Building Sources of Zero Entropy: Rescaling and Inserting Delays

Ali Akhavi ✉
GREYC, Université de Caen-Normandie, France
LIPN, Université Sorbonne Paris Nord, France

Fréderic Paccaut ✉
LAMFA CNRS UMR 7352, Université de Picardie Jules Verne, France

Brigitte Vallée ✉
GREYC, CNRS and Université de Caen-Normandie, France

Abstract

Most of the natural sources that intervene in Information Theory have a positive entropy. They are well studied. The paper aims in building, in an explicit way, natural instances of sources with zero entropy. Such instances are obtained by slowing down sources of positive entropy, with processes which rescale sources or insert delays. These two processes – rescaling or inserting delays – are essentially the same; they do not change the fundamental intervals of the source, but only the "depth" at which they will be used, or the "speed" at which they are divided. However, they modify the entropy and lead to sources with zero entropy. The paper begins with a "starting" source of positive entropy, and uses a natural class of rescalings of sublinear type. In this way, it builds a class of sources of zero entropy that will be further analysed. As the starting sources possess well understood probabilistic properties, and as the process of rescaling does not change its fundamental intervals, the new sources keep the memory of some important probabilistic features of the initial source. Thus, these new sources may be thoroughly analysed, and their main probabilistic properties precisely described. We focus in particular on two important questions: exhibiting asymptotical normal behaviours à la Shannon-MacMillan-Breiman; analysing the depth of the tries built on the sources. In each case, we obtain a parameterized class of precise behaviours. The paper deals with the analytic combinatorics methodology and makes a great use of generating series.

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

Keywords and phrases Information Theory, Probabilistic analysis of sources, Sources with zero-entropy, Analytic combinatorics, Dirichlet generating functions, Transfer operator, Trie structure, Continued fraction expansion, Rice method, Quasi-power Theorem

Digital Object Identifier 10.4230/LIPIcs.AofA.2022.1

Category Invited Talk

Funding This work was supported by the Agence Nationale de la Recherche through the project CODYS (ANR-18-CE40-0007).

1 Introduction

General context. A source is one of the main objects of Information Theory. A source is a probabilistic process which emits a digit from a given alphabet Σ, one at each discrete time. Very often, a source $\mathcal{P}$ is defined on the unit interval $\mathcal{I}$ and associates with $x \in \mathcal{I}$ an infinite word $M(x) = (a_1, a_2, \ldots, a_n, \ldots)$ where the successive symbols $a_i = a_i(x)$ belong to Σ. This infinite word is the expansion of x in "base" $\mathcal{P}$. Some sources are directly defined in relation with some concepts of Information Theory (memoryless sources, Markov chains), whereas other ones are related to objects of Number Theory, via numeration systems, for instance.

When both the input x and a depth k are fixed, one considers the reals y for which the word $M(y)$ has the same prefix of length k as $M(x)$. In a quite general setting, this defines an interval denoted as $\mathcal{I}_k(x)$ of length $I_k(x)$. When the unit interval is endowed with some density, and for any fixed k, the random variable $x \mapsto -\log I_k(x)$ plays a central role. The source admits the entropy h (in Shannon's meaning) when $\mathbb{E}[-\log I_k(x)] \sim h \cdot k$. Most of the classical sources (in Information Theory contexts) have a finite entropy.

Here, we focus on sources of zero entropy. Adapting the philosophy described in [3], we say that a source admits a Shannon weight f if $\mathbb{E}[-\log I_k(x)] \sim f(k)$ (when $k \to \infty$). In fact, we deal with a slightly stronger notion, and consider the notion of exponential weight (see Definition 3). Papers [2] and [3] study various sources of zero entropy that arise in Number Theory contexts, and notably two sources, the Stern-Brocot source, and the Sturm source. In [2], the authors analyze the tries built on these two sources; they exhibit in [3] their weight (however not à la Shannon, see definition 2.1 in [3]), and prove that the weight of the Stern-Brocot source is of order $\Theta(k/\log k)$ whereas the weight of the Sturm Source is of order $\Theta(\log k)$. There do not appear, in this Number Theory context, other sublinear weights, for instance of square root type $\sqrt{k}$.

Rescaling or inserting delays. We wish to build a class of sources which appear in a natural way, and admit weights (à la Shannon) of various sublinear growth. We use rescaling processes, that do not really change the source –namely the intervals $\mathcal{I}_k(x)$– but modify the depth where such an interval will be used. Here, rescaling processes, via rescaling functions g, "slow down" an initial source $\mathcal{P}$ of positive entropy and transform it into a source $\mathcal{P}_{\langle g \rangle}$ of zero entropy. Another intuitive process, for slowing down the source, consists in inserting in $\mathcal{P}$ waiting times γ (called here delays) between two symbols of $\mathcal{P}$. One obtains in this way a source $\mathcal{P}^{\langle \gamma \rangle}$ with "delays". The two processes coincide for a convenient choice of the pair (g, γ), and provide a double point of view that will be used in the sequel of the paper.

Our strategy is as follows: We first choose as initial sources $\mathcal{P}$ three types of sources in our favorite set (see Definition 13), and a "natural" class of delays, described with two parameters (a, b) and defined in Eqn (18); we then insert delays $\gamma_{a,b}$ inside each source $\mathcal{P}$ and obtain a class of sources $\mathcal{P}_{a,b}$. The Shannon weights $g_{a,b}$ of $\mathcal{P}_{a,b}$ –except for $(a, b) = (1, 0)$, where we recover the initial source – are all sublinear (and may be of various types: see Proposition 15). As we start from the same initial source $\mathcal{P}$, with nice probabilistic properties, and the rescaling process does not modify the fundamental intervals of the source, (only the depth at which they are used), we expect the source $\mathcal{P}_{a,b}$ to keep the memory of some important features of the source $\mathcal{P}$. Indeed, as the present paper shows it, such a source may be precisely analysed, and its main probabilistic properties exhibited; we focus on two phenomena: the precise behaviour of the function $x \mapsto \log I_k(x)$ (à la Shannon-MacMillan-Breimann), and the probabilistic analysis of the tries built on words emitted from $\mathcal{P}_{a,b}$.

Methods and results. Our methods are inspired by analytic combinatorics described in the book of Flajolet and Sedgewick [15]. In this context of Information Theory, the generating series of a source introduced in [22] are Dirichlet series, here called the Λ series of the source, and may be defined and a priori used for any source: paper [22] exhibits their importance in the analysis of the main probabilistic properties of the source. Later on, various papers [6, 4, 12], using the Rice method, relate the analysis of the shape of a trie built on the source with the $\Lambda(s)$ series. However, until now, except in [2], these series were only used in the context of sources of positive entropy.

Here, we use these Λ series in the wider context of sources with a possible zero entropy. We start with sources of positive entropy (gathered in a favorite set), and we distort the starting sources into sources of zero entropy. This distortion is translated into a relation between the two Λ series –attached respectively to the new source, and to the initial source– (see Proposition 8). The Λ series of sources in our favorite set are precisely described, notably from the tameness point of view (see Definition 12 and Proposition 14). Using a delay $\gamma_{a,b}$ first modifies the nature of the dominant pole of the initial $\Lambda(s)$: the parameter a moves its location, and the parameter b increases its order (See Proposition 17). Theorem 18 then describes some sufficient conditions – both on the initial tameness and parameters (a, b))– under which the tameness of the new source $\mathcal{P}_{a,b}$ may be proven.

We obtain two types of results. A first result (see Theorem 4 and Lemma 16) deals with the lengths $J_k(x)$ of fundamental intervals of the new source and proves phenomena à la Shannon-MacMillan-Breimann, described by the asymptotic normality of $x \mapsto \log J_k(x)$, adjusted with the "speed" $g_{a,b}(k)$ attached to the delay $\gamma_{a,b}$. This result is a straightforward consequence of the Quasi-Power theorem of Hwang [16]. A second result describes the expectation of the trie depth and is more subtle. Theorem 5 first relates, in a very general context, the tameness of a source (of possible zero entropy) and the expectation $\mathbb{E}[D_n]$ of its trie depth. This result (not really new) gathers many various results, some of them being classical, and some of them having not been yet considered. Using this general result and applying it to our sources $\mathcal{P}_{a,b}$ provides in Theorem 19 the analysis of the average depth $\mathbb{E}[D_n]$ for tries built on $\mathcal{P}_{a,b}$. As previously, the parameters (a, b) intervene in the behaviour of the trie, with two regimes, respectively obtained for $a = 1$ and $a > 1$.

Plan of the paper. Section 2 describes the general context. Section 3 explains the two slowing down processes which transform a source with positive entropy into a source of zero entropy. Section 4 describes the favorite set which gathers all the sources that will be used as starting points. Finally, Section 5 describes the class of sources $\mathcal{P}_{a,b}$ and exhibits their main probabilistic properties (asymptotic log-normality à la Shannon-MacMillan-Breimann and estimates for the average depth trie).

2 Sources, weights, generating series

We first describe a general source $\mathcal{Q}$ –of possible zero entropy– related to partitions in Section 2.1 and introduce its generating functions. The following sections then explain how these series intervene in two questions of interest: the asymptotic normality of $x \mapsto \log I_k(x)$ relative to the source $\mathcal{Q}$ (in Section 2.2) and the asymptotic behaviour of the average trie depth, when the trie is built on the words emitted by the source $\mathcal{Q}$ (in Section 2.3).

2.1 General sources associated with partitions

A source $\mathcal{Q}$ is a probabilistic process which emits a digit from a given alphabet Σ of cardinality r (possibly infinite denumerable), one at each discrete time. A source $\mathcal{Q}$ is very often defined on the unit interval $\mathcal{I}$ and associates with $x \in \mathcal{I}$ an infinite word $M(x) = (a_1, a_2, \ldots, a_n, \ldots)$ where the successive digits $a_i = a_i(x)$ belong to Σ. This infinite word is called the expansion of x in "base" $\mathcal{Q}$. Some sources are directly defined in relation with some concepts of Information Theory (memoryless sources, Markov chains), whereas other ones are related to objects of Number Theory, via numeration systems, for instance.
For a fixed depth k, one considers the reals y for which the word $M(y)$ begins with a given prefix w of length k. When both the input x and a depth k are fixed, one also considers the reals y for which the word $M(y)$ has the same prefix of length k as $M(x)$. In a quite general

setting[1], this defines (up to a denumerable subset of $\mathcal{I}$) two families of intervals, respectively denoted as $\mathcal{I}_w$ and $\mathcal{I}_k(x)$. The interval $\mathcal{I}_w$ for $w \in \Sigma^\star$, of length p_w, is the fundamental interval associated with the prefix w; its length p_w is the probability that the word $M(x)$ begins with prefix w; the interval $\mathcal{I}_k(x)$, of length $I_k(x)$, is the k-th coincidence interval of x. When the unit interval is endowed with some density, the mapping $x \mapsto \log I_k(x)$ is, for any fixed k, a random variable of great interest, whose asymptotics (when $k \to \infty$) is widely studied. In particular, the paper [3] defines two notions of weights (related to the asymptotics of this random variable) that extend the notions of entropy (almost everywhere, in probability) introduced in [9].
Here, we consider two weights of different flavour: first, now, the Shannon weight; then, in the next Section, the exponential weight.

▶ **Definition 1.** *A source $\mathcal{Q}$ has a Shannon weight $g(k) \geq 0$, and, resp. a Shannon entropy $h > 0$, if the sequence $\mathbb{E}[-\log I_k]$ satisfies the following (for $k \to \infty$),*

$$(1/g(k))\,\mathbb{E}[-\log I_k] \to 1\,, \qquad (1/k)\,\mathbb{E}[-\log I_k] \to h\,.$$

The notion of Shannon weight thus extends the Shannon entropy, and sources with zero Shannon entropy have a sublinear Shannon weight $g(k)$ for which $g(k)/k \to 0$. The second weight (exponential weight) is described via the generating series of the source. These generating series were first introduced and used in [22].

▶ **Definition 2.** *For $s \in \mathbb{C}$, the Λ generating series of the source involve the fundamental probabilities p_w that a word begins with a prefix w, for a given depth k, or for all depths k,*

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

These generating series are Dirichlet series that satisfy $|\Lambda_k(s)| \leq \Lambda_k(\Re s)$, $|\Lambda(s)| \leq \Lambda(\Re s)$.

2.2 Exponential weight and asymptotic normality of $\log I_k(x)$.

The exponential weight is defined via the (possible) quasi-power behaviour of the $\Lambda_k(s)$ series. We will relate it later with the Shannon weight (Theorem 4*a*).

▶ **Definition 3.** *The source $\mathcal{Q}$ has an exponential weight f if its Λ_k series satisfies the following: There exist a real number A (with $-\infty \leq A < 0$) and two analytic functions u and v defined on a complex neighborhood $\mathcal{V}$ of the real half-line $s > A$, for which, for any real $c > A$, there exists a complex neighborhhod $\mathcal{V}_c$ of c and a function ϵ_c for which the following estimate for Λ_k,*

$$\Lambda_k(s) = v(s)\cdot u(s)^{f(k)}\cdot[1 + O(\epsilon_c(k))]\,, \qquad \epsilon_c(k) \to 0 \quad (k\to\infty) \tag{2}$$

holds (uniformly) for $s \in \mathcal{V}_c$. The function u is called a base function.

The series $\Lambda_k(s)$ is closely related to the moment generating function $M_k(s)$ of the variable $(-\log I_k(x))$. Indeed, as the random variable $x \mapsto I_k(x)$ is a staircase function that has value p_w on the interval fundamental $\mathcal{I}_w$ ($w \in \mathcal{A}^k$), the moment generating function $M_k(s) = \mathbb{E}[I_k^{-s}(x)]$ is expressed in terms of the family (p_w) for $w \in \mathcal{A}^k$,

$$M_k(s) = \mathbb{E}[I_k^{-s}(x)] = \sum_{w\in\mathcal{A}^k} p_w^{-s}\cdot p_w = \sum_{w\in\mathcal{A}^k} p_w^{1-s} = \Lambda_k(1-s)\,. \tag{3}$$

[1] This is the case when there is an underlying sequence of partitions, in the sense of [9] or [2]. This will be the case here (Section 4).

The importance of exponential weight is related to the Quasi-Power Theorem due to Hwang [16] that deals with the moment generating function, when it has a quasi-power form as in (2). Due to Relation (3), it leads to an asymptotic Gaussian law for the mapping $x \mapsto \log I_k(x)$. The occurrence of a Gaussian law is a refinement of the Shannon-MacMillan-Breimann Property, which usually deals with the almost everywhere behaviour of $\log I_k(x)$.

▶ **Theorem 4.** *Consider a source $\mathcal{Q}$ and the length $I_k(x)$ of its k-th coincidence intervals. If the source $\mathcal{Q}$ admits an exponential weight f with base u, the random variable $x \mapsto \log I_k(x)$ satisfies the following:*

(a) *Its expectation and its variance admit the following asymptotic estimates for $k \to \infty$,*

$$\mathbb{E}[\log I_k(x)] = u'(1) \cdot f(k) + O(1), \quad \mathrm{Var}[\log I_k(x)] = (u''(1) - u'(1)^2) \cdot f(k) + O(1)\,.$$

In particular, the source $\mathcal{Q}$ has a Shannon weight equal to $|u'(1)| \cdot f(k)$.

(b) *If the function $u(s)$ is strictly log-convex, the variable $x \mapsto \log I_k(x)$ asymptotically follows a gaussian law*

$$\Pr\left[x \,\middle|\, \frac{\log I_k(x) - u'(1) f(k)}{[(u''(1) - u'(1)^2) f(k)]^{1/2}} \le x\right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt\,.$$

Usually, classical applications of the Quasi-Power Theorem exhibit asymptotic gaussian laws for which the expectation and the variance are (most of the time)[2] of order $\Theta(k)$ or $\Theta(\log k)$ (See [16]). The present paper will provide natural instances of applications of the Quasi-Power Theorem with various behaviours of the expectation and the variance (See Section 5.2 and Theorem 16).

2.3 Role of the Λ series in the analysis of tries

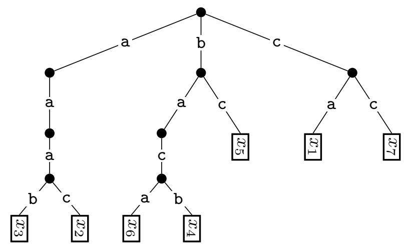
We are also interested in a probabilistic analysis of the shape of a trie $\texttt{Trie}(\boldsymbol{x})$ built on a sequence $\boldsymbol{x}$ of (infinite) words that are independently drawn from the source $\mathcal{Q}$ of alphabet Σ. $\texttt{Trie}(\boldsymbol{x})$ is a tree that is recursively defined via the cardinality $N(\boldsymbol{x})$ of the sequence $\boldsymbol{x}$:

(a) If $N(\boldsymbol{x}) = 0$, then $\texttt{Trie}(\boldsymbol{x}) = \emptyset$;
(b) If $N(\boldsymbol{x}) = 1$, with $\boldsymbol{x} = (x)$, then $\texttt{Trie}(\boldsymbol{x})$ is a leaf labeled by x;
(c) If $N(\boldsymbol{x}) \ge 2$, then $\texttt{Trie}(\boldsymbol{x})$ is formed with an internal node and r subtries[a] resp. equal to

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

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.

[a] r is the cardinality of the alphabet Σ.



Here, we focus on a particular parameter of $\texttt{Trie}(\boldsymbol{x})$, defined as the length of a *random* branch and called the trie depth. When $N(\boldsymbol{x}) = n$, it is denoted by $D_n(\boldsymbol{x})$.

[2] There is a notable exception in the study of partitions [17] where expectations and variance of order k^d for any $d \in]0, 1[$ occur.

The trie is central in text algorithmics [11]. The study of its geometric parameters, notably its depth, is thus an important subject in analysis of algorithms, where two different strategies have been introduced: the paper [13] uses the Rice method, whereas the book [21] is based on Depoissonization tools. (See [23] for a comparison of these methodologies). Here, we choose the Rice method, which was introduced by Nörlund [18], [19] and widely used in analytic combinatorics since the seminal papers of Flajolet and Sedgewick [13] [14]. The Rice method is indeed dedicated to the study of sequences of the form Eqn (4), which exhibit the main role played by the function $s\Lambda(s)$, provided it *be tame at* $s = c$ for $c < 2$.

Tameness is a central notion in analysis of algorithms which has been only recently introduced in various works (see [4], [7], [5], [2] for instance). It is defined by a position, an order and a shape. Tameness will be precisely described later in Definition 12. For the moment, we give an informal definition:

The function $\Lambda(s)$, that is analytic on $\{\Re s > c\}$ is tame *at $s = c$, with order $b \geq 0$, if there exists a region $\mathcal{R}_c \supset \{\Re s \geq c\}$ where the function Λ is meromorphic, has a sole possible pole of order $b + 1$ ($b \geq 0$) at $s = c$ and is of polynomial growth there as $|\Im s| \to +\infty$. Moreover, this region $\mathcal{R}_c$ has one of the three following possible shapes, described in Definition 12 and in Figure 1:* a periodic shape (P), an hyperbolic shape (H), or a strip shape (S) .

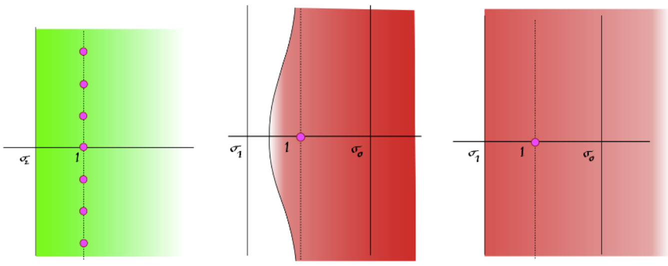


Figure 1 Tameness regions, drawn at $c = 1$, and their possible shapes (from left to right) : Periodic shape (P) – Hyperbolic shape (H) – and Strip shape (S).

The next result relates the analysis of the average depth $\mathbb{E}[D_n]$ and the tameness of the series $\Lambda(s)$. The proof is given in the annex (Section B.1). The result has not been stated before in its full generality, because the study of sources with positive entropy only relies on the case ($c = 1$, $b = 0$), and the possible strip shape only recently "discovered". Only two other particular cases ($c = 1$, $b = 1$) and ($c = 3/2$, $b = 0$) have been already studied in [2].

▶ **Theorem 5.**

(i) If the series $\Lambda(s)$ is well-defined for $s \geq 2$*(and thus for $\Re s \geq 2$), the expectation $\mathbb{E}[D_n]$ of the trie depth is expressed as an alternating sum which involves $\Lambda(\ell)$, for $\ell \geq 2$,*

$$n\mathbb{E}[D_n] = \sum_{\ell=2}^{n} (-1)^{\ell} \binom{n}{\ell} \ell\, \Lambda(\ell)\,. \tag{4}$$

(ii) If the series $s\Lambda(s)$ is tame at $s = c$ for $1 \leq c < 2$ with order b and a tameness region $\mathcal{R}_c$ delimited by a frontier δ_c, *the following holds:*

(a) *the expectation $n\mathbb{E}[D_n]$ involves the Rice kernel $L_n(s)$ and decomposes as*

$$n\,\mathbb{E}[D_n] = \mathrm{Res}\left[L_n(s) \cdot s\Lambda(s); s = c\right] + \int_{\delta_c} L_n(s) \cdot s\Lambda(s) ds \quad \textit{with} \;\; L_n(s) := \frac{\Gamma(n+1)\Gamma(-s)}{\Gamma(n+1-s)}\,.$$

(b) *The general estimate holds,*

$$n\mathbb{E}[D_n] = \begin{cases} n^c P_b(\log n) & +O(n^{c-\delta}), & (S) \text{ shape}; \\ n^c P_b(\log n) & +n^c O\left(\exp[-(\log n)^\rho]\right), & (H) \text{ shape}; \\ n^c G_b(\tau \log n) & +O(n^{c-\delta}), & (P) \text{ shape}. \end{cases} \tag{5}$$

The dominant term $\mathrm{Res}\,[L_n(s) \cdot s\Lambda(s); s = c]$ *involves a polynomial* P_b*, or a combination* G_b *between polynomials and a periodic function* Π*, together with the boolean* $[\![c = 1]\!]$*,*

$$P_b(u) = [\![c = 1]\!]\, \alpha_{b+1}\, u^{b+1} + \left[\sum_{i=0}^{b} \alpha_i u^i\right], \quad G_b(u) = [\![c = 1]\!]\, \alpha_{b+1}\, u^{b+1} + \left[\sum_{i=0}^{b} \alpha_i u^i\right] \Pi(\{u\}).$$

The remainder term $\int_{\delta_c} L_n(s) \cdot s\Lambda(s)ds$ *involves the tameness shapes of Definition 12; in particular, in case of the* (H) *shape with exponent* ν*, the exponent* ρ *satisfies* $\rho < \frac{1}{1+2\nu}$*. For the other two shapes,* δ *is the strip width.*

The two main theorems (Theorems 4 and 5) exhibit the importance of the two generating series $\Lambda_k(s)$ and $\Lambda(s)$ of the source $\mathcal{Q}$: the Λ_k series intervenes via the exponential weight of the source in the study of the map $x \mapsto \log I_k(x)$; the Λ series intervenes via its tameness in the study of the average trie depth. The strategy of the paper is then as follows: we start with a given source $\mathcal{P}$ of positive entropy, with a well-known behaviour of its generating series $\Lambda_k(s)$ and $\Lambda(s)$. Section 4 describes the choice of the "starting" sources. Then, we distort the source $\mathcal{P}$ into another source $\mathcal{Q}$ (of zero entropy), in a way $\mathcal{Q}$ keeps the memory of some important features of $\mathcal{P}$. The next Section describes the distortion process.

3 Change of sources via rescaling or inserting delays

This section describes two processes for slowing down a given source: first, the rescaling process (in Sections 3.1 and 3.2), then the insertion of delays (in Section 3.3). The two processes are essentially the same, and they are compared in Section 3.4. Then we describe in Section 3.5 the particularities of these new sources when using tries. The Section finally establishes (in 3.6) a transfer of fundamental importance, between the Λ series of the initial source and the Λ series of the new source.

3.1 Rescaling sources and change of weights

The following process creates another source $\mathcal{Q}$ from a given source $\mathcal{P}$: It keeps the same fundamental intervals, but it changes the depth where they are used. It thus modifies the "speed" of the subdivision of the related partitions.

▶ **Definition 6.**

(i) *A function* g *is a rescaling function if it is defined on* $\{0\} \cup \mathbb{R}_{\geq 1}$ *and satisfies the following:*
  - **(a)** *the equalities* $g(0) = 0$*,* $g(1) = 1$ *hold ;*
  - **(b)** g *is a strictly increasing continuous map* $g : \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 1}$*, with* $\lim_{x\to\infty} g(x) = \infty$ *;*
  - **(c)** *the inverse function* g^{-1} *has integer values at integer points.*

(ii) *With a source* $\mathcal{P}$*, its coincidence intervals* $\mathcal{I}_k(x)$*, together with a rescaling function* g*, one associates the source* $\mathcal{Q}$ *where the coincidence intervals* $\mathcal{J}_k(x)$ *satisfy*

$$\mathcal{J}_k(x) := \mathcal{I}_{\lfloor g(k) \rfloor}(x) \,. \tag{6}$$

The source $\mathcal{Q}$ *is the rescaled source of* $\mathcal{P}$ *via rescaling* g*. It is denoted as* $\mathcal{P}_{\langle g \rangle}$*.*

Rescaling acts here on the depth of fundamental intervals in a uniform[3] way: it changes the depth in the same way for all the intervals of the same initial depth. Remark that Item (c) entails the equality $\mathcal{I}_k = \mathcal{J}_{g^{-1}(k)}$. Using the notions of the rescaled source (Definition 6) and various weights (given in Definitions 1 and 3), this leads to the following result:

▶ **Proposition 7.** *Consider a rescaling g. If the source $\mathcal{P}$ has weight f (exponential, resp. Shannon), then the source $\mathcal{P}_{\langle g\rangle}$ has a weight (exponential, resp. Shannon) equal to $f \circ \lfloor g \rfloor$.*

Rescaling changes the weight; starting with a source $\mathcal{P}$ of entropy $h > 0$ (i.e., a Shannon weight $f(k) = h \cdot k$), the source $\mathcal{P}_{\langle g\rangle}$ has a Shannon weight $h \cdot \lfloor g(k) \rfloor$.

Then, using rescaling g for which $g(k)/k \to 0$, the source $\mathcal{P}_{\langle g\rangle}$ is of zero entropy.

3.2 Rescaled source

We associate with a rescaling g its sequence of differences,

$$\delta(g)(k) := \lfloor g(k) \rfloor - \lfloor g(k-1) \rfloor \qquad k \geq 1\,. \tag{7}$$

When $g(k) = k$ for any k, the differences $\delta(g)(k)$ are all equal to 1. For other functions g, there may exist integers k for which $\delta(g)(k)$ may be 0, or strictly larger than 1. We now describe the influence of rescaling on the expansion of x in base $\mathcal{P}$,

$$M^{(\mathcal{P})}(x) = a_1\, a_2\, a_3 \ldots a_k, \ldots\,, \qquad (a_i \in \Sigma)\,; \tag{8}$$

we consider the generalized digits $b_1\, b_2\, \ldots, b_k, \ldots b_n, \ldots$, where each b_k is a block of initial digits of length $\delta(g)(k)$, and the expansion of x in base $\mathcal{P}_{\langle g\rangle}$ is

$$M^{(\mathcal{P}_{\langle g\rangle})}(x) = b_1\, b_2\, b_3 \ldots b_k \ldots\,, \qquad b_k = \left[a_{\lfloor g(k-1)\rfloor+1} \ldots a_{\lfloor g(k)\rfloor}\right], \ldots \tag{9}$$

When the k-th block is empty (i.e., $\delta(g)(k) = 0$ or $\lfloor g(k) \rfloor = \lfloor g(k-1) \rfloor$), the equality $\mathcal{J}_k(x) = \mathcal{J}_{k-1}(x)$ holds between two successive coincidence intervals, and the source $\mathcal{P}_{\langle g\rangle}$ does not emit any digit from Σ at time k. In this situation, we decide to emit a fictive symbol $[\cdot]$ at time k; in this way, we get a proper coding for the words emitted by the new source, and remember that it does not emit any digit from Σ at time k. Letting $\Sigma^0 := \{[\cdot]\}$, the generalized digit b_k always belongs to $\Sigma^{\delta(k)}$.

3.3 Inserting delays in a source $\mathcal{P}$.

The rescaled source $\mathcal{P}_{\langle g\rangle}$ may emit generalized digits as soon as $\delta(g)(k) \geq 2$. We now decide to always deal with a rescaling g for which all the differences defined in (7) satisfy $\delta(g)(k) \in \{0, 1\}$. In this case, the source $\mathcal{P}_{\langle g\rangle}$ only emits digits from the alphabet $\overline{\Sigma} = \{[\cdot]\} \cup \Sigma$. In such a source $\mathcal{P}_{\langle g\rangle}$, we view the symbol $[\cdot]$ as a waiting symbol, describing the situation where we wait for emitting the next symbol and do not emit any symbol from Σ. In this way, the length $\gamma(\ell) \geq 1$ of the plateau between the two successive indices (where symbols $a_{\ell-1} \in \Sigma$ and $a_\ell \in \Sigma$ are emitted) measures the ℓ-th waiting time between two times when symbols from Σ are emitted. In computer science contexts, the waiting time $\gamma(\ell)$ is called the ℓ-th "delay".

We are then led to another process for slowing down a source $\mathcal{P}$ defined on Σ: starting with source $\mathcal{P}$, where the expansion of x in base $\mathcal{P}$ is given in (8), we insert, for each ℓ, a delay $\gamma(\ell)$ between $a_{\ell-1}$ and a_ℓ, and obtain a new expansion, on the alphabet $\overline{\Sigma} = \{[\cdot]\} \cup \Sigma$,

$$M^{(\mathcal{P}^{\langle\gamma\rangle})}(x) = a_1\, [\cdot]^{\gamma(2)-1}\, a_2 \ldots [\cdot]^{\gamma(\ell)-1}\, a_\ell\, \ldots\,. \tag{10}$$

---

[3] See the conclusion for a possible non-uniform definition.

3.4 Comparing the two points of view

Starting from a source $\mathcal{P}$ on Σ, we have thus introduced two new sources on $\overline{\Sigma} = \{[\cdot]\} \cup \Sigma$,
(a) $\mathcal{P}_{\langle g\rangle}$ is rescaled from $\mathcal{P}$ via a rescaling g with differences $\delta(g) \in \{0, 1\}$ (described in (9)).
(b) $\mathcal{P}^{\langle\gamma\rangle}$ is obtained from $\mathcal{P}$ via inserting delays $\gamma \geq 1$ (described in (10)).

First, the equivalence holds: $\delta(g)(k) \in \{0, 1\} \iff \gamma(\ell) \geq 1$
Second, in this case, comparing the two expansions described in (9) and (10), and using Item (c) of Definition 6 proves the coincidence of the two sources $\mathcal{P}_{\langle g\rangle}$ and $\mathcal{P}^{\langle\gamma\rangle}$ when the pair (g, γ) satisfies

$$\text{Relation} \quad (R) \qquad \sum_{k=1}^{\ell} \gamma(k) = g^{-1}(\ell) \qquad \text{for any integer } \ell \geq 1\,. \tag{11}$$

3.5 Sources $\mathcal{P}^{\langle\gamma\rangle}$ in Information Theory contexts.

We deal with infinite words emitted by $\mathcal{P}^{\langle\gamma\rangle}$, described in (10). As the inserted delays γ are the same for any word of the source, the words of $\mathcal{P}^{\langle\gamma\rangle}$ are very particular, and easy recognizable: they are obtained with the insertion of a deterministic process (the delays) at deterministic indices into a probabilistic one (the source $\mathcal{P}$).

We are interested in sorting a sequence $\boldsymbol{x}$ of words emitted by $\mathcal{P}^{\langle\gamma\rangle}$. A first idea is to use the trie $\texttt{Trie}(\boldsymbol{x})$. In fact, there is a close connection between the two tries: $\texttt{Trie}(\boldsymbol{x})$ and $\texttt{Trie}(\boldsymbol{y})$ built on the compressed sequence $\boldsymbol{y}$, formed with the compressed words of $\boldsymbol{x}$, where the delays are removed. In the annex, (see Section D.1), we compare the branches of the two tries, and their lengths, and prove the following important property which will be the starting point for proving Proposition 20:

When the branches of the compressed $\texttt{Trie}(\boldsymbol{y})$ *have length* $k_i(\boldsymbol{y})$*, the branches of* $\texttt{Trie}(\boldsymbol{x})$ *have length* $g^{-1}(k_i(\boldsymbol{y}))$*, where* g *is the rescaling associated with* γ *via Relation* (R)*.*

When we use $\texttt{Trie}(\boldsymbol{x})$ for sorting the sequence $\boldsymbol{x}$, there are three cases: the first case arises when the "trie-woman" knows that she deals with a source of type $\mathcal{P}^{\langle\gamma\rangle}$, with a precise knowledge of the delays. In this case, as she is not stupid, she probably does not build the trie $\texttt{Trie}(\boldsymbol{x})$: she first compresses the sequence $\boldsymbol{x}$ (i.e. removes the delays), obtains a sequence $\boldsymbol{y}$, and then uses the trie $\texttt{Trie}(\boldsymbol{y})$ –built on the source $\mathcal{P}$– to sort the sequence $\boldsymbol{x}$. However, there are other two cases, that are called "blind cases" : [(i) the "trie-woman" knows that she deals with a source of type $\mathcal{P}^{\langle\gamma\rangle}$, without knowing the form of the delay or (ii) she does not know that she deals with a $\mathcal{P}^{\langle\gamma\rangle}$ source]. In these blind cases, she builds the trie $\mathcal{T}(\boldsymbol{x})$ and uses it. We assume here that the "trie woman" is always blind.

3.6 Influence of rescaling or inserting delays on the Λ series.

The following result describes the influence of rescaling –or inserting delays– on the Λ series.

▶ **Proposition 8.** *The* Λ_k *series of the source* $\mathcal{P}_{\langle g\rangle}$ *obtained by rescaling* $\mathcal{P}$ *with* g *is*

$$\Lambda_k^{(\mathcal{P}_{\langle g\rangle})}(s) = \Lambda_{\lfloor g(k)\rfloor}^{(\mathcal{P})}(s)\,. \tag{12}$$

The Λ *series of the source* $\mathcal{P}^{\langle\gamma\rangle}$ *obtained by inserting delays* γ *in* $\mathcal{P}$ *is*

$$\Lambda^{(\mathcal{P}^{\langle\gamma\rangle})}(s) = \sum_{k\geq 0} \Lambda_k^{(\mathcal{P}^{\langle\gamma\rangle})}(s) = \sum_{\ell\geq 0} \gamma(\ell)\, \Lambda_\ell^{(\mathcal{P})}(s)\,. \tag{13}$$

This easy result, of fundamental importance, shows that each point of view –rescaling or inserting delays– is of interest: The Λ_k series of a rescaled source $\mathcal{P}_{\langle g\rangle}$, via a rescaling g, will be studied via g. The Λ series of a source $\mathcal{P}^{\langle\gamma\rangle}$, with delays γ, will be studied via γ.

4 The favorite set

The paper starts with a well understood source of positive entropy, and transforms it with rescaling or inserting delays. This section is devoted to the "starting" sources. We will use dynamical sources (see Section 4.1), and we need these sources to be good (see Section 4.2) and tame (see Section 4.3). We then describe in Section 4.4 our favorite set which gathers the "starting" sources we choose: they are all good and tame, with various tameness shapes.

4.1 Dynamical sources

Dynamical sources, introduced by Vallée in [22], are related to dynamical systems of the interval $\mathcal{I} := [0,1]$. One starts with a topological partition $\{\mathcal{I}_\sigma\}$ of $\mathcal{I}$ indexed by symbols $\sigma \in \Sigma$, a coding map $\tau : \mathcal{I} \to \Sigma$ which equals σ on $\mathcal{I}_\sigma$, and a shift map $T : \mathcal{I} \to \mathcal{I}$. The mapping T is defined by its branches $T_\sigma : \mathcal{I}_\sigma \to \mathcal{I}$ that are assumed to be *surjective*, of class $\mathcal{C}^2$, and strictly monotonic. The source produces on the real x the word $M(x)$ that encodes the trajectory $(x, Tx, T^2x \ldots)$ via the coding map τ, namely, $M(x) = (\tau(x), \tau(Tx), \tau(T^2x), \ldots)$ ..
When the input x is randomly drawn from $\mathcal{I}$, this becomes a probabilistic process.
Each (local) inverse of T^k is associated with a prefix $w \in \Sigma^k$ and denoted as h_w. As each branch of T is surjective, each inverse branch h_w is defined on $\mathcal{I}$, and the image of h_w is the interval $\mathcal{I}_w = [h_w(0), h_w(1)]$ which gathers all the reals x for which $M(x)$ begins with w. There is thus an underlying sequence $(\mathcal{P}_k)_k$ of partitions, with $\mathcal{P}_k := \{\mathcal{I}_w \mid w \in \Sigma^k\}$ which will be inherited for rescaled sources.

All memoryless sources are dynamical sources, associated with a increasing piecewise linear shift. A main instance is the standard binary system, obtained by $T(x) = \{2x\}$ (where $\{\cdot\}$ is the fractional part). Another dynamical source plays a central role here: the source `CF`, related to continued fraction, and associated with the non-linear shift $T(x) = \{1/x\}$.

In the context of analytic combinatorics, the importance of dynamical sources relies on the following fact: their Λ series are themselves generated by the secant transfer operator:

▶ **Lemma 9.** *The secant transfer operator $\mathbf{G}_s$ of the source involves the inverses h_σ of the branch T_σ and acts on functions $F : \mathcal{I}^2 \to \mathbb{C}$, as follows:*

$$\mathbf{G}_s[F](x,y) = \sum_{\sigma\in\Sigma} \left| \frac{h_\sigma(x) - h_\sigma(y)}{x-y} \right|^s F(h_\sigma(x), h_\sigma(y))\,. \tag{14}$$

The k-th iterate $\mathbf{G}_s^k$ involves the inverse branches h_w of the shift T^k

$$\mathbf{G}_s^k[F](x,y) = \sum_{w\in\Sigma^k} \left| \frac{h_w(x) - h_w(y)}{x-y} \right|^s F(h_w(x), h_w(y))\,.$$

The fundamental relations hold: $\quad \Lambda_k(s) = \mathbf{G}_s^k[1](0,1), \quad \Lambda(s) = (I - \mathbf{G}_s)^{-1}[1](0,1)\,.$

4.2 Good sources

The Good Class gathers the dynamical sources for which there is an iterate T^n of the shift T that is strictly expansive. This class contains, together with all the memoryless sources, many other sources, as the Continued Fraction Source.

▶ **Definition 10.** *A dynamical source is good if it satisfies the following:*
(i) *The constant ρ defined as follows satisfies $\rho < 1$,*

$$\rho = \limsup_{k\to\infty} (\sup_{w\in\Sigma^k} p_w)^{1/k} .$$

(ii) *There is a constant $B > 0$, for which, for any $w \in \Sigma$, one has $|h''_w| \le B|h'_w|$.*
(iii) *There exists $A < 1$ for which the series $\sum_{w\in\Sigma} p_w^s$ converges on $\Re s > A$.*

Item (i) is the most important; items (ii) and (iii) are only useful for sources on infinite alphabets Σ, and the real A in (iii) is the convergence abscissa of the series $\Lambda_1(s)$.
When the source is good, the secant transfer operator, defined in (14), when acting on the functional space $\mathcal{C}^1([0,1]^2)$, has, for any real $s > A$, a unique dominant eigenvalue $\lambda(s)$, separated from the remainder of the spectrum by a spectral gap. The function $\lambda(s)$ (which depends analytically on s) is called the dominant eigenvalue of the source. It satisfies $\lambda(1) = 1$, $\lambda'(1) = -h$ where $h > 0$ is the entropy of the source. Moreover, there exists a (complex) neighborhood $\mathcal{V}$ of the real axis ($s > A$) on which the series $\Lambda_k(s)$ has a quasi-power behaviour that involves functions $v(s)$ and $c_k(s)$, that are analytic on $\mathcal{V}$, under the form,

$$\Lambda_k(s) = v(s)\lambda(s)^k + c_k(s) = v(s)\lambda(s)^k \left[1 + \frac{1}{v(s)}\frac{c_k(s)}{\lambda(s)^k}\right] = v(s)\lambda(s)^k \left[1 + O(\rho^k)\right] . \tag{15}$$

The last estimate indeed holds, due to the Spectral Radius Theorem and the existence of a spectral gap. Comparing with Eqn (2) leads to the following result:

▶ **Proposition 11.** *A good source has a "dominant eigenvalue" denoted as $\lambda(s)$. It has an exponential weight equal to k, with a base $u(s) = \lambda(s)$. Its Shannon entropy is $|\lambda'(1)|$.*

Moreover the pressure function $L(s) = \log\lambda(s)$ admits convexity properties : it is always convex and the two conditions are equivalent (see for instance [1]):
(a) the source is conjugated[4] to an unbiased memoryless source;
(b) the function L is affine.
For a memoryless source (p, q), the function $\lambda(s)$ is defined on the whole plane $\mathbb{C}$, and

$$\lambda(s) = (p^s + q^s) , \qquad \Lambda(s) = \frac{1}{1-\lambda(s)} . \tag{16}$$

The three conditions are equivalent (See Section C.1 in the annex).
(a) There exists $\tau > 0$ for which $\lambda(1 + 2i\pi\tau) = 1$;
(a) The function λ is periodic with period $2i\pi\tau$;
(a) The ratio $(\log q)/(\log p)$ is rational.
In this case, the memoryless source is periodic with period τ. Otherwise, it is aperiodic.

4.3 Tameness and tameness shapes

With Eqn (15), the Λ series satisfies

$$\Lambda(s) = v(s)\frac{1}{1-\lambda(s)} + c(s), \qquad c(s) = \sum_k c_k(s) , \tag{17}$$

[4] Two sources are conjugated if their shifts are conjugated in $\mathcal{C}^2([0,1])$.

where $v(s)$ and $c(s)$ are analytic for $s \in \mathcal{W} := \mathcal{V} \cap \{\Re s > \tau\}$ for some $\tau < 1$. As the inequality $|\Lambda(s)| \le \Lambda(\Re s)$ holds on the hyperplane $\Re s > 1$, the function $\Lambda(s)$ is meromorphic on the domain $\{\Re s > 1\} \cup \mathcal{W}$ with an only simple pole at $s = 1$. However, in order to apply the Rice formula and to be able to study tries (see Theorem 5), we need to know a region on the left of the vertical line $\Re s = 1$ where the Λ series is *tame*, i.e, meromorphic and of polynomial growth when $|\Im s| \to \infty$. We now give a general definition of tameness, where the point of tameness may be any $c \ge 1$. This definition is already used in the statement of Theorem 5.

▶ **Definition 12.**

(i) *A function $M(s)$ analytic on $\{\Re s > c\}$ (c real) is* tame *at c of order $b \ge 0$, if there exists a region $\mathcal{R}_c \supset \{\Re s \ge c\}$ where the function M is meromorphic, has a sole possible pole of order $b+1$ at $s = c$ and is of polynomial growth as $|\Im s| \to +\infty$. Moreover, this region $\mathcal{R}_c$ has one of the three following possible*[5] *shapes (See Fig. 1):*

(S) Strip shape: $\mathcal{R}_c = \{s \mid \Re s > c - \delta\}$ *for some* $\delta > 0$.

(H) Hyperbolic shape (of exponent ν): $\mathcal{R}_c = \mathcal{R}_c^+ \cup \mathcal{R}_c^-$ *(with positive constants A, B)*

$$\mathcal{R}_c^+ := \{s = \sigma + it \mid |t| \ge B,\ \ \sigma > c - \tfrac{A}{|t|^\nu}\ ,\ \ \mathcal{R}_c^- := \{s = \sigma + it \mid |t| \le B\ \ \sigma > c - \tfrac{A}{B^\nu}, \}\,.$$

(P) Periodic[6] shape (of period τ): $\mathcal{R}_c = \{s \mid \Re s > c - \delta\} \setminus \bigcup \{s_k\}$ *where the points s_k, of the form $s_k = c + 2ki\pi\tau$ (for $k \in \mathbb{Z}$, $k \neq 0$) are poles of Γ of order at most $b+1$.*

(ii) *A source $\mathcal{Q}$ is tame at c of order b with a tameness shape in $\{S, H, P\}$ if its Λ series is tame at c of order b with the given tameness shape.*

(iii) *A good source is tame if its Λ series is tame at $s = 1$ with order 0.*

4.4 Instances of tame sources. Our favorite set of starting sources

These various prescribed tameness shapes (that may appear at a first glance somewhat artificial) indeed intervene in possible behaviors of classical sources. Here, we choose particular sources $\mathcal{P}$ of positive entropy as starting points: we wish good sources, proven to be tame, with various tameness shapes in $\{S, H, P\}$. This leads us to our favorite set of starting sources:

▶ **Definition 13.** *The favorite set of starting sources contains three types of sources:*

(a) *the class $\mathcal{MP}$ of all the periodic memoryless sources ;*

(b) *the class $\mathcal{MA}$ of all the aperiodic memoryless sources related to a pair $(p, q = 1-p)$ for which the ratio $\beta = (\log q)/(\log p)$ is irrational with a finite irrationality exponent;*

(c) *the Continued Fraction source, denoted as* `CF`, *associated with the shift $T(x) = \{1/x\}$.*

The following theorem first shows that our favorite set is well-chosen. Then, using Theorem 14 inside Theorem 5 leads to the analysis of the trie depth for each "favorite" source.

▶ **Theorem 14.** *All the starting sources of the favorite set are tame. Their shape is*

(P) *for any source of the class $\mathcal{MP}$*

(H) *for any source of the class $\mathcal{MA}$: the exponent ν of the tameness region is related to the irrationality exponent μ of $\beta = \log p / \log q$ via the equality $\nu = 2(\mu + 1)$.*

(S) *for the* `CF` *source.*

---

[5] To the best of our knowledge, these shapes are the only ones to occur in "classical" analyses.

[6] This means here that $M(s)$ is of polynomial growth on a family of horizontal lines $t = t_k$ with $t_k \to \infty$, and on vertical lines $\Re(s) = c - \delta'$ with some $\delta' < \delta$.

The proof of the previous result is not trivial (see a sketch of proof in the annex). The first assertion is classical, but the second and the third assertions are based on subtle properties. For aperiodic memoryless sources, the geometry of the tameness region has not been precisely described before the works of [12] and [20], which relate it to the irrationality exponent. This is why the associated remainder term in previous trie analyses was not precise, too. The third assertion is based on a seminal work of Dologopyat [10], extended by Baladi and Vallée [1] to the case of the CF dynamical system, then to the secant operator by Cesaratto and Vallée [4].

Sketch of the proof. See Section C in the annex. ◀

5 Sources $\mathcal{P}_{a,b}$ associated with a source $\mathcal{P}$ from the favorite set.

The sources to be analyzed are defined in Section 5.1 by the insertion of a delay $\gamma_{a,b}$ described in (18). Then, after computing the rescaling $g_{a,b}$ which is associated with the delay $\gamma_{a,b}$, we obtain in Section 5.2 our first result about asymptotic normality à la Shannon-MacMillan-Breiman (Theorem 16). Then Section 5.3 is devoted to the description of Λ series of the sources $\mathcal{P}_{a,b}$. Section 5.4 "inserts" the results of Section 5.3 inside Theorem 5 and provides the final result (Theorem 19). Section 5.5 provides a direct comparison between the tries of the two sources (the initial one, and the one with delays). Finally, Section 5.6 focuses on a particular class of delays, the computational delays, of main algorithmic interest.

5.1 The sources $\mathcal{P}_{a,b}$.

We start with a source $\mathcal{P}$ of our favorite set and consider the source $\mathcal{P}_{a,b}$ obtained from $\mathcal{P}$ by inserting in $\mathcal{P}$ delays $\gamma_{a,b}$ that satisfy, for any integer $\ell \geq 1$,

$$\gamma_{a,b}(\ell) = \lfloor a^\ell \rfloor \cdot \ell^b\,, \quad (a > 1, \quad b \text{ integer}, \quad b \geq 0)\,. \tag{18}$$

The particular form of these delays is interesting from four points of view

- **(a)** The initial source $\mathcal{P}$ is the source $\mathcal{P}_{1,0}$. The parameters (a, b) describe the distortion the delays $\gamma_{a,b}$ impose to the initial source $\mathcal{P}$;
- **(b)** This class contains the computational delays, defined in Section 5.6 ;
- **(c)** Proposition 8 provides an expression for the $\Lambda(s)$ series of source $\mathcal{P}_{a,b}$, from which we exhibit its main singularity (in Proposition 17) and then describe its possible tameness (in Proposition 18). This will be central for applying Theorem 5 and obtain estimates for the expected depth $\mathbb{E}[D_n^{\langle a,b\rangle}]$ of the trie built on the source $\mathcal{P}_{a,b}$;
- **(d)** Proposition 8 provides an expression for the $\Lambda_k(s)$ series of source $\mathcal{P}_{a,b}$ that will be used for applying Theorem 4 and obtain gaussian laws for $\log I_k^{\langle a,b\rangle}(x)$, *provided that* the source $\mathcal{P}_{a,b}$ be viewed as a rescaled source via an explicit rescaling $g_{a,b}$ associated with delay $\gamma_{a,b}$ via Relation (R). We study this last point in the next Section.

5.2 Rescaling $g_{a,b}$ and asymptotic gaussian laws for $\log I_k(x)$.

Relation (R) described in (11) does not lead (generally speaking) to an explicit expression of $g_{a,b}$; but, as we deal with exponential weights, we only need the principal[7] part $g_{a,b}^{(0)}$ of $g_{a,b}$, for which the following lemma provides an explicit expression.

[7] We recall that the principal part $g^{(0)}$ of g is given by the decomposition

$$g(x) = g^{(0)}(x) + A + \epsilon(x), \qquad \epsilon(x) \to 0, \quad g^{(0)}(1/x) \to 0 \quad (x \to \infty)\,. \tag{19}$$

▶ **Lemma 15.** *Consider a delay $\gamma_{a,b}$ defined in* (18)*. The principal part $g^{(0)}_{a,b}$ of the rescaling $g_{a,b}$ associated with $\gamma_{a,b}$ via Relation* (R) *satisfies*

$$g^{(0)}_{1,b}(x) = (b{+}1)^{1/(b+1)}\, x^{1/(b+1)}\,, \quad (b \geq 0)\,; \quad g^{(0)}_{a,b}(x) = \log_a \left[\frac{x}{(\log_a(x))^b}\right]\,, \quad (a > 1,\, b \geq 0)\,.$$

Proof. Given in the annex. (Section B.2). ◀

Using Proposition 8 and Definition 3 leads to the first important result of the paper:

▶ **Theorem 16.** *The source $\mathcal{P}_{a,b}$ has an exponential weight equal to $\lfloor g^{(0)}_{a,b}\rfloor$ defined in Lemma 15, with the eigenvalue function $\lambda(s)$ of the initial source $\mathcal{P}$ as a base function. Applying Theorem 4 to the source $\mathcal{P}_{a,b}$ with $g := \lfloor g^{(0)}_{a,b}\rfloor$ provides asymptotical normality à la Shannon-MacMillan-Breimann for the source $\mathcal{P}_{a,b}$.*

This result is a straightforward application of the Quasi-Power Theorem of Hwang [16]. It leads to asymptotical normality phenomena related to a quite general behaviour of expectations and variances. For more effective results, –in particular, in analyses related to tries–, we have limited ourselves to rescalings $g = g_{a,b}$ that are associated with delays $\gamma = \gamma_{a,b}$ defined in (18). However, with Proposition 7, the previous result holds for any rescaling g that fulfills Definition 6 and satisfies $g(k)/k \to 0$.

We thus provide "natural" instances of asymptotic normality phenomena, with expectations and variances of order $\Theta(g(k))$ where g is any increasing function with $g(k)/k \to 0$.

5.3 Study of the $\Lambda(s)$ series of the $\mathcal{P}_{a,b}$ source.

The delays $\gamma_{a,b}$ involve the integer part $\lfloor a^\ell \rfloor$. Using Proposition 8 and decomposition $\lfloor a^\ell \rfloor = a^\ell - \{a^\ell\}$, the Λ series of the source $\mathcal{P}_{a,b}$ decomposes as a sum,

$$\left[\sum_\ell a^\ell \ell^b \Lambda_\ell(s)\right] - \left[\sum_\ell \{a^\ell\} \ell^b \Lambda_\ell(s)\right]\,. \tag{20}$$

We then mainly study the first series, which provides the "main" behaviour of the total $\Lambda(s)$ series (its dominant singularity, its possible tameness), and finally, for non integer values of a, we have to restrict ourselves to the half plane $\Re s > 1$, due to the occurrence of the second series that is (only) analytic on $\Re s > 1$ and bounded on each half-plane $\Re s \geq 1 + \rho > 1$.

We first prove that the dominant singularity of the Λ series attached to the source $\mathcal{P}_{a,b}$ is a pole of order $(b+1)$ located at σ_a, where $\sigma_a \geq 1$ is defined by the equation $a\lambda(s) = 1$.

▶ **Proposition 17** (Dominant singularity). *Consider any source $\mathcal{P}$ from the favorite set, with a dominant eigenvalue $\lambda(s)$, and the modified source $\mathcal{P}_{a,b}$ obtained from $\mathcal{P}$ by inserting delays of the form $\gamma_{a,b}$ with $a \geq 1$. Consider the real σ_a defined by the relation $a\lambda(\sigma_a) = 1$.*
The Λ series of the source $\mathcal{P}_{a,b}$ has a pole of order $b+1$ located at σ_a, and is meromorphic on the domain $\{\Re s > \sigma_a\} \cup \mathcal{W}_a$; with $\mathcal{W}$ defined in Section 4.3, there are two cases for $\mathcal{W}_a$: $\mathcal{W}_a = \mathcal{W}$ for $a = 1$. If $a > 1$, there exists $\tau_a < \sigma_a$ for which $\mathcal{W}_a = \mathcal{W} \cap \{\Re s > \max(1, \tau_a)\}$.

Proof. Given in the annex. (Section B.3). ◀

We now describe conditions under which the source $\mathcal{P}_{a,b}$ is proven to be tame at $s = \sigma_a$.

▶ **Proposition 18** (Tameness). *Consider any source $\mathcal{P}$ from the favorite set. The source $\mathcal{P}_{a,b}$ is tame at σ_a in the following cases:*

(i) $[a = 1$, any $\mathcal{P}]$: *The source $\mathcal{P}_{a,b}$ is tame at $s = 1$, of order b, with the same shape as $\mathcal{P}$;*

(ii) $[a > 1$ and $\mathcal{P}$ is memoryless and tame$]$: *The source $\mathcal{P}_{a,b}$ is tame at $s = \sigma_a$ of order b with a tameness region $\mathcal{R}_{\sigma_a}$ of the same shape as $\mathcal{R}_1$, but limited to the half plane $\Re s > 1$;*

(iii) $[\mathcal{P} = \texttt{CF}$ and $a \neq 1$ is close enough to $1]$: *The source $\mathcal{P}_{a,b}$ is tame at $s = \sigma_a$ with a tameness strip defined from the tameness strip $\mathcal{S}$ of $\mathcal{P}$ as $\mathcal{S} \cap \{\Re s > 1\}$.*

▶ Remark. For $\mathcal{P} = \texttt{CF}$ and a not close enough to 1, we do not know the tameness of the source $\mathcal{P}_{a,b}$ at $s = \sigma_a$. The case σ_a close enough to 1 is particular from this point of view.

Proof. It is done in Section C of the annex. ◀

5.4 Average depth for tries built on the source $\mathcal{P}_{a,b}$

We now obtain our second main result. We first limit the real σ_a to be less than 2 (i.e., $a < 1/\lambda(2)$, where $\lambda(s)$ is the eigenvalue of $\mathcal{P}$), so that Condition (a) of Theorem 5 is fullfilled.

▶ **Theorem 19.** *Consider any source $\mathcal{P}$ from the favorite set (with an eigenvalue λ), a delay $\gamma_{a,b}$, the source $\mathcal{P}_{a,b}$ obtained with inserting a delay $\gamma_{a,b}$ in $\mathcal{P}$, and the blind*[8] *trie built on words of the source $\mathcal{P}_{a,b}$. Define σ_a by the relation $a\lambda(\sigma_a) = 1$ and assume the parameter a be less than $1/\lambda(2)$.*
With notations of Theorem 5, the estimate holds for the average depth of the blind trie,

$$\mathbb{E}[D_n^{\langle a,b\rangle}] = \begin{cases} n^{\sigma_a - 1} P_b(\log n) & [1 + O(n^{-\delta})]\,, & \mathcal{P} = \texttt{CF},\ a \text{ close to } 1\,; \\ n^{\sigma_a - 1} P_b(\log n) & [1 + O\left(\exp[-(\log n)^{\rho}]\right)]\,, & \mathcal{P} \in \mathcal{MA},\ 1 \le a \le 1/\lambda(2)\,; \\ n^{\sigma_a - 1} G_b(\tau \log n) & [1 + O(n^{-\delta})\,, & \mathcal{P} \in \mathcal{MP},\ 1 \le a \le 1/\lambda(2)\,. \end{cases}$$

This theorem exhibits a class of various behaviours that may occur in the average trie depth $\mathbb{E}[D_n^{\langle a,b\rangle}]$, both in the dominant term and the remainder term. In particular, a term n^d with $d \in]0,1[$ arises in the dominant term as soon as parameters d and a are related by the equality $\quad d = \lambda^{[-1]}(1/a) - 1, \quad$ where $\lambda^{[-1]}$ is the inverse of the eigenvalue mapping. For the unbiased memoryless source $(1/2, 1/2)$, the relation between a and d is $d = \log_2 a$.

5.5 A direct estimate for the average depth $\mathbb{E}[D_n^{\langle a,b\rangle}]$.

We have mentioned in Section 3.5 a direct connection between the two tries

(i) $\texttt{Trie}(\boldsymbol{x})$ built on a sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ of words emitted by the $\mathcal{P}^{\langle\gamma\rangle}$ source;

(ii) $\texttt{Trie}(\boldsymbol{y})$ built on the compressed sequence $\boldsymbol{y}$ obtained by removing all the delays in the words x_i of the sequence $\boldsymbol{x}$. $\texttt{Trie}(\boldsymbol{y})$ is thus built on the source $\mathcal{P}$.

This relation leads to the following result (proven in Section D.1 of the annex.)

▶ **Proposition 20.** *Consider any source $\mathcal{P}$ of the favorite set, and the source $\mathcal{P}_{a,b}$ built with insertion of delays $\gamma_{a,b}$. The following inequality holds between the average trie depth of the $\mathcal{P}_{a,b}$ source and the average trie depth of the initial source $\mathcal{P} = \mathcal{P}_{1,0}$:*

$$\mathbb{E}[D_n^{\langle a,b\rangle}] \ge g_{a,b}^{-1}\left(\mathbb{E}[D_n^{\langle 1,0\rangle}]\right)\,.$$

The estimate is sharp for $a = 1$. For $a > 1$, the right member involves a term $n^{(1/h)\log a}$ whereas the left member involves a term $n^{\sigma_a - 1}$ (see Theorem 19). The equality between the two exponents holds for some $a > 1$ only if the starting source $\mathcal{P}$ is memoryless unbiased.

[8] This notion is defined in Section 3.5.

This is an interesting result which applies to sources $\mathcal{P}_{a,b}$ (even if they are not proven to be tame) provided they be associated with a tame source $\mathcal{P}$; this is the case for the CF source when the parameter a is not close to 1 and varies in the whole interval $]1, 1/\lambda(2)[$. We obtain

$$\mathbb{E}[D_n^{\langle a,b\rangle}] \geq n^{(1/h)\log a}\, P_b(\log n)\,[1 + O(n^{-\delta})]\,, \qquad \text{with} \qquad 1/h = (6\log 2)/\pi^2\,.$$

5.6 Computational version of a source from the favorite set

The *exact* computation of the word $M(x)$ given in (8) is an important algorithmic issue for the initial source $\mathcal{P}$. See for instance [8] for a description in the case when when $\mathcal{P} = \mathtt{CF}$. This leads to two important questions: – the first one about the *needed precision* $\pi(\ell)$ for computing *exactly* the ℓ-th digit of the source $\mathcal{P}$ when the first $(\ell-1)$ digits have been already computed, – the second one, about the *cost* (in arithmetical complexity) for computing exactly the ℓ-th digit of the source $\mathcal{P}$ when the first $(\ell-1)$ digits have been already computed. As all the favorite sources have a positive entropy, the needed precision $\pi(\ell)$ is $\Theta(\ell)$ and the main operations that are performed are multiplications (on integers with $\Theta(\ell)$ digits). Then, the computational cost is $\Theta(\ell^2)$. This defines a (particular) delay $\widehat{\gamma}$, called the computational delay, that is equal here to $\gamma_{1,2}$ for any reference source. The source $\widehat{\mathcal{P}}$ attached to the delay $\widehat{\gamma}$ "replaces" the theoretical source $\mathcal{P}$ in any complexity issues (related in particular to tries), and is called the computational source. (See Section D.2 in the annex for precisions).

With Lemma 15, the principal part of the scaling $g_{1,2}$ is equal to $(3x)^{1/3}$. Applying the main Theorems (Theorems 16 and 19) in this particular case leads to the following result:

▶ **Theorem 21.** *Consider the computational source $\widehat{\mathcal{P}}$ associated with a reference source $\mathcal{P}$.*

(a) *Except in the case when the reference source is memoryless unbiased, the random variable $x \mapsto \log \widehat{I}_k(x)$ asymptotically follows a gaussian law*

$$\Pr\left[x \left| \frac{\log \widehat{I}_k(x) - |\lambda'(1)| \cdot \lfloor (3k)^{1/3} \rfloor}{\left[|\lambda''(1) - \lambda'(1)^2| \cdot \lfloor (3k)^{1/3} \rfloor\right]^{1/2}} \leq u \right.\right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt\,.$$

(b) *The general estimate holds for the average depth of the blind trie, with the notations of Theorem 5 and $b = 2$, for which P_2 and G_2 are of degree 3,*

$$\mathbb{E}[\widehat{D}_n] = \begin{cases} P_2(\log n) & [1 + O(n^{-\delta})]\,, & \mathcal{P} = \mathtt{CF}\,; \\ P_2(\log n) & [1 + O\left(\exp[-(\log n)^{\rho}]\right)]\,, & \mathcal{P} \in \mathcal{MA}\,; \\ G_2(\tau \log n) & [1 + O(n^{-\delta})\,, & \mathcal{P} \in \mathcal{MP}\,. \end{cases}$$

6 Conclusion and open problems

The paper introduces sources with delays that provide "natural" instances of sources with zero entropy: here, the ℓ-th delay (that is inserted between the symbols $a_{\ell-1} = a_{\ell-1}(x)$ and $a_\ell = a_\ell(x)$) only depends on the depth ℓ where it is inserted, and is uniform (i.e., the same for any input x). In this sense, this is a "toy-model", that however "shows the path" for future research. We may indeed think about more realistic models where the (non uniform) ℓ-th delay may depend on the prefix $M_\ell(x)$ of length ℓ of the word $M(x)$, that is itself defined in terms of interval $\mathcal{I}_\ell(x)$. Such delays intervene in the modelling of two classical sources in Number Theory: the Stern-Brocot source is viewed as a source with delays from the CF source, and the Sturm source as a source with delays from the Stern-Brocot source. Such generalized delays also intervene in the analysis of VLMC sources (VLMC = Variable Length Markov Chains). In this sense, using generalized (i.e., non uniform) delays may be of great interest in the modelling and analysis of a large variety of sources of zero entropy.

References

1 Viviane Baladi and Brigitte Vallée. Euclidean Algorithms are gaussian. *Journal of Number Theory, Volume 110, Issue*, pages 331–386, 2005.

2 Valérie Berthé, Eda Cesaratto, Frédéric Paccaut, Pablo Rotondo, Martín D. Safe, and Brigitte Vallée. Two Arithmetical Sources and their associated Tries. In Michael Drmota and Clemens Heuberger, editors, *31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, AofA 2020, June 15-19, 2020, Klagenfurt, Austria (Virtual Conference)*, volume 159 of *LIPIcs*, pages 4:1–4:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.AofA.2020.4.

3 Valérie Berthé, Eda Cesaratto, Pablo Rotondo, and Martin Dario Safe. Lochs-type Theorems beyond positive entropy. arXiv preprint, 2022. doi:10.48550/arXiv.2202.04008v1.

4 Eda Cesaratto and Brigitte Vallée. Gaussian Distribution of Trie depth for Strongly Tame sources. *Comb. Probab. Comput.*, 24(1):54–103, 2015. doi:10.1017/S0963548314000741.

5 Julien Clément, James Allen Fill, Thu Hien Nguyen Thi, and Brigitte Vallée. Towards a realistic analysis of the Quickselect algorithm. *Theory Comput. Syst.*, 58(4):528–578, 2016. doi:10.1007/s00224-015-9633-5.

6 Julien Clément, Philippe Flajolet, and Brigitte Vallée. Dynamical Sources in Information Theory: A General Analysis of Trie structures. *Algorithmica*, 29(1):307–369, 2001. doi:10.1007/BF02679623.

7 Julien Clément, Thu Hien Nguyen Thi, and Brigitte Vallée. Towards a realistic analysis of some popular sorting algorithms. *Comb. Probab. Comput.*, 24(1):104–144, 2015. doi:10.1017/S0963548314000649.

8 Robert M Corless. Continued fractions and chaos. *American Mathematical Monthly*, 99:203–215, 1992.

9 Karma Dajani and A. Howard Fieldsteel. Equipartition of interval partitions and an application to number theory. In *Proceedings of the American Mathematical Society*, volume 129, pages 3453–3460, 2001.

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

11 Philippe Flajolet. The Ubiquitous Digital Tree. In *STACS 2006, 23rd Annual Symposium on Theoretical Aspects of Computer Science, Marseille, France, February 23-25, 2006, Proceedings*, volume 3884 of *Lecture Notes in Computer Science*, pages 1–22. Springer, 2006. doi:10.1007/11672142_1.

12 Philippe Flajolet, Mathieu Roux, and Brigitte Vallée. Digital Trees and Memoryless sources: from Arithmetics to Analysis. *Proceedings of AofA'10, DMTCS, proc AM*, pages 231–258, 2010. URL: http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/article/view/dmAM0117/3289.

13 Philippe Flajolet and Robert Sedgewick. Digital Search Trees revisited. *SIAM J. Comput.*, 15(3):748–767, 1986. doi:10.1137/0215054.

14 Philippe Flajolet and Robert Sedgewick. Mellin Transforms and Asymptotics: Finite differences and Rice's integrals. *Theor. Comput. Sci.*, 144(1&2):101–124, 1995. doi:10.1016/0304-3975(94)00281-M.

15 Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009. URL: http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521898065.

16 Hsien-Kuei Hwang. On Convergence Rates in the central limit theorems for Combinatorial Structures. *Eur. J. Comb.*, 19(3):329–343, 1998. doi:10.1006/eujc.1997.0179.

17 Hsien-Kuei Hwang. Limit Theorems for the Number of Summands in Integer Partitions. *J. Comb. Theory, Ser. A*, 96(1):89–126, 2001. doi:10.1006/jcta.2000.3170.

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

19 Niels Erik Nörlund. *Vorlesungen über Differenzenrechnung.* Chelsea Publishing Company, New York, 1954.

20 Mathieu Roux and Brigitte Vallée. Information theory: Sources, Dirichlet series, and realistic analyses of data structures. In Petr Ambroz, Stepan Holub, and Zuzana Masáková, editors, *Proceedings 8th International Conference Words 2011, Prague, Czech Republic, 12-16th September 2011*, volume 63 of *EPTCS*, pages 199–214, 2011. doi:10.4204/EPTCS.63.26.

21 Wojtek Szpankowski. *Average case analysis of algorithms on sequences.* Interscience series in Discrete Mathematics and Optimization. Wiley, 2001.

22 Brigitte Vallée. Dynamical Sources in Information Theory: Fundamental Intervals and Word Prefixes. *Algorithmica*, 29(1):262–306, 2001. doi:10.1007/BF02679622.

23 Brigitte Vallée. The Depoissonisation Quintet: Rice-Poisson-Mellin-Newton-Laplace. In James Allen Fill and Mark Daniel Ward, editors, *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, AofA 2018, June 25-29, 2018, Uppsala, Sweden*, volume 110 of *LIPIcs*, pages 35:1–35:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.AofA.2018.35.

A Annex

The annex is organized into three Sections. The first Section (Section B) gathers proofs of Theorem 5, Lemma 15, Proposition 17. The second Section (Section C) gathers the material which is related to tameness. It is first used in the proof of Theorem 14 but also in its extension in Theorem 18. Finally the third Section (Section D) is devoted to the proofs of the last two results that are stated in Section 5: direct comparison of the tries (stated in Section 5.5) and computational version of the source (stated in Section 5.6).

B Proofs of Theorem 5, Lemma 15, Proposition 17

B.1 Proof of Theorem 5

We summarize (Part I) the general estimates that are needed. Then, we describe the dominant parts (Part II) and finally the remainder terms (Part III)

Part I. Needed estimates. Along the proof, we compare the two functions, the Rice kernel $L_n(s)$ and the function $n^s\Gamma(-s)$ along a vertical (or an hyperbolic) line. The comparisons are based on the estimates provided in the annex of [4], and summarized as follows. Note that the exponential decreasing of the Γ function along vertical (or hyperbolic) lines plays a central role.

▶ **Lemma 22.** Consider a curve γ, that is a vertical line $\gamma = \{\Re s = c\}$ or an hyperbolic curve, and, for some $\tau > 0$, the infinite set of points $c_k = c + 2i\pi k\tau$ for $k \in \mathbb{Z}$. The curve γ is decomposed into two parts : the central part $\gamma_n^- := \{s \in \gamma \mid |\Im s| \le \sqrt{n}\}$ and the exterior part $\gamma_n^+ := \{s \in \gamma \mid |\Im s| > \sqrt{n}\}$. In the same way, the set $\mathbb{Z}$ is decomposed into two parts, the central part $\mathbb{Z}_n^- := \{k \mid 2\pi|k|\tau \le \sqrt{n}\}$ and the exterior part $\mathbb{Z}_n^+ := \{k \mid 2\pi|k|\tau > \sqrt{n}\}$. The following holds:

(i) The series $\sum_{k\in\mathbb{Z}} c_k\Gamma(-c_k)$ and the integral $\int_\gamma s\Gamma(-s)$ are absolutely convergent.

(ii) The following integrals or sums are $O(1/n)$ for $n \to \infty$,

$$\int_{\gamma_n{}^+} n^s\Gamma(-s)ds\,, \quad \int_{\gamma_n{}^+} L_n(s)ds\,, \quad \sum_{k\in\mathbb{Z}_n^+} n^{c_k} c_k\Gamma(-c_k)\,, \quad \sum_{k\in\mathbb{Z}_n^+} L_n(c_k)\,,$$

(iii) The following estimates holds between the residues $A_n^{(1)}(c) = \mathrm{Res}[sL_n(s)) \cdot \Lambda(s); s = c]$ and $A_n^{(0)}(c) = \mathrm{Res}[n^s s\Gamma(-s)\Lambda(s); s = c]$,

$$A_n^{(1)}(c) = A_n^{(0)}(c)\left[1 + O\left(\frac{1}{n}\right)\right]. \tag{21}$$

(iv) The following estimates hold between the series of residues,

$$\Sigma_n^{(1)} := \sum_{k\in\mathbb{Z}} A_n^{(1)}(c_k), \qquad \Sigma_n^{(0)} := \sum_{k\in\mathbb{Z}} A_n^{(0)}(c_k) \qquad \Sigma_n^{(1)} = \Sigma_n^{(0)} \cdot \left[1 + O\left(\frac{1}{n}\right)\right]. \tag{22}$$

(v) The two integrals satisfy

$$\int_\gamma L_n(s)ds = \left[\int_\gamma n^s\Gamma(-s)ds\right]\left[1 + O\left(\frac{1}{n}\right)\right].$$

Part II. Study of the dominant parts. The dominant terms are brought by residues. For a meromorphic function F with a pole of order $b+1$ at $s = c$, the term $\mathrm{Res}[n^s sF(s); s = c]$ is computed from the singular expansion of $n^s \cdot F(s)$ at $s = c$, using the analytic expansion of n^s at $s = c$,

$$n^s = e^{s\log n} = n^c \sum_{i\geq 0} \frac{1}{i!}(s-c)^i \log^i n.$$

Then, the residue $\mathrm{Res}[n^s sF(s); s = c]$ is written as $n^c \cdot Q(\log n)$ with a polynomial Q of degree b. There are four cases, according to the shape: in cases (S) and (H), there is only one pole at c, whereas in case (P), all the points $c_k = c + 2i\pi k\tau$ are also poles. Moreover, the boolean $[[c = 1]]$ intervenes due to a supplementary pole brought by $\Gamma(-s)$ at $s = 1$.

Case (S) or (H). The two residues $A_n^{(1)}(c)$ and $A_n^{(0)}(c)$ are compared with (21) and:

(i) $c \neq 1$. As c is a pole of order $b+1$ of $\Lambda(s)$ at $s = c$, the function $sL_n(s)) \cdot \Lambda(s)$ has a pole c of order $b+1$, and $A_n^{(0)}(c) = n^c c\,\Gamma(-c)\,Q(\log n)$, with a polynomial Q of degree b.

(ii) $c = 1$. As $\Gamma(-s)$ has a pole at $s = 1$, the function $\Gamma(-s)\cdot\Lambda(s)$ has now a pole at $s = 1$ of order $b+2$, and $A_n^{(0)}(c) = n^c c\,Q(\log n)$, with a polynomial Q of degree $b+1$.

Case (P). The function Λ admits poles $c_k = c + 2i\pi k\tau$ (for $k \in \mathbb{Z}$) of order $b+1$. Then, there arise the two series $\Sigma_n^{(1)}$ and $\Sigma_n^{(0)}$ that are compared in (22).

(i) Case $c \neq 1$. In the series $\Sigma_n^{(0)}$, each pole c_k brings a term $A_n^{(0)}(c_k)$, where the polynomial Q is the same for each c_k (because $\Lambda(s)$ is periodic), and

$$\Sigma_n^{(0)} = \sum_{k\in\mathbb{Z}} n^{c_k} c_k\,\Gamma(-c_k)\,Q(\log n) = n^c Q(\log n)\Pi(n), \quad \Pi(n) := \left[\sum_{k\in\mathbb{Z}} n^{2i\pi k\tau} c_k\,\Gamma(-c_k)\right].$$

The function Π is absolutely convergent, due to exponential decreasing of the Γ function. Moreover, with the equality $n^{2i\pi k\tau} = e^{2i\pi k\tau\log n}$, the function Π is a periodic function (of period 1) of the variable $\tau\log n$, and finally the series of residues is

$$\Sigma_n^{(0)} = n^c\,Q(\log n)\,\Pi(\{\tau\log n\}), \qquad \Pi \text{ periodic of period } 1, \ \ Q \text{ polynomial of degree } b.$$

(ii) The pole $c = 1$ is now a pole of order $b+2$, whereas the other poles c_k (for $k \neq 0$) are of order $b+1$. We consider the pole $c = 1$ separately, together the expansion it brings, then the other poles for $k \neq 0$, and we obtain

$$\mathrm{Res}[n^s s\Gamma(-s)\Lambda(s); s = c] + \Sigma_n^{(0)} = n\left[R(\log n) + Q(\log n)\,\Pi(\{\tau\log n\})\right],$$

where R and Q are polynomials of resp. degree $b+1$ and b.

Finally, we check, in each of the four cases, the form of the dominant terms that is stated in Theorem 5.

Part III. Study of the remainder terms. The remainder terms are easily obtained in the cases (S) or (P) where the curve δ_c is a vertical line. In case (H), the remainder term is not precisely studied in the litterature except in the case of a memoryless aperiodic source, where the study is done in [12] and [4]. We recall the study here. On the hyperbolic curve δ, and letting $s = \sigma + it$ and $L := \log n$, the following estimates hold

$$|n^s| = n^\sigma = n^c \exp[-ALt^{-\nu}], \qquad |\Lambda(s)\Gamma(-s)| \le \exp[-Kt]\,,$$

and entails the bound $\quad |n^s\Lambda(s)\Gamma(-s)| \le n^c \exp[-Kt - ALt^{-\nu}]\,.$

When n (and then L) is fixed, the minimum of the function $t \mapsto Kt + ALt^{-\nu}$ is reached for $t^{\nu+1} = \nu L/K$ and the maximum of the function is of order $n^c \exp[-(\log n)^\beta]$ with $\beta < 1/(1+\nu)$. Using the same principles as in Laplace's method, this entails the estimate

$$\int_\delta L_n(s)\Lambda(s) = n^c O\left(\exp[-(\log n)^\beta]\right)\,.$$

B.2 Proof of Lemma 15

Consider a delay $\gamma_{a,b}$ defined in (18)*. The associated rescaling $g_{a,b}$ via Relation* (R) *and its inverse $g^{-1}_{a,b}$ satisfy the following:*

(a) *the inverse g^{-1} satisfies*

$$g^{-1}_{1,b}(x) = \frac{1}{b+1}x^{b+1}\left[1 + O\left(\frac{1}{\ell}\right)\right], \quad (a=1)\,; \quad g^{-1}_{a,b}(x) = \frac{a}{a-1}a^\ell \ell^b\left[1 + O\left(\frac{1}{\ell}\right)\right], \quad (a>1)\,.$$

(b) *the function $g_{a,b}$ and the principal part $g^{(0)}_{a,b}$ satisfy*

$$g^{(0)}_{1,b}(x) = (b+1)^{1/(b+1)}\, x^{1/(b+1)}, \quad (a=1)\,; \qquad g^{(0)}_{a,b}(x) = \log_a\left[\frac{x}{(\log_a(x))^b}\right], \quad (a>1)\,;$$

$$\textit{with} \quad g_{1,b}(x) = g^{(0)}_{1,b}(x) + L_b + O\left(\frac{1}{x}\right); \qquad g_{a,b}(x) = g^{(0)}_{a,b}(x) + K_a + O\left(\frac{\log\log x}{\log x}\right).$$

Proof. The proof has two steps.

Proof of Item (a). Expression de $\Gamma_{a,b}(\ell) = \sum_{k\le\ell}\gamma_{a,b}(k)\,.$

$$\text{When } a=1, \text{ one has} \quad \Gamma_{1,b}(\ell) := \sum_{k=1}^{\ell} k^b = \frac{1}{b+1}\ell^{b+1}\left[1 + O\left(\frac{1}{\ell}\right)\right]. \tag{23}$$

For the general case $a > 1$, we consider the sum, first without integer parts, then with integer parts. First, without integer parts,

$$\widehat{\Gamma}_{a,b}(\ell) := \sum_{k=1}^{\ell} a^k k^b = a^\ell \ell^b \cdot \Theta_{a,b}(\ell) \quad \text{with} \quad \Theta_{a,b}(\ell) = \sum_{k=0}^{\ell-1} a^{-k}\left(1 - \frac{k}{\ell}\right)^b.$$

Now, $\Theta_{a,b}(\ell)$ has a limit (when $\ell \to \infty$) equal to the geometric sum $\sum_{k\ge0} a^{-k} = a/(1-a)$. The difference indeed decomposes as

$$\sum_{k\ge0} a^{-k} - \Theta_{a,b}(\ell) = \sum_{k=0}^{\ell-1} a^{-k}\left[1 - \left(1 - \frac{k}{\ell}\right)^b\right] + \sum_{k\ge\ell} a^{-k}\,.$$

The second term tends to 0 for $\ell \to \infty$ (with exponential speed). As each term of the first term is less than $(k/\ell) \cdot b$, the first term is itself less than

$$\frac{b}{\ell}\sum_{k=0}^{\ell-1} k a^{-k} \le \frac{b}{\ell}\sum_{k=0}^{\infty} k a^{-k}, \qquad \sum_{k=0}^{\infty} k a^{-k} < \infty\,,$$

and is $O(1/\ell)$ (for $\ell \to \infty$). Finally,

$$\sum_{k\ge 0} a^{-k} - \Theta_{a,b}(\ell) = O\left(\frac{1}{\ell}\right), \qquad \widehat{\Gamma}_{a,b}(\ell) = \frac{a}{a-1} a^\ell \ell^b \left[1 + O\left(\frac{1}{\ell}\right)\right].$$

We now consider the sums with integer parts. The difference with the previous case is

$$\sum_{k=1}^{\ell} \{a^k\} k^b = O(\ell^{b+1}) = a^\ell \ell^b O\left(\frac{\ell}{a^\ell}\right) = a^\ell \ell^b O\left(\frac{1}{\ell}\right),$$

and finally for $a > 1$
$$\Gamma_{a,b}(\ell) := \sum_{k=1}^{\ell} \lfloor a^k \rfloor k^b = \frac{a}{a-1} a^\ell \ell^b \left[1 + O\left(\frac{1}{\ell}\right)\right]. \tag{24}$$

Proof of Item (b). Expression for $g^{(0)}_{a,b}$.
With Relation (R) described in (11) and the estimates (24) and (23), we have to determine the principal part $g^{(0)}_{a,b}(x)$ of the function $g_{a,b}(x)$ for which the inverse $g^{-1}_{a,b}(x)$ admits the estimate

$$g^{-1}_{a,b}(x) = \frac{a}{a-1} a^x x^b \left[1 + O_b\left(\frac{1}{x}\right)\right] \quad (a > 1), \qquad g^{-1}_{1,b}(x) = \frac{1}{b+1} x^{b+1} \left[1 + O_b\left(\frac{1}{x}\right)\right].$$

We begin with the easy case $a = 1$, and obtain, for some constant L_b,

$$g_{1,b}(x) = (b+1)^{1/(b+1)} x^{1/(b+1)} + L_b + O\left(x^{-1/(b+1)}\right).$$

When $a > 1$, we check the estimate $\quad g_{a,b}(x) = \log_a(x) - b \log_a(\log_a x)) - \log_a\left(\frac{a}{a-1}\right),$

with the following remainder terms $\quad O\left(\frac{\log\log x}{\log x}\right) \quad (b \ne 0), \quad O\left(\frac{1}{\log x}\right) \quad (b = 0)$. ◀

B.3 Proof of Proposition 17

Consider a source $\mathcal{P}$ of the Good Class, and the modified source $\mathcal{P}_{a,b}$ obtained from $\mathcal{P}$ by inserting delays of the form $\gamma_{a,b}$ with $a \ge 1$. Consider the real $\sigma_a \ge 1$ defined by the equation $a\lambda(s) = 1$ so that $\sigma_1 = 1$.
The Λ series of the source $\mathcal{P}_{a,b}$ has a pole of order $b+1$ located at σ_a, and is meromorphic on a neighborhood $\mathcal{W}_a$ of the real axis. With $\mathcal{W}$ defined in Section 4.3, there are two cases for $\mathcal{W}_a$: for $a = 1$, $\mathcal{W}_a = \mathcal{W}$. If $a > 1$, there exists $\tau_a < \sigma_a$ for which $\mathcal{W}_a = \mathcal{W} \cap \{\Re s > \max(1, \tau_a)\}$.

Proof. We will deal in the paper with the delay $\gamma_{a,b}$ already defined in (18), but we first study a delay $\pi_{a,b}$ closely related to $\gamma_{a,b}$ and defined as

$$\pi_{a,b}(\ell) = \lfloor a^\ell \rfloor \cdot \pi_b(\ell), \qquad \text{with} \quad \pi_b(\ell) := (\ell+1)(\ell+2)\ldots(\ell+b). \tag{25}$$

We will return at the end of the proof to the delay $\gamma_{a,b}$ with the relation

$$\ell^b = \sum_{c\le b} \alpha_c \pi_c(\ell) \qquad \gamma_{a,b}(\ell) = \lfloor a^\ell \rfloor \ell^b = \sum_{c \le b} \alpha_c \pi_{a,c}(\ell), \qquad \alpha_b = 1. \tag{26}$$

We first deal with the modified series (associated with π instead of γ),

$$\Lambda^{\langle a,b\rangle}(s) = \sum_\ell \pi_{a,b}(\ell)\Lambda_\ell(s) = \widehat{\Lambda}^{\langle a,b\rangle}(s) + \sum_\ell \{a^\ell\}\pi_b(\ell)\Lambda_\ell(s), \quad \widehat{\Lambda}^{\langle a,b\rangle}(s) = \sum_\ell \lfloor a^\ell \rfloor \pi_b(\ell)\Lambda_\ell(s).$$

Case $a = 1$. With the definition of delay π, the equality holds, with (17), and the real neighborhood $\mathcal{W}$ defined there,

$$b!\left[\frac{1}{1-\lambda(s)}\right]^{b+1} = \sum_{\ell\geq 0}\pi_b(\ell)\lambda(s)^\ell, \qquad \sum_{\ell}\pi_b(\ell)\Lambda_\ell(s) = b!v(s)\left[\frac{1}{1-\lambda(s)}\right]^{b+1} + d(s)\,, \quad (27)$$

where $d(s) := \sum_\ell \pi_b(\ell)c_\ell(s)$ is analytic on $\mathcal{W}$. The series $\Lambda^{\langle 1,b\rangle}(s)$ is then meromorphic on $\{\Re s > 1\}\cup\mathcal{W}$ and has a pole of order $b+1$ at $s = 1$.

Case $a > 1$. The series $\widehat{\Lambda}^{\langle a,b\rangle}(s)$ is written as

$$\sum_\ell a^\ell\pi_b(\ell)\Lambda_\ell(s) = b!v(s)\left[\frac{1}{1-a\lambda(s)}\right]^{b+1} + d_a(s), \qquad \text{with} \quad d_a(s) = \sum_\ell a^\ell\pi_b(\ell)c_\ell(s)\,.$$

It has a dominant singularity (a pole of order $b+1$) at $s = \sigma_a$ where σ_a is the real s for which $a\lambda(s) = 1$. Near the real axis, using the neighborhood $\mathcal{W}$, there exists $\tau_a < \sigma_a$ for which $d_a(s)$ is analytic and $\widehat{\Lambda}^{\langle a,b\rangle}(s)$ is meromorphic on $\mathcal{W}\cap\{\Re s > \tau_a\}$. As the second series (with fractional parts) is analytic for $\Re s > 1$, the series $\Lambda^{\langle a,b\rangle}(s)$ is meromorphic on the domain $\{\Re s > \sigma_a\}\cup\mathcal{W}_a$ with $\mathcal{W}_a := \mathcal{W}\cap\{\Re s > \max(1,\tau_a)\}$ with a unique pole of order $b+1$ at $s = \sigma_a$.

We now return to the Λ series defined with delays $\gamma_{a,b}$ with (26). ◀

C Tameness

This section is devoted to tameness studies. First, in Section C.1, we focus on the memoryless case and recall the dichotomy between periodic memoryless sources and aperiodic ones. Then, Section C.2 describes the aperiodic memoryless case in the general case $a \geq 1$. The first case $a = 1$ was studied in works [12] and [20] and used in the proof of Theorem 14. The second case $a > 1$ is new (however in the same spirit as the case $a = 1$), and used in the proof of Theorem 18. Finally, Section C.3 studies the tameness of the source $\mathcal{P} = \texttt{CF}$, first for the parameters ($a = 1$, $b = 0$) (where the Dolgopyat-Baladi-Cesaratto-Vallée result is recalled and used in the proof of Theorem 14) then for a close enough to 1 (where the previous study is extended and used in the proof of Theorem 18).

C.1 Generalities for memoryless sources

When the source $\mathcal{P}$ is memoryless, the Λ series of the source $\mathcal{P}_{a,b}$ decomposes as in (20). The second series (that does not appear for integer values of a) is analytic on $\Re s > 1$ and bounded on each vertical line $\Re s = 1+\rho$ with $\rho > 0$. We focus on the first series (without integer parts), here in the memoryless case. We use the coefficients $\pi_{a,b}(\ell)$ defined in (25), the polynomial decomposition (26) and the identity (27). This gives in the memoryless case

$$\widehat{\Lambda}^{(\mathcal{P}_{a,b})}(s) = \sum_\ell a^\ell\ell^b\Lambda_\ell(s) = \sum_\ell a^\ell\sum_{c\leq b}\alpha_c\pi_c(\ell)\lambda(s)^\ell = \sum_{c\leq b}\alpha_c\, c!\left[\frac{1}{1-a\lambda(s)}\right]^{c+1}. \qquad (28)$$

For a real $a \geq 1$, we consider the value σ_a defined by the implicit equation $a\lambda(s) = 1$.

Periodicity condition. We now recall conditions under which a memoryless source (p, q) is periodic. Assume that there is $\tau > 0$ for which the two equations hold

$$[p + q = 1, \quad p^{1+2i\pi\tau} + q^{1+2i\pi\tau} = 1] .$$

Then the triangular inequality holds

$$1 = \left|p^{1+2i\pi\tau} + q^{1+2i\pi\tau}\right| \le \left|p^{1+2i\pi\tau}\right| + \left|q^{1+2i\pi\tau}\right| = p + q = 1 ,$$

and becomes an equality. This may occur only if there exists a real θ for which the equality holds, $q^{1+2i\pi\tau} = \theta \cdot p^{1+2i\pi\tau}$. This leads to the sequence of equalities

$$1 = p^{1+2i\pi\tau} + q^{1+2i\pi\tau} = (1 + \theta) \cdot p^{1+2i\pi\tau} = (1 + \theta) \cdot p = 1 ,$$

which entails $\quad p^{2i\pi\tau} = q^{2i\pi\tau} = 1 \quad$ and then $\quad \left[(2i\pi\tau) \log p = 2i\pi k, \quad (2i\pi\tau) \log q = 2i\pi\ell\right]$.
Finally, $(\log q/ \log p) = (k/\ell)$ for some pair of integers (k, ℓ) for which $\gcd(k, \ell) = 1$. Now, the equality satisfied by $e^{1/\tau}$, namely $1 = p + q = (e^{1/\tau})^k + (e^{1/\tau})^\ell$ proves that $e^{1/\tau}$ is an algebraic integer.

Periodic memoryless sources. If the source $\mathcal{P}$ is periodic, then the function $s \mapsto \lambda(s)$ is periodic of period $2i\pi\tau$ with some $\tau > 0$. This entails that the function $s \mapsto a\lambda(s)$ is periodic too with the same period, and the poles of $\widehat{\Lambda}^{\mathcal{P}_{a,b}}$ are located at points $\sigma_a + 2i\pi k\tau$. The second series (with fractional parts) has no poles on the hyperplane $\Re s > 1$. Then, the poles of $\Lambda^{\mathcal{P}_{a,b}}$ are located at points $\sigma_a + 2i\pi k\tau$.

C.2 Aperiodic memoryless sources

The pair (p, q) is here denoted as (p_1, p_2). We consider together the cases $a = 1$ and $a > 1$. In the case $a = 1$, the proof of [12] studies the set $\mathcal{Z} := \{s \mid \lambda(s) = 1, \quad s \neq 1\}$, and exhibits its hyperbolic shape defined via the irrationality exponent μ of $\beta = (\log p_2)/(\log p_1)$. We consider here, for any $a \ge 1$ the set

$$\mathcal{Z}_a := \{s \mid \lambda(s) = 1/a, \quad s \neq \sigma_a\} , \qquad \lambda(\sigma_a) = p_1^{\sigma_a} + p_2^{\sigma_a} = 1/a , \tag{29}$$

which gathers the poles of $\widehat{\Lambda}^{(\mathcal{P}_{a,b})}$ defined in (28). As previously, these poles will be the only poles of $\Lambda^{\mathcal{P}_{a,b}}$ on $\Re s > 1$. We wish to determine a region on the left of the vertical line $\Re s = \sigma_a$ but close to the vertical line $\Re s = \sigma_a$ where $\Lambda^{\mathcal{P}_{a,b}}$ is analytic: we look for a (non empty) region $\mathcal{R}_{\sigma_a}$ located "between" $\mathcal{Z}_a$ and the vertical line $\Re s = \sigma_a$, and defined by an equation of hyperbolic type as in Definition 12. We follow the general scheme of [12] given for $a = 1$ that we easily extend to a general $a \ge 1$.

An implicit equation. We let $\log p_1 := w_1$, $\log p_2 = w_2$, and consider rational approximations of the ratio w_2/w_1 (that do not depend –of course– on the real a), and thus integers q for which there exists an integer q_2 such that

$$v = q\frac{w_2}{w_1} - q_2 \quad \text{is small} \qquad \text{and thus} \quad w_2 \left(2i\pi \frac{q}{w_1}\right) = 2i\pi v + 2i\pi q_2 \qquad (v \quad \text{small}) . \tag{30}$$

We focus on the points of $\mathcal{Z}_a$ close enough to the vertical line $\Re s = \sigma_a$ with an imaginary part close to a multiple of $(2\pi)/w_1$ of the form $q(2\pi)/w_1$. We then let :

$$s \in \mathcal{Z}_a, \qquad s = \sigma_a + \Delta_a + 2i\pi \frac{q}{w_1} ,$$

and focus on complex numbers s for which the complex number Δ_a has both a small real part and a small imaginary part. With (30), the two equalities hold,

$$\begin{cases} p_1^s &= \exp\left[w_1(\sigma_a + \Delta_a + 2i\pi\frac{q}{w_1})\right] &= p_1^{\sigma_a+\Delta_a} \\ p_2^s &= \exp\left[w_2(\sigma_a + \Delta_a + 2i\pi\frac{q}{w_1})\right] &= p_2^{\sigma_a+\Delta_a}\cdot \exp[2i\pi v] \end{cases} \tag{31}$$

and entail the equality $$p_1^{\sigma_a+\Delta_a} + p_2^{\sigma_a+\Delta_a}\cdot \exp[2i\pi v] = \frac{1}{a}\,. \tag{32}$$

Then $\Delta_a = \Delta_a(v)$ is implicitely determined (via Eqn (32)) as a function of v, and Eqn (29) entails the equality $\Delta_a(0) = 0$. With the complex version of the implicit theorem, the first two derivatives of Δ_a at $v = 0$ involve two real numbers $\delta_a^{(1)}$ and $\delta_a^{(2)} > 0$ under the form,

$$\Delta_a'(0) = i\delta_a^{(1)}, \qquad \Delta_a''(0) = -\delta_a^{(2)}\,,$$

and there exists, for any $q \in \mathbb{Z}$, for each (small) real number v, a point $s = \sigma + it$ of $\mathcal{Z}_a$, with

$$\sigma = \sigma_a - \delta_a^{(2)}v^2 + O(v^4), \qquad t = 2i\pi\frac{q}{w_1} + i\delta_a^{(1)}v + O(v^3)\,. \tag{33}$$

Diophantine approximations. We consider now the reals v which are associated with a pair (q, q_2) as in (30), and focus on rationals q_2/q which are a best diophantine approximation of $\beta = w_2/w_1$: this means that, for every rational number q_2'/q' different from q_2/q with $0 < q' \le q$, one has

$$|q\beta - q_2| < |q'\beta - q_2'|\ .$$

We recall that ν is an irrationality exponent for β if, for any $\epsilon > 0$, there is a finite number of rationals q_2/q for which

$$|q\beta - q_2| \le q^{1-\nu-\epsilon}\,.$$

Continued fractions theory proves the existence of an infinite number of rationals q_2/q for which $|q\beta - q_2| \le q^{-1}$. Then an irrationality exponent is at least equal to 2.
The irrationality exponent of β is the smallest possible irrationality exponent of β; it is denoted as $\mu(\beta)$. The inequality $\mu(\beta) \ge 2$ always holds.

Two curves. As in [12], we consider a ratio $\beta = w_2/w_1$ with an irrationality exponent $\mu(\beta)$, and we deal with v as in (30), now associated with a rational q_2/q which is a best diophantine approximation of $\beta = w_2/w_1$. Then the point $s = s(q) \in \mathcal{Z}_a$ in (33) is "close" to a curve

$$s = \sigma_a - C^{(a)}t^{-2\mu-2}\,, \qquad \text{for some constant } C_a\,.$$

More precisely, as in [12], the following holds, with the irrationality exponent $\mu = \mu(\beta)$:

(i) *For any $\nu > \mu$, there exists $B_\nu^{(a)} > 0$, for which all elements $s = \sigma + it$ of the set $\mathcal{Z}_a \cap \{\Im s \ge 1\}$ satisfy $\sigma \le \sigma_a - B_\nu^{(a)}t^{-2\nu-2}$.*

(ii) *For any $\theta < \mu$, there exist $A_\theta^{(a)} > 0$, and an infinite set of elements $s = \sigma + it$ of $\mathcal{Z}_a$ such that $\sigma \ge \sigma_a - A_\theta^{(a)}t^{-2\theta-2}$.*

Part (i) of the result provides the free of poles region $\mathcal{R}_{\sigma_a}$ we look for, with its frontier. Part (ii) says that this region is in a sense optimal.

Tameness. Consider $\nu' = \nu(1+\epsilon)$ with $\nu > \mu$ and the curve $\mathcal{C}$ defined by the equation $s = \sigma_a - B_\nu^{(a)} t^{-2\nu'-2}$. Any point $s = \sigma + it$ of this curve is at a distance δ of $\mathcal{Z}_a$, with

$$\delta > B_\nu^{(a)} \left[\frac{1}{t^{2\nu+2}} - \frac{1}{t^{2\nu'+2}} \right] = B_\nu^{(a)} \frac{t^{2(\nu'-\nu)} - 1}{t^{2\nu'+2}} \geq \frac{1}{t^{2\nu'+2}}$$

for $|t|$ large enough. Moreover, for any $s = \sigma + it$ that is at a distance δ from a point of $\mathcal{Z}_a$, the following inequality holds and proves that Λ is of polynomial growth on the curve $\mathcal{C}$.

$$|\Lambda(s)| \leq \frac{1}{|\lambda'(\sigma_a)|} \frac{1}{\delta} = \Theta(t^{2\nu'+2})$$

C.3 Tameness in the CF case

We begin with the decomposition (20). We first focus on the first series and use the Dolgopyat-Baladi-Cesaratto-Vallée results:

For $\mathcal{P} = \mathtt{CF}$, there is a "truncated" vertical strip

$$\mathcal{S} := \{s = \sigma + it \mid |\sigma - 1| \leq \theta, \quad |t| \geq t_0, \quad t_0 > 0\},$$

where the $\Lambda_\ell(s)$ series satisfy for any $\ell \geq 1$, the bound

$$|\Lambda_\ell(s)| \leq \rho^\ell \cdot |t|^\xi \text{ for some } \rho < 1 \text{ and some } \xi > 0.$$

We first give sufficient conditions on a under which σ_a belongs to the strip S. As the relation holds,

$$(a-1)/a = |\lambda(1) - \lambda(\sigma_a)| \sim |\sigma_a - 1||\lambda'(1)|\,,$$

a sufficient condition for a under which σ_a belong to S is $a < 1/(1 - \theta|\lambda'(1)|)$.
When s belongs to the vertical strip S, and as soon as $a\rho \leq \delta < 1$, one has

$$|\widehat{\Lambda}(s)| \leq \sum_\ell a^\ell \ell^b |\Lambda_\ell(s)| \leq \left(\sum_\ell \ell^b a^\ell \rho^\ell \right) \cdot |t|^\xi \leq \left(\sum_\ell \ell^b \delta^\ell \right) \cdot |t|^\xi,$$

and using a similar decomposition as in (28) one obtains

$$\widehat{\Lambda}(s) \leq A_\delta \cdot |t|^\xi \qquad \text{with} \quad A_\delta := \left(\sum_\ell \ell^b \delta^\ell \right) = \sum_{c \leq b} |\alpha_c| \cdot c! \left(\frac{1}{1-\delta} \right)^{c+1}.$$

Finally, when a satisfies the two following conditions, for some $\delta < 1$, that involve the geometric characteristics on the "free of poles" strip of the source $\mathcal{P}$,

$$a \leq \min\left(\frac{\delta}{\rho}, \frac{1}{1 - \theta|\lambda'(1)|} \right),$$

the real σ_a belongs to S, the series $\widehat{\Lambda}(s)$ is tame at σ_a, with a tameness strip

$$\{s = \sigma + it \mid 1 - \theta < \sigma < 1 + \theta\}.$$

However, as $a \neq 1$, the second series (with fractional parts) appears: it is analytic on $\Re s > 1$ and bounded on each vertical line $\Re s = 1 + \kappa$ with $\kappa > 0$. Finally, a tameness strip for the total series $\Lambda(s)$ is $\quad \{s \mid 1 + \kappa \leq \sigma < 1 + \theta\}, \qquad$ with $\quad 1 + \kappa < \sigma_a < 1 + \theta$.
We have to compare the two terms, the principal term n^{σ_a} and the remainder term $n^{1+\kappa} = n^{\sigma_a} \cdot n^{1-\sigma_a+\kappa}$ where κ is any positive number such that $1 + \kappa \in]1, \sigma_a[$. With barycentric coordinates, we have $1 + \kappa = \mu + (1-\mu)\sigma_a$ with $\mu \in]0, 1[$ so that

$$1 - \sigma_a + \kappa = \mu + (1+\mu)\sigma_a - \sigma_a = \mu(1 - \sigma_a) = -\mu(\sigma_a - 1)\,.$$

Then, the multiplicative remainder term is $n^{1-\sigma_a+\kappa} = n^{-\mu(\sigma_a-1)}$, for any $\mu \in]0, 1[$.

D Comparisons between tries. Computational version of a source

D.1 Direct comparison between tries of the two sources

We first prove an inequality related to a conjecture which appears at the end of [3]. This conjecture states the equality in (34). We will exhibit in the following cases when the equality does not hold.

▶ **Proposition 23.** *Consider a delay γ, the associated rescaling g via Relation (R), (see Eqn* (11)*) and its inverse g^{-1}. If the function g^{-1} is convex, the following inequality holds between the average trie depth $\mathbb{E}[D_n^{\langle\gamma\rangle}]$ of a trie built on the source $\mathcal{P}_{\langle\gamma\rangle}$ and the average trie depth $\mathbb{E}[D_n]$ of a trie built on the source $\mathcal{P}$,*

$$\mathbb{E}[D_n^{\langle\gamma\rangle}] \geq g^{-1}\left(\mathbb{E}[D_n]\right) . \tag{34}$$

Proof. For a sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{I}^n$, we consider the two sequences of expansions

$$M(\boldsymbol{x}) = (M(x_1), M(x_2) \ldots M(x_n)), \quad M^{\langle\gamma\rangle}(\boldsymbol{x}) = (M^{\langle\gamma\rangle}(x_1), M^{\langle\gamma\rangle}(x_2) \ldots M^{\langle\gamma\rangle}(x_n)) .$$

We wish to compare the two tries $\texttt{Trie}\,(M(\boldsymbol{x}))$ and $\texttt{Trie}\,(M^{\langle\gamma\rangle}(\boldsymbol{x}))$, and notably their depths

$$D_n(\boldsymbol{x}) = D_n\left[\texttt{Trie}(M(\boldsymbol{x}))\right], \qquad D_n^{\langle\gamma\rangle}(\boldsymbol{x}) = D_n\left[\texttt{Trie}\,(M^{\langle\gamma\rangle}(\boldsymbol{x}))\right] .$$

Due to the definition of the source $\mathcal{P}^{\langle\gamma\rangle}$ in terms of the source $\mathcal{P}$, and definition of $\texttt{Trie}\,(M^{\langle\gamma\rangle}(\boldsymbol{x}))$, the branches of the two tries are related, as we now see: we denote by $M_k(x)$ the prefix of the expansion $M(x)$ of length k, and by $M_k^{\langle\gamma\rangle}(x)$ the prefix of the expansion $M^{\langle\gamma\rangle}(x)$ of length k. Remark first that a comparison between two words of $\mathcal{P}^{\langle\gamma\rangle}$ always terminates on a symbol in Σ. Remark also that the prefixes of $M^{\langle\gamma\rangle}(x)$ that end with a symbol in Σ are exactly those of length equal to $g^{-1}(k)$ for some k. Then, for $i \in [1..n]$, the i-th branch of the second trie $\texttt{Trie}\,(M^{\langle\gamma\rangle}(\boldsymbol{x}))$ is thus written for some $k_i = k_i(\boldsymbol{x})$ as $M^{\langle\gamma\rangle}_{g^{-1}(k_i)}(x_i)$ (and has a length $g^{-1}(k_i(\boldsymbol{x}))$) whereas the corresponding i-th branch of the first trie $\texttt{Trie}\,(M(\boldsymbol{x}))$ is $M_{k_i}(x_i)$ (and has a length $k_i(\boldsymbol{x})$). Then the depths of the two tries are respectively

$$D_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} k_i(\boldsymbol{x}), \qquad D_n^{\langle\gamma\rangle}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} g^{-1}(k_i(\boldsymbol{x})) .$$

The depth D_n of a trie for which the lengths of branches are $D_n^{(i)}$ is defined via its law

$$\Pr[D_n \geq k] = \frac{1}{n}\sum_{i=1}^{n} \Pr[D_n^{(i)} \geq k], \qquad \text{so that} \quad \mathbb{E}[D_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[D_n^{(i)}] .$$

The expectations of the two depths are thus

$$\mathbb{E}[D_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[k_i(\boldsymbol{x})] \qquad \mathbb{E}[D_n^{\langle\gamma\rangle}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[g^{-1}(k_i(\boldsymbol{x})] .$$

When g^{-1} is a convex function, the following inequalities hold and entail the stated bound,

$$g^{-1}\left(\mathbb{E}[D_n]\right) = g^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[k_i(\boldsymbol{x})]\right) \leq \frac{1}{n}\left(\sum_{i=1}^{n} g^{-1}\left(\mathbb{E}[k_i(\boldsymbol{x})]\right)\right)$$

$$g^{-1}\left(\mathbb{E}[k_i(x)]\right) \leq \mathbb{E}[g^{-1}(k_i(\boldsymbol{x}))] .$$

◀

The present paper exhibits a class of delays of the form $\gamma_{a,b}(\ell) = a^\ell \ell^b$ for which the inverse of the weight is

$$g^{-1}(x) = \begin{cases} g_{1,b}^{-1}(x) & = \Theta\left(x^{b+1}\right) \quad (a = 1)\,, \\ g_{a,b}^{-1}(x) & = \Theta\left(a^x x^b\right) \quad (a > 1)\,, \end{cases} \tag{35}$$

and is convex for $x \geq 1$.
As our starting sources are tame sources $\mathcal{P}$ for which Theorem 5 applies with $(c = 1,\, b = 0)$, the depth $\mathbb{E}[D_n]$ satisfies, with $h = |\lambda'(1)|$,

$$\mathbb{E}[D_n] = A_n\left[1 + O(\epsilon_n)\right] \qquad A_n := \frac{1}{h}\log n\,, \quad \epsilon_n = \frac{1}{\log n}\,;$$

For $n \to \infty$ and, for any $g := g_{a,b}$, the estimates $\mathbb{E}[D_n] \sim A_n$ lead to the estimates $g^{-1}(\mathbb{E}[D_n]) = \Theta\left(g^{-1}(A_n)\right)$. Then, using the equality $a^{(1/h)\log n} = n^{(1/h)\log a}$), one obtains

$$g^{-1}(\mathbb{E}[D_n]) = \Theta\left(g^{-1}(A_n)\right) = \begin{cases} \Theta\left((\log n)^{b+1}\right) & (a = 1)\,, \\ \Theta\left(n^{(1/h)\log a}(\log n)^b\right) & (a > 1)\,. \end{cases} \tag{36}$$

Theorem 19 of the present paper exhibits the following estimates for $\mathbb{E}[D_n^{\langle\gamma\rangle}]$ that involve the real σ_a given by the equation $a\lambda(\sigma_a) = 1$,

$$\mathbb{E}[D_n^{\langle\gamma\rangle}] = \begin{cases} \Theta\left((\log n)^{b+1}\right) & (a = 1)\,, \\ \Theta\left(n^{\sigma_a - 1}\right)(\log n)^b\big) & (a > 1)\,. \end{cases} \tag{37}$$

Then, in the case $a > 1$, there are two exponents (a priori) distinct:
(i) an exponent equal to $(1/h)\log a$ in the estimate of $g^{-1}(\mathbb{E}[D_n])$
(ii) an exponent $\sigma_a - 1$ in the estimate of $E[D_n^{\langle\gamma\rangle}]$.

With Proposition 23, we prove (in an indirect way) the inequality $\sigma_a - 1 \geq \frac{1}{h}\log a$. For an unbiased memoryless source of cardinality r with entropy $\log r$, the equality holds, $[\sigma_a - 1 = \log a/\log r]$, and the two exponents are the same. We now directly compare the two exponents :

▶ **Lemma 24.** *The following inequality holds between the two exponents,*

$$\sigma_a - 1 \geq \frac{1}{h}\log a\,,$$

and the equality holds only in the case of a source conjugated to an unbiased memoryless source.

Proof. Inside the proof, we use the notation $\sigma_a = \sigma(a)$. The function $x \mapsto \sigma(x)$ is defined through the implicit equation $x\lambda(\sigma(x)) = 1$ that involves the dominant eigenvalue $s \mapsto \lambda(s)$. The derivative $\sigma'(x)$ satisfies $\lambda(\sigma(x)) + x\sigma'(x)\lambda'(\sigma(x)) = 0$ and thus

$$\sigma'(x) = \frac{1}{x}\left[-\frac{\lambda(\sigma(x))}{\lambda'(\sigma(x))}\right] = \frac{1}{x}\left[\frac{-1}{L'(\sigma(x))}\right]\,,$$

where $L(s)$ is the pressure function equal to $L(s) = \log\lambda(s)$. The function L is always convex, so that $u \mapsto L'(u)$ is increasing. Then, the function $u \mapsto -1/L'(u)$ is thus increasing, too, and as the function $\sigma(x)$ is increasing, this entails that the function $\tau : x \mapsto \frac{-1}{L'(\sigma(x))}$ is also increasing, and satisfies, for $u \geq 1$ the inequality $\tau(u) \geq \tau(1)$ with $\tau(1) = \frac{-1}{L'(1)} = \frac{1}{h}$. Thus the inequality holds,

$$\sigma(a) - 1 = \int_1^a \frac{1}{u}\tau(u)du \geq \frac{1}{h}\int_1^a \frac{1}{u}du = \frac{1}{h}\log a\,.$$

The inequalities become equalities only in the case when L' is linear, which occurs only when the source is conjugated to an unbiased memoryless source. ◀

D.2 Computational version of a good source

The computation of the successive symbols of the word $M(x)$ indeed involves the computation of the intervals $(\mathcal{I}_1(x), \mathcal{I}_2(x) \ldots, \mathcal{I}_\ell(x), \ldots)$. The computation of the ℓ-th digit of the word $M(x)$, when the first $\ell - 1$ digits (which form the prefix w) have been already computed, continues inside the interval $\mathcal{I}_{\ell-1}(x) = \mathcal{I}_w$. We already know the prefix w of length $\ell - 1$, and, as this interval $\mathcal{I}_w$ is the union of intervals $\mathcal{I}_{w\cdot\sigma}$ for $\sigma \in \Sigma$, we have to compare x and the end points of the possible intervals $\mathcal{I}_{w\cdot\sigma}$, in order to output the new digit σ for which x belongs to $\mathcal{I}_{w\cdot\sigma}$.

We have then to deal with the convenient precision $\pi(\ell)$ (on the real x and on the end points $a_\ell(x)$ and $b_\ell(x)$ of $\mathcal{I}_\ell(x)$) that is needed for an exact comparison between the end points and the input x. We deal in the classical model where there is an oracle which *freely* gives, for each real y and a given precision π, a rational approximation of y whose numerator p and denominator q are integers, with a number π of binary digits, for which $|y - p/q| \le 2^{-\pi}$. We freely ask the oracle, but, when the rational approximations are given by the oracle, we have to compute with the integers p, q we are given; now, the computation is not free, and depends on the binary size of the rationals, and thus on π.

The main questions are: first, how is the integer $\pi(\ell)$ compared to the depth ℓ? second, what are the main operations on the integers (numerators and denominators) to be done? For our reference sources, of positive entropy, the needed precision is $\pi(\ell) = \Theta(\ell)$ and the main operations used in the computation of the shift T are additions and multiplications. Then, for our reference sources, the arithmetic complexity for computing a_ℓ when $(a_1, a_2, \ldots, a_{\ell-1})$ is already computed is of the form $\Theta(\pi(\ell)^b) = \Theta(\ell^b)$ with $b = 2$.