pole of order $d + 1$. Then the polynomial P has degree d, with a dominant term equal to $[1/\Gamma(d + 1)]\,|a_d|$. We remark that the factor $[1 + O(1/n)]$ equals 1 in the case $c = 1$.

(ii) If c is not an integer, the product $\Gamma(-s) \cdot \psi(s)$ has a pole of order d, and $P(t)$ has degree $d - 1$, with a dominant term equal to $[\Gamma(-c)/\Gamma(d)]\,|a_d|$.

4 Tameness of the generating functions. Main result

We thus study the function $s \mapsto s\Lambda(s)$ for each source. We wish to prove that it is tame at $s = c$ with $c > 2$ and some order d. We have to make precise the location of the dominant pole $s = c$, its order, and prove the polynomial growth of $s\Lambda(s)$ on a half-plane $\Re s > c - \delta$. We then compute the residue $A_n[s\Lambda(s)]$ defined in (10) following the principles of the previous Step 3. Remembering the division by n in Eq. (8), we finally obtain our main result.

4.1 The Stern-Brocot source

The iterate $\mathbf{A}_s^n$ of the operator $\mathbf{A}_s$ is written as

$$\mathbf{A}_s^n[F](x, y) = \left| \left(\frac{1}{1 + nx}\right) \left(\frac{1}{1 + ny}\right) \right|^s F\left(\frac{x}{1 + nx}, \frac{y}{1 + ny}\right)\,.$$

Then, in particular, when $(x, y) = (0, 1)$, the quasi-inverse writes as

$$(I - \mathbf{A}_s)^{-1}[L](0, 1) = L(0, 1) + \sum_{n \geq 1} \left(\frac{1}{n + 1}\right)^s L\left(0, \frac{1}{n + 1}\right)\,. \tag{11}$$

With Proposition 3, this is applied to the DGF $\Lambda(s)$,

$$\Lambda(s) = L_s(0, 1) + \sum_{n \geq 1} \left(\frac{1}{n + 1}\right)^s L_s\left(0, \frac{1}{n + 1}\right), \qquad \text{with} \quad L_s := (I - \mathbf{G}_s)^{-1}[1]\,. \tag{12}$$

As L_s belongs to $\mathcal{C}^1([0, 1]^2)$ (see Proposition 10), one deals with $M_s : y \mapsto (\partial/\partial y)L_s(0, y)$, and the following estimate holds:

$$\Lambda(s) = \zeta(s)L_s(0, 0) + O\left(\zeta(s + 1)\right) ||M_s||_0\,, \qquad ||F||_0 := \sup\{|F(x, y)| \mid (x, y) \in [0, 1]^2\}\,. \tag{13}$$

We now use deep results due to Dolgopyat in [6], that have been adapted by Baladi and Vallée in [1] to the plain quasi-inverse $(I - G_s)^{-1}$, then extended by Cesaratto and Vallée [4] to the quasi-inverse of the secant operator $\mathbf{G}_s$. They prove the following:

▶ **Proposition 10** (Dolgopyat, Baladi, Cesaratto, Vallée). *The mapping $s \mapsto L_s := (I - \mathbf{G}_s)^{-1}[1]$ viewed as a mapping from $\mathbb{C}$ to $\mathcal{C}^1([0,1]^2)$ is analytic for $\Re s > 1-\delta$, except at $s = 1$, where it has a simple pole, and is of polynomial growth in the half-plane $\Re s > 1-\delta$ for $|\Im s| \to \infty$ for some $\delta > 0$. Moreover, for s close to 1*

$$L_s(x,y) \sim_{s\to 1} \frac{1}{\mathcal{E}}\frac{1}{s-1}\Phi(x,y), \qquad M_s(x,y) := \frac{\partial}{\partial y}L_s(x,y) \sim_{s\to 1} \frac{1}{\mathcal{E}}\frac{1}{s-1}\frac{\partial}{\partial y}\Phi(x,y)\,. \quad (14)$$

Here, Φ is the extension[3] *of the Gauss density $\phi(x) = (1/\log 2)(1/(1+x))$ and satisfies $\Phi(0,0) = \phi(0) = 1/(\log 2)$ and $\mathcal{E}$ is the entropy of the Gauss map, equal to $(1/\log 2)\zeta(2)$.*

The expression (12) and the estimate (13) prove that $\Lambda(s)$ is analytic for $\Re s > 1$. Then, with (12) and (14), we see that $s = 1$ is a pole of order 2 for $\Lambda(s)$, with the estimate

$$\Lambda(s) \sim_{s\to 1} \zeta(s)\left[\frac{1}{\mathcal{E}}\frac{1}{s-1}\Phi(0,0)\right] \sim_{s\to 1} \frac{1}{\zeta(2)}\left(\frac{1}{s-1}\right)^2.$$

We now study the tameness of $\Lambda(s)$ at $s = 1$. The function $\zeta(s)$ is tame at $s = 1$, with a tameness width equal to 1, as it will be recalled in Lemma 12. The functions L_s and M_s are tame at $s = 1$, as it was recalled in Proposition 10, with a tameness width $\delta < 1$. Finally:

▶ **Proposition 11.** *The DGF $\Lambda(s)$ of the Stern-Brocot source is analytic on $\Re s > 1-\delta$ (for some $\delta > 0$), except at $s = 1$, where it has a pole of order 2. It is of polynomial growth on $\Re s \geq 1-\delta$ for $|\Im s| \to \infty$. Moreover, the following estimates hold:*

$$s\Lambda(s) \sim_{s\to 1} \frac{1}{\zeta(2)}\left(\frac{1}{s-1}\right)^2 \qquad \Gamma(-s)\cdot s\Lambda(s) \sim_{s\to 1} \frac{6}{\pi^2}\left(\frac{1}{s-1}\right)^3.$$

Using then Step 3 of Section 3.2, the equality holds for the residue defined in (10)

$$A_n[s\cdot\Lambda(s)] = n\,P_2(\log n), \qquad P_2(t) = \frac{3}{\pi^2}\,t^2 + b_1\,t + b_0 \qquad \text{for some constants } b_1, b_0\,. \quad (15)$$

4.2 The Sturm source

With the expression of $\Lambda(s)$ given in (4), we need properties of the zeta functions (plain or double), together with its inverse $1/\zeta(s)$. They are recalled in the following Lemma.

▶ **Lemma 12.** *The following holds for the functions ζ and $1/\zeta$:*

(a) *For any $a_0 > 0$, the function $\zeta(s)$ is meromorphic on the half-plane $\Re s > 2a_0$ with only a simple pole at $s = 1$ and is of polynomial growth on $\Re s \geq 2a$, with $a > a_0$ for $|\Im s| \to \infty$.*

(b) *For any $b_0 > 0$, the function $1/\zeta(s)$ is analytic on the half-plane $\Re s > 1 + 2b_0$ and its modulus is less than $\zeta(1+2b)$ on any half-plane $\Re s \geq 1+2b$ with $b > b_0$.*

Proof. Assertion (a) is classical and proven for instance in [21], Chapter II.3, Theorem 7.

Assertion (b) is a consequence of Mertens' inequality recalled in Chapter II.3, Corollary 8.1 of [21], that provides an upper bound for $1/\zeta(s)$

$$|\zeta(\sigma+i\tau)|^{-4} \leq \zeta(\sigma)^3|\zeta(\sigma+2i\tau)| \qquad \text{for } \sigma \geq 1+2b > 1,\, b > 0\,.$$

Using the inequality $|\zeta(\sigma+2i\tau)| \leq \zeta(\sigma)$, we obtain $|\zeta(\sigma+i\tau)|^{-1} \leq \zeta(\sigma) \leq \zeta(1+2b)$. ◀

[3] precisely described in [22, Théorème 5, Eq. (48)].

We now return to the double zeta function. The following estimate holds on $\Re s > 1$, and relates the double zeta function and the plain zeta function

$$\zeta(s, s-1) = \frac{1}{s-1}\zeta(2s-2) + O(\zeta(2s-1))\,.$$

Then, Assertion (a) of Lemma 12 shows that $\zeta(s, s-1)$ is analytic on $\Re s > 1 + a_0$ $(a_0 > 0)$ with a pole only at $s = 3/2$, and a residue equal to 1. It is of polynomial growth for $|\Im s| \to \infty$ on $\Re s \geq 1 + a$ for $a > a_0$. Furthermore, with Assertion (b) of Lemma 12, the inverse $1/\zeta(2s-1)$ is analytic on $\Re s > 1+b_0$ $(b_0 > 0)$ and of polynomial growth on $\Re s \geq 1+b$ for $b > b_0$. Choosing $a_0 = b_0$ provides a tameness width $\delta > 1/2 - \epsilon$ for any $\epsilon > 0$. Finally:

▶ **Proposition 13.** *For any $a_0 > 0$, the DGF $\Lambda(s)$ of the Sturm source is analytic on $\Re s > 1 + a_0$, except at $s = 3/2$ where it admits a simple pole. It is of polynomial growth on $\Re s \geq 1 + a$ for any $a > a_0$. Moreover, the following estimate holds:*

$$s\Lambda(s) \sim_{s\to 3/2} \frac{36}{\pi^2}\left(\frac{1}{2s-3}\right) \qquad \Gamma(-s)\cdot s\Lambda(s) \sim_{s\to 3/2} \frac{36}{\pi^2}\,\Gamma(-3/2)\left(\frac{1}{2s-3}\right)\,.$$

Using then Step 3 of Section 3.2, and the value $\Gamma(-3/2) = (4/3)\sqrt{\pi}$, the following estimate holds for the residue defined in (10):

$$A_n[s\cdot\Lambda(s)] = \frac{24}{\pi^{3/2}}n^{3/2}\left[1 + O(1/n)\right]\,. \tag{16}$$

4.3 Statement of the main result

Using Step 3 of Section 3.2, with the estimates of the residue $A_n[s\Lambda(s)]$ obtained in (15) and (16), together with the remainder term associated with the tameness strip, and remembering the division by n in Eq. (8), we obtain our final result.

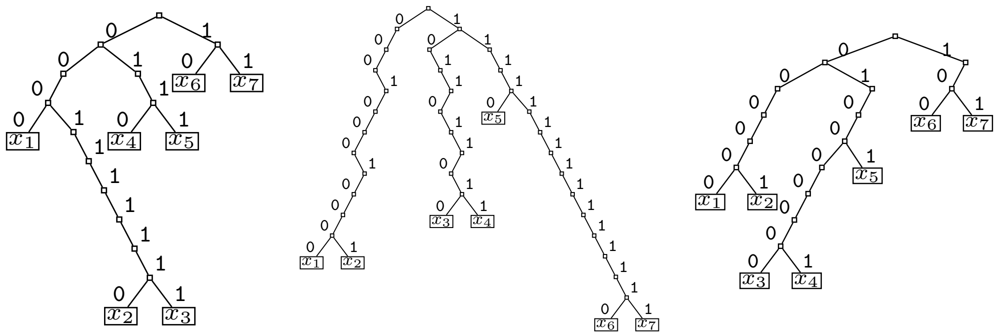


Figure 3 Instances of tries built on seven words emitted from each source of interest: the Stern-Brocot source (on the left), the Sturm Source (in the middle). As the value $n = 7$ is small, and the moments $\mathbb{E}[D_n^2]$ are infinite, there does not really exist a "typical trie". The third trie (on the right) is built on seven words emitted by the Farey dynamical source mentioned in the Conclusion.

▶ **Theorem 14.** *Consider, for each source, a trie built on n words independently drawn from the source. Then, the mean value of the trie depth grows as $\Theta(\log^2 n)$ for the Stern-Brocot source whereas it grows as $\Theta(n^{1/2})$ for the Sturm source,*

$$[\text{Stern-Brocot case}] \qquad \mathbb{E}[D_n] = \frac{3}{\pi^2}\log^2 n + b_1\log n + b_0 \; + O(n^{-\delta}) \quad \textit{for some } \delta > 0\,;$$

$$[\text{Sturm case}] \qquad \mathbb{E}[D_n] = \frac{24}{\pi^{3/2}}\,n^{1/2} + O(n^a) \quad \textit{for any } a > 0\,.$$

Figure 3 clearly exhibits some important features of each source. This explains –in an experimental way– why the trie is a good tool for studying the characteristics of a source.

5 Conclusions and further work

This paper appears as (one of) the first study dedicated to sources of zero Shannon entropy, and performed with Analytic Combinatorics tools. It focuses on a particular parameter of the source, the trie depth. We wish to extend this first study in several directions.

We wish to use Analytic Combinatorics tools, notably singularity analysis, to directly derive estimates of Prop. 6, that are presently obtained via fine Number Theory arguments. It is probably possible to directly deal with the bivariate DGF $\Lambda(s, v)$, whose expression seems closely related (in both cases) to generalized versions of the polylogarithm. We then hope using the methods of Flajolet in [7], dedicated to singularity analyses of the polylogarithm. This will be a first step towards the distribution of the coincidence C_n defined in Section 2.2.

The VLMC sources (VLMC = Variable Length Markov Chain) are the simplest sources where the dependency from the past is unbounded. The paper [3] deeply studies this model and analyzes the depth of associated suffix tries in some particular cases.
We wish to focus on a whole natural sub-class of VLMC sources, related to the intermittency phenomenon. We consider the binary case, assume the equality $\Pr[Y_0 = 1] = 1$, and focus on the events $\mathcal{S}_k :=$ [the prefix finishes with a sequence of exactly k occurrences of 0].
A VLMC is *intermittent of exponent* $a > 0$ when the following conditional probability distribution holds: $\Pr[0 \mid \mathcal{S}_0] = (1/2), \quad \Pr[0 \mid \mathcal{S}_k] = (k/(k+1))^a, \quad (k \geq 1)$.
Then, the series $\Lambda(s)$ involves two functions of Riemann ζ type, and strongly depends on the parameter a. We wish to perform a complete analysis of the trie depth for this precise class, exhibiting the dependence with respect to parameter a.

Fig 3 exhibits an instance of a trie built on the Farey dynamical source. As recalled in Section 2.4, the Farey DS admits the Stern-Brocot partition as a generating partition. Moreover, the Farey DS admits, as an invariant density, the density $1/t$ whose integral is infinite. Then, the fundamental probabilities of the two sources [Stern-Brocot and Farey] are not clearly related. This strongly differs from the framework of the papers [5, 11, 4] which deal with ergodic dynamical sources, whose invariant measure is absolutely continuous with respect to the Lebesgue measure. Then, the analysis of trie depth for the Farey source will be both a natural and difficult question.

References

1 V. Baladi and B. Vallée. Euclidean algorithms are Gaussian. *J. Number Theory*, 110(2):331–386, 2005. doi:10.1016/j.jnt.2004.08.008.

2 V. Berthé. Fréquences des facteurs des suites sturmiennes. *Theoret. Comput. Sci.*, 165(2):295–309, 1996. doi:10.1016/0304-3975(95)00224-3.

3 P. Cénac, B. Chauvin, F. Paccaut, and N. Pouyanne. Uncommon suffix tries. *Random Structures Algorithms*, 46(1):117–141, 2015. doi:10.1002/rsa.20500.

4 E. Cesaratto and B. Vallée. Gaussian distribution of trie depth for strongly tame sources. *Combin. Probab. Comput.*, 24(1):54–103, 2015. doi:10.1017/S0963548314000741.

5 J. Clément, P. Flajolet, and B. Vallée. Dynamical sources in information theory: a general analysis of trie structures. *Algorithmica*, 29(1-2):307–369, 2001. doi:10.1007/BF02679623.

6 D. Dolgopyat. On decay of correlations in Anosov flows. *Ann. of Math. (2)*, 147(2):357–390, 1998. doi:10.2307/121012.

7 P. Flajolet. Singularity analysis and asymptotics of Bernoulli sums. *Theoret. Comput. Sci.*, 215(1-2):371–381, 1999. doi:10.1016/S0304-3975(98)00220-5.

8 P. Flajolet and R. Sedgewick. Digital search trees revisited. *SIAM J. Comput.*, 15(3):748–767, 1986. doi:10.1137/0215054.

9 P. Flajolet and R. Sedgewick. Mellin transforms and asymptotics: finite differences and Rice's integrals. *Theoret. Comput. Sci.*, 144(1-2):101–124, 1995. doi:10.1016/0304-3975(94)00281-M.

10 P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009. doi:10.1017/CBO9780511801655.

11 P. Flajolet and B. Vallée. Continued fractions, comparison algorithms, and fine structure constants. In *Constructive, experimental, and nonlinear analysis (Limoges, 1999)*, volume 27 of *CRC Math. Model. Ser.*, pages 53–82. CRC, Boca Raton, FL, 2000.

12 G. H. Gonnet and R. Baeza-Yates. *Handbook of Algorithms and Data Structures: in Pascal and C.* Addison-Wesley, Reading, MA, second edition, 1991.

13 R. R. Hall. A note on Farey series. *J. London Math. Soc. (2)*, 2:139–148, 1970. doi:10.1112/jlms/s2-2.1.139.

14 S. Kanemitsu, R. Sitaramachandra Rao, and A. Siva Rama Sarma. Some sums involving Farey fractions. I. *J. Math. Soc. Japan*, 34(1):125–142, 1982. doi:10.2969/jmsj/03410125.

15 A. Ya. Khinchin. *Continued fractions.* The University of Chicago Press, Chicago, Ill.-London, 1964.

16 N. Moshchevitin and A. Zhigljavsky. Entropies of the partitions of the unit interval generated by the Farey tree. *Acta Arith.*, 115(1):47–58, 2004. doi:10.4064/aa115-1-4.

17 N. E. Nörlund. *Lecons sur les équations linéaires aux différences finies.* Collection de monographies sur la théorie des fonctions. Gauthier-Villars, Paris, 1929.

18 N. E. Nörlund. *Vorlesungen über Differenzenrechnung.* Chelsea Publishing Company, New York, 1954.

19 T. Prellberg and J. Slawny. Maps of intervals with indifferent fixed points: thermodynamic formalism and phase transitions. *J. Statist. Phys.*, 66(1-2):503–514, 1992. doi:10.1007/BF01060077.

20 W. Szpankowski. *Average case analysis of algorithms on sequences.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2001. doi:10.1002/9781118032770.

21 G. Tenenbaum. *Introduction à la théorie analytique et probabiliste des nombres*, volume 1 of *Cours Spécialisés [Specialized Courses]*. Société Mathématique de France, Paris, second edition, 1995.

22 B. Vallée. Opérateurs de Ruelle-Mayer généralisés et analyse en moyenne des algorithmes d'Euclide et de Gauss. *Acta Arith.*, 81(2):101–144, 1997. doi:10.4064/aa-81-2-101-144.

23 B. Vallée. Dynamical sources in information theory: fundamental intervals and word prefixes. *Algorithmica*, 29(1-2):262–306, 2001. doi:10.1007/BF02679622.

24 B. Vallée. The depoissonisation quintet: Rice-Poisson-Mellin-Newton-Laplace. In *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 110 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 35, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018. Long version arXiv:1802.04988. doi:10.4230/LIPIcs.AofA.2018.35.

A Appendix

We now give two proofs of the fact that the Shannon entropy $\mathcal{E}$ of the Stern-Brocot source is zero.

A.1 About the entropy of the Stern-Brocot source. Analytical proof

We follow the approach of Prellberg and Slawny [19] that we adapt to our framework, and there are three steps in the proof. The first two steps deal with the case $s < 1$, and the third step lets $s \to 1$.

Step 1. We study the operator $\mathbf{G}_{s,v}$ for a pair (s,v), with s real, $s \in]1/2,1[$ and v complex. For $|v|<1$, this operator acts on the space $\mathcal{C}^1([0,1]^2)$. When v is real, with $0<v<1$, it admits a unique dominant eigenvalue denoted as $\lambda(s,v)$ that depends analytically on the pair (s,v). There is an inequality which relates the spectral radii, $r[\mathbf{G}_{s,v}] \le r[\mathbf{G}_{s,|v|}] = \lambda(s,|v|)$.

For $v \to 0$, the eigenvalue $\lambda(s,v)$ tends to 0, whereas it coincides at $v=1$ with the dominant eigenvalue $\lambda(s)$ of $\mathbf{G}_{s,1} = \mathbf{G}_s$ that is strictly larger than 1 for $s<1$. There thus exists, for any $s<1$, a real number $v=v(s)$ for which the operator $\mathbf{G}_{s,v(s)}$ has a dominant eigenvalue $\lambda(s,v(s))$ equal to 1.

Moreover, for any pair (v_1,v_2), with $v_1<v_2$, the inequality $\lambda(s,v_2) \ge \lambda(s,v_1)\,[v_2/v_1]$ holds, and entails the following:

(i) for any $v<v(s)$, the strict inequality $\lambda(s,v)<1$

(ii) the inequality $\lambda'_v(s,v(s))>0$.

The Implicit Function Theorem can be applied, and it defines a real analytic function $v :]1/2,1[\to]0,1[$ that satisfies the equation

$$\lambda'_s(s,v(s)) + v'(s)\lambda'_v(s,v(s)) = 0\,. \tag{17}$$

All these remarks entail, that, for any $s \in]1/2,1[$, there exists $w(s)>v(s)$ for which the quasi-inverse $v \mapsto (I-\mathbf{G}_{s,v})^{-1}[1]$ is meromorphic for $|v|<w(s)$ with only a (simple) pole at $v=v(s)$, with the estimate,

$$(I-\mathbf{G}_{s,v})^{-1}[1] \sim_{v\to v(s)} \frac{\lambda(s,v(s))}{1-\lambda(s,v(s))}\Psi_{s,v}e_{s,v(s)}[1]\,,$$

so that the residue at $v=v(s)$ is

$$a(s) = \frac{1}{\lambda'_v(s,v(s))}\Psi_{s,v(s)}\,e_{s,v(s)}[1] \tag{18}$$

and defines an analytical map $s \mapsto a(s)$ for $s \in]1/2,1[$.

Step 2. We return to the operator $\mathbf{H}_s$ (and its quasi-inverse). For any $\Psi \in \mathcal{C}^1([0,1]^2)$, the mapping $v \mapsto (I-v\mathbf{A}_s)^{-1}[\Psi](0,1)$ is well defined and analytic for $v<1$. Then, with Proposition 3, the previous properties can be transfered to the quasi-inverse $v \mapsto (I-v\mathbf{H}_s)^{-1}[1](0,1)$: it is meromorphic for $v<u(s) := \min(1,w(s))$ with a unique pole (simple) at $v=v(s)$, and we remark the strict inequality $u(s)>v(s)$ for $s \in]1/2,1[$. Then, with singularity analysis of meromorphic functions, [for instance Theorem IV.10 p. 258 in [10]], we obtain

$$\mathbf{H}_s^k[1](0,1) = v(s)^{-k}\cdot a(s)\left[1+O\left(\frac{u(s)}{v(s)}\right)^{-k}\right], \qquad \text{for } k\to\infty\,,$$

where the coefficient $a(s)$ is related to dominant spectral properties of $\mathbf{G}_{s,v(s)}$ and is strictly positive (see (18)). Now, the analytical dependence with respect to s entails that, on any closed interval $[s_0,s_1]$ with $1/2<s_0<s_1<1$, the ratio $v(s)/u(s)$ is bounded by a constant $b<1$, whereas $|a(s)|$ and $|v'(s)|$ admit strictly positive lower bounds and $s \mapsto |a'(s)|$ an upper bound. One has there

$$e_k(s) := \sum_{|w|=k} p_w^s = \mathbf{H}_s^k[1](0,1) = v(s)^{-k}\cdot a(s)\left[1+O(b^k)\right] \qquad \text{for } k\to\infty\,.$$

One then takes the derivative with respect to s,

$$\frac{1}{k}e_k'(s) = \frac{1}{k}\sum_{|w|=k} p_w^s \log p_w = -v'(s)\, v(s)^{-k-1}\, a(s)\, \left[1 + \frac{1}{k}\frac{a'(s)}{a(s)}\frac{v(s)}{v'(s)}\right] [1 + O(b^k)]$$

$$= -v'(s)\, v(s)^{-k-1}\, a(s)\, \left[1 + O\left(\frac{1}{k}\right)\right] \qquad \text{for } k \to \infty\,.$$

As this is true on any interval $[s_0, s_1] \subset]1/2, 1[$, we deduce the asymptotic behaviour for any $s \in]1/2, 1[$,

$$\lim_{k\to\infty} \frac{1}{k}\left[\frac{e_k'(s)}{e_k(s)}\right] = -\frac{v'(s)}{v(s)}\,. \tag{19}$$

Step 3. When $s \to 1$. Due to the equality $\mathbf{G}_{s,1} = \mathbf{G}_s$, the operator $\mathbf{G}_{s,1}$ has a dominant eigenvalue equal to 1, and the function $s \mapsto v(s)$ may be extended at $s = 1$ via the equality $\lim_{s\to 1^-} v(s) = v(1) = 1$. Moreover, as $s \to 1^-$, the derivative $\lambda_s'(s, v(s))$ has a limit (equal to the derivative of the dominant eigenvalue $\lambda'(s)$ at $s = 1$), and thus the second term in (17) has also a limit: as we have already seen, the derivative $\lambda_v'(s, 1)$ is closely related to $\zeta(2s - 1)$ and tends to ∞ for $s \to 1$, and this entails that $\lim_{s\to 1^-} v'(s) = 0$. Finally the right member of (19) has a limit when $s \to 1^-$, and thus

$$\lim_{s\to 1}\lim_{k\to\infty} \frac{1}{k}\left[\frac{e_k'(s)}{e_k(s)}\right] = 0\,.$$

We are interested in the following limit (if it exists)

$$\lim_{k\to\infty} \frac{1}{k}\left[\frac{e_k'(1)}{e_k(1)}\right] = \lim_{k\to\infty} \frac{1}{k} \lim_{s\to 1}\left[\frac{e_k'(s)}{e_k(s)}\right]\,,$$

and we thus wish to exchange the limits. This is possible if we have uniform convergence of the derivatives. Uniform convergence holds in the context of monotonic functions whose (simple) limit is continuous. Here, this is the case: as the functions $s \mapsto e_k(s)$ are log concave for any k, thus, for any k, the quotient $s \mapsto e_k'(s)/e_k(s)$ defines a decreasing mapping of s, whereas the map $s \mapsto v'(s)/v(s)$ is continuous on $]1/2, 1[$ and extended with its limit when $s \to 1^-$. This legitimates the exchange of limits, and

$$\lim_{k\to\infty} \frac{1}{k}\left[\frac{e_k'(1)}{e_k(1)}\right] = \lim_{s\to 1}\lim_{k\to\infty} \frac{1}{k}\left[\frac{e_k'(s)}{e_k(s)}\right] = 0\,.$$

A.2 About the entropy of the Stern-Brocot source. Ergodic proof

Step 1. Two sources and their fundamental intervals. Both sources, the Stern-Brocot source and the Farey source, are associated with the binary coding, corresponding to the choice of the inverse branches

$$a\colon x \mapsto \frac{x}{1+x}, \quad \text{when } x < 1/2, \qquad b\colon x \mapsto \frac{1}{1+x}, \quad \text{when } x > 1/2.$$

Any Farey fundamental interval J_w is associated with a binary word $w \in \{a, b\}^\star$. Due to the equality $\{a, b\}^\star = \{a^\star b\}^\star \cdot a^\star$, any binary word $w \in \{a, b\}^k$ is written in a unique way as

$$[a^{n_1} b] \cdot [a^{n_2} b] \cdot \ldots \cdot [a^{n_\ell} b] \cdot a^r, \qquad \ell + r + \sum_{j=1}^{\ell} n_j = k, \quad r \ge 0, \quad n_j \ge 0 \;\; (\forall j \in [1, \ell])\,. \tag{20}$$

The integer m_j defined for $j \in [1 \dots \ell]$ as $m_j := n_j + 1$ satisfies $m_j \geq 1$, and the Farey LFT $a^{m-1} \circ b$ coincides with the `Cfe` LFT $g_m : x \mapsto 1/(m+x)$. Third, the LFT a^r is of the form $a^r : x \mapsto x/(1+rx)$. Then, the Farey LFT h_w associated with the word w is related to the sequence $\boldsymbol{m} = (m_1, m_2, \dots, m_\ell)$ via the `Cfe` LFT $g_{\boldsymbol{m}} := g_{m_1} \circ g_{m_2} \dots \circ g_{m_\ell}$ and the relation $h_w = g_{\boldsymbol{m}} \circ a^r$ holds.
Finally, the Farey fundamental interval $J_w = [h_w(0), h_w(1)]$ associated with the word w coincides with the interval $[g_m(0), g_m(1/(1+r))]$. Then its length p_w involves r together the coefficients of the LFT $g_m(x) := (ax+b)/(cx+d)$. The equality $|ad - bc| = 1$ holds; moreover the denominator coefficients satisfy $0 \leq c \leq d$, and the coefficient d coincides with the continuant $q(\boldsymbol{m})$ relative to the sequence $\boldsymbol{m} = (m_1, m_2, \dots m_\ell)$. Then, one has

$$p_w = \left| g_{\boldsymbol{m}}(0) - g_{\boldsymbol{m}}\left(\frac{1}{1+r}\right) \right| = \frac{1}{d(c + d(1+r))} \geq \frac{1}{d^2(r+2)}\,, \qquad \frac{1}{p_w} \leq q(\boldsymbol{m})^2 (r+2)\,. \tag{21}$$

Step 2. Changing base. For $\boldsymbol{m} \in \mathbb{N}^\star$, we denote by $\ell(\boldsymbol{m})$ the number of components of $\boldsymbol{m}$, by $c(\boldsymbol{m})$ the sum of the components of $\boldsymbol{m}$, by $q(\boldsymbol{m})$ the continuant associated with $\boldsymbol{m}$. We now fix a length k (that will tend to ∞ later). With $x \in [0,1]$, we associate the word $w_{\langle k \rangle}(x)$ of length k produced by the Stern-Brocot source on x. With (20), it defines almost everywhere a pair $(\boldsymbol{m}(x), r(x))$ with $\boldsymbol{m}(x) \in \mathbb{N}^\star$ and $r(x) \geq 0$, that depends on x and the depth k. It is thus denoted as $(\boldsymbol{m}_{\langle k \rangle}(x), r_{\langle k \rangle}(x))$ and the equality $c(\boldsymbol{m}_{\langle k \rangle}(x)) + r_{\langle k \rangle}(x) = k$ holds. It is clear (but important) to remark the following: as x belongs to the `Cfe` interval relative to $\boldsymbol{m}_{\langle k \rangle}(x)$, then the sequence $\boldsymbol{m}_{\langle k \rangle}(x)$ provides the beginning of the `Cfe` of x. Of course, this production may be quite slow (when $\ell_{\langle k \rangle}(x)$ is much smaller than k), and this is why the entropy of the Stern-Brocot source will be zero.
We then define three random variables on the unit interval that depend on k and relate the `Cfe` of x together its Farey expansion of depth k,

$$\ell\langle k \rangle(x) := \ell(\boldsymbol{m}_{\langle k \rangle}(x)), \quad c_{\langle k \rangle}(x) := c(\boldsymbol{m}_{\langle k \rangle}(x)) = m_1(x) + \dots + m_{\ell_{\langle k \rangle}(x)}(x), \tag{22}$$

$$q_{\langle k \rangle}(x) := q(\boldsymbol{m}_{\langle k \rangle}(x)) = q(m_1(x), \dots, m_{\ell_{\langle k \rangle}(x)}(x))\,. \tag{23}$$

By definition of the process, for each k, the inequality $\boldsymbol{\ell}_{\langle k \rangle}(x) \leq \boldsymbol{c}_{\langle k \rangle}(x) \leq k$ holds.

Step 3. Entropy of the Stern-Brocot source. Denote by $\pi_{\langle k \rangle}(x)$ the measure of the fundamental Farey interval of depth k the input x belongs to. Then, the entropy of the Stern-Brocot source is the limit (if it exists) of the sequence $e(k)$,

$$e(k) = \frac{1}{k}\mathbb{E}\Big[|\log \pi_{\langle k \rangle}(x)|\Big] = \frac{1}{k} \sum_{w \in \{a,b\}^k} p_w \cdot |\log p_w|\,.$$

Using (21), and applying it to the pair $(\boldsymbol{m}_{\langle k \rangle}(x), r_{\langle k \rangle}(x))$, one obtains

$$\frac{1}{\pi_{\langle k \rangle}(x)} \leq q_{\langle k \rangle}(x)^2 \cdot (r_{\langle k \rangle}(x) + 2)\,; \qquad |\log \pi_{\langle k \rangle}(x)| \leq 2 \log q_{\langle k \rangle}(x) + \log(r_{\langle k \rangle}(x) + 2)\,. \tag{24}$$

As the bound $r_{\langle k \rangle}(x) \leq k$ holds, this entails the inequality

$$e(k) \leq \frac{1}{k}\log(k+2) + 2d(k), \qquad d(k) = \mathbb{E}\left[\frac{1}{k}\log q_{\langle k \rangle}(x)\right]. \tag{25}$$

As the first term in (25) tends to 0 for $k \to \infty$, we then focus on the second term $d(k)$.

We first remark that the sequence $k \mapsto \ell_{\langle k \rangle}(x)$ is increasing (not strictly in general). Then, there are two cases: it is bounded (and then stationary) or it tends to ∞. For an input x, the sequence $\ell_{\langle k \rangle}(x)$ is stationary if and only if the Farey word produced on x finishes by an infinite sequence of a or by a b. This arises if and only if x is rational. Then, almost everywhere, the increasing sequence $k \mapsto \ell_{\langle k \rangle}(x)$ tends to ∞.

Step 4. Application of the Dominated Convergence Theorem. We now show that the sequence $d(k)$ tends to 0 with the Dominated Convergence Theorem. We thus use two inequalities for the random variable $f_k(x) := (1/k) \log q_{\langle k \rangle}(x)$. One has always, for any $m \in \mathbb{N}^\star$,

$$q(\boldsymbol{m}) \le \prod_{i=1}^{\ell(m)} (m_i + 1) \qquad \text{and} \qquad \log q(\boldsymbol{m}) \le \sum_{i=1}^{\ell(m)} \log(m_i + 1) \le \sum_{i=1}^{\ell(m)} m_i = c(\boldsymbol{m}) \,. \quad (26)$$

Applied to $\boldsymbol{m} := \boldsymbol{m}_{\langle k \rangle}(x)$, this proves the domination :

$$\frac{1}{k} \log q_{\langle k \rangle}(x) \le \frac{c_{\langle k \rangle}(x)}{k} \le 1 \qquad \text{for almost every } x \in [0, 1] \ .$$

Moreover, for any $\boldsymbol{m} \in \mathbb{N}^\star$, using again (26) in a more precise way, we derive the bound

$$\frac{1}{k} \log q(\boldsymbol{m}) \le \left[\frac{c(\boldsymbol{m})}{k}\right] \left[\frac{\ell(\boldsymbol{m})}{c(\boldsymbol{m})}\right] \left[\frac{1}{\ell(\boldsymbol{m})} \sum_{i=1}^{\ell(m)} \log(m_i + 1)\right] ,$$

that holds in particular for $\boldsymbol{m} = \boldsymbol{m}_{\langle k \rangle}(x)$.

The first factor is now at most 1 for $x \in [0, 1]$. Furthermore, almost everywhere, the sequence $\ell_{\langle k \rangle}(x) \to \infty$ when $k \to \infty$. Then, with the Ergodic Theorem, the third factor tends almost everywhere to a finite limit C. Furthermore, using a result due to Khinchin described in [15, Theorem 35], together the Ergodic Theorem applied to the sequence $\min(M, m_i)$ (for any given constant M), the second factor also tends to 0 almost everywhere. Finally, the sequence $x \mapsto (1/k) \log q_{\langle k \rangle}(x)$ tends almost everywhere to 0.

We now apply the dominated convergence theorem to the sequence of random variables $f_k(x) = (1/k) \log q_{\langle k \rangle}(x)$ that is bounded by 1 and converges to 0 almost everywhere. This proves that $d(k)$ tends to 0. This is the same for the initial entropy sequence $e(k)$.