HAL open science

Digital Trees and Memoryless Sources: from Arithmetics to Analysis

Philippe Flajolet, Mathieu Roux, Brigitte Vallée

To cite this version:

Philippe Flajolet, Mathieu Roux, Brigitte Vallée. Digital Trees and Memoryless Sources: from Arithmetics to Analysis. 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10) , Jun 2010, Vienne, Austria. pp.231-258. hal-01083405v1

HAL Id: hal-01083405

https://hal.science/hal-01083405v1

Submitted on 17 Nov 2014 (v1), last revised 20 Aug 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Digital Trees and Memoryless Sources: from Arithmetics to Analysis

Philippe Flajolet[1], Mathieu Roux[2,3], and Brigitte Vallée[2]

[1] *Algorithms Project, INRIA Rocquencourt, F-78153 Le Chesnay (France)*
[2] *GREYC, Université de Caen, 14032 Caen Cedex (France)*
[3] *LMNO, Université de Caen, 14032 Caen Cedex (France)*

Digital trees, also known as *"tries"*, are fundamental to a number of algorithmic schemes, including radix-based searching and sorting, lossless text compression, dynamic hashing algorithms, communication protocols of the tree or stack type, distributed leader election, and so on. This extended abstract develops the asymptotic form of expectations of the main parameters of interest, such as tree size and path length. The analysis is conducted under the simplest of all probabilistic models; namely, the *memoryless source*, under which letters that data items are comprised of are drawn independently from a fixed (finite) probability distribution. The precise asymptotic structure of the parameters' expectations is shown to depend on fine singular properties in the complex plane of a ubiquitous *Dirichlet series*. Consequences include the characterization of a broad range of asymptotic regimes for error terms associated with trie parameters, as well as a classification that depends on specific *arithmetic properties*, especially irrationality measures, of the sources under consideration.

Keywords: Analysis of algorithms, complex asymptotics, digital tree, Mellin transform, memoryless sources, trie

Introduction

Digital trees, also known as *"tries"*, serve to represent finite collections of words over some finite alphabet: each subtree stemming directly from the root is associated with the subcollection of words starting with a given letter; each subtree at level two corresponds to a given prefix of length two, and so on. Such trees can then be used to implement dynamic dictionaries, to represent an index in a dynamic hashing scheme, to provide a simple combinatorial description of radix-based sorting methods, and so on. As a matter of fact, the simplicity and flexibility of digital trees is at the root of many important algorithms of computer science [10, 18, 34, 35, 36]; see also the partial listing in our abstract.

The interest of digital tree structures entirely rests upon their average-case and probabilistic properties, as their worst-case behaviour is unbounded. These properties depend on the probabilistic features of the process that emits words, which is called a *source*: a source on the alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$ is completely determined by the family of probabilities $\{\pi_w\}$, where π_w is the probability that a word begins with the finite prefix $w \in \mathcal{A}^\star$. As noted early [7, 8, 18, 36], quantifying the main parameters of the digital tree is strongly dependent upon the location of poles in the complex plane of the *fundamental*

1365–8050 © 2005 Discrete Mathematics and Theoretical Computer Science (DMTCS), Nancy, France

Dirichlet series associated with the source:

$$\Lambda(s) := \sum_{w \in \mathcal{A}^\star} \pi_w^s. \tag{1}$$

This dependency is revealed by an analysis based on either Mellin transforms or the theory of Nörlund–Rice integrals [11, 12, 36]. The location of poles, especially those near the vertical line $\Re(s) = 1$, has been, for instance, recognized to dictate the asymptotic form of expected *size* (the number of internal nodes), expected path length, and the depth profile of the tree, as the number n of items stored in the tree gets large.

The simplest probabilistic source is the *memoryless source*, where letters of words are drawn *independently* according to a fixed probability distribution. The model is then determined by the basic letter probabilities, $p_j := \mathbb{P}(a_j) \equiv \pi_{a_j}$, with $p_1 + \cdots + p_r = 1$. The Dirichlet series in this case admits the simple form

$$\Lambda(s) = \frac{1}{1 - \lambda(s)}, \qquad \text{with} \quad \lambda(s) := p_1^s + \cdots + p_r^s, \tag{2}$$

which leads us to study the set of the roots of the equation $\lambda(s) = 1$. Even though the memoryless source is the simplest of all models, the rich geometry of this set turns out to be conditioned by arithmetic properties of the p_j to be examined in Section 2.

Here is a representative sample of results relative to the expected size $\overline{S}_n$, in the case of a memoryless source over a binary alphabet ($r = 2$; $p_1 + p_2 = 1$). In the *unbiased case*, $p_1 = p_2 = 1/2$, Knuth discovered in the mid 1960s that

$$\overline{S}_n = \frac{n}{\log 2} + n\, P(\log_2 n) + O\left(n^{1/2}\right); \tag{3}$$

see the account in [18, p. 131]. Here, $P(u)$ is an oscillatory function of a minute amplitude, about 10^{-5}. The surprise here is that $\overline{S}_n/n$ does *not* converge to a limit as $n \to +\infty$. In fact, such periodicity phenomena arise precisely in those cases where $(\log p_1)/(\log p_2)$ is a rational number, a property which seems to have been first enunciated by Fayolle *et al.* [8] in 1986.

By contrast, we shall see that, in *most* cases, there exists a real number θ, with $\theta > 1$, which depends on the probability vector $\mathfrak{p} = (p_1, \ldots, p_r)$, such that

$$\overline{S}_n = \frac{n}{H} + O\left(n \exp\left(-(\log n)^{1/\theta}\right)\right), \qquad H := -p_1 \log p_1 - p_2 \log p_2. \tag{4}$$

An error bound of this form is, in particular, smaller than $O(n(\log n)^{-a})$, but larger than any $O(n^{1-a})$, for any $a > 0$. The exponent θ is conditioned by diophantine approximation properties of the p_j spelled out in Section 2. For instance, in the case $p_1 = 1/3$, $p_2 = 2/3$, we can adopt $\theta = 8.25$, thanks to number-theoretic works of Rhin, Toffin, Wu, and Salikhov [30, 31, 32, 41]; in addition, a bound of the form (4) holds for any probability vector $\mathfrak{p}$ with rational components that satisfies the simple "aperiodicity" condition of Section 2. We shall also establish that, except for a set of memoryless sources of Lebesgue measure 0, an exponent $\theta > 3$, which only depends on the dimension, is sufficient. In addition, for special values of $(\log p_1)/(\log p_2)$, known as Liouville numbers, the error bounds can come arbitrarily close to $O(n)$. (Preliminary results not reported here indicate that our upper bounds are essentially tight.)

Plan of the paper. First, we recall in Section 1 a few classical facts relative to the data structures under scrutiny and their analysis. This, via Mellin transforms, leads to the Dirichlet series $\Lambda(s)$ of (2), which is the central character of our study. We then investigate the geometry of the poles of $\Lambda(s)$ in Section 2. Our treatment here develops points left pending in [8], and is inspired by the book by Lapidus and van Frankenhuijsen [20], dedicated to the *complex dimension of fractal strings*. We then develop an abstract asymptotic discussion of Mellin transforms involving $\Lambda(s)$ in Section 3 and finally reap the crop and draw conclusions regarding trie parameters in Section 4. The present study is furthermore directly relevant to the analysis of the Quicksort algorithm [9], equivalently, of path length in binary search trees, under a "symbol-complexity" model, as discussed in [39], where the rôle of irrationality measures in such contexts seems to have been first discussed.

History. The saga of the average-case analysis of digital trees starts in the mid-sixties, with works of Knuth (under some coaching by De Bruijn), who gives an excellent account in volume 3 of his *magnum opus* [18], first published in 1973: see especially the discussion of radix-exchange sort on pp. 130–134, with the "right" periodic fluctuations in the binary unbiased case ($p_1 = p_2 = 1/2$). There is also the interesting exercise Ex. 5.2.2.53 (p. 138), whose solution (pp. 637–638) describes the set of poles as *"difficult to analyze"* in general, and where a non-trivial periodic case $p_1 = 2/(1+\sqrt{5})$ is explicitly given.

The problem surfaced again independently around 1980 in the performance evaluation community, when attempting to determine the characteristics of communication protocols of the tree (or stack) type—the Capetanakis–Tsybakov–Mikhailov (CTM) Protocol. See Longo's book [21], and especially Massey's survey [21, pp. 73–137]. The presence of fluctuations in trie size has an impact on the variance, hence eventually on the stability properties of the CTM Protocol. Some authors who did not notice the fluctuations, albeit tiny, were then led to wrong conclusions. The situation was gradually corrected in the early 1980s, especially after the connection with tries and Knuthian analyses could be established: see the special issue [25] edited by Massey and the contemporary study [8].

The paper by Fayolle *et al.* [8] seems to have been the first to conduct (in the binary case) a detailed discussion of the geometry of poles and related integration contours, with the "periodicity criterion" explicitly enunciated (cf Theorem 1). As it was recognized in subsequent years, largely by Jacquet, Louchard, and Szpankowski (see, e.g., [17, 23]), digital tree analyses can serve as the basis of a remarkably precise understanding of the Lempel and Ziv schemes for data compression. In the way, some authors occasionally lost sight of the difficulties arising from the complicated geometry of poles of $\Lambda(s)$, so that a few relatively minor errors somehow crept into the literature; e.g., the error terms in the aperiodic case in "Fact 1" of [22, p. 481] and Theorem 1A of [17, p. 165]. (Our own work is not exempt from criticism: for instance, Theorem 6 of [7] tacitly requires growth assumptions on what is the equivalent of our $\Lambda(s)$ here.) These would be corrected shortly after, thanks, largely, to the rigorous and thorough discussion offered by Schachinger, for instance, in the Appendix to his paper [33]. See also typically Lemma 8.22 in Szpankowski's reference text [36]. Some of these questions have been recently revisited by Mohamed and Robert [26, 27]: in their original perspective, the periodic versus aperiodic dichotomy appears to reflect the lattice and non-lattice cases of renewal theory.

Interestingly enough, problems of a similar nature arise in a seemingly unrelated domain, that of "fractal strings". In Lapidus' terms, the general question is "Can one hear the shape of a fractal drum?". (In more dignified words: does the spectrum of the Laplacian associated to a region determine the shape of the boundary of that region, when this boundary is fractal?) In the one-dimensional case, a highly informative reference is the monograph [20] by Lapidus and van Frankenhuijsen, which did inspire several of our developments. Similar comments apply to renewal theory and dynamical systems theory, where the periodicity–aperiodicity dichotomy (Section 2) plays a rôle: we refer to the works of Pollicott [29, p. 143], as well as Baladi, Cesaratto, and Vallée [3, 4, 6, 38] for a dynamical discussion.

Acknowledgements. Thanks to Cecilia Holmgren for rekindling our interest in these questions and inciting us to make explicit the bounds (4), which were needed in her analysis of the "deconstruction" of digital trees [16, p. 32]. Thanks also to Tanguy Rivoal and Nicolas Brisebarre for communicating valuable information regarding irrationality

$$X : \left\{ \begin{array}{lcl} X_1 & = & \mathtt{babbc}\cdots \\ X_2 & = & \mathtt{abcba}\cdots \\ X_3 & = & \mathtt{abbca}\cdots \\ X_4 & = & \mathtt{ccbaa}\cdots \\ X_5 & = & \mathtt{cccaa}\cdots \end{array} \right\}; \qquad \mathtt{Trie(X)} =$$

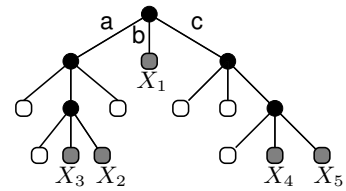


Fig. 1: A collection of five infinite words (left) over the alphabet $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ and the associated trie (right).

measures of logarithms and related questions. Thanks to Wojtek Szpankowski, Philippe Jacquet, and Philippe Robert for useful comments on these questions. We are also grateful to the referees of AOFA'10 for offering many useful comments and suggestions.

1 Digital trees, parameters, and models

What is fixed throughout is a finite *alphabet* $\mathcal{A} = \{a_1, \ldots, a_r\}$ of cardinality $r \geq 2$, the elements of which are *letters*. We consider the basic set $\mathbf{W} := \mathcal{A}^\infty$ of infinite *words* (or "sequences") over $\mathcal{A}$. The product set $\mathbf{W}^n$ then represents the collection of all n-tuples of such words.

Tries (digital trees). Given a collection $X \in \mathbf{W}^n$ of words, assumed to consist of distinct elements of $\mathbf{W}$, we can associate to it a *digital tree* or *"trie"*, denoted by $\mathtt{Trie(X)}$, that is defined as follows. First, introduce the infinite *perfect tree* T^∞ of branching degree r and consider the jth edge stemming from a node as labelled by the letter $a_j \in \mathcal{A}$; a finite or infinite branch in T^∞ is then identified by a word, either finite or infinite. A node in the infinite perfect tree T^∞ will also be referred to as a *place*. An n-collection of distinct elements then becomes a collection of n infinite branches in the tree, which themselves determine an infinite subtree of T^∞, called the "pre-trie". The digital tree, $\mathtt{Trie(X)}$, associated with $X \in \mathbf{W}^n$ is the minimal finite subtree of the pre-trie that separates from one another all the elements of X. An internal node corresponds to a place, which is "occupied" by at least two elements of X; an external node either is empty or corresponds to exactly one element of X. See Figure 1 or refer to the classical literature [18, 24, 35] for details.

Models. Our purpose is to analyse tries when words of $\mathbf{W}$ are produced by certain information *sources*. This amounts to endowing $\mathbf{W}$ with a probability distribution. A general source is completely defined by the set of probabilities π_w, for $w \in \mathcal{A}^\star$, where π_w is the probability that a random element of $\mathbf{W}$ starts with the prefix w. The probabilities π_w are called the *fundamental probabilities*.

This induces a product probability structure on $\mathbf{W}^n$, under which all elements of a sequence $X \in \mathbf{W}^n$ are distinct with probability 1 (see [28] for measure-theoretic definitions). A parameter, such as the *size* of $\mathtt{Trie(X)}$, then becomes a random variable, which is finite with probability 1. Note that the product space structure signifies that the n components of a random $X \in \mathbf{W}^n$ are chosen *independently*.

In addition to the fixed-size model just described, it is of advantage to consider the *Poisson model* of parameter x, under which the number N of elements placed in the trie is itself a random variable with the distribution $\mathbb{P}(N = n) = e^{-x}x^n/n!$. If ξ is a parameter defined over each $\mathbf{W}^n$, with expectation $\mathbb{E}_{\mathbf{W}^n}[\xi]$, then, its expectation under the Poisson model of parameter x is

$$\mathbb{E}_{\mathcal{P}(x)}[\xi] = e^{-x} \sum_{n \geq 0} \mathbb{E}_{\mathbf{W}^n}[\xi] \frac{x^n}{n!}. \tag{5}$$

Thus, for large real x, the Poisson expectation is a weighted average of the fixed-n expectations with a preponderance of those cases for which $x = n \pm O(\sqrt{n})$.

Analysis. We now restrict attention to the trie-size parameter S, and consider the expectation $\overline{S}(x) := \mathbb{E}_{\mathcal{P}(x)}[S]$. We refer to [7, 36] for details. Briefly, the size of $\texttt{Trie}(\texttt{X})$ is the number of places in the tree which are occupied by at least two different words X_j. If the prefix w identifies the place, then, under the Poisson model, that place w is occupied by at least two words with probability $1 - (1 + \pi_w\, x)e^{-\pi_w\, x}$, where π_w is the probability that a random element of $\mathbf{W}$ starts with the prefix w. Summing over all possible places, we find that the poissonized expectation of size is a sum over the set $\mathcal{A}^\star$ of all finite words w, of the form

$$\overline{S}(x) = \sum_{w \in \mathcal{A}^\star} \left[1 - (1 + \pi_w x)e^{-\pi_w x}\right], \tag{6}$$

This formula is our starting point. Note also that (5) implies an explicit form of $\overline{S}_n := \mathbb{E}_{\mathbf{W}^n}[S]$.

Asymptotics. As it is well known [11, 36], a royal road to the asymptotic analysis of $\overline{S}(z)$ as given by (6) is the *Mellin transform*. In the case at hand, it is

$$S^\star(s) = \int_0^\infty \overline{S}(x) x^{s-1}\, dt. \tag{7}$$

Then the harmonic sum property of Mellin transforms [11, 13, 36] yields:

$$S^\star(s) = -(s+1)\Gamma(s)\Lambda(-s); \qquad \Lambda(s) \equiv \sum_{w \in \mathcal{A}^\star} \pi_w^s \tag{8}$$

Analogous developments yield the expectation of *path length* in tries [7, 24, 36], defined as the sum of the distances of the root to all (non-empty) terminal vertices. In addition, a similar treatment is applicable to the *path length of binary search trees*, and to the *Quicksort algorithm*, when the total cost of comparisons between symbols is taken into account [9, 39]: in this case, we are dealing with a cost function, which relies on suitable characteristics of prefixes(i) and is consequently of the "trie type".

What we have just done amounts to lifting to the level of a general source the main algebraic arguments to be found in the classical literature [7, 9, 18, 24, 36]. These are often only developed in the case of the simplest of all source models, the *memoryless source*, which assumes that letters of words are produced independently: if p_j is the probability of letter a_j, the probability π_w is then equal to the product of the probabilities of symbols, which the prefix w is comprised of. Such a memoryless source is then completely specified by the vector $\mathfrak{p} = (p_1, \ldots, p_r)$, with $p_j \in (0, 1)$ and the condition $\sum p_j = 1$. In this case, the Dirichlet series $\Lambda(s)$ takes the form of a "quasi-inverse", as in (2). We state:

Proposition 1 *The Mellin transforms of the Poissonized expectations of trie size* (S)*, trie path length* (P)*, and symbol-comparison cost of Quicksort* (Q) *are expressible in terms of the Dirichlet series* $\Lambda(s)$*:*

$$\begin{cases} S^\star(s) &= -(s+1)\,\Gamma(s)\,\Lambda(-s) & \text{(trie size, } S\text{);} \\ P^\star(s) &= -s\,\Gamma(s)\,\Lambda(-s) & \text{(trie path length, } P\text{);} \\ Q^\star(s) &= \dfrac{2}{s(s+1)}\,\Gamma(s)\,\Lambda(-s) & \text{(Quicksort cost, } Q\text{).} \end{cases} \tag{9}$$

(i) In this analysis, each infinite word is associated with a "parameter" v and the computation involves $\mathbb{E}_{\mathcal{P}(x)}[2/(N_{[u,t]} + 2)]$, where $N_{[u,t]}$ is the number of words whose parameter v belongs to the interval $[u, t]$.

In the case of a memoryless source with probability vector $\mathfrak{p} = (p_1, \ldots, p_r)$*, the series* $\Lambda(s)$ *admits the simple form* $\Lambda(s) = 1/(1 - \lambda(s))$ *with* $\lambda(s) := p_1^s + \cdots + p_r^s$.

Note: there are similar expressions for variants of tries, such as Patricia tries and digital search trees. All involve $\Lambda(-s)$ modulated by a Gamma factor and an elementary function; see, e.g., [24, 36].

It is then well known (via Mellin inversion and the Residue Theorem) that singularities of a Mellin transform such as (9) provide, under suitable conditions, the asymptotic expansion of the original function, here, as $x \to \infty$. In this way, all parameters listed in Proposition 1 (and many more!) can be precisely analysed, once enough information has been gathered regarding the complex-analytic structure of $\Lambda(s)$. We are going to carry out the programme just outlined in what follows, starting with a detailed investigation of the structure of poles of the fundamental Dirichlet series $\Lambda(s)$ in the next section.

2 Geometry of the poles of $\Lambda(s)$

The purpose of this section is to gather information on the set $\mathcal{Z}$ of solutions to the equation

$$\lambda(s) = 1, \qquad \text{where} \quad \lambda(s) = p_1^s + p_2^s + \cdots + p_r^s,$$

with s the unknown. The family of probabilities $\mathfrak{p} = (p_1, \ldots, p_r)$ will be conventionally indexed in decreasing order, $p_1 \geq \cdots \geq p_r$. Since $\Lambda(s) = (1 - \lambda(s))^{-1}$, the set $\mathcal{Z}$ coincides with the set of poles of $\Lambda(s)$. In addition, by conjugacy, it is sufficient to study $\mathcal{Z}_+ := \mathcal{Z} \cap \{\Im(z) > 0\}$.

Definition 1 *If* all *the ratios* $(\log p_j)/(\log p_k)$ *are rational numbers, then the family of probabilities* $\mathfrak{p}$ *is said to be* periodic. *If* at least one *such ratio is irrational, the family of probabilities is said to be* aperiodic.

It is easily recognized that periodicity of $\mathfrak{p}$ is equivalent to the existence of a positive real number a such that each p_j is of the form a^{e_j}, for some $e_j \in \mathbb{Z}_{>0}$. In that case, the entire function $\lambda(s)$ admits the complex period $i\tau$, where $\tau := 2\pi/|\log a|$, so that the roots of the equation $\lambda(s) = 1$ are invariant under the translation $s \mapsto s + i\tau$. Also, clearly, $s = 1$ is the only real root of the equation $\lambda(s) = 1$. This, supplemented by the strong triangle inequality yields the following well-known statement.

Theorem 1 *The following conditions are equivalent:* (a) *the family of probabilities* $\mathfrak{p}$ *is periodic in the sense of Definition 1;* (b) *the set* $\mathcal{Z} \cap \{\Re(s) = 1\}$ *contains a point* $s \neq 1$*;* (c) *the set* $\mathcal{Z} \cap \{\Re(s) = 1\}$ *is of the form* $1 + i\tau\mathbb{Z}$*, for some* $\tau > 0$*;* (d) *the function* $\lambda(s)$ *admits the imaginary period* $i\tau$*, for some* $\tau > 0$.

Thus families such as $(1/2, 1/2), (1/2, 1/4, 1/4), (1/3, 1/3, 1/3)$, and $(\varphi^{-1}, \varphi^{-2})$, with φ the golden ratio, are periodic. A periodic family being associated with an algebraic number a (since $\sum a^{e_j} = 1$), the class of periodic families is denumerable, hence of measure 0: *periodic families are exceptional*.

The remainder of this section is devoted to the structure of the set $\mathcal{Z}$ in the *aperiodic case*. The following notations will be employed throughout:

$$w_i := |\log p_i|; \qquad \alpha_{k,j} := \frac{w_j}{w_k} \equiv \frac{\log p_j}{\log p_k}; \qquad \alpha^{(k)} := (\alpha_{k,j})_{j \in [1..r]}. \tag{10}$$

Aperiodicity in this context means that, for each k, *at least* one $\alpha_{k,j}$ is irrational.

We first offer a brief informal discussion of what goes on. With $s = \sigma + it$, the equation $\lambda(s) = 1$ becomes

$$p_1^{\sigma+it} + \cdots + p_r^{\sigma+it} = 1, \qquad p_1 + \cdots + p_r = 1. \tag{11}$$

For Mellin asymptotics, we are interested in those complex roots that are close to the line $\Re(s) = 1$, that is, $1 - \sigma$ is close to 0. It is then easily realized that the only way to satisfy the equation in (11) is to have all the quantities p_j^{it} *simultaneously* close to 1. Accordingly, we must have, *simultaneously*[ii]

$$t \approx \frac{2\pi q_1}{|\log p_1|}, \quad \cdots, \quad t \approx \frac{2\pi q_r}{|\log p_r|}, \tag{12}$$

for some $q_j \in \mathbb{Z}$ (here, $A \approx B$ means that $A - B$ is close to 0). In other words, *each* rational number q_j/q_1 should be a "good" approximation to the corresponding ratio $(\log p_j)/(\log p_1) = \alpha_{1,j}$. Note here the rôle of the *"common denominator"* q_1. Accordingly, *the search for elements of $\mathcal{Z}$ close to $\Re(s) = 1$ involves "good" simultaneous rational approximations to the quantities $\alpha_{1,j}$* (of which one at least is an irrational number). Note that, in the case $r = 2$, the *best* rational approximants are well-described by continued fraction theory.

The next step borrows material from important works of Lapidus and van Frankenhuijsen. The *quality* of the approximations in (12) is essentially determined by the quantities $(v_1, \ldots, v_r)$, where

$$v_j := q_1 \frac{\log p_j}{\log p_1} - q_j. \tag{13}$$

We first discuss the case of a binary alphabet, $r = 2$ and write $v := v_2$. For a root $s = \sigma + it$, set now

$$s = 1 + \frac{2i\pi}{w_1} q_1 + \Delta, \tag{14}$$

so that $\Re(\Delta) = \sigma - 1$, where we consider $|\Delta|$ small. The original equation (11) becomes, upon taking into account the simplifications $e^{2i\pi} = 1$:

$$e^{-w_1} e^{-w_1 \Delta} + e^{-w_2} e^{-2i\pi v} e^{-w_2 \Delta} - 1 = 0. \tag{15}$$

The main idea then consists in regarding this equation as implicitly defining Δ. By the analytic version of the *Implicit Function Theorem*, this determines Δ as an analytic function of v, and one finds

$$\Delta(v) = d_1 iv - d_2 v^2 + \sum_{j=3}^{\infty} d_j (iv)^j, \tag{16}$$

where the expansion holds for all v satisfying $|v| < R_0$ (for some $R_0 > 0$), and the computable constants d_1, d_2 are real, with $d_2 > 0$. An immediate consequence is the *existence* of a root of $\lambda(s) = 1$ associated with any sufficiently good approximation q_2/q_1 to $\alpha_{1,2}$ satisfying $|v| < R_0$. Precisely, there is a root of $\lambda(s) = 1$ at

$$s = \sigma + it, \quad \sigma = 1 - d_2 v^2 + O(v^4), \quad t = \frac{2\pi}{w_1} q_1 + d_1 v + O(v^3). \tag{17}$$

The imaginary part is very close to an integer multiple of $2\pi/w_1$; the closeness of the real part to 1 is dictated by the quality of the approximation, as measured by (13) (the last quantity is itself conditioned by fine diophantine properties of $\alpha_{1,2}$). Figure 2 illustrates the usefulness of (17), when $p_1 = 2/3$, $p_2 = 1/3$.

[ii] In this discussion and later, we use the notation "$A \approx B$" to indicate informally that quantities A and B are "*approximately equal*"; that is, up to additive terms or multiplicative factors that are inessential in context.

q_2/q_1	v	$2\pi q_1/w_1$	$\Im(s_0)$	$\Re(s_0)$
$65/24$	$+2.8 \cdot 10^{-2}$	371.90	371.81	0.9977
$84/31$	$-5.1 \cdot 10^{-3}$	480.383	480.400	0.999925
$485/179$	$+2.5 \cdot 10^{-3}$	2773.8272	2773.8189	0.99998222
$1054/389$	$-1.0 \cdot 10^{-4}$	6028.03801	6028.03836	0.9999999675

Fig. 2: For $\mathfrak{p} = (2/3, 1/3)$, the first few best (continued fraction) approximants q_2/q_1, the corresponding values of the "quality factor" v and $2\pi q_1/w_1$, and the associated poles s_0 found near $\Re(s) = 1$, in accordance with (17).

The general case $r \geq 2$ proceeds along similar lines. A reasoning once more based on the Implicit Function Theorem (now in its multivariate version), with the change of variables (14), implies that elements of $\mathcal{Z}$ close to $\Re(s) = 1$ can be associated to good simultaneous approximations. The distance $1 - \Re(s)$ of such a pole s to the line $\Re(s) = 1$ is then governed by a *quadratic form* in the components v_j of the quality vector; the imaginary part is essentially proportional to the common denominator q_1.

As the foregoing discussion suggests, properties of the geometry of poles of $\Lambda(s)$ crucially depend on *simultaneous approximation* properties of the ratios $|\log p_j|/|\log p_k|$ supplemented by analytic arguments related to the *Implicit Function Theorem*.

2.1 Simultaneous approximations, BSADs, and approximation functions

Let $\lfloor x \rceil$ represent the *nearest integer function*, $\lfloor x \rceil := \lfloor x + \frac{1}{2} \rfloor$, and $\{x\} = x - \lfloor x \rceil$ the *centred fractional part*. For a vector $x = (x_1, \ldots, x_r) \in \mathbb{R}^r$, we define $\{x\}$ to be the vector $\{x\} = (\{x_1\}, \ldots, \{x_r\})$. We consider a norm on $\mathbb{R}^r$ denoted by $\|.\|$.

Approximation function and irrationality exponent. The integer Q is a *best simultaneous approximation denominator* of the vector $\beta \in \mathbb{R}^r$, *BSAD* for short, if

$$\|\{Q\beta\}\| < \|\{q\beta\}\|, \qquad \text{for all integers } q \text{ with } 0 < q < Q. \tag{18}$$

The existence, for $\beta \in \mathbb{R}^r \setminus \mathbb{Q}^r$, of arbitrarily good approximations is a basic fact of number theory; it implies that the set of BSADs is infinite [15, Th. 200].

Consider a vector $\beta \notin \mathbb{Q}^r$. The *approximation function* of β is the staircase function $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, which is constant on each interval $[Q^-, Q^+)$ formed with two successive BSAD and satisfies

$$f(q) = \frac{1}{\|\{Q^-\beta\}\|}, \qquad \text{for any } q \in [Q^-, Q^+), \qquad \text{so that} \quad f(q) \geq \frac{1}{\|\{q\beta\}\|}, \qquad \text{for all } q \in \mathbb{Z}_{>0}$$

A real ν is called an *approximation exponent*[iii] of $\beta \notin \mathbb{Q}^r$ if there exists a function $g(t) = A_\nu t^{\nu-1}$, with $A_\nu > 0$, which satisfies $g(q) \geq f(q)$, where f is the approximation function. The *irrationality exponent* $\mu(\beta)$ of the vector β is the infimum of the approximation exponents. The irrationality exponent $\mu(\beta)$ is infinite if there does not exist any approximation function with polynomial growth. When $\mu(\beta)$ is finite, the vector β is said to be *diophantine*; otherwise it is said to be *Liouvillean*. Many scalars β (such as $\pi, \tan(1), \exp(-2), \zeta(3), \log(5)$) and a few special vectors (such as those formed with logarithms of suitably independent rational or algebraic numbers [2, Th. 3.1]) are known to have finite irrationality

[iii] Since all norms on $\mathbb{R}^r$ are equivalent, this notion is independent of the choice of the norm.

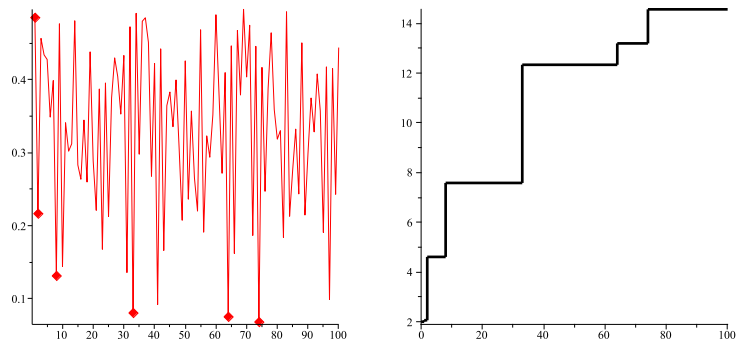


Fig. 3: Approximation function of $p_1 = \cot(1)$, $p_2 = e^{-2}$, and $p_3 = 1 - p_1 - p_2$. *Left*: The quantities $\|\{q\alpha^{(1)}\}\|$, corresponding to $\alpha^{(1)} = (1, w_2/w_1, w_3/w_1)$ and denominators $q \in [1\,..\,100]$, with BSADs rendered by diamonds. *Right*: the corresponding approximation function f_1 with jumps at the BSADs. (Data are for the sup-norm $\|\cdot\|_\infty$.)

exponent. The existence of such an exponent indicates that β cannot be too well approximated by rational vectors with a common denominator.

Approximation function and irrationality exponent of a family of probabilities. In an aperiodic family of probabilities (p_i), each row of the matrix $\mathcal{A} := (\alpha_{k,\ell})$, written as $\alpha^{(k)} := (\alpha_{k,1}, \ldots, \alpha_{k,r})$, has its diagonal element equal to 1, has at least one irrational component, and admits an approximation function f_k. The following related definition will be of use later.

Definition 2 [Global approximation function] *For each $k \in [1\,..\,r]$, consider a norm (which may depend on k) and the approximation function f_k of $\alpha^{(k)}$ relative to this norm. The functions F_-, F_+ defined as*

$$F_-(t) := \min_k \{f_k(\rho_k t)\}, \qquad F_+(t) := \max_k \{f_k(\rho_k t)\}, \quad \textit{with} \quad \rho_k := w_k/(2\pi). \tag{19}$$

are the lower, respectively, upper, global approximation function *relative to the family of norms.*

It can be proved elementarily (Proposition 3 in Appendix A) that these approximation functions f_k are in a suitable sense similar, and so are F_-, F_+. As a consequence, if there exists k for which $\alpha^{(k)}$ has a finite irrationality exponent, then any $\alpha^{(\ell)}$ has the same irrationality exponent: this is the *irrationality exponent of the family of probabilities* $\mathfrak{p} = (p_1, p_2, \cdots, p_r)$. Several facts are known about irrationality exponents. We quote here two important results, which will be useful later.

Proposition 2 *The irrationality exponent $\mu(\mathfrak{p})$ of a family $\mathfrak{p} = (p_1, p_2, \cdots, p_r)$ satisfies the following: (i) it is finite, as soon as all the probabilities p_i are rational and form an aperiodic system; (ii) it satisfies* almost surely*, in the sense of Lebesgue measure*[iv]*, $\mu(\mathfrak{p}) = r/(r-1)$.*

Proof: (i) This is a direct consequence of Baker's general theorem on linear form in logarithms [2, Th. 3.1], which implies in particular that any irrational ratio $\rho := (\log p_j)/(\log p_k)$ admits an (effectively computable) irrationality exponent. See also Rhin [30] for non-astronomical values of bounds on such exponents.

(ii) We rely on a classical result [5, Ch. VII] in the metric theory of diophantine approximation, which generalizes Khinchin's Theorem in the metric theory of continued fractions. It asserts the following, with $|\{x\}|$, the distance to the

[iv] Namely, the Lebesgue measure on the simplex consisting of the set of $(p_1, \ldots, p_r)$ such that $p_j > 0$ and $p_1 + \cdots + p_r = 1$.

nearest integer: *for a decreasing function $\psi(q)$, the set of solutions to the simultaneous inequalities, $|\{qy_j\}| < \psi(q)$, for $1 \le j \le n$, is almost surely (with respect to Lebesgue measure in $(y_1, \ldots, y_n) \in [0,1]^n$) infinite if $\sum \psi(q)^n$ diverges; the set of solutions is almost surely finite if $\sum \psi(q)^n$ converges.* Here, due to the fact that $\alpha_{k,k} = 1$, our problem is in fact an $(r-1)$ dimensional one, so $n = r-1$. Note that $\sum q^{-n\gamma}$ converges or diverges, according as $n\gamma > 1$ or $n\gamma < 1$. This means that, *almost surely*, we can adopt $\nu := 1 + 1/(r-1)$ as the irrationality exponent of the vector $\mathfrak{p}$. □

2.2 Statement of the main result

As a consequence of Theorem 1, in the aperiodic case, there does not exist any point of $\mathcal{Z}$ located on the vertical line $\Re(s) = 1$. The following result makes precise the geometry of the set $\mathcal{Z}$ *near* the line $\Re(s) = 1$, and describes how this geometry depends on approximation properties of the family of probabilities. It involves a particular family of norms (the canonical family), which is itself attached to the family $\mathfrak{p}$ and will be defined in Subsection 2.4, together with the function Δ.

Theorem 2 *Consider an aperiodic family of probabilities $(p_1, p_2, \ldots, p_r)$. There is a family of norms for which the global approximation functions F_-, F_+ of Definition 2 satisfy the following.*

For any $\eta > 0$, there exists an $\epsilon > 0$ and t_0 such that

(i) all *the elements $s = \sigma + it$ of the set $\mathcal{Z}_+ \cap \{\Im(s) > t_0\}$ satisfy*

$$\sigma \le \sigma_+(t) := 1 - \frac{1-\eta}{F_-(t+\epsilon)^2}\,;$$

(ii) *there are* infinitely many *elements $s = \sigma + it$ of $\mathcal{Z}_+$, which satisfy*

$$\sigma \ge \sigma_-(t) := 1 - \frac{1+\eta}{F_+(t-\epsilon)^2}.$$

Moreover, $\epsilon \to 0$ (and $t_0 \to \infty$), as $\eta \to 0$.

Theorem 2 specializes immediately to the case of finite irrationality exponents.

Theorem 3 *If the aperiodic family (p_i) has a finite irrationality exponent μ, then, the following holds:*

(i) *For any $\nu > \mu$, there exists $B_\nu > 0$, for which* all *elements $s = \sigma + it$ of the set $\mathcal{Z}_+ \cap \{\Im(s) \ge 1\}$ satisfy*

$$\sigma \le 1 - B_\nu t^{2-2\nu}. \tag{20}$$

(ii) *For any $\theta < \mu$, there exist $C_\theta > 0$, and an* infinite set *of elements $s = \sigma + it$ of $\mathcal{Z}_+$ such that*

$$\sigma \ge 1 - C_\theta t^{2-2\theta}, \qquad t \ge 1.$$

Part (i) of this theorem, is essential to the subsequent sections, as it guarantees a pole-free region for $\Lambda(s)$, under the assumption of a finite irrationality exponent. Part (ii) indicates that the corresponding bounds are essentially best possible. See Figure 4 for an illustration.

The proof of Theorem 2 (and its specialization, Theorem 3) is in two phases: we first show that poles of $\Lambda(s)$ can only be present in definite regions, "ladders", that are associated with "good" (including best) approximants (Subsection 2.3). We then show, conversely, that any sufficiently good approximant is indeed attached to a pole (Subsection 2.4), this by methods markedly inspired by the treatment of Lapidus and van Frankenhuijsen in [20, §3.5]. We can then conclude with the proof of Theorem 2 in Subsection 2.5.

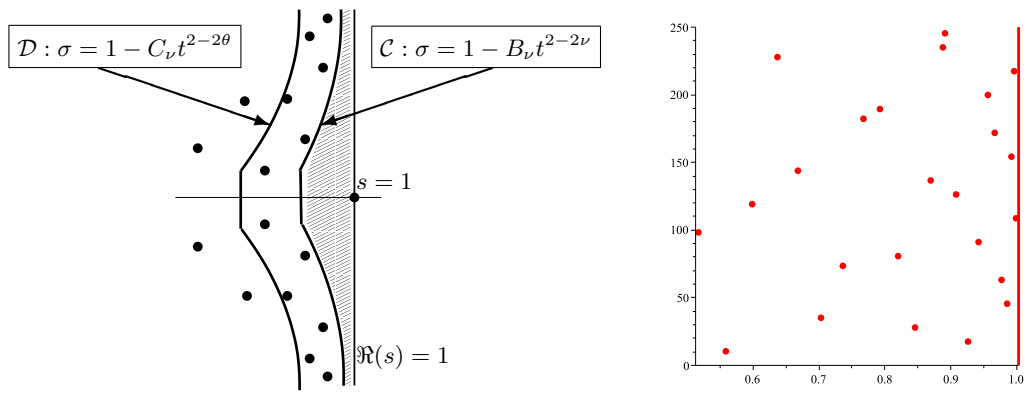


Fig. 4: *Left*: Excluded and guaranteed regions for poles $s = \sigma + it$ of $\Lambda(s)$, according to Theorem 3: the region right of the curve $\mathcal{C}$ (hashed) is pole-free (except $s = 1$); the region between the two curves $\mathcal{C}, \mathcal{D}$ is guaranteed to contain infinitely many poles. *Right*: The poles of $\mathcal{Z}_+$ with $\Re(s) > \frac{1}{2}$ and $\Im(s) < 250$, for $\mathfrak{p} = (\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$.

2.3 Ladders and poles

We shall deal repeatedly with *vertical strips* of the form

$$\mathcal{B}(\delta) := \{s = \sigma + it;\ \sigma \in]1 - \delta, 1]\}, \tag{21}$$

our purpose being to locate elements of $\mathcal{Z}$ in such sufficiently narrow strips. This is made precise by Lemmas 1–3 below, with *proofs to be found in Appendix* A. First the notion of *ladder* is essential.

Definition 3 [Ladders] *Denote by $\mathcal{R}(\delta, \epsilon)$ the rectangle*

$$\mathcal{R}(\delta, \epsilon) := \{s = \sigma + it;\ \sigma \in]1 - \delta, 1],\ \ |t| \le \epsilon w\},$$

where the scaling parameter is $w := 1/w_r$. For each index $k \in [1 \,.\,.\, r]$, the translate $\mathcal{R}_k(q, \delta, \epsilon)$ of $\mathcal{R}(\delta, \epsilon)$ by the complex $2i\pi q/w_k$ gives rise, when q varies in $\mathbb{Z}$, to a regular ladder*, denoted by $\mathcal{L}_k(\delta, \epsilon)$. The intersection $\bigcap \mathcal{L}_k(\delta, \epsilon)$ for $k \in [1 \,.\,.\, r]$ defines the* joint ladder *$\mathcal{L}(\delta, \epsilon)$ relative to parameters (δ, ϵ).*

First, the geometry of a joint ladder is closely related to approximability properties. Precisely, a step of the joint ladder exists only if it is associated with good approximation denominators.

Lemma 1 *For any ϵ, with $0 < \epsilon < \pi/2$, the step of the joint ladder $\mathcal{L}(\delta, \epsilon)$ relative to the sequence of integers $(q_1, q_2, \ldots, q_r)$ is nonempty, if and only if, for any $k \in [1 \,.\,.\, r]$, the vector $\{q_k \alpha^{(k)}\}$ lies in the set*

$$\mathcal{U}(\epsilon) := \left\{(u_1, \ldots, u_r) \in \mathbb{R}^r; \quad |u_\ell| \le \frac{\epsilon}{\pi} w w_\ell\right\}. \tag{22}$$

Next, the elements of $\mathcal{Z}$ close enough to the line $\Re(s) = 1$ *must* belong to a joint ladder.

Lemma 2 *For any ϵ, with $0 < \epsilon < \pi/2$, there exists $\delta(\epsilon) > 0$ for which the inclusion $\mathcal{Z} \cap \mathcal{B}(\delta) \subset \mathcal{L}(\delta, \epsilon)$ holds for any $\delta \le \delta(\epsilon)$. Furthermore, $\delta(\epsilon)$ is $O(\epsilon^2)$, as $\epsilon \to 0$. A pair (ϵ, δ) with $\delta \le \delta(\epsilon)$ is called* compatible.

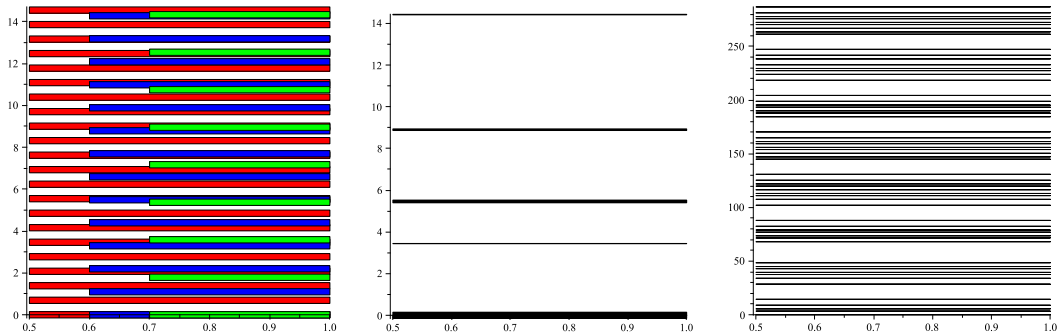


Fig. 5: Ladders associated with the probabilities $\mathfrak{p} = (1/2, 1/3, 1/6)$, when $\epsilon w = 0.15$. *Left*: the three regular ladders $\mathcal{L}_1(\delta_1, \epsilon), \mathcal{L}_2(\delta_1, \epsilon), \mathcal{L}_3(\delta_3, \epsilon)$, with $\delta_1 = 0.5$, $\delta_2 = 0.6$, $\delta_3 = 0.7$ (different δ are adopted for readability). *Middle:* The joint ladder $\mathcal{L}(0.5, \epsilon)$ resulting from the intersection of the three regular ladders on the left, for $\Im(s) \leq 15$. *Right*: the joint ladder $\mathcal{L}(0.5, \epsilon)$, for $\Im(s) < 300$. (Data are here relative to the sup-norm.)

Finally, for well chosen parameters (ϵ, δ), each step of the joint ladder contains *at most* one pole of Λ.

Lemma 3 *There exists a compatible pair (ϵ, δ), for which all the elements of $\mathcal{Z}$ located in the vertical strip $\mathcal{B}(\delta)$ belong to the joint ladder $\mathcal{L}(\delta, \epsilon)$ and any step of the ladder contains at most one element of $\mathcal{Z}$.*

Figuratively: *poles close enough to $\Re s = 1$ are somewhere to be found among steps of sufficiently narrow ladders*.

2.4 The Δ–function and the canonical family of norms

We now establish the converse property that poles systematically arise in association with sufficiently good simultaneous approximations. In accordance with the indications above (e.g., Equation (15)), the proof crucially relies on the existence of an analytic (effectively computable) mapping Δ, which describes, for well chosen parameters (ϵ, δ), the unique possible root of $\mathcal{Z}$ contained in each step of the joint ladder. *The proofs of Lemmas 4 and 5 are to be found in Appendix* A.

Lemma 4 *There exist a compatible pair (ϵ, δ), and an analytic function Δ, defined on the hypercube $\mathcal{U}(\epsilon)$ of* (22)*, for which the only possible root of $\mathcal{Z}$ contained in the step of the joint ladder $\mathcal{L}(\delta, \epsilon)$ of index $(q_1, q_2, \ldots, q_r)$ can be written, for any $k \in [1 \,.\,.\, r]$, as*

$$z = 2i\pi \frac{q_k}{w_k} + \Delta(\{q_k \alpha^{(k)}\}). \tag{23}$$

The key idea from [20, §3.5] is to start from a multivariate extension of $\lambda(s)$

$$\lambda(s, u) := \sum_{\ell=1}^{r} p_\ell^s \exp\left(-2\pi i u_\ell\right), \tag{24}$$

where $u = (u_1, \ldots, u_r)$ (see Appendix A for details). The function $\Delta(u)$ is then defined as the multivariate analytic function that expresses the solution in s of the equation $\lambda(s, u) = 1$, whose existence

is granted by the *Implicit Function Theorem*. Then, one can determine a canonical family of norms, attached to the mapping Δ (Equations (49) and (50) of Appendix A), which prove to be well adapted to the geometry of the poles of Λ.

Lemma 5 *Let $S := 1 - \Re(\Delta)$ and denote by T the Hessian quadratic form of S at the origin. Each restriction of T to a hyperplane $\{u_k = 0\}$ is positive definite and defines a norm denoted by T_k. Then, for any $\eta > 0$, there exists $\epsilon > 0$, such that, for any u that belongs to $\mathcal{U}(\epsilon)$, with $u_k = 0$ for some $k \in [1\,..\,r]$, one has*

$$(1-\eta)\|u\|_{T_k}^2 \le S(u) \le (1+\eta)\|u\|_{T_k}^2 \tag{25}$$

2.5 Proof of Theorem 2

Lemmas 1–5 now make it possible to conclude with the proof of Theorem 2.

Proof: [(i) *of Theorem 2*]. Consider an aperiodic family of probabilities, and the family of norms (T_k). For any $\eta > 0$, there exists $\epsilon_0 > 0$ such that the inequality (25) holds on $\mathcal{U}(\epsilon_0)$. Fix a pair (ϵ, δ) as in Lemma 4, with $\epsilon < \epsilon_0$. By this lemma, any root $s = \sigma + it$ of $\lambda(s) = 1$ in the vertical strip $\mathcal{B}(\delta)$, distinct from 1, is written for any $k \in [1\,..\,r]$ as in (23) for some $q_k \in \mathbb{Z}, q_k \neq 0$, with $\{q_k\alpha^{(k)}\} \in \widetilde{\mathcal{U}}(\epsilon)$. Then, the inequality (25) applies to each $\{q_k\alpha^{(k)}\}$, and

$$\delta > 1 - \sigma = 1 - \Re[\Delta(\{q_k\alpha^{(k)}\})] \ge (1-\eta)\|\{q_k\alpha^{(k)}\}\|_{T_k}^2 \ge \frac{1-\eta}{f_k(q_k)^2}, \tag{26}$$

where f_k is the approximation function of $\alpha^{(k)}$ relative to the norm T_k. Furthermore, since $s = \sigma + it$ belongs to $\mathcal{R}_k(q_k, \delta, \epsilon)$, one has

$$\left| t - 2\pi\frac{q_k}{w_k} \right| \le \epsilon w, \qquad \text{and thus} \quad q_k \le \rho_k(t + \epsilon w).$$

Then, any root $s = \sigma + it$ of the vertical strip $\mathcal{B}(\delta)$, distinct of $s = 1$, satisfies, for any k,

$$\delta > 1 - \sigma \ge \frac{1-\eta}{f_k(q_k)^2}, \qquad \text{and thus} \quad 1 - \sigma \ge \frac{1-\eta}{F_-(t+\epsilon w)^2},$$

where $F_-(t) := \min\{f_k(\rho_k t)\}$ is the (canonical) lower global approximation function. Then, assertion (i) holds as long as t_0 is chosen such that $F_-(t_0 + \epsilon w) > (1-\eta)/\delta$.

[(ii) *of Theorem 2*]. For any $\eta > 0$, there exists $\epsilon_0 > 0$ such that the inequality (25) holds on $\mathcal{U}(\epsilon_0)$. Fix a pair (ϵ, δ) as in Lemma 4. Consider now a BSAD q_k of $\alpha^{(k)}$. Since q_k is a BSAD large enough, then the vector $\{q_k\alpha^{(k)}\}$ is close enough to $\underline{0}$, and the complex number $\Delta(\{q_k\alpha^{(k)}\})$ belongs to $\mathcal{R}(\delta, \epsilon)$. The root $s = \sigma + it$ associated to this BSAD q_k of $\alpha^{(k)}$ satisfies

$$1 - \sigma \le (1+\eta)\,\|\{q_k\alpha^{(k)}\}\|_{T_k}^2 = \frac{1+\eta}{f_k(q_k)^2}$$

where f_k is the approximation function of $\alpha^{(k)}$ relative to the norm T_k. Thus, in the same vein as previously, the inequality $q_k \ge \rho_k(t - \epsilon w)$ holds and entails, for $t > \epsilon w$,

$$1 - \sigma \le \frac{1+\eta}{f_k(q_k)^2} \qquad \text{and thus} \quad 1 - \sigma \le \frac{1+\eta}{F_+(t-\epsilon w)^2},$$

where $F_+(t) := \max\{f_k(\rho_k t)\}$ is the (canonical) upper global approximation function. □

Note. Our purpose has been to locate poles in as precise a way as seems possible, so as to supplement some of the results of [20]. Weaker estimates can be obtained in a simpler way, by considering only the approximation function f_1, taken relative to any of the classical norms, for instance the sup-norm. In that case, the resulting boundaries of the two regions of Theorem 2 become of the (less tight) form $1 - c_1F(c_2t)^{-2}$ and $1 - d_1F(d_2t)^{-2}$, for some $c_1, c_2, d_1, d_2 > 0$, with here $F \equiv f_1$.

3 Mellin analysis and the function $\Lambda(s)$

We can now start to exploit the results of Section 2, which provide a precise description of the location of the poles of $\Lambda(s)$. This only involves a routine use of Mellin transform theory [11]. It is convenient to develop a somewhat abstract analytic framework, which covers the applications of Section 1 as well as many more of algorithmic interest. We shall then easily specialize the discussion and treat our three pilot examples relative to trie parameters in the next section, Section 4.

We show here that any exclusion region for the poles of $\Lambda(s)$ provides an estimate of (an upper bound on) the error term in asymptotic forms associated with a natural class of functions that includes the trie parameters of Section 1. See Theorem 4 below, with bounds of the form $\exp(-(\log x)^{1/\theta})$, with $\theta > 1$, in the case of a finite irrationality exponent. Direct consequences of Theorem 4 —and the deep literature on diophantine approximation— are that such bounds, $\exp(-(\log x)^{1/\theta})$, with $\theta > 1$ hold for all aperiodic sources involving *rational probabilities*, see Corollary 1; they also hold for *almost all probability vectors* $\mathfrak{p}$, in a measure-theoretic sense, see Corollary 2.

3.1 Error bounds

First we formalize the *analytic* context[(v)] common to problems such as size and path length of tries, or the symbol-complexity of quicksort.

Definition 4 *A function $F(x)$ and its Mellin transform assumed to be of the form $F^\star(s) = \Lambda(-s)g(s)$ are said to* satisfy Assumption A *if the following conditions hold:*

(i) *for some $h > 0$, the open fundamental strip of $F^\star(s)$ intersects the substrip $-1 < \Re(s) < -1 + h$;*

(ii) *the function $g(s)$ is meromorphically continuable to the strip $-1 - h < \Re(s) < -1 + h$, with at most a pole at $\Re(s) = -1$, and it satisfies, as $|s| \to \infty$ in that strip, the following conditions, uniformly with respect to $\Re(s)$,*

$$|g(s)| = O\left(e^{-K|\Im(s)|}\right), \qquad -1 - h < \Re(s) < -1 + h, \quad |s| \to \infty, \tag{27}$$

for some constant $K > 0$.

For instance[(vi)], $g(s) = \Gamma(s)$ satisfies (27) with $h = 1/2$ and any $K < \pi/2$.

Regarding $\Lambda(s)$, we gather here a few properties to be of use. These are based on Theorems 2 and 3, as well as on the proofs of the supporting Lemmas (Section 2 and Appendix A).

There exists a vertical strip $\mathcal{B}(\delta)$ and a parameter ϵ, such that the following holds.

- (i) *Poles $\chi \in \mathcal{Z}_+$ are uniformy separated from one another.*
- (ii) *Outside of the ladder $\mathcal{L}(\delta, \epsilon)$, the function $\Lambda(s)$ is uniformly bounded.*
- (iii) *Inside the ladder, the function $\Lambda(s)$ remains uniformly bounded, provided that s be uniformly far from the set $\mathcal{Z}_+$.*
- (iv) *The values of the derivative $\lambda'(\chi)$ at poles χ are uniformy bounded away from* 0.

(v) Other scenarios are possible and various growth assumptions imply various asymptotic scales: see typically [20, Ch. 7] for error terms in the prime orbit function of suspended flows.

(vi) See Item 6.1.44 of [1], after which $|\Gamma(x + iy)| \sim \sqrt{2\pi}e^{-\pi|y|/2}|y|^{x-1/2}$ (Stirling's formula!), as $|y| \to \infty$.

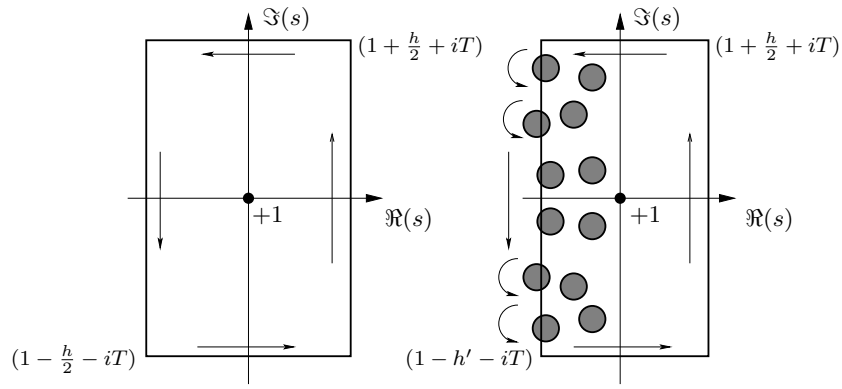


Fig. 6: The original integration contour $\mathcal{C}_0(T)$ [left]; the final contour $\mathcal{C}_1(T)$ [right].

For convenience, we shall take the inverse Mellin integral under the form

$$F(x) = \frac{1}{2i\pi}\int_{1+h/2-i\infty}^{1+h/2+i\infty} g(-s)\Lambda(s)x^s\,ds, \tag{28}$$

so that poles farther left contribute lesser terms in asymptotic expansions. Under Assumption A, the contribution of the pole $s=1$ of the integrand is of the form $xP(\log x)$, where $\deg(P)=m$ and m is the multiplicity of the pole of $g(-s)$ at $s=1$ (with $m=0$ in case $g(-s)$ is analytic at $s=1$).

Lemma 6 *Under Assumption A, for an aperiodic probability vector* $\mathfrak{p}$*, the estimate*

$$F(x) = xP(\log x) + x\Phi(x) + O\left(x^{1-h'}\right), \qquad x\to+\infty, \tag{29}$$

holds for some $h'\in\mathbb{R}_{>0}$ *with* $h'\le h$*, where* $P(\log x) := \mathrm{Res}(g(-s)\Lambda(s)x^{s-1}; s=1)$ *is a polynomial in* $\log x$*, and*

$$\Phi(x) := \sum_{\chi}\frac{g(-\chi)}{\lambda'(\chi)}x^{\chi-1}, \tag{30}$$

with the sum being extended to all the poles χ *of* $\Lambda(s)$ *such that* $1-h'\le\Re(\chi)<1$.

Note. This formula is, in its own small way, a pendant of "explicit formulae" in the theory of prime numbers. (We believe that it can be extended to a sum over *all* poles of $\Lambda(s)$.) It does have some merit for the understanding of the non-asymptotic regime of finite values of x: see Appendix B.

Proof: The proof is, as usual, based on contour integration and the residue theorem (cf [8]). We start by considering the integral of (28) taken along the rectangular contour $\mathcal{C}_0(T)$ defined by its corners at $1-h/2-iT$ and $1+h/2+iT$; see Figure 6 (left). By suitably narrowing the strip with h', we may assume that all the results of the previous section apply, to the effect that *poles are separated by a distance at least* D in the strip $1-2h'<\Re(s)<1$.

We shall later let $T\to+\infty$, but first restrict values of T in such a way that T lies outside of a fixed ladder $\mathcal{L}(\delta_1,\epsilon_1)$ satisfying the conditions of Lemma 3 in the previous section. In this way, we guarantee that $|1-\lambda(s)|$ stays uniformly bounded away from 0 on the horizontal segments of the contour. (See Theorem 2 and Equation (42) of Appendix A.) and use $|1-\lambda(s)|>|\Re(1-\lambda(s))|$.)

We then set $d := D/10$ (say) and deform $\mathcal{C}_0(T)$ into a new contour $\mathcal{C}_1(T)$ obtained as follows. Draw around each pole χ a disc of radius d. If the vertical segment on the left of the contour $\mathcal{C}_0$ intersects one of the discs, then replace the intersection of the segment and the disc by an arc of the boundary of the disc, keeping the contour connected (Figure 6). We shall also choose that arc of the boundary, which lies to the left of $\Re(s) = 1 - h'$. In this way, all points of $\mathcal{C}_1(T)$ are at distance at least d of a pole and, on the left boundary of the contour $\mathcal{C}_1(T)$, we have $\Re(s) \leq 1 - h'$.

We can now argue as follows. Along the left boundary of $\mathcal{C}_1(T)$, with $s = \sigma + it$, either some of the quantities $\cos(w_j t)$ are sufficiently bounded away from 1, which ensures that $\Re(\lambda(s))$ is bounded away from 1; or, else, by Lemma 4 of the previous section, there must be a root of $\lambda(s) = 1$. The latter case is precisely avoided by the contour's deformation. Thus, the inverse Mellin integral (28) taken along this left boundary of the contour is in absolute value bounded from above by a quantity of the form $K_1 x^{1-h'}$, for some $K_1 > 0$.

Finally, let $T \to \infty$. The contribution of the two horizontal segments of $\mathcal{C}_1(T)$ tends to 0 (see previous remarks). The inverse Mellin integral of (28) taken along $\mathcal{C}_1(\infty)$ can then be classically evaluated in two different ways: (i) as a sum of the contributions on the left and right sides, which gives $F(x) + O(x^{1-h'})$; (ii) as a sum of residues, which gives $xP(\log x) + x\Phi(x)$. This completes the proof. □

Theorem 4 *Consider a function $F(x)$ and its Mellin transform that satisfy Assumption A. Let the probability vector $\mathfrak{p}$ have a* finite irrationality exponent *μ and take any $\nu > \mu$. Then, $F(x)$ admits the asymptotic form* (29)*, with the correction term Φ satisfying*

$$\Phi(x) = O\left(\exp\left(-(\log x)^{1/(2\nu-1)}\right)\right), \qquad x \to \infty. \tag{31}$$

Proof: As before, the set of poles of $\Lambda(s)$ is denoted by $\mathcal{Z}$, and we let $\mathcal{Z}_+$ be the subset of poles χ such that $\Im(\chi) > 0$. We use here $c_1, c_2, \ldots$ to represent positive constants, which need not be made explicit.

First a preliminary remark. By developments of the previous section, the quantities $1/|\lambda'(\chi)|$ are bounded by an absolute constant, so that we have

$$\Phi(x) = O\left(\sum_{\chi \in \mathcal{Z}_+} \left|g(-\chi)x^{\chi-1}\right|\right). \tag{32}$$

From Theorem 2, all the poles $s = \sigma + it$ of $\mathcal{Z}_+$ satisfy

$$\sigma < 1 - \frac{c_1}{G(c_2 t)^2}, \tag{33}$$

where G is any upper bound for the lower approximation function F_-. In the case of a *finite* approximation exponent μ, we choose (momentarily) some $\widetilde{\nu}$ such that $\mu < \widetilde{\nu} < \nu$ and we may accordingly adopt $G(t) = c_3 t^{\widetilde{\nu}-1}$.

Let $\chi_1, \chi_2, \ldots$ be the sequence of poles $\chi \in \mathcal{Z}_+$ ordered by increasing values of $\Im(\chi)$ and write $\chi_j = \sigma_j + it_j$. Thanks to the (vertical) uniform separation property, we have, the lower bound $t_j > c_4 j$, for some $c_4 > 0$, so that $|g(-\chi_j)|$ is bounded from above:

$$|g(-\chi_j)| \leq e^{-Kc_4 j}. \tag{34}$$

On the other hand, we have $t_j < c_5 j$ since the number of poles in a vertical strip is bounded from above by the number of steps of any sufficiently narrow ladder $\mathcal{L}(\delta, \epsilon)$; see the previous section. This last upper bound on t_j then implies

$$\sigma_j - 1 < -\frac{c_6}{j^{2\widetilde{\nu}-2}}. \tag{35}$$

In summary, thanks to (32), (34), and (35), we have obtained, for some $c', K' > 0$, the estimate

$$\Phi(x) = O\left(\sum_{j\geq 1} \exp\left(U(j)\right)\right), \qquad \text{with} \quad U(j) := -K'j - c'j^{2-2\widetilde{\nu}}\,\log x. \tag{36}$$

The function $j \mapsto U(j)$ is unimodal with a peak at $j_0 \equiv j_0(x)$, which is determined by cancelling a derivative. With $L := \log x$, it is found to satisfy $j_0 \sim c_7 L^{1/(2\widetilde{\nu}-1)}$, to the effect that $U(j_0) \sim c_8 L^{1/(2\widetilde{\nu}-1)}$, We can then simply bound the terms $j \leq j_0$ of the sum in (36) by the maximal term $e^{U(j_0)}$; we can also take advantage of the geometric decay of the remaining terms, corresponding to $j > j_0$ (due to the $g(-\chi)$ factor), see (34). In this way, we find

$$|\Phi(x)| \leq c_9 j_0\, \exp U(j_0) + O\left(\exp(U(j_0))\right) = O\left(\exp\left(-c_{10} L^{1/(2\widetilde{\nu}-1)}\right)\right).$$

This last estimate implies the statement, since $\widetilde{\nu} > \nu$. □

Note on the Liouvillean case (infinite irrationality exponent). Let $G(t)$ be an upper bound of the global approximation function in the sense of the previous section and of (33) above. We abbreviate the discussion and, for convenience, assume enough differentiability and convexity properties of G. Also, we disregard inessential constants. The case of a superpolynomial function G corresponds to the *Liouvillean case* of number theory, where the irrationality exponent is infinite and the ratios w_i/w_j are (simultaneously) *extremely well approximated* by rationals. (For the binary case $r = 2$, corresponding probability vectors $\mathfrak{p}$ can be constructed by means of continued fraction theory.) Here, we take $f \approx g$ to mean that f and g are asymptotically equivalent up to a polynomial transformation. (i.e., $\log\log f \sim \log\log g$). Then, the computations of the previous section generalize: the index j_0 of the maximum term satisfies $G(j_0)^3/G'(j_0) \approx L$, with $L = \log x$. Note that for smooth fast growing functions, such as towers of exponentials, we have $G' \approx G$, so that $G(j_0) \approx \sqrt{L}$. It then suffices to solve asymptotically for j_0 and bound the terms as before. For instance, we have the following approximate correspondences,

$$\begin{aligned} G(t) \approx e^t \quad &\Longrightarrow \quad \Phi(x) \approx O\left(\exp(-\log\sqrt{\log x})\right) \approx \frac{1}{\log x} \\ G(t) \approx e^{e^t} \quad &\Longrightarrow \quad \Phi(x) \approx O\left(\exp(-\log\log\sqrt{\log x})\right) \approx \frac{1}{\log\log x}, \end{aligned} \tag{37}$$

and so on. In this way, *probability vectors can be constructed such that the upper bound* (31) *on* $|\Phi(x)|$ *tends to* 0 *arbitrarily slowly.*

3.2 Rational probabilities and metric aspects

Here are two consequences of the foregoing developments and of Proposition 2 of the previous section.

Corollary 1 *Let* $\mathfrak{p} = (p_1, \ldots, p_r)$ *be an aperiodic probability vector all of whose components are rational numbers. Then there exists an effectively computable number* $\theta > 1$ *such that the error term* $\Phi(x) \equiv \Phi_{\mathfrak{p}}(x)$ *satisfies*

$$\Phi(x) = O\left(\exp\left(-(\log x)^{1/\theta}\right)\right).$$

The statement also holds in the case where the p_j *are algebraic.*

Corollary 2 *Fix* r *and take any* $\theta > (r+1)/(r-1)$. *In the sense of Lebesgue measure,* almost all *probability vectors* $\mathfrak{p} = (p_1, \ldots, p_r)$ *are such that the error term* $\Phi(x) \equiv \Phi_{\mathfrak{p}}(x)$ *satisfies*

$$\Phi(x) = O\left(\exp\left(-(\log x)^{1/\theta}\right)\right).$$

In the binary case, the bound is, for any $\epsilon > 0$, almost surely *of the form* $\Phi(x) = O\left(\exp\left(-(\log x)^{1/3-\epsilon}\right)\right)$.

Parameter	*Mellin transform*	*multiplicity*	*asymptotics*
Trie size	$S^\star(-s) = (s-1)\Gamma(-s)\Lambda(s)$	$r = 1$	$S(x) \sim \frac{1}{h_1} x$
Trie path length	$P^\star(-s) = s\Gamma(-s)\Lambda(s)$	$r = 2$	$P(x) \sim \frac{1}{h_1} x \log x$
Quicksort cost	$Q^\star(-s) = \frac{2}{s(s-1)}\Gamma(-s)\Lambda(s)$	$r = 3$	$Q(x) \sim \frac{1}{h_1} x (\log x)^2$

Fig. 7: The three reference Mellin transforms, $S^\star(-s), P^\star(-s), Q^\star(-s)$. The last two columns indicate the multiplicity r of the pole at $s = 1$ and the main asymptotic term relative to the original functions, $S(x), P(x), Q(x)$. The quantity h_1 is the natural entropy of the source.

4 Asymptotic analysis of tries

We can finally conclude regarding the fine structure of trie parameters. For ease of reference, Figure 7 tabulates the three Mellin transforms that appear in Proposition 1 of Section 1, and are relative to trie size (S), trie path length (P), and the symbol-comparison cost of Quicksort (Q). (Since we make use of the non-standard form of Mellin inversion (28), these are expressed here in terms of $-s$.)

The function $\Gamma(-s)$ has a simple pole at $s = 1$, where $\Gamma(-s) \sim 1/(s-1)$ and higher terms involve Euler's constant as well as Riemann zeta values at the integers [40]. The function $\Lambda(s)$ being $\Lambda(s) = \frac{1}{1-\lambda(s)}$, where $\lambda(s) = \sum_{j=1}^r p_j^s$, it can be expanded near its simple pole at $s = 1$, based on

$$1 - \lambda(s) = \sum_{m=1}^{\infty} h_m (-1)^{m-1} \frac{(s-1)^m}{m!}, \qquad h_m = \sum_{j=1}^{r} p_j |\log p_j|^m,$$

where the h_j represent *generalized entropies*. Accordingly, $\Lambda(s)$ has a simple pole at $s = 1$, with residue equal to the (natural) *entropy* h_1, also often denoted by H.

All the ingredients are in place to compute (automatically even, under the symbolic system MAPLE) the dominant contributions in the asymptotic expansion of $S(x), P(x), Q(x)$, which are of the form $xA(\log x)$ where A is a polynomial: the quantities $A(\log x)$ are given by $\operatorname{Res}\left(f^\star(-s)x^{s-1}\right)_{s=1}$, with $f^\star(-s)$ any one of the three functions of Figure 7. This gives us, with $L = \log x$, the polynomials relative to the three parameters of interest,

$$\left\{ \begin{array}{lllllll} A^{(S)}(L) & = & \frac{1}{h_1}, & A^{(P)}(L) & = & \frac{1}{h_1} L + \frac{1}{h_1}\left(\gamma + \frac{h_2}{2h_1}\right), \\ A^{(Q)}(L) & = & \frac{1}{h_1} L^2 + \frac{1}{h_1}\left(2\gamma - 4 + \frac{h_2}{h_1}\right) L + c_0, & & & \end{array} \right. \tag{38}$$

where c_0 involves γ, $\zeta(2) \equiv \pi^2/6$, and h_1, h_2, h_3.

It is then a trivial task to transpose Theorem 4 (set $a := h'$) and state our final result.

Theorem 5 *Consider the three parameters S, P, Q of tries built from a random number Poisson(x) of infinite words drawn independently from a fixed memoryless source with probability vector $\mathfrak{p} = (p_1, \ldots, p_r)$. Assume that the system $\mathfrak{p}$ is* aperiodic *with a finite irrationality exponent μ and take ν to be any number*

larger than μ. Then, each parameter $\xi \in \{S, P, Q\}$ has an expectation that satisfies asymptotically

$$\mathbb{E}_{\mathcal{P}(x)}[\xi] = xA^{(\xi)}(\log x) + x\Phi^{(\xi)}(x) + O\left(x^{1-a}\right), \qquad x \to \infty, \tag{39}$$

for some $a > 0$, where $A^{(\xi)}$ is polynomial given by (38) *and $\Phi^{(\xi)}$ satisfies in each case the estimate*

$$\Phi^{(\xi)}(x) = O\left(\exp\left(-(\log x)^{1/(2\nu-1)}\right)\right). \tag{40}$$

Appendix B illustrates some of the subtle numerical phenomena at stake in the aperiodic case. By contrast, in the *periodic case*, it is well known [18, 36] that the asymptotic expansion (39) assumes a simpler form: for some $a > 0$,

$$\mathbb{E}_{\mathcal{P}(x)}[\xi] = xA^{(\xi)}(\log x) + x\Phi^{(\xi)}(x) + O\left(x^{1-a}\right),$$

where $\Phi(x)$ is plainly a periodic function of $\log x$.

Notes. All previous remarks apply. Namely:

1. A bound of the form (40) holds as soon as the probabilities p_j are all *rational* and satisfy the *aperiodicity* condition (Corollary 1). Also, such a bound holds *almost surely*, with an exponent that then only depends on the dimension (Corollary 2).

2. The knowledge of the order of growth of a best approximation function implies a whole range of upper bounds, which are weaker than (40), and can come arbitrarily close to x; see (37). We believe that lower bounds of a similar shape can be developed.

3. The results above apply to many algorithms involving variants of tries, such as digital search trees and Patricia tries, Lempel–Ziv data compression, tree or stack communication protocols, the Bentley–Sedgewick Ternary Search Tries, and so on [24, 36].

4. Similar results can be obtained for the *fixed-size model*, where the number of elements in the trie is fixed at n, asymptotically in n. This fact owes to the complex *Wahlverwandtschaften* between Mellin transforms and the Nörlund–Rice integrals of difference calculus [12].

The research just presented here fits into a broader scheme, that of "dynamical sources" [7, 14, 37, 38], themselves further extended as "general sources" in [39]; see especially the discussion at the end of [39, §3]. Recent work of Cesaratto and Vallée [6] for instance shows that, for a large class of such sources, *not* including memoryless sources, the limit distribution of the average depth in tries is asymptotically normal. It is piquant to note that the conceptually simplest of all information sources are, in certain ways, far from being the ones exhibiting the simplest behaviour.

References

[1] ABRAMOWITZ, M., AND STEGUN, I. A. *Handbook of Mathematical Functions*. Dover, 1973. A reprint of the tenth National Bureau of Standards edition, 1964.

[2] BAKER, A. *Transcendental number theory*, second ed. Cambridge University Press, Cambridge, 1990.

[3] BALADI, V., AND VALLÉE, B. Euclidean algorithms are Gaussian. *Journal of Number Theory 110* (2005), 331–386.

[4] BALADI, V., AND VALLÉE, B. Exponential decay of correlations for surface semi-flows without finite Markov partitions. *Proceedings of the American Mathematical Society 133*, 3 (2005), 865–874.

[5] CASSELS, J. W. S. *An Introduction to Diophantine Approximation*. Cambridge University Press, 1957.

[6] CESARATTO, E., AND VALLÉE, B. Gaussian distribution of trie depth for dynamical sources. Manuscript,, 2010.

[7] CLÉMENT, J., FLAJOLET, P., AND VALLÉE, B. Dynamical sources in information theory: A general analysis of trie structures. *Algorithmica 29*, 1/2 (2001), 307–369.

[8] FAYOLLE, G., FLAJOLET, P., AND HOFRI, M. On a functional equation arising in the analysis of a protocol for a multiaccess broadcast channel. *Advances in Applied Probability 18* (1986), 441–472.

[9] FILL, J. A., AND JANSON, S. The number of bit comparisons used by Quicksort: An average-case analysis. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA04)* (2004), pp. 293–300.

[10] FLAJOLET, P. The ubiquitous digital tree. In *STACS 2006* (2006), B. Durand and W. Thomas, Eds., vol. 3884 of *Lecture Notes in Computer Science*, pp. 1–22. Proceedings of 23rd Annual Symposium on Theoretical Aspects of Computer Science, Marseille, February 2006.

[11] FLAJOLET, P., GOURDON, X., AND DUMAS, P. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science 144*, 1–2 (June 1995), 3–58.

[12] FLAJOLET, P., AND SEDGEWICK, R. Mellin transforms and asymptotics: finite differences and Rice's integrals. *Theoretical Computer Science 144*, 1–2 (June 1995), 101–124.

[13] FLAJOLET, P., AND SEDGEWICK, R. *Analytic Combinatorics*. Cambridge University Press, 2009. Also available electronically from the authors' home pages.

[14] FLAJOLET, P., AND VALLÉE, B. Continued fractions, comparison algorithms, and fine structure constants. In *Constructive, Experimental, and Nonlinear Analysis* (Providence, 2000), M. Théra, Ed., vol. 27 of *Canadian Mathematical Society Conference Proceedings*, American Mathematical Society, pp. 53–82.

[15] HARDY, G. H., AND WRIGHT, E. M. *An Introduction to the Theory of Numbers*, fifth ed. Oxford University Press, 1979.

[16] HOLMGREN, C. *Split Trees, Cutings and Explosions*. PhD thesis, Department of Mathematics, Uppsala University, February 2010. In *Uppsala Dissertations in Mathematics*, vol. **67**.

[17] JACQUET, P., AND SZPANKOWSKI, W. Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees. *Theoretical Computer Science 144*, 1–2 (1995), 161–197.

[18] KNUTH, D. E. *The Art of Computer Programming*, 2nd ed., vol. 3: Sorting and Searching. Addison-Wesley, 1998.

[19] LAGARIAS, J. C. Best simultaneous Diophantine approximations I: Growth rates of best approximation denominators. *Transactions of the American Mathematical Society 272*, 2 (1982), 545–554.

[20] LAPIDUS, M. L., AND VAN FRANKENHUIJSEN, M. *Fractal Geometry, Complex Dimensions and Zeta Functions: Geometry and Spectra of Fractal Strings*. Springer, 2006.

[21] LONGO, G., Ed. *Multi–User Communication Systems*, vol. 265 of *CISM Courses and Lecture Notes*. Springer Verlag, 1981.

[22] LOUCHARD, G., AND SZPANKOWSKI, W. Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm. *IEEE Transactions on Information Theory 41*, 2 (Mar. 1995), 478–488.

[23] LOUCHARD, G., AND SZPANKOWSKI, W. On the average redundancy rate of the Lempel-Ziv code. *IEEE Transactions on Information Theory 43*, 1 (1997), 2–8.

[24] MAHMOUD, H. M. *Evolution of Random Search Trees*. John Wiley, 1992.

[25] MASSEY, J. L., Ed. *Special Issue on Random-Access Communications* (Mar. 1985), vol. IT-31 of *IEEE Transactions on Information Theory*.

[26] MOHAMED, H., AND ROBERT, P. A probabilistic analysis of some tree algorithms. *Annals of Applied Probability 15*, 4 (2005), 2445–2471.

[27] MOHAMED, H., AND ROBERT, P. Dynamic tree algorithms. *Annals of Applied Probability 20*, 1 (2010), 26–51.

[28] PITTEL, B. Asymptotical growth of a class of random trees. *Annals of Probability 13*, 2 (1985), 414–427.

[29] POLLICOTT, M. A complex Ruelle–Perron–Frobenius theorem and two counterexamples. *Ergodic Theory and Dynamical Systems 4* (1984), 135–146.

[30] RHIN, G. Approximants de Padé et mesures effectives d'irrationalité. In *Séminaire de théorie des nombres, Paris 1985-86* (1988), C. Goldstein, Ed., Birkhäuser, pp. 155–164.

[31] RHIN, G., AND TOFFIN, P. Approximants de Padé simultanés de logarithmes. *Journal of Number Theory 24*, 3 (1986), 284–287.

[32] SALIKHOV, V. K. On the irrationality measure of $\ln 3$. *Doklady Mathematics 76*, 3 (2007), 955–957. Translated from *Doklady Akademii Nauk* **417** (2007), no. 6, pp. 753–755.

[33] SCHACHINGER, W. Limiting distributions for the costs of partial match retrievals in multidimensional tries. *Random Structures & Algorithms 17*, 3–4 (2000), 428–459.

[34] SEDGEWICK, R. *Algorithms in C, Parts 1–4*, third ed. Addison–Wesley, Reading, Mass., 1998.

[35] SEDGEWICK, R., AND FLAJOLET, P. *An Introduction to the Analysis of Algorithms*. Addison-Wesley Publishing Company, 1996.

[36] SZPANKOWSKI, W. *Average-Case Analysis of Algorithms on Sequences*. John Wiley, 2001.

[37] VALLÉE, B. Dynamical sources in information theory: Fundamental intervals and word prefixes. *Algorithmica 29*, 1/2 (2001), 262–306.

[38] VALLÉE, B. Euclidean dynamics. *Discrete and Continuous Dynamical Systems 15*, 1 (2006), 281–352.

[39] VALLÉE, B., CLÉMENT, J., FILL, J. A., AND FLAJOLET, P. The number of symbol comparisons in QuickSort and QuickSelect. In *ICALP 2009, Part I* (2009), S. A. *et al.*, Ed., vol. 5555 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 750–763. Proceedings of the 36th International Colloquium on Automata, Languages and Programming.

[40] WHITTAKER, E. T., AND WATSON, G. N. *A Course of Modern Analysis*, fourth ed. Cambridge University Press, 1927. Reprinted 1973.

[41] WU, Q. On the linear independence measure of logarithms of rational numbers. *Mathematics of Computation 72* (2003), 901–911.

A Proofs relative to Section 2 (Geometry of poles)

A.1 Invariance of the irrationality exponent

We establish here the fact, asserted in Subsection 2.1, that the irrationality exponent is well-defined, independently of the ordering of the probabilities (p_j). We recall that $\alpha^{(k)}$, defined in (10), belongs to the hyperplane $\{u_k = 0\}$. For any integer q_k, the sequence of integers $(q_\ell)_{\ell \neq k}$ defined by $q_\ell := \lfloor q_k\ \alpha_{k,\ell} \rceil$ is called the sequence of numerators of $\alpha^{(k)}$ relative to denominator q_k.

Proposition 3 *Fix an aperiodic family of probabilities $(p_i)_{i\in[1..r]}$ and, for any $k \in [1..r]$, a norm $\|.\|_{(k)}$ defined on the hyperplane $\{u_k = 0\}$. Then*

- (i) *Consider for some $k \in [1\,..\,r]$, an integer q_k for which the inequalities $|\{q_k\alpha_{k,\ell}\}| \leq (w_\ell)/(4w_r)$ hold for any ℓ; associate the sequence $(q_\ell)_{\ell\neq k}$ of numerators of $\alpha^{(k)}$ relative to denominator q_k. Then, for any $\ell \in [1..r]$, the integers q_ℓ satisfy the inequalities $|\{q_\ell\alpha_{\ell,j}\}| \leq (w_j)/(2w_r)$ for any j, and the sequence $(q_j)_{j\neq\ell}$ is the sequence of numerators of $\alpha^{(\ell)}$ relative to denominator q_ℓ.*
- (ii) *There exists a constant c (that depends on the family of norms $\|\|_{(k)}$) such that any quadruple $(f_k, f_\ell, q_k, q_\ell)$, formed with the approximation function f_k of $\alpha^{(k)}$ relative to norm $\|\|_{(k)}$, the approximation function f_ℓ of $\alpha^{(\ell)}$ relative to norm $\|\|_{(\ell)}$, a BSAD q_k of $\alpha^{(k)}$ relative to $\|\|_{(k)}$, and a numerator q_ℓ of $\alpha^{(k)}$ relative to denominator q_k, satisfies $f_k(q_k) \leq cf_\ell(q_\ell)$.*
- (iii) *If there exists k for which $\alpha^{(k)}$ has a finite irrationality exponent equal to μ, then, any $\alpha^{(\ell)}$ has the same finite irrationality exponent.*

Proof: Here, it proves convenient to operate with the norm $\|.\|_\star$ defined by $\|u\|_\star = \sup \frac{|u_\ell|}{w_\ell}$.
Assume that q_1 is a good enough approximation denominator of $\alpha^{(1)}$, so that $\{q_1\alpha^{(1)}\}$ is small enough, namely $\|\{q_1\alpha^{(1)}\}\|_\star \leq \eta$, with η to be fixed later. Denote by $(q_1, q_2, \ldots, q_r)$ the numerators of the approximation. For any pair (k, ℓ) of indices, let

$$\delta(k,\ell) := \left|\frac{q_k}{w_k} - \frac{q_\ell}{w_\ell}\right|, \qquad A(k,\ell) := \left|q_k\frac{w_\ell}{w_k} - q_\ell\right|.$$

(i) Since $q_1, \ldots, q_r$ are the numerators of the approximation of $\alpha^{(1)}$, the following equalities hold

$$A(1,\ell) = |\{q_1\alpha_{1,\ell}\}|, \qquad \|\{q_1\alpha^{(1)}\}\|_\star = \sup \delta(1,i).$$

On the other hand, the relations between the $\delta(k,\ell)$ and the $A(k,\ell)$ imply

$$\frac{A(k,\ell)}{w_\ell} = \delta(k,\ell) \leq \delta(1,k) + \delta(1,\ell) \leq 2\|\{q_1\alpha^{(1)}\}\|_\star$$

Then, if $\eta \leq w/4$, with $w := 1/w_r$, then $A(k,\ell)$ equals $|\{q_k\alpha_{k,\ell}\}|$. This proves finally, for any $k \in [1\,..\,r]$, the inequalities $\|\{q_k\alpha^{(k)}\}\|_\star \leq 2\|\{q_1\alpha^{(1)}\}\|_\star$.

(ii) Since all norms are equivalent, the last bound yields the second Assertion of the statement.

(iii) Suppose now that q_1 is a BSAD for $\alpha^{(1)}$ relative to the norm $\|.\|_\star$, with $\|\{q_1\alpha^{(1)}\}\|_\star \leq \eta$. For the pair (f_k, f_1) of approximation functions, relative to the norm $\|.\|_\star$, f_k for $\alpha^{(k)}$, f_1 for $\alpha^{(1)}$, the relations

$$\frac{1}{f_k(q_k)} \leq \|\{q_k\alpha^{(k)}\}\|_\star \leq 2\|\{q_1\alpha^{(1)}\}\|_\star = \frac{2}{f_1(q_1)}$$

entail the inequality $f_1(q_1) \leq 2f_k(q_k)$, for any BSAD q_1 of $\alpha^{(1)}$.

This is true in particular when q_k is a BSAD for $\alpha^{(k)}$. If now $\alpha^{(k)}$ has an irrationality exponent equal to $\mu(\alpha^{(k)})$, then, for any $\nu > \mu$, we have

$$f_1(q_1) \leq A_\nu q_k^{\nu-1} \leq A_\nu \left(\frac{q_k}{q_1}\right)^{\nu-1} q_1^{\nu-1} \leq B_\nu q_1^{\nu-1}$$

This last inequality implies that the irrationality exponent $\mu(\alpha^{(1)})$ is at least equal to $\mu(\alpha^{(k)})$. Consequently, all the irrationality exponents are equal. □

A.2 Proofs relative to ladders and poles (§2.3)

Proof: [of Lemma 1]. We first consider a family of intervals of the real line $\mathcal{I}(a_i, \epsilon)$ with centers a_i (for $1 \leq i \leq n$) and the same radius ϵ and use the fact that this family has a nonempty intersection if and only if any intersection of two intervals of the family is nonempty. (Order the centers a_i in increasing order, and observe that, as the inequality $|a_1 - a_n| < 2\epsilon$ holds, the middle a of the interval $[a_1, a_n]$ satisfies $|a_i - a| < \epsilon$ for any i, with $1 \leq i \leq n$.) We then apply this property to the vertical intervals that define a step of the joint ladder: this proves that the step of the ladder relative to the sequence of integers $(q_1, q_2, \ldots, q_r)$ is nonempty if and only if, for any pair (k, ℓ),

$$\mathcal{R}_k(q_k, \delta, \epsilon) \cap \mathcal{R}_\ell(q_\ell, \delta, \epsilon) \neq \emptyset. \tag{41}$$

Then, for any pair (k, ℓ) of indices, the following three assertions are equivalent to Assertion (41),

$$2\pi \left| \frac{q_k}{w_k} - \frac{q_\ell}{w_\ell} \right| \leq 2\epsilon w, \qquad \left| q_k \frac{w_\ell}{w_k} - q_\ell \right| \leq \frac{\epsilon}{\pi} w_\ell w, \qquad |\{q_k \alpha_{k,\ell}\}| \leq \frac{\epsilon}{\pi} w_\ell w,$$

since $0 < \epsilon < \pi/2$. □

Proof: [of Lemma 2]. First, for any ϵ with $0 < \epsilon < \pi/2$, the continuity of the map $\sigma \mapsto \cos(\epsilon)\, p_k^\sigma + \sum_{j \neq k} p_j^\sigma$ together with the equality

$$\cos(\epsilon)\, p_k + \sum_{j \neq k} p_j = 1 - [1 - \cos(\epsilon)] p_k < 1$$

imply the following property: *Denote by* $\alpha_- = \min \alpha_{j,k} = \alpha_{r,1}$. *For any* k, *and any* ϵ *with* $0 < \epsilon < \pi/2$, *there exists* $\delta_k(\epsilon) > 0$ *such that*

$$\cos(\epsilon\alpha_-)\, p_k^\sigma + \sum_{j \neq k} p_j^\sigma < 1 \qquad \text{for any } \sigma > 1 - \delta_k(\epsilon). \tag{42}$$

Assume now (by contradiction) that there exists $\epsilon > 0$, such that, for any $\delta > 0$, there is an index $k \in [1 \,..\, r]$ and a root $s = \sigma + it$ of $\lambda(s) = 1$ that satisfy

$$1 - \delta < \sigma < 1, \qquad \text{and} \quad \forall q \in \mathbb{Z}, \quad \left| t - 2q \frac{\pi}{w_k} \right| > \epsilon w.$$

This implies the inequality $\cos(t w_k) < \cos(\epsilon\alpha_-)$, and, by Equation 42, the existence of $\delta := \min \delta_k(\epsilon)$ for which, for any $\sigma > 1 - \delta$,

$$1 = \Re\lambda(s) < \sum_{j \neq k} p_j^\sigma + p_k^\sigma \cos(\epsilon\alpha_-) < 1, \tag{43}$$

which provides a contradiction. □

Proof: [of Lemma 3]. Consider an element $s = \sigma + it$ which belongs to the joint ladder $\mathcal{L}(\delta, \epsilon)$. Then, the inequalities $\cos(tw_j) \geq \cos(\epsilon w w_j) \geq \cos\epsilon$ and $p_j^\sigma > p_j$ hold and imply

$$|\Re(\lambda'(s))| = \sum_{j=1}^{r} w_j p_j^\sigma \cos(tw_j) > (\cos\epsilon) \left[\sum_{j=1}^{r} w_j p_j\right] = (\cos\epsilon)|\lambda'(1)|.$$

We have proved: *For any $\epsilon > 0, \delta > 0$, the derivative $\lambda'(s)$ at any point of the ladder $\mathcal{L}(\delta, \epsilon)$ satisfies*

$$|\Re(\lambda'(s))| > (\cos\epsilon)|\lambda'(1)|. \tag{44}$$

Consider now two roots s and $s + h$ of the equation $\lambda(s) = 1$, located in the vertical strip $\mathcal{B}(\delta)$. Then, the Taylor expansion of λ at $s + h$ yields the inequality

$$|\lambda(s + h) - \lambda(s) - h\lambda'(s)| = |h\lambda'(s)| \leq \frac{|h|^2}{2} \sup\{|\lambda''(z);\ \ z \in \mathcal{B}(\delta)\} \tag{45}$$

If $z \in \mathcal{B}(\delta)$, the inequality $|\lambda''(z)| \leq \lambda''(1 - \delta)$ holds, and, then, by Lemma 2 supplemented by Inequalities (44) and (45), the distance $|h|$ between the two roots satisfies

$$|h| \geq (2\cos\epsilon)\frac{\lambda'(1)}{\lambda''(1 - \delta)}.$$

Since the lower bound tends to $2|\lambda'(1)|/\lambda''(1)$ when ϵ (thus δ) tends to 0, it is possible to choose ϵ (thus δ), so that the inequality

$$\left[\delta^2 + 4\epsilon^2 w^2\right]^{1/2} < (2\cos\epsilon)\frac{|\lambda'(1)|}{\lambda''(1 - \delta)} \tag{46}$$

holds. In this case, each rectangle $\mathcal{R}_k(q, \delta, \epsilon)$ contains at most one element of $\mathcal{Z}$. □

A.3 Proofs relative to the Δ–function (§2.4)

Proof: [of Lemma 4]. We take a compatible pair (ϵ, δ), as granted by Lemma 2. The only possible element z of $\mathcal{Z}$ contained in a (nonempty) step of the joint ladder, relative to the sequence $(q_1, q_2, \ldots, q_r)$, is written, for any $k \in [1\,..\,r]$, as $z = 2i\pi q_k/w_k + s_k$, where $s := s_k$ belongs to $\mathcal{R}(\delta, \epsilon)$, and it satisfies

$$p_k^{s_k} + \sum_{\ell \neq k} p_\ell^{s_k} \exp\left(-2\pi i q_k \alpha_{k,\ell}\right) = \sum_{\ell=1}^{r} p_\ell^{s_k} \exp\left(-2\pi i \{q_k \alpha_{k,\ell}\}\right) = 1. \tag{47}$$

The main idea [20, §3.5] is now to consider the more general equation, $\lambda(s, u) = 1$, for $u = (u_1, u_2, \ldots, u_r) \in \mathbb{C}^r$, where $\lambda(s, u)$ is as defined by (24). Each solution $z \in \mathcal{Z}$ belonging to the step of label $(q_1, q_2, \ldots, q_r)$ of the ladder $\mathcal{L}(\delta, \epsilon)$ leads to r solutions of the form $(s_k, \{q_k \alpha^{(k)}\})$, which satisfy, by Lemma 1,

$$\forall \ell, k, \quad \Re s_k = \Re s_\ell, \qquad \Im(s_k - s_\ell) = 2\pi\left(\frac{q_\ell}{w_\ell} - \frac{q_k}{w_k}\right) \qquad \{q_k \alpha^{(k)}\} \in \mathcal{U}(\epsilon),$$

where $\mathcal{U}(\epsilon)$ is defined in Lemma 1. The pair $(1, \underline{0})$, together with all the pairs $(s_k, \{q_k \alpha^{(k)}\})$, belong to the set

$$\underline{\mathcal{Z}}(\epsilon, \delta) := \{(s, u) \in \Sigma(\delta, \epsilon); \quad \lambda(s, u) = 1\}, \qquad \text{with} \quad \Sigma(\delta, \epsilon) = \mathcal{R}(\delta, \epsilon) \times \mathcal{U}(\epsilon).$$

In order to describe the set $\underline{\mathcal{Z}}(\epsilon, \delta)$, we apply the Implicit Function Theorem to the equation $\lambda(s, u) = 1$ near the point $(1, \underline{0})$. This is possible since the function $\lambda : \Sigma(\delta, \epsilon) \to \mathbb{C}$ is of class $\mathcal{C}^1$ (it is indeed analytic), and the derivative

$\partial\lambda/\partial s(1,\underline{0})$, which is equal to $\lambda'(1)$, is nonzero. This implies the existence of a neighborhood $\mathcal{V} \subset \mathcal{U}(\epsilon)$ of $\underline{0}$ in $\mathbb{R}^r$, a complex neighborhood $\mathcal{W} \subset \mathcal{R}(\delta,\epsilon)$ of 1, and a function Δ defined on $\mathcal{V}$ real-analytic such that, any solution (s,u) of the equation $\lambda(s,u)=1$ with $(s,u) \in \mathcal{W}\times\mathcal{V}$ can be written as $s=\Delta(u)$. This means:

$$(\mathcal{W}\times\mathcal{V})\bigcap \underline{\mathcal{Z}}(\epsilon,\delta) = \{(s,u)\in\mathcal{W}\times\mathcal{V};\ \ s=\Delta(u)\}. \tag{48}$$

Choose first $\epsilon' < \epsilon, \delta'' < \delta$ such that the two inclusions $\mathcal{U}(\epsilon') \subset \mathcal{V}$, $\mathcal{R}(\delta'',\epsilon') \subset \mathcal{W}$ hold. Then, the choice $\delta' = \min(\delta(\epsilon'),\delta'')$ leads to a compatible pair (ϵ',δ'), and the equality (48) implies

$$\underline{\mathcal{Z}}(\epsilon',\delta') = \left\{(s,u)\in\mathcal{R}(\delta',\epsilon')\times\mathcal{U}(\epsilon');\ \ s=\Delta(u)\right\}.$$

We can now return to the set $\mathcal{Z}\cap\mathcal{B}(\delta')$, via all the special elements of $\underline{\mathcal{Z}}(\epsilon',\delta')$ of the form $(s_k,\{q_k\alpha^{(k)}\})$.

We have proved that the only possible element of $\mathcal{Z}$ in a nonempty step of label $(q_1,q_2\ldots,q_r)$ of the ladder $\mathcal{L}(\delta,\epsilon)$ can be written as in (23). This element does exist as soon as the quantities $\Delta(\{q_k\alpha^{(k)}\})$ lie in $\mathcal{R}(\delta,\epsilon)$. □

Proof: [of Lemma 5]. The first derivative $\Delta'(\underline{0})$ (which is nonzero, by the Implicit Function Theorem) and the second derivative $\Delta''(\underline{0})$ of Δ at $\underline{0}$ are easily computed. Furthermore, the function Δ can be extended to a complex neighborhood of $\underline{0}$, on which it is an analytic function of complex numbers $(z_1,z_2,\ldots z_r)$. Near $\underline{0}$, the values $\Delta(iu)$ are real for $u\in\mathbb{R}^r$. This implies that the components of $\Delta'(\underline{0})$ belong to $i\mathbb{R}$, whereas all the coefficients of the Hessian matrix of $\Delta''(\underline{0})$ are real,

$$\frac{1}{4\pi^2}\Delta''(\underline{0})_{j,k} = \left(\frac{\lambda''(1)}{\lambda'(1)^3}+\frac{w_j+w_k}{\lambda'(1)^2}\right)p_jp_k + p_k\delta_{j,k}\left(\frac{1}{\lambda'(1)}\right) \tag{49}$$

(here, $\delta_{j,k}$ is the Kronecker symbol) and define the quadratic form

$$-T(u) := \frac{1}{2}\,{}^t u\Delta''(\underline{0})u \tag{50}$$

which is real for real r-uples. (Generally, the coefficients of the derivative of order k are real for even k, and purely imaginary for odd k.) The Taylor expansion of Δ at $u\in\mathcal{U}(\epsilon)$,

$$\Delta(u) = 1 + u\cdot\Delta'(\underline{0}) + \frac{1}{2}\,{}^t u\Delta''(\underline{0})u + O(\|u\|^3) \tag{51}$$

entails the two relations, for any $u\in\mathcal{U}(\epsilon)$,

$$S(u) := 1-\Re\Delta(u) = T(u)+O(\|u\|^4), \qquad \Im\Delta(u) = -iu\cdot\Delta'(\underline{0})+O(\|u\|^3).$$

Observe that the second component $\{q_k\alpha^{(k)}\}$ of the solutions $(s_k,\{q_k\alpha^{(k)}\})$ of interest belongs to the union of hyperplanes of the form $\{u_\ell=0\}$, and consider the restriction of Δ to

$$\widetilde{\mathcal{U}}(\epsilon) := \mathcal{U}(\epsilon)\ \cap\ \bigcup\{u_\ell=0\}. \tag{52}$$

We now prove that the restriction of the quadratic form $u\mapsto T(u)$ to each hyperplane $\{u_k=0\}$ is (real) positive definite . We use a more general result:

Lemma 7 *Any point $(s,u)\in\underline{\mathcal{Z}}(\epsilon,\delta)$ satisfies $\Re s\le 1$. The only point (s,u) of $\underline{\mathcal{Z}}(\epsilon,\delta)$ for which $\Re s=1$ and $u_k=0$ for some $k\in[1\,..\,r]$ is the point $(1,\underline{0})$.*

Proof: Note that all the points $(1+it, u)$ defined by the equations $2\pi u_k = -tw_k$ are points of $\underline{\mathcal{Z}}(\epsilon, \delta)$ which satisfy $\Re s = 1$ with $u \neq 0$. In fact, for any element (s, u) of $\underline{\mathcal{Z}}(\epsilon, \delta)$, the equality

$$\Re\lambda(s, u) = \sum_{k=1}^{r} p_k^{\sigma} \cos(tw_k + 2\pi u_k) = 1 \tag{53}$$

holds and implies that $\Re s \leq 1$ for any (s, u) in $\underline{\mathcal{Z}}(\epsilon, \delta)$. Consider now a point $(s, u) \in \underline{\mathcal{Z}}(\epsilon, \delta)$ with $\sigma = 1$ and a component $u_k = 0$, for some $k \in [1 \,.\,.\, r]$. For this index k, the equality $\cos(tw_k + 2\pi u_k) = 1$ implies the equality $t = 0$, and thus, the other equalities $\cos(tw_\ell + 2\pi u_\ell) = 1$, for $\ell \neq k$ imply the equalities $u_\ell = 0$. □

To conclude, Lemma 7 implies that the restriction T_ℓ of the quadratic form $u \mapsto T(u)$ to each hyperplane $\{u_\ell = 0\}$ is positive definite. Denote by $\|u\|_{T_\ell}$ the norm relative to the positive definite form T_ℓ. Then, for any $\eta > 0$, there exists $\epsilon > 0$ such that the inequalities (25) hold for any u in $\widetilde{\mathcal{U}}(\epsilon)$ defined in (52). □

B Numerical aspects

Here is a campaign of experiments conducted with $p_1 = 1/3$, $p_2 = 2/3$, corresponding to an entropy $H \doteq 0.636514168$. We consider the poissonized mean trie-size $S(x)$ of Section 1 and examine the *correction function*

$$\kappa_1(x) = S(x) - \left(\frac{x}{H} - 1\right),$$

which measures the *error* committed when approximating $S(x)$ by the sum of its main asymptotic term (n/H) and the contribution (-1) induced by the pole at 0 of the Mellin transform $S^\star(s)$. Thus $\kappa_1(x) = x\Phi(x) + O(x^{1-a})$ in the notations of Section 3

For moderately large values of x, this correction seems to *decay* steadily to 0, as suggested by the plots of $\kappa_1(x)$ in Figure 8 (left, centre).

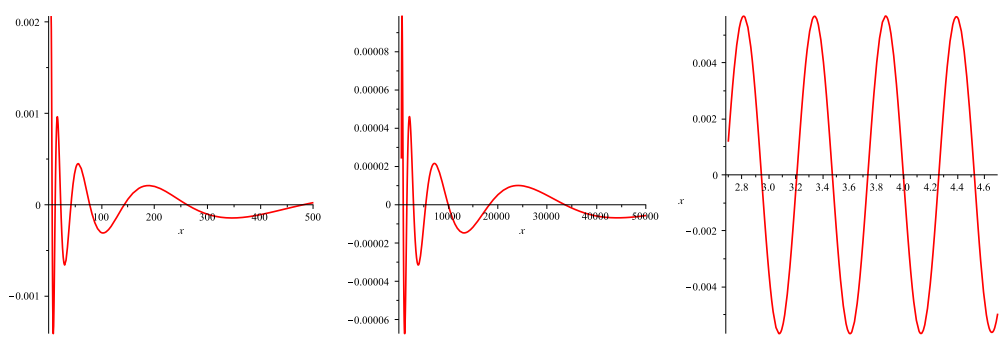


Fig. 8: The correction function $\kappa_1(x)$. *Left, centre*: plot of $\kappa_1(x)$, for $x \in [5, 5\cdot 10^2]$ (where $|\kappa_1(x)| < 2\cdot 10^{-3}$) and $x \in [5\cdot 10^2, 5\cdot 10^4]$ (where $|\kappa_1(x)| < 10^{-4}$). *Right*: plot of $\kappa_1(x)\cdot x^{0.62686}$ against $\log_{10} x$, for $5\cdot 10^2 \leq x \leq 5\cdot 10^4$.

This smallness of $\kappa_1(x)$ continues further and, for instance, the values of $\kappa_1(x)$ appear to stay in the range $[-10^{-5}, 10^{-5}]$ for all x with $10^4 \leq x \leq 10^7$. Indeed, the behaviour of $\kappa_1(x)$ in this whole range is remarkably well accounted for by a term of the form $x^{-0.62686}$ (modulated by oscillations), as is illustrated by Figure 8 (right). Here, the exponent 0.62686 is dictated by the two "low" singularities of $\Lambda(s)$ at $s = 0.62686 \pm 5.17927i$

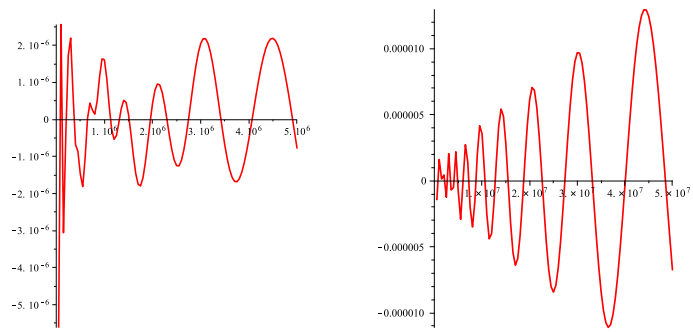


Fig. 9: The correction function $\kappa_1(x)$. *Left:* for x near 10^6. *Right:* for x in $x \in [5 \cdot 10^5, 5 \cdot 10^7]$.

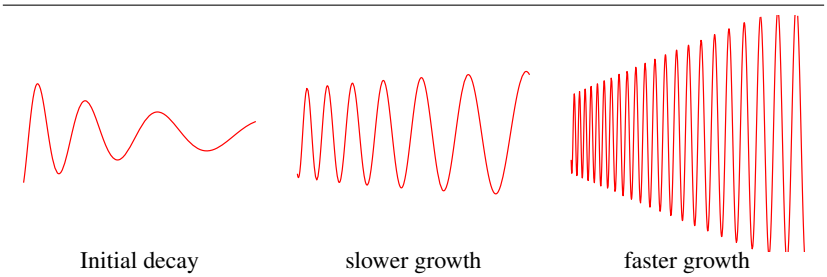
and the corresponding coefficient is small since $|\Gamma(s)| \doteq 9 \cdot 10^{-4}$. The regularity and the smooth periodic aspect of the curve are striking.

As a matter of fact, in spite of the former numerical data, our analytic results predict that the function $\kappa_1(x)$ *must tend to infinity*, rather than decaying to 0, and it is even the case that $\kappa_1(x)$ should exceed infinitely often *any* function $x^{1-\epsilon}$, for *any* $\epsilon > 0$. This is illustrated by the two diagrams of Figure 9: the one on the left exemplifies a first "*phase transition*" region near $x = 10^6$; the one on the right corresponding to $x \in [5 \cdot 10^5, 5 \cdot 10^7]$ shows the beginning of the region where $\kappa_1(x)$ actually starts to *grow*:

For higher values of x, the increase becomes more and more pronounced, as evidenced by the following table:

x :	10^{10}	10^{11}	10^{12}	10^{13}	10^{14}
$\kappa_1(x)$:	$6.0 \cdot 10^{-5}$	$-9.8 \cdot 10^{-4}$	$-1.2 \cdot 10^{-2}$	$-1.0 \cdot 10^{-1}$	$-7.4 \cdot 10^{-1}$

It can be seen that, amongst the poles of $\Lambda(s)$, only the ones nearest to $\Re(s) = 1$ matter: these are "record" poles in the sense that there is no other pole in their south-east quadrant. Then, a term $T_j(x)$ in $x\Phi(x)$ that corresponds to such a pole χ_j has a certain *region of prevalence*, something like $[x_j, y_j]$, for at least doubly exponentially growing sequences (x_j), (y_j) (this last fact that can be deduced from general results of Lagarias [19]). Under mild additional assumptions (namely, that the intervals $[x_\ell, y_\ell]$ be of "limited overlap"), in a region $[x_j, y_j]$, the term $T_j(x)$ dominates in amplitude the earlier $T_k(x)$ (those with $k < j$), since these have too small an exponent of x; it also dominates the later $T_m(x)$ (those with $m > j$), since these are still too much tamed by the smallness of their Gamma factor. Here is then a simplified depiction of the succession of regimes of $\kappa_1(x)$, as x increases:



Initial decay	slower growth	faster growth
$[1, y_1]$	$[x_{j-1}, y_{j-1}]$	$[x_j, y_j]$
	lower frequency	higher frequency.

Note. A somewhat related phenomenon, but one that involves the complex Riemann zeros, is discussed in [14] in relation with the size of tries built on continued fraction representations of uniform $[0, 1]$ real numbers.