INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Dynamical Sources in Information Theory: A General Analysis of Trie Structures

Julien Clément, Philippe Flajolet, Brigitte Vallée

N° 3645
Mars 1999

THÈME 2

Rapport de recherche

ISSN 0249-6399

Dynamical Sources in Information Theory: A General Analysis of Trie Structures

Julien Clément, Philippe Flajolet, Brigitte Vallée

Thème 2 — Génie logiciel
et calcul symbolique
Projet Algo

Rapport de recherche n° 3645 — Mars 1999 — 63 pages

Abstract: Digital trees, also known as tries, are a general purpose flexible data structure that implements dictionaries built on sets of words. An analysis is given of three major representations of tries in the form of array-tries, list tries, and bst-tries ("ternary search tries"). The size and the search costs of the corresponding representations are analysed precisely in the average case, while a complete distributional analysis of height of tries is given. The unifying data model used is that of dynamical sources and it encompasses classical models like those of memoryless sources with independent symbols, of finite Markov chains, and of nonuniform densities. The probabilistic behaviour of the main parameters, namely size, path length, or height, appears to be determined by two intrinsic characteristics of the source: the entropy and the probability of letter coincidence. These characteristics are themselves related in a natural way to spectral properties of specific transfer operators of the Ruelle type.

Key-words: Information theory, dynamical sources, analysis of algorithms, digital trees, tries, ternary search tries, transfer operator, continued fractions.

(Résumé : tsvp)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

Sources dynamiques en théorie de l'information: une analyse générale des arbres digitaux

Résumé : Les arbres digitaux, également connus sous le nom de "tries" sont une structure de donée générique et flexible qui permet d'implanter des dictionnaires construits sur des ensembles de mots. Nous donnons une analyse de troies représentations principales de ces arbres, les arbres-tableaux, les arbres-listes, et les arbres ternaires de recherche. La taille et les coûts de recherche de ces représentations sont analysés précisément en moyenne, tandis qu'une analyse en distribution de la hauteur est obtenue. Le modèle unificateur d'analyse est celui des "sources dynamiques", lesquelles recouvrent les modèles classiques comme les sources sans mémoire (à symboles indépendants), les chaines de Markov finies, et les densités initiales non uniformes. Les propriétés probabilistes des principaux paramètres de taille, longueur de cheminement et hauteur apparaissent liées à deux caractéristiques fondamentales de la source: l'entropie et la probabilité de coincidence. Ces caractéristiques se trouvent elle-mêmems reliées aux propriétés spectrales d'opérateurs de transfert du type introduit par Ruelle.

Mots-clé : Théorie de l'information, sources dynamiques, analyse d'algorithmes, arbres digitaux, tries, arbres ternaires de recherche, operateur de transfert, fractions continues.

DYNAMICAL SOURCES IN INFORMATION THEORY: A GENERAL ANALYSIS OF TRIE STRUCTURES

JULIEN CLÉMENT[1,2], PHILIPPE FLAJOLET[2], BRIGITTE VALLÉE[1,2,*]

[1] GREYC, Université de Caen, F-14032 Caen (France)
[2] Algorithms Project, INRIA-Rocquencourt, F-78150 Le Chesnay (France)
— February 27, 1999 —

Abstract. Digital trees, also known as tries, are a general purpose flexible data structure that implements dictionaries built on sets of words. An analysis is given of three major representations of tries in the form of array-tries, list tries, and bst-tries ("ternary search tries"). The size and the search costs of the corresponding representations are analysed precisely in the average case, while a complete distributional analysis of height of tries is given. The unifying data model used is that of dynamical sources and it encompasses classical models like those of memoryless sources with independent symbols, of finite Markov chains, and of nonuniform densities. The probabilistic behaviour of the main parameters, namely size, path length, or height, appears to be determined by two intrinsic characteristics of the source: the entropy and the probability of letter coincidence. These characteristics are themselves related in a natural way to spectral properties of specific transfer operators of the Ruelle type.

Introduction

Tries. Digital trees, usually called *tries*, are both an abstract structure and a data structure that can be superimposed on a set of words produced by some source. As an abstract structure, tries are based on a splitting according to symbols encountered in words. Consider a fixed alphabet $\mathcal{M} = \{a_1, \ldots, a_r\}$, and let $Y \subset \mathcal{M}^\infty$ be any finite set of infinite words over $\mathcal{M}$. The trie associated to Y is defined recursively by the rule,

$$\mathrm{trie}(Y) = \langle \mathrm{trie}(Y \setminus a_1), \ldots, \mathrm{trie}(Y \setminus a_r) \rangle,$$

where $Y \setminus \alpha$ means the subset of Y consisting of strings that start with α stripped of their initial symbol α, with recursion being halted as soon as Y contains less than two elements. The advantage of the trie is that it only maintains the minimal prefix set of characters that is necessary to distinguish all the elements of Y. In their abstract versions, tries are thus essentially equivalent to prefix trees in the theory of variable length coding.

Clearly the tree $\mathrm{trie}(Y)$ supports the search of any word w in the set Y by following an access path dictated by the successive symbols of w. Similarly, it may be used to implement insertions and deletions, so that it is a fully dynamic dictionary data type. In addition, tries efficiently support set-theoretic operations like union and intersection [57], as well as partial match queries or interval search [47], while suitable

*Corresponding author. E-mail address: `Brigitte.Vallee@info.unicaen.fr`.

adaptations make them a method of choice for complex text processing tasks [22, Ch. 7]. This variety of applications justifies considering the trie structure as one of the central general-purpose data structures of computer science [22, 34, 39, 51].

When it comes to implementation, several options are possible depending on the decision structure chosen to guide descent into subtrees at each node. Three major choices present themselves:

— The "array-trie" uses an *array* of pointers to access subtrees directly.
— The "list-trie" relies upon (sorted) *linked list* traversals.
— The "bst-trie" uses *binary search trees* (bst) as subtree access method.

The array-trie thus constitutes a direct realization of the abstract trie structure; we occasionally employ the term of "standard trie" to refer to situations where parameters are shared by the abstract trie and the array trie (this is for instance the case for the height, size, and path length parameters). The list-trie and the bst-trie combine the abstract trie structure with a subtree access method and are globally referred to as "hybrid tries".

Our original motivation for considering hybrid trie structures came from a recent paper of Bentley and Sedgewick [3] who, following early ideas of Clampett [6], developed an elegant implementation of bst-tries, under the name of *ternary search trie*. The basic idea of [3, 6] is to represent the bst-trie as a ternary tree where search on symbols is conducted like in a standard binary search tree over the alphabet set $\mathcal{M}$, while trie descent is performed by following an escape pointer whenever equality of symbols is detected. This structure is briefly discussed in the recent edition of Knuth's treatise [34, p. 512] and complete code is detailed in the latest edition of Sedgewick's book [51, Sec. 15.4]. The code is especially compact and, in simulations, the implementation constants appear to be particularly small. Bentley and Sedgewick report that, in practical situations, their data structure can be more efficient than hashing while offering considerably wider functionality. Our goal, as analysts, is to examine this claim and precisely quantify what goes on.

For each implementation, we analyse the parameters of path length and size (i.e., number of nodes). Path length determines search costs while size characterizes the memory requirements of the data structure. In addition, we also analyse the height of the abstract trie, which provides a valuable measure of extremal search performance. The analysis is then expected to provide useful guidance as to the choice of representation that is suitable for any particular application.

Sources. In information theory contexts, the two simpler models of sources are memoryless sources, where symbols in words are each emitted independently of the previous ones, and Markov chains, where the probability of emitting a symbol depends solely on a bounded part of the past history. However, data on which tries are built often arise from sources that may involve intricate dependencies. Our analyses are carried out within a general framework related to dynamical systems theory that encompasses memoryless and Markov sources as well as nonuniform density models. This model of probabilistic dynamical sources has been introduced and thoroughly developed by Vallée in [61], and it can describe nonmarkovian phenomena where, precisely, the dependency on past history is unbounded. A high level of generality is thus attained by the model.

A probabilistic dynamical source is defined by two objects: a symbolic mechanism and a density.

— The mechanism associates to a real number x of the $[0, 1]$ interval an infinite word $M(x)$ over the alphabet $\mathcal{M}$. (The approach is obviously related to symbolic dynamics.) Such a mechanism generalizes numerations system, the binary expansion of a real x or the continued fraction expansion of the real x being well-known instances.
— A probability density f over the $[0, 1]$ interval allows values drawn over that interval to be nonuniform.

Previous authors have studied separately the effect of nonuniform densities and of a few specific mechanisms. With the notable exception of [11], all analyses carried out so far have been relative to a uniform density, $f(x) \equiv 1$. The first analyses were in the context of memoryless sources. For the additive parameters of size and path length, they were performed by De Bruijn and Knuth around 1965 and reported in the first edition of [34] published in 1973.. Height was later analysed under Poisson and Bernoulli models in a series of papers [16, 20, 46, 63]. Asymmetric memoryless sources and Markov chain models were then treated systematically by Pittel, as well as Szpankowski and his collaborators: see for instance [26, 27, 43, 54, 55]. Devroye [11] has been the first to consider the effect on tries of a nonuniform density but only in conjunction with standard binary representations. In this paper, we allow concomitantly a nonuniform density *and* a mechanism that is capable of modelling probabilistic bias on individual digits.

Operator methods. The methods used in earlier works to analyse standard tries under simpler source models are rather diverse and range from probabilistic to analytic. We feel that the present paper unifies most of what was known before regarding average-case analysis of basic parameters like size and path length, as well as distributional analysis of height, while extending at the same time the analysis to hybrid tries like list-tries and bst-tries. As we shall see, all the estimates appear to involve two crucial characteristics of the source: the entropy and the probability of symbol coincidence. The first quantity intervenes in the expectation of size and path length, the second one in the expectation of height. For memoryless sources, generating functions are classically used in conjunction with Mellin transforms. However, when the source has memory, classical generating function techniques are no longer immediately applicable; for instance, under the Markov chain model, Jacquet and Szpankowski [26] resort to a notion of alignment in conjunction with a suitable inclusion-exclusion principle.

In the present setting of probabilistic dynamical sources, the main tool is the *Ruelle transfer operator* [2, 49]. Classically, it is used as a "generating operator" since it can easily generate some of the objects that are essential in the analysis. Here, the analysis of tries involves the *prefixes* of the words: all the source words which start with the same prefix "come from" an interval of $[0, 1]$ that is called a fundamental interval. In this context, the classical Ruelle operator is no longer sufficient. In a previous paper on which the present study is largely based, Vallée [61] has already introduced a generalization of the Ruelle operator, based on a *secant construction*, that acts on functions of two variables and suitably generates fundamental intervals. In the case of hybrid tries, we need additionally to generate simultaneously several fundamental intervals. For this purpose, we devise further generalizations of the Ruelle operator, based on *multi-secant constructions*, acting on functions of three and four variables. As in their classical version, these generalized Ruelle operators depend on a complex parameter s and they make it possible to express intervening *Dirichlet series of pairs of fundamental intervals* that play a central rôle throughout the paper. The Mellin transform is also heavily used, and, although poles of the Dirichlet series cannot always be made explicit, they can at least be pretty well localized in the vicinity of a frontier line, so that precise asymptotic estimates of parameters are possible.

Furthermore, positivity properties of the (generalized) Ruelle operators (for real values of parameter s) entail the existence of dominant spectral objects. In particular, we prove the existence of the *dominant eigenvalue function* $\lambda(s)$ defined in the neighbourhood of the real axis. This function turns out to be ubiquitous in the analysis of trie parameters. The main intrinsic characteristics of the source, its *entropy* $h(\mathsf{S})$ and its *coincidence probability* $c(\mathsf{S})$, are proven to be independent of the initial density f on the unit interval, being dependent only on the mechanism of the source through the relations

$$h(\mathsf{S}) = -\lambda'(1) \qquad c(\mathsf{S}) = \lambda(2).$$

Plan of the paper and results. Section 1 describes the general framework of dynamical sources, and defines fundamental intervals, as well as the two basic characteristics of the source, the entropy and the coincidence probability. Sections 2 and 3 are devoted to the specifications of hybrid tries and to the basic algebraic part of the analysis. All analyses reduce to a few Dirichlet series of fundamental measures. Section 4 introduces the generalized Ruelle operators and shows how they generate the Dirichlet series needed. In Section 5, we then transfer the properties of these operators to various Dirichlet series, and relate dominant spectral objects of the Ruelle operator to the entropy and the coincidence probability. In Sections 6 and 7, we come back to average-case analysis, and obtain precise estimates of size, path lengths, and height. The last section, Section 8 concludes with examples that include memoryless sources, Markov chain models, continued fraction representations, as well as nonuniform initial densities.

The following results are established for a random trie built on n items.

(i) The height of a standard trie has expectation of order $\log n$ and its probability distribution is asymptotically of the doubly exponential type with sharp tail decay,

$$\mathrm{E}[h_n] \sim \frac{2}{|\log c(\mathsf{S})|} \log n$$
$$\lim_{n\to\infty} \sup_{k\geq 0} \left|\Pr\{h_n \leq k\} - \exp\left[-\rho c(\mathsf{S})^k n^2\right]\right| = 0,$$

where ρ is a positive constant relative to source S and initial density f.

(ii) The average size of the trie is, up to possible small fluctuations, well approximated by a quantity of order n

$$S(n) \approx \frac{1}{h(\mathsf{S})}\, n.$$

This result in particular generalizes and improves some of the results obtained by Devroye [11] obtained under $\mathcal{L}^2$–conditions on the density f.

(*iii*) The average path length depends on the hybrid implementation under consideration. Let $\langle A\rangle$ represent an array-trie, $\langle L\rangle$ a list-trie, and $\langle B\rangle$ a bst-trie. For any source built on a finite alphabet, the average path lengths are invariably of order $n\log n$,

$$P_A(n)\sim\frac{1}{h(\mathsf{S})}n\log n,\qquad P_L(n)\sim\frac{K_L(\mathsf{S})}{h(\mathsf{S})}n\log n,\qquad P_B(n)\sim\frac{K_B(\mathsf{S})}{h(\mathsf{S})}n\log n,$$

with explicit constants $K_L(\mathsf{S}), K_B(\mathsf{S})$ that only depend on the mechanism of the source.
For an infinite (denumerable) alphabet, the array-trie is not meaningful anymore as a data structure. In that case, the average path lengths relative to list-tries or bst-tries, may happen to be of different orders. (See Prop. 11; such is for instance the case for the continued fraction source discussed in Section 8.)

These results are first established under a Poisson model that is technically easier to deal with, then transferred to the fixed-size model, called the Bernoulli model, by means of depoissonization techniques. The paper is built around eight theorems that cover (in order!) all combinations

$$\left\{\begin{array}{c}\text{Algebraic analysis}\\ \text{Asymptotic analysis}\end{array}\right\}\left\{\begin{array}{c}\text{Poisson model}\\ \text{Bernoulli model}\end{array}\right\}\left\{\begin{array}{c}\text{Size \& path lengths}\\ \text{Height}\end{array}\right\}.$$

An extended abstract of some of these results restricted to uniform density models in the memoryless and Markov chain cases has been presented at SODA'98 [7]. We refer to this paper for a comparison between theoretical predictions and empirical studies based on large textual data. It is justified there that, in practice, bst-tries make trie searching about three times faster[1] than their binary counterparts. The present paper is a companion to the general study of dynamical sources presented in [61], from which we adapted freely parts of our Sections 1, 4, and 5. As regards methodology, it is also related to approaches followed by one of us in the analysis of gcd-like algorithms in computational number theory [60, 62].

1. Probabilistic dynamical sources

Here, we describe the general framework of probabilistic *dynamical sources*. First, we introduce *symbolic* dynamical sources, with two types of mechanism, either basic or Markovian. Such mechanisms are related to *dynamical systems* defined from expanding analytic maps of the unit interval. (The reader may wish to consult [2, 37, 49] as general background references.) Then, upon endowing the unit interval with some (analytic) density, we define the concept of a *probabilistic* dynamical source. Finally, we present the notion of fundamental intervals and fundamental measures, then introduce the two basic characteristics of the source, the entropy and the coincidence probability.

Fundamentals of dynamical sources are studied in [61] and related notions play an important rôle for the analysis of gcd-like algorithms in computational number theory [58, 60, 62].

[1] For instance, the present manuscript was efficiently spell-checked by a modified version of the `epelle` program based on an implementation of bst-tries by J. Clément [8].

1.1. **Basic symbolic dynamical sources.** In information theory contexts, a source is a mechanism which produces infinite words on some alphabet $\mathcal{M}$. We are first interested in sources that are associated to *basic dynamical systems*, where the mechanism is the same at each step. The reader who is unfamiliar with the theory of expanding maps and symbolic dynamics may wish to examine the definition that follows in the light of the usual binary number representation. In that case, the alphabet is $\mathcal{M} = \{0,1\}$, the shift is $T(x)$ is $\{2x\}$ ($\{\cdot\}$ is the fractional part), the partition is $\{\mathcal{I}_0, \mathcal{I}_1\} = \{(0,\frac{1}{2}), (\frac{1}{2},1)\}$, and the encoding mapping is $\sigma(\mathcal{I}_0) = 0$, $\sigma(\mathcal{I}_1) = 1$.

Definition 1 (Basic symbolic dynamical source)**.** *A* basic dynamical source *is defined by four elements:*

- (a) *an* alphabet $\mathcal{M}$ *included in* $\mathbb{N}$*, that is finite or denumerable;*
- (b) *a* topological partition *of* $\mathcal{I} :=]0,1[$ *with disjoint open intervals* $\mathcal{I}_m, m \in \mathcal{M}$*, i.e.,* $\overline{\mathcal{I}} = \bigcup_{m\in\mathcal{M}} \overline{\mathcal{I}}_m$
- (c) *an* encoding *mapping* σ *which is constant and equal to* m *on each* $\mathcal{I}_m$*;*
- (d) *a* shift *mapping* T *whose restriction to each* $\mathcal{I}_m$ *is a real analytic bijection from* $\mathcal{I}_m$ *to* $\mathcal{I}$*. Let* h_m *be the local inverse of* T *restricted to* $\mathcal{I}_m$ *and* $\mathcal{H}$ *be the set* $\mathcal{H} := \{h_m,\ m \in \mathcal{M}\}$*. It is assumed that there exists a fixed complex neighbourhood* $\mathcal{V}$ *of* $\overline{\mathcal{I}}$ *on which the set* $\mathcal{H}$ *satisfies the following:*
  - (d_1) *the mappings* h_m *extend to holomorphic maps on* $\mathcal{V}$*, mapping* $\mathcal{V}$ *strictly inside* $\mathcal{V}$ *(i.e.,* $h(\overline{\mathcal{V}}) \subset \mathcal{V}$*);*
  - (d_2) *the mappings* $|h'_m|$ *extend to holomorphic maps* $\widetilde{h}_m$ *on* $\mathcal{V}$ *and there exist numbers* $\delta_m < 1$ *for which* $0 < |\widetilde{h}_m(z)| \le \delta_m$ *for* $z \in \mathcal{V}$*;*
  - (d_3) *there exists some* $\gamma < 1$ *for which the series* $\sum_{m\in\mathcal{M}} {\delta_m}^s$ *converges on* $\mathrm{Re}(s) > \gamma$*.*

The words emitted by the source are then produced as follows: The mapping $T : \mathcal{I} \to \mathcal{I}$ (that is almost everywhere defined) is used for iterating the process, as a shift mapping; the mapping $\sigma : \mathcal{I} \to \mathcal{M}$ is used for encoding. The word $M(x)$ of $\mathcal{M}^\infty$ associated to a real $x \in \mathcal{I}$ is then formed with the symbols

$$M(x) := (M_1(x), M_2(x), \ldots, M_k(x), \ldots), \tag{1}$$

where the k-th component $M_k(x)$ of $M(x)$ is equal to $\sigma(T^{k-1}x)$. The number of branches of T equals the cardinality of the alphabet, and the alphabet is used for coding the distinct branches of T, denoted by $T_{[m]}$, or the distinct inverse branches of T denoted by h_m. Here, h_m is a bijection from $\mathcal{I}$ to $\mathcal{I}_m$, which coincides with the inverse of the restriction $T_{[m]}$ of T to $\mathcal{I}_m$. Note that the domain of definition of $T_{[m]}$ is $\mathcal{I}_m$.

Remarks. We call such a dynamical source *basic* because of the equalities $T(\mathcal{I}_m) = \mathcal{I}$. Elsewhere in the literature, it is only asked for a dynamical system that the image $T(\overline{\mathcal{I}}_m)$ be a union of some elements $\overline{\mathcal{I}}_j$ of the partition. The conditions (d_1) and (d_2) express that the inverse branches h_m are contractions, or equivalently that T is expansive. The condition (d_3) automatically holds for a finite alphabet (with $\gamma = -\infty$), so that it is only useful in the case of infinite alphabets. In fact, in order for our treatment to apply, it is sufficient that some fixed iterate of T should satisfy conditions (d). For instance, the shift mapping T associated to the continued fraction source does not fulfill conditions (d_1) and (d_2) but its second iterate T^2 does.

Memoryless sources. All memoryless sources can be described in the basic dynamical framework. A source is said to be memoryless when the random variables M_k are independent and follow the same law. The memoryless source associated to a probability system $P = (p_m)_{m\in\mathcal{M}}$ (either finite or denumerable) is the source where all the components M_k are independent and follow a multinomial law[2] of parameters $(p_m)_{m\in\mathcal{M}}$. The corresponding topological partition of $\mathcal{I}$ is then defined by

$$\mathcal{I}_m :=]q_m, q_{m+1}[, \qquad \text{where } q_m = \sum_{j<m} p_j,$$

and the restriction of T to $\mathcal{I}_m$ is the affine mapping defined by $T(q_m) = 0$ and $T(q_{m+1}) = 1$.

Special cases of importance are the *b-ary expansion* transformations that are defined by

$$T(x) = \{bx\}, \qquad \sigma(x) = \lfloor bx \rfloor,$$

where $\lfloor u \rfloor$ is the integer part of u and $\{u\} = u \bmod 1 = u - \lfloor u \rfloor$ is the fractional part of u. These transformations give rise to the b-ary expansions of x in base b and are associated to symmetric memoryless sources (i.e., memoryless sources where all p_j's are equal).

Continued fraction expansions. This general framework may also create rather different types of sources with memory. This arises as soon as a mapping T with (some) nonlinear branches is used. (In a sense, it is the derivative $T'(x)$ that keeps memory of the previous history.) The *continued fraction* transformation is an example of this situation. The alphabet is then $\mathbb{N}$, the topological partition of $\mathcal{I}$ is defined by $\mathcal{I}_m :=]1/(m+1), 1/m[$, and the restriction of T to $\mathcal{I}_m$ is the decreasing linear fractional transformation $T(x) := (1/x) - m$,

$$T_{CF}(x) = \{\frac{1}{x}\}, \qquad \sigma_{CF}(x) = \lfloor \frac{1}{x} \rfloor.$$

When iterated, this transformation gives rise to the continued fraction expansion of x.

The inverse branches are all the linear fractional transformations h_m defined by $h_m(x) := 1/(x+m)$. The first branch h_1 does not satisfy (d_1) and (d_2), but the set of the linear fractional transformations $\{h_m \circ h_n\}$ clearly satisfies conditions (d) so that, as noted already, the theory of Sections 4 and 5 applies to this case.

Homoclinal or heteroclinal? Our definition of a probabilistic dynamical source does not preclude the situation where some branches are increasing, some others are decreasing. We introduce the term of *homoclinal* to refer to the case when all branches are simultaneously of the same type, and *heteroclinal* for the other case. Binary and continued fraction representations are homoclinal, while the familiar binary reflected code (also known as Gray code),

$$0, 1, 11, 10, 110, 111, 101, 100, 1100, \ldots$$

[2]Memoryless sources are thus binomial or multinomial probability distributions and it is common practice to refer to them as "Bernoulli sources". In this paper, we qualify as Bernoulli model *any* source model, not necessarily memoryless, where the number of items considered is fixed (as opposed to Poisson models).

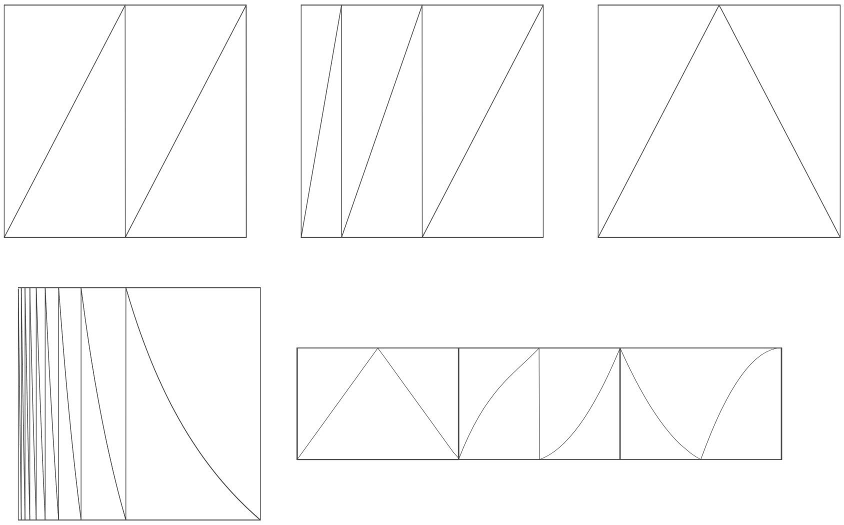

FIGURE 1. Five dynamical sources. From top to bottom: (a) binary representations; (b) a memoryless source with probabilities $(\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$; ($c$) the binary reflected code; (d) continued fraction representations; (e) a Markov source $(\mathsf{S}_0, \mathsf{S}_1, \mathsf{S}_2)$ that switches between two different representations. The plots represents the graphs of the associated shifts T.

corresponding to

$$T(x) = 2x \text{ if } x \in [0, \tfrac{1}{2}[, \qquad T(x) = 2 - 2x \text{ if } x \in [\tfrac{1}{2}, 1[,$$

is heteroclinal.

1.2. **Markov symbolic dynamical sources.** Until now, the shift T used at each stage has been always defined according to some unique rule. Very often, the modelling of real-life sources leads to introducing dependencies in the form of a shift that depends on the last symbol emitted. This gives rise to what we name Markov sources of which classical Markov chains are only a special case.

Definition 2 (Markov symbolic dynamical sources)**.** *Let $\mathcal{M}$ be a finite alphabet of cardinality r, and let $\mathsf{S} = (\mathsf{S}_0, \mathsf{S}_1, \mathsf{S}_2, \ldots, \mathsf{S}_i, \ldots, \mathsf{S}_r)$ be a set of $r+1$ different basic dynamical systems, all defined on the same alphabet $\mathcal{M}$. The basic dynamical system S_0 is used to begin with, and the dynamical system S_j is chosen whenever the previously emitted symbol is j.*

We now describe more precisely the mechanism of the source. One associates to a real x of $\mathcal{I}$ an infinite word $M(x)$ on alphabet $\mathcal{M}$, as in (1)

$$M(x) := (M_1(x), \ldots, M_k(x), \ldots),$$

together with the sequence of the iterates of the real x

$$(T^{<1>}(x), T^{<2>}(x), \ldots T^{<k>}(x), \ldots),$$

that are now defined by the initial conditions $M_1(x) := \sigma_0(x)$, $T^{<1>}(x) := T_0(x)$, and the recurrence relations

(2)

$$\text{if } M_k(x) = j, \text{ then } T^{<k+1>}(x) := T_j(T^{<k>}(x)), \text{ and } M_{k+1}(x) := \sigma_j(T^{<k>}(x)).$$

Each shift T_j is then associated to a topological partition $(\mathcal{I}_{i|j}), (1 \le i \le r)$ of the unit interval $\mathcal{I}$, and it has to satisfy hypotheses (d) of Definition 1. We denote by $(T_{[i|j]}), (1 \le i \le r)$ the branches of T_j, so that $T_{[i|j]}$ is a real analytic bijection from $\mathcal{I}_{i|j}$ to $\mathcal{I}$, that is asked to be expansive. The inverse branches of T_j are denoted by $h_{i|j}$, so that $h_{i|j}$ is a real analytic bijection from $\mathcal{I}$ to $\mathcal{I}_{i|j}$ that extends to a holomorphic map on $\mathcal{V}$, mapping $\mathcal{V}$ strictly inside $\mathcal{V}$ (i.e., $h_{i|j}(\overline{\mathcal{V}}) \subset \mathcal{V}$ for $1 \le i \le r$, $0 \le j \le r$).

The classical model of Markov chains of order 1 is then simply the case when all the S_j's are memoryless. More precisely, if the system S_j is a memoryless system of parameters $\Pi_j := (p_{i|j})_{i \le r}$, the transition matrix Π of the Markov chain is the $r \times r$ matrix

$$\Pi := (p_{i|j}) \quad 1 \le i, j \le r,$$

and the initial probability system is the vector Π_0.

Relation between Markov sources and general dynamical systems. Any Markovian source can be associated to a dynamical system that is no longer basic. We take $r+1$ copies of $\mathcal{I}$, for instance $\mathcal{I}_0 := \mathcal{I} =]0,1[$ and $\mathcal{I}_j :=]j, j+1[$. Denoting by Φ_m the translation $\Phi_m(x) := x + m$, we then define, for $1 \le i \le r$ and $0 \le j \le r$

$$\mathcal{I}_{i,j} := \Phi_j(\mathcal{I}_{i|j}), \qquad T_{i,j} := \Phi_i \circ T_{[i|j]} \circ \Phi_j^{-1},$$

so that $T_{i,j}$ is now a bijection from $\mathcal{I}_{i,j}$ on $\mathcal{I}_i$. The system S associated to partition $\mathcal{I}_{i,j}$ of $]0, r+1[$ and to branches $T_{i,j}$ is a general dynamical system.

One can use both interpretations of a Markovian source, but, here, we prefer to stay in the unit interval and we adopt the first formalism that is closer to the intuition underlying Markov chains. This is the point of view that has been adopted by Ruelle himself in [48].

1.3. **Fundamental intervals and prefixes.** We consider now the k-th iterate of the shift. In the case of a basic source, this is plainly the k-th iterate of T taken in the usual sense. In the case of a Markovian source, this is the iterated shift $T^{<k>}$ in the sense of definition (2). Each branch (or each inverse branch) of the k-th iterate of the shift is called a branch (or an inverse branch) of depth k. The depth of the inverse branch h is denoted by $|h|$. A branch or an inverse branch of depth k is then associated in a unique way to a finite word $\mathbf{w} = (m_1, m_2, \dots, m_k)$ of length k that keeps track of past choices. In the basic case, each branch and each inverse branch of depth k associated to $\mathbf{w} = (m_1, m_2, \dots, m_k)$ is of the form

$$T_{[\mathbf{w}]} = T_{[m_k]} \circ T_{[m_{k-1}]} \circ T_{[m_1]}, \quad h_{\mathbf{w}} = h_{m_1} \circ h_{m_2} \circ \cdots \circ h_{m_k}, \tag{3}$$

where $T_{[i]}$ and h_i denote the i-th branch or inverse branch of T. In the Markov case, each branch and each inverse branch of depth k associated to $\mathbf{w} = (m_1, m_2, \dots, m_k)$ is of the form

$$\begin{aligned} T_{[\mathbf{w}]} &= T_{[m_k|m_{k-1}]} \circ T_{[m_{k-1}|m_{k-2}]} \circ T_{[m_1|0]} \\ h_{\mathbf{w}} &= h_{m_1|0} \circ h_{m_2|m_1} \circ \cdots \circ h_{m_k|m_{k-1}}. \end{aligned} \tag{4}$$

For a finite alphabet of cardinality r, there are r^k branches of depth k. We denote by $\mathcal{H}_k$ the set of branches of depth k. Cyclic branches, i.e., branches for which the associated word starts and ends with the same symbol, play an important rôle in the case of Markov sources. We denote by $\mathcal{C}$ and $\mathcal{C}[i]$ the set of cyclic branches and the set of cyclic branches that start and end with symbol i. In the same vein, $\mathcal{C}_k, \mathcal{C}_k[i]$ denote the same objects relative to depth k.

We now present one of the main objects of the paper.

Definition 3 (Fundamental intervals). *The* fundamental interval *relative to the inverse branch h is the transform $\mathcal{I}_h := h(\mathcal{I})$ of the unit interval $\mathcal{I}$ by the inverse branch h. Its depth is the depth $|h|$ of h. The fundamental intervals of depth 1 are thus exactly the intervals of the initial partition. A fundamental interval $\mathcal{I}_h$ of depth k is formed with all the real numbers x of $\mathcal{I}$ which produce a word $M(x)$ whose prefix $\mathbf{w}$ of length k is exactly the finite word associated to h. It is also denoted by $\mathcal{I}_{\mathbf{w}}$.*

1.4. **Probabilistic dynamical sources.** In the sequel, we are interested in *probabilistic dynamical sources*, where the words are emitted according to a source mechanism as previously described, and we also allow for a prescribed initial distribution that is determined by a density function on the unit interval.

Definition 4 (Probabilistic dynamical sources). *Let S be a dynamical source (basic or Markovian) and let f be a real analytic probability density on interval $\mathcal{I}$ that extends to an analytic function on $\mathcal{V}$. Let $F(z) = \int_0^z f(t)\,dt$ be the associated distribution function. The pair (S, F) is called a probabilistic dynamical source. The set $\mathcal{M}^\infty$ of the words produced by the dynamical probabilistic source (S, F) is the set $M(\mathcal{I})$ endowed with the probability induced from f by M.*

In this context, the measure u_h of the fundamental interval $\mathcal{I}_h$ associated to an inverse branch h as defined in (3) or (4) is

$$u_h := |F(h(0)) - F(h(1))|. \tag{5}$$

This quantity plays an especially important rôle, since it equals the probability that a source word starts with the prefix $\mathbf{w}$ of $\mathcal{M}^*$ relative to h. It is called the *fundamental measure* relative to h and it is also denoted by $u_{\mathbf{w}}$.

1.5. **Dirichlet series of fundamental intervals, entropy, and coincidence probability.** The *entropy* $h(\mathsf{S}, F)$ relative to a probabilistic dynamical source (S, F) is defined as the limit, if it exists, of a quantity that involves the fundamental measures u_h,

$$h(\mathsf{S}, F) := \lim_{k\to\infty} \frac{-1}{k} \sum_{|h|=k} u_h \log u_h. \tag{6}$$

In the same vein, the probability that two independent words have the same prefix of length k equals $\sum_{|h|=k} u_h^2$. In general, this quantity appears to decrease exponentially with k, which leads to define the *coincidence probability* $c(\mathsf{S}, F)$ as the corresponding rate,

$$c(\mathsf{S}, F) := \lim_{k\to\infty} \Big(\sum_{|h|=k} u_h^2 \Big)^{1/k}, \tag{7}$$

provided it exists. The previous two definitions involve the *series of fundamental measures of depth k*

$$\Lambda_k(F,s) := \sum_{|h|=k} u_h^s = \sum_{|h|=k} |F(h(0)) - F(h(1))|^s, \tag{8}$$

and one has

$$h(\mathsf{S},F) := \lim_{k\to\infty} \frac{-1}{k}\frac{d}{ds}\,\Lambda_k(F,s)\big|_{s=1}, \qquad c(\mathsf{S},F) := \lim_{k\to\infty} [\Lambda_k(F,2)]^{1/k}. \tag{9}$$

In the sequel, we show that, in a precise sense, the quantities $\Lambda_k(F,s)$ defined in (8) behave asymptotically as k-th powers of a certain function,

$$\Lambda_k(F,s) \approx \lambda(s)^k \tag{10}$$

for a well-defined $\lambda(s)$ that is analytic near the real axis (Prop. 5). Entropy and coincidence probability depend only on the mechanism S, being independent of the distribution F. They can then be expressed solely in terms of the function $\lambda(s)$ in a way consistent with (9) and (10):

$$h(\mathsf{S}) = -\lambda'(1), \qquad c(\mathsf{S}) = \lambda(2).$$

These two characteristics of a source play an important rôle in all subsequent analyses of tries.

2. General tries and models of analysis

We describe here the standard trie and the companion hybrid trie implementations along with the main parameters that are relevant from a complexity standpoint.

2.1. **Definition of tries.** Consider the problem of comparing n infinite words that, within our analytical framework, are taken to be independently produced by a common dynamical source. Proceeding by elementary comparisons between their symbols yields a tree, called a trie [22, 34, 39, 51]. Let X be a sequence of reals of the unit interval, $X = (x_1, \dots, x_n) \in \mathcal{I}^n$. One considers the sequence of words $M(X)$ produced by the dynamical source S,

$$M(X) := (M(x_1), M(x_2), \dots, M(x_n)).$$

We also need to consider the sequence $\sigma(X)$ formed with the first symbol $\sigma(x_i)$ of each word $M(x_i)$; it is called the first "slice" of $M(X)$,

$$\sigma(X) := (\sigma(x_1), \sigma(x_2), \dots, \sigma(x_n)).$$

Thus, two distinct kinds of collections of symbols will intervene in the analysis: the infinite words produced by the source (represented as vertical words on Figure 2) as well as the finite sequences (that appear as horizontal slices on Figure 2).

In order to build the tree structure in a global fashion, we start from the root. First one groups together all the words which start with the same first symbol m, along a branch labelled by m, so that the corresponding subtree collects all the words starting with symbol m and stripped of their initial symbol. In our model, stripping a word $M(x)$ of its initial symbol is equivalent to shifting x, that is, considering $T(x)$. Note also that the reals associated to the group of words whose

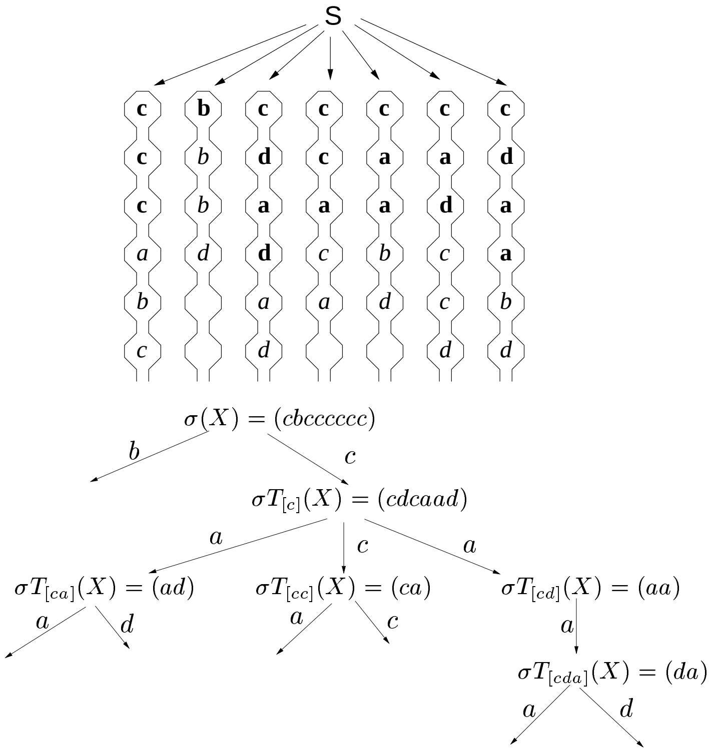

FIGURE 2. A source S emits independently infinite words that are represented "vertically" (top). In contrast, at any internal node of a trie (bottom), only "slices", appearing "horizontally", are to be considered.

first symbol equals m belong to the same fundamental interval $h_m(\mathcal{I})$. In this case, the suffixes of words are associated to the shifted sequence

$$T_{[m]}(X) := (T_{[m]}(x_1), T_{[m]}(x_2), \ldots, T_{[m]}(x_n)),$$

where $T_{[m]}(x)$ is only defined when the first symbol $\sigma(x)$ equals m and then equals $T(x)$. This process of splitting will continue until all words have been separated from each other.

To summarize, in the language of shifts, one associates to X a digital tree, called a trie and denoted by $\text{trie}(X)$, that is defined by the following recursive rules:

(R_0) If $X = \varnothing$, then $\text{trie}(X)$ is the empty tree.
(R_1) If $X = (x)$ has cardinality equal to 1, then $\text{trie}(X)$ consists of a single external node that contains the word $M(x)$.
(R_2) If X has cardinality $|X|$ at least equal to 2, then $\text{trie}(X)$ is an internal node represented by $\circ$ to which are attached the subtries built on the sets $T_{[m]}(X)$

for all possible values of m. Then trie(X) is defined by

$$\text{trie}(X) = \left\langle \circ, \left\{\text{trie}(T_{[m]}(X))\right\}_{m\in\mathcal{M}} \right\rangle . \tag{11}$$

Such a tree structure underlies classical radix sorting methods. It can be built by following the recursive rules R_0, R_1, R_2. Any prefix $\mathbf{w}$ which is common to at least two words of $M(X)$ is associated to an internal node of the trie, itself associated to a fundamental interval $\mathcal{I}_{\mathbf{w}}$ containing at least two elements of X.

2.2. **Implementations of tries and hybrid tries.** In the abstract tree structure representing the trie, each internal node is connected by edges to its children. We assume that the alphabet is totally ordered. Then, there are three natural implementations of such a node based on the classical data structures of array, ordered list, and binary search tree. There result "hybrid tries" which combine the abstract trie structure with a data structure that governs access to children in nodes.

(a) The simplest implementation uses arrays whose cardinality equals the size of the alphabet. Then one accesses children directly through an array of pointers. Note that this solution is meaningless practically for infinite alphabets, and space-wasting for large alphabets (with too many null pointers being allocated), but is quite adequate when the cardinality of the alphabet is small (typically, for binary words).

(b) The "list-trie" structure remedies the high storage cost of array-tries by linking sister subtrees at the expense of replacing direct array access by a (sorted) *linked list* traversal.

(c) The "bst-trie" uses *binary search trees* (bst) as subtree access method, with the goal of combining advantages of array-tries in terms of time cost, and list-tries in terms of storage cost. As noted in the introduction, the hybrid trie obtained is strictly equivalent to the *ternary search trie* structure proposed recently in [3], see also [6]. Indeed, the bst-trie can be viewed as a ternary tree where search on symbols is conducted like in a standard binary search tree (straight links on Figure 3) over the alphabet set $\mathcal{M}$, while trie descent is performed by following an escape pointer (curved links on Figure 3) whenever equality of symbols of detected.

On Figure 3, three implementations of the same "abstract" trie are drawn. For the standard trie, in either its abstract trie or array-trie version, size and path length are respectively 6 and 21. For the hybrid versions, the pointers of the data structures at nodes induce an extra path length that is 15 for the list-trie and 8 for the bst-trie.

2.3. **Parameters.** Let us first consider standard tries, that is, either the abstract structure or the array-trie implementation. The *level* of a node in a trie is the number of edges that connect it to the root. The *height* of the trie is the maximum level of any external node. It represents a measure of the distance between the two closest elements of $M(X)$ since it equals the minimum number of comparisons required to separate any pair $(M(x_i), M(x_j))$ of elements of $M(X)$. The *path length* of the standard trie is the sum of the levels of all (nonempty) external nodes. Path length thus equals the total number of symbols that need to be examined in order to distinguish all elements of $M(X)$. (Once divided by the number of elements, it gives the mean cost of a positive search, that is, the search cost averaged over all

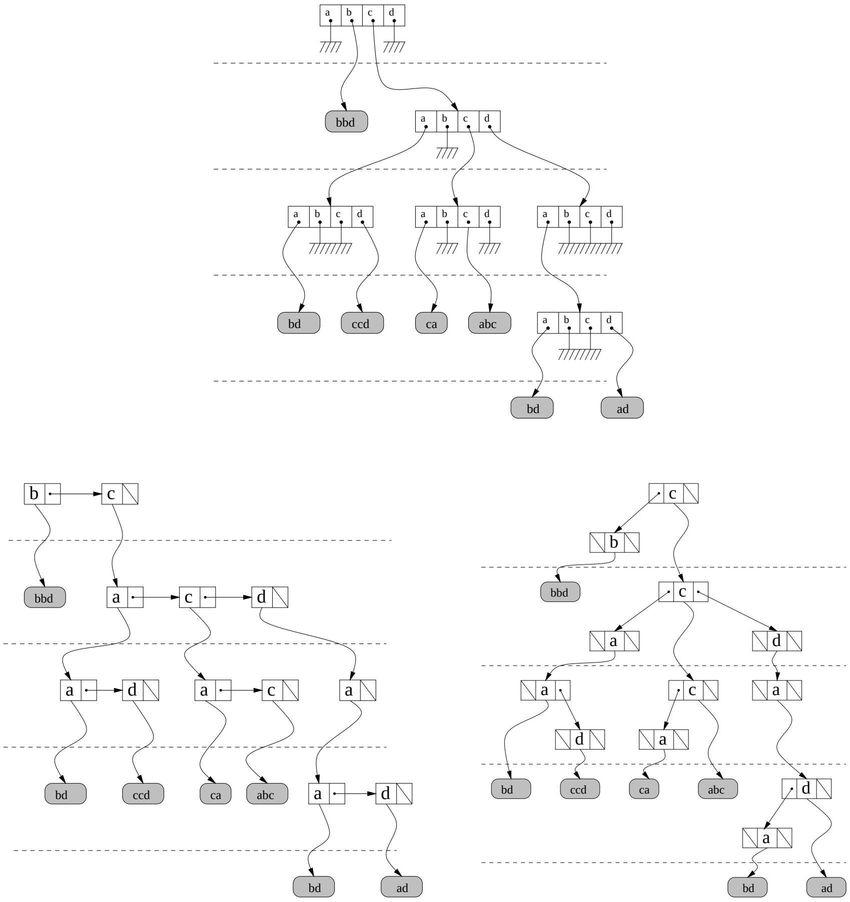

FIGURE 3. Three representations of a trie built from the sequence of words $(\mathbf{ccc}abc, \mathbf{b}bbd, \mathbf{cdad}ad, \mathbf{cca}ca, \mathbf{caa}bd, \mathbf{cad}ccd, \mathbf{cdaa}bd)$ over the alphabet $\mathcal{A} = \{a, b, c, d\}$: the array-trie (top; the null pointers are represented as electrically grounded), the list-trie (bottom left) and the bst-trie (bottom right).

items for an item that is present in the trie). The *size* of the tree is the number of its internal nodes. Adding to the size the cardinality of X gives the number of fundamental intervals necessary to isolate all elements of $M(X)$. Clearly, size determines the storage requirements of hybrid trie implementations.

In a hybrid trie, path length decomposes as a sum of two components: the first one arises from the underlying trie structure; the second one is the additional cost incurred by the traversal of internal node structures. It is this overhead which is analysed here for hybrid tries. (If one is interested in a global external path length, it suffices to combine additively path length of the abstract trie with this additional path length.)

2.4. **Poisson and Bernoulli models.** The Bernoulli model considers a sequence of n infinite words independently produced by the same dynamical source. This sequence is of the form $M(X)$, has cardinality n, and is thus obtained by n independent drawings $x_1, x_2, \ldots, x_n$ in the interval $\mathcal{I}$, with the interval being endowed with the probability density f (F denotes the associated distribution function). The Bernoulli model of index n relative to the source (S, F) is denoted by $(\mathcal{B}_n, \mathsf{S}, F)$.

Rather than fixing the cardinality n of the set X, it proves technically convenient to consider that the sequence X has a variable number N of elements that obeys a Poisson law of parameter z, namely,

$$\Pr\{N = k\} = e^{-z}\frac{z^k}{k!}.$$

This model is called the Poisson model of rate z. When relative to the dynamical source (S, F), it is denoted by $(\mathcal{P}_z, \mathsf{S}, F)$. From the definition, the expectations of a random variable Y under the Poisson and Bernoulli models are related by

$$\mathrm{E}[Y; \mathcal{P}_z, \mathsf{S}, F] = e^{-z}\sum_{n=0}^{\infty} \mathrm{E}[Y; \mathcal{B}_n, \mathsf{S}, F]\frac{z^n}{n!}. \tag{12}$$

A similar relation holds for probabilities of events as they are always representable as expectations of indicator variables. The interest of the Poisson model is that there is complete independence on what happens in disjoint subintervals of $\mathcal{I}$. In particular, the number of elements that fall into any interval of measure u is itself distributed as a Poisson variable of rate zu. Such a property holds notably for the fundamental intervals associated to source (S, F), whose measure $u_{\mathbf{w}}$ is given by (5).

The strong independence property of the Poisson model provides an easy access to the expectation of basic parameters. It becomes then necessary to return to the Bernoulli model —the one of interest for average-case analysis commonly parametrized by the number n of data items. The process of translating from a Poisson to a Bernoulli model is called depoissonization and several strategies are available. Here are are the ones used in this paper.

1. *Algebraic depoissonization* relies on the fact that, by Equation (12), a Poisson quantity is, up to a factor of e^{-z} the exponential generating function of its Bernoulli counterparts:

$$\mathrm{E}[Y; \mathcal{B}_n, \mathsf{S}, F] = n!\,[z^n]\,e^{z}\,\mathrm{E}[Y; \mathcal{P}_z, \mathsf{S}, F], \tag{13}$$

   where $[z^n]h(z)$ represents the coefficient of z^n in the expansion of $h(z)$ at 0. This technique is the basis of all our exact results under the Bernoulli model: Theorems 2, 4 are derived in this way as counterparts of Theorems 1, 3.
2. *Asymptotic depoissonizations.* In the Poisson model, N is narrowly concentrated near its mean z with a high probability, so that the Poisson rate z plays a rôle much similar to the Bernoulli cardinality n. It is then reasonable to expect that average values of parameters under both models should be asymptotically equivalent. Generally, the problem is of a Tauberian character. Asymptotic depoissonizations are ways of establishing such an equivalence, based on some supplementary assumptions.

   2.1. *Dirichlet depoissonization* relies on the existence of Mellin transforms and an approximation principle between the Dirichlet series of Poisson and

Bernoulli models. This technique is used here for the average-case analysis of the additive parameters of size and path length under the Bernoulli model: Theorem 6 is derived in this way from its Poisson version, Theorem 5.

2.2. *Saddle-point depoissonization* relies on estimates of Poisson averages (or, equivalently, generating functions) for "nonprobabilistic" values of the rate z taken in the complex domain, this in conjunction with a saddle point analysis. This technique is used here for the distributional analysis of the multiplicative parameter of height under the Bernoulli model: Theorem 8 is derived in this way from its Poisson version, Theorem 7.

The net result for us is that dominant asymptotic estimates established initially under the Poisson model remain eventually valid under the Bernoulli model.

3. Algebraic analysis of trie parameters

We show here that the expectations of the main parameters of height, size, and various forms of path length in standard or hybrid tries can all be expressed as sums that involve the measures of fundamental intervals.

3.1. **Additive parameters.** The standard trie built on a *sequence* X of elements of $\mathcal{I}^n$ depends in fact only on the *set* that underlies X. However, the construction cost of the list-trie and even the shape of the bst-trie depend on the relative order of elements in the sequence X. What is common to all implementations is that the structure of the node indexed by $\mathbf{w}$ is fully determined by the corresponding slice,

$$\sigma T_{[\mathbf{w}]}(X) := (\sigma T_{[\mathbf{w}]}(x_1), \sigma T_{[\mathbf{w}]}(x_2), \dots, \sigma T_{[\mathbf{w}]}(x_n)).$$

First, the root of the trie is determined by the slice $\sigma(X)$ and the collection of root subtrees is determined by the collection of shifts $T_{[m]}(X)$. In this perspective, the decomposition (11) is synonymous to

$$\text{trie}(X) = \left\langle \sigma(X), \left\{\text{trie}(T_{[m]}(X))\right\}_{m\in\mathcal{M}} \right\rangle. \tag{14}$$

Consider now an "additive" parameter γ on trie(X) defined recursively by the rule

$$\begin{aligned} \gamma[\text{trie}(X)] &= 0 \quad \text{if} \quad |X| \le 1 \\ \gamma[\text{trie}(X)] &= \delta(\sigma(X)) + \sum_{m\in\mathcal{M}} \gamma[\text{trie}(T_{[m]}(X))] \quad \text{if} \quad |X| \le 2. \end{aligned}$$

The parameter δ is sometimes called the "toll". The recurrence relation can be unwound, leading to

$$\gamma[\text{trie}(X)] = \sum_{\mathbf{w}\in\mathcal{M}^*} \delta[\sigma T_{[\mathbf{w}]}(X)], \tag{15}$$

provided that $\delta(\mathbf{s})$ is zero on slices $\mathbf{s}$ that contain either 0 or 1 symbol.

We now describe the probabilistic model that is induced by the Poisson model at each possible node of the trie determined by a prefix $\mathbf{w}$. Since the probability that a word starts with prefix $\mathbf{w}$ is equal to the fundamental measure $u_{\mathbf{w}}$ defined in (5), the probability that the next symbol emitted is m equals

$$p_{\mathbf{w},m} = \frac{u_{\mathbf{w}\cdot m}}{u_{\mathbf{w}}}. \tag{16}$$

Here, the notation $\mathbf{w}{\cdot}m$ denotes the concatenation of sequence $\mathbf{w}$ and of symbol m. Since all elements of X are independently drawn, at the internal node labelled by $\mathbf{w}$, symbols are then emitted by the memoryless source $\mathsf{B}_{\mathbf{w}}$ that is defined by the probabilities $\{p_{\mathbf{w},m}\}_{m\in\mathcal{M}}$.

Moreover, if the cardinality of X is a random Poisson variable of rate z, the length of the slice $\sigma T_{[\mathbf{w}]}(X)$ is also a random Poisson variable of rate $zu_{\mathbf{w}}$. It follows from (15) that the expectation of parameter γ is a sum of expectations of parameter δ,

$$\mathrm{E}[\gamma;\mathcal{P}_z,\mathsf{S},F] = \sum_{\mathbf{w}\in\mathcal{M}^*} \mathrm{E}[\delta;\mathcal{P}_{zu_{\mathbf{w}}},\mathsf{B}_{\mathbf{w}}], \tag{17}$$

where $\mathbf{w}$ ranges over all (finite) sequences and $\mathsf{B}_{\mathbf{w}}$ denotes the memoryless source defined in (16).

3.2. **Search costs at nodes.** We consider here the four additive parameters of interest, namely, size, and the three path lengths relative to each kind of hybrid trie, and evaluate the expectations of the corresponding tolls.

First, the toll δ_S associated to size equals 1 provided that the internal node indexed by $\mathbf{w}$ exists or equivalently that the slice $\sigma T_{[\mathbf{w}]}(X)$ has at least two symbols:

$$\delta_S(\mathbf{s}) = \begin{cases} 1 & \text{if } |\mathbf{s}| \geq 2 \\ 0 & \text{otherwise.} \end{cases}$$

In the same vein, the toll δ_A for path length of an array-trie is simply

$$\delta_A(\mathbf{s}) = \begin{cases} |\mathbf{s}| & \text{if } |\mathbf{s}| \geq 2 \\ 0 & \text{otherwise.} \end{cases}$$

The parameters $\delta_L(\mathbf{s})$ and $\delta_B(\mathbf{s})$ that are relative to path length of list-tries and of bst-tries are exactly traversal costs of node structures built over a slice $\mathbf{s}$: The symbols of $\mathbf{s}$ (where repetition is allowed) are inserted in order in a structure (a list or a binary search tree), and then the toll is the cost incurred by the retrieval of each occurrence of each symbol of $\mathbf{s}$.

The proposition that follows is the key step in the algebraic part of the treatment of additive trie parameters. In substance, the evaluation of the expected tolls $\mathrm{E}[\delta]$ corresponding to list-tries and bst-tries is analogous to the (easy) analysis of lists and of the (harder) analysis of binary search trees on a finite domain with a nonuniform probability distribution; see for instance [1, 5] for related developments. Our approach in this paper relies on a symbolic description of parameters by *generating functions* and is, perhaps, of independent interest.

Proposition 1 (Toll parameters)**.** *Let* B *be a memoryless source relative to the set of probabilities* $\{p_i\}_{i\in\mathcal{M}}$ *and* $\mathcal{P}_z$ *the Poisson model of rate* z*. Then, in the model* $(\mathcal{P}_z,\mathsf{B})$*, expectations of the toll parameters relative to the size of a trie and the path length of an array-trie are respectively*

$$\mathrm{E}[\delta_S;\mathcal{P}_z,\mathsf{B}] = 1-(1+z)e^{-z}, \qquad \mathrm{E}[\delta_A;\mathcal{P}_z,\mathsf{B}] = z(1-e^{-z}).$$

In the model $(\mathcal{P}_z, \mathsf{B})$, the expectations of traversal costs for ordered lists and binary search trees are respectively

$$\begin{aligned}
\mathrm{E}[\delta_L; \mathcal{P}_z, \mathsf{B}] &= \sum_{j \in \mathcal{M}} P_{[>j]}\ z\ (1 - e^{-p_j z}), \\
\mathrm{E}[\delta_B; \mathcal{P}_z, \mathsf{B}] &= 2 \sum_{\substack{(i,j) \in \mathcal{M}^2 \\ i<j}} \frac{p_i p_j}{{P_{[i,j]}}^2} \left[e^{-zP_{[i,j]}} - 1 + zP_{[i,j]} \right],
\end{aligned}$$

where $P_{[i,j]} = \sum_{k=i}^{j} p_k$ *and* $P_{[>j]} = \sum_{k>j} p_k$.

Proof. We consider an ordered alphabet $\mathcal{M} = \{a_1 < a_2 < \cdots < a_r\}$ where the i-th symbol is denoted by a_i. For any set $\mathcal{L} \subseteq \mathcal{M}^*$, the ordinary generating function (ogf) and exponential generating function (egf) relative to a parameter δ over $\mathcal{L}$ are defined as

$$\begin{aligned}
F(z, u, x_1, \ldots, x_r) &= \sum_{\mathbf{s} \in \mathcal{L}} z^{|\mathbf{s}|} u^{\delta(\mathbf{s})} {x_1}^{|\mathbf{s}|_1} \cdots {x_r}^{|\mathbf{s}|_r} \\
\widehat{F}(z, u, x_1, \ldots, x_r) &= \sum_{\mathbf{s} \in \mathcal{L}} \frac{z^{|\mathbf{s}|}}{|\mathbf{s}|!} u^{\delta(\mathbf{s})} {x_1}^{|\mathbf{s}|_1} \cdots {x_r}^{|\mathbf{s}|_r}.
\end{aligned} \tag{18}$$

Here $|\mathbf{s}|$ and $|\mathbf{s}|_i$ denote respectively the total length of $\mathbf{s}$ and the number of occurrences of a_i in $\mathbf{s}$. Formally, the variables z and u "mark" the length of the sequence $|\mathbf{s}|$ and the value of the parameter δ, while the variable x_i records the occurrences of the symbol a_i.

Formal languages and generating functions are intimately related. In the course of the analysis, some operations on languages are better translated in terms of egf's, whereas others are better expressed by means of ogf's. The two types of generating functions are related by the combinatorial Laplace transform that is defined by

$$\mathcal{L}[\sum_n f_n \frac{z^n}{n!}] = \sum_n f_n z^n. \tag{19}$$

For instance, the generating functions relative to the whole set $\mathcal{M}^*$ are related by

$$\mathcal{L}\left(e^{z(x_1 + \cdots + x_r)}\right) = \frac{1}{1 - z(x_1 + \cdots + x_r)}.$$

When the symbols of $\mathcal{M}$ are emitted by a memoryless source B relative to probabilities $\{p_i\}$, the expectation of δ in the model $(\mathcal{P}_z, \mathsf{B})$ is

$$\mathrm{E}[\delta; \mathcal{P}_z, \mathsf{B}] = e^{-z} \frac{\partial}{\partial u}\ \widehat{F}(z, u, p_1, \ldots, p_r)\Big|_{u=1}, \tag{20}$$

where $\widehat{F}$ is the exponential generating function associated to parameter δ defined in (18).

1. The expectations of the first two parameters δ_S, δ_A are direct consequences of properties of the Poisson process, but we develop them within the generating function framework as they serve to introduce basic principles. The decomposition,

$$\mathcal{M}^* = (\epsilon + \mathcal{M}) + \sum_{i \geq 2} \mathcal{M}^k,$$

once translated into egf's yields

$$1+z(x_1+\cdots+x_r)+u(e^{z(x_1+\cdots+x_r)}-1-z(x_1+\cdots+x_r)),$$
$$1+z(x_1+\cdots+x_r)+(e^{zu(x_1+\cdots+x_r)}-1-zu(x_1+\cdots+x_r)),$$

as egf's relative to δ_S and δ_A. An application of (20) then gives the form of $\mathrm{E}[\delta_A]$.

2. The next parameter δ_L requires the shuffle ('ш') decomposition [17, 38],

$$\mathcal{M}^* = {a_1}^* ш \cdots ш {a_r}^*.$$

meaning that each word decomposes into (possibly void) subwords which are repetitions (a_i^*) of the same symbol a_i, shuffled in all possible ways. It is a known and easy result that, in all generality, a shuffle of languages over disjoint alphabets corresponds to the product of corresponding egf's. The search cost of a key a_α in a slice $\mathbf{s}$ is then equal to the number of distinct symbols $i<\alpha$ in $\mathbf{s}$. The corresponding exponential generating function is

$$\widehat{f}_\alpha(z,u,x_1,\ldots,x_r)=\prod_{i<\alpha}(u(e^{zx_i}-1)+1)\prod_{i\geq\alpha}e^{zx_i}.$$

It now suffices to sum over all $\alpha\in\mathcal{M}$ and apply the formula (20) in order to get $\mathrm{E}[\delta_L]$.

3. The last parameter δ_B needs a more subtle approach. We proceed by stages and describe the search for a symbol α in $\mathrm{bst}(\mathbf{s})$ by: a search of the maximum along the rightmost branch of $\mathrm{bst}(\mathbf{s}_{\leq\alpha})$, where $\mathbf{s}_{<\alpha}$ means $\mathbf{s}$ restricted to elements of index smaller than α; a dual search of the minimum along the leftmost branch of $\mathrm{bst}(\mathbf{s}_{\geq\alpha})$. Each one-sided search is described by a regular expression and corresponding multivariate rational functions. The combination is achieved by a shuffle product that involves formal Laplace transforms.

(i) *Extrema analysis.* The first problem to be solved is thus the analysis of length of the rightmost branch in a tree built on random words, or equivalently the analysis of left-to-right maxima (also called "records"). Given the alphabet $\mathcal{M}$, the regular expression decomposition

$$\mathcal{M}^*=\prod_{j=1}^{r}\left(\epsilon+a_j\cdot(a_1+a_2+\cdots+a_j)^*\right)$$

expresses precisely all the possible decompositions of words by sets of left-to-right maxima. By general principles, concatenation translates as an ordinary product of ogf's. Accordingly, the ordinary multivariate generating function

$$N_{\max}(z,u,x_1,x_2,\ldots,x_r)=\prod_{j=1}^{r}\left(1+\frac{zux_j}{1-z(x_1+\cdots+x_j)}\right),$$

has its coefficient $[z^nu^kx_1^{n_1}\cdots x_r^{n_r}]$ equal to the number of words of length n having k maxima and n_j occurrences of symbol j. Dually, the multivariate ogf for minima is

$$N_{\min}(z,u,x_1,x_2,\ldots,x_r)=\prod_{j=1}^{r}\left(1+\frac{zux_j}{1-z(x_j+\cdots+x_r)}\right).$$

Similar decompositions have been used by Prodinger [45] in the study of the maximum of geometrically distributed random variables.

(*ii*) *Search costs.* Consider next the search cost c_α of some fixed symbol α in a bst. The shuffle (' $\sqcup\!\sqcup$ ') decomposition

$$(\mathcal{M} \setminus \{\alpha\})^* = (a_1 + \cdots + a_{\alpha-1})^* \sqcup\!\sqcup\ (a_{\alpha+1} + \cdots + a_r)^*,$$

expresses the fact that each word decomposes into subwords $< a_\alpha$ and $> a_\alpha$, shuffled in all possible ways. Let ' $\sqcup\!\sqcup$ ' also denote the operation on ogf's that translates the shuffle product of languages over disjoint alphabets. As noted before, shuffles of languages correspond to products of egf's, while the Laplace transform relates egf's and ogf's. Thus, one has

$$f(z) \sqcup\!\sqcup\, g(z) = \mathcal{L}[\mathcal{L}^{-1}[f(z)] \cdot \mathcal{L}^{-1}[g(z)]].$$

Equipped with this operation, we can express the ogf relative to c_α as

$$\begin{aligned} C_\alpha(z, u, x_1, \ldots, x_r) &= [N_{\max}(z, u, x_1, \ldots, x_{\alpha-1}) \sqcup\!\sqcup\, N_{\min}(z, u, x_{\alpha+1}, \ldots, x_r)] \\ &\quad \cdot \left[1 + \frac{z x_\alpha}{1 - z(x_1 + \cdots + x_r)}\right], \end{aligned} \tag{21}$$

where the last factor takes into account trailing sequences that may contain α.

(*iii*) *Explicit forms.* Eq. (21) describes an ogf that condenses all the information on costs, including the full distribution. Then the egf is obtained by taking the formal inverse Laplace transform. The average cost is as usual obtained by differentiating (21) with respect to u and setting $u = 1$ following the principle of (20). The rest of the computation (details omitted) is carried out by means of Laplace transforms, partial fraction expansions, and logarithmic derivatives, using the obvious relations

$$\mathcal{L}[e^{az}] = \frac{1}{1 - az}, \qquad \mathcal{L}[z e^{az}] = \frac{z}{(1 - az)^2}, \qquad \frac{1}{1 - az} \sqcup\!\sqcup\, \frac{1}{1 - bz} = \frac{1}{1 - (a + b)z}.$$

□

3.3. **Size and path length.** The form of the recurrence (17), the form of the probabilities at each node (16), and the expressions obtained in Proposition 1 somewhat simplify upon the unwinding of the recursion (15). As a consequence, the expectations of the four additive parameters can be solely expressed with fundamental measures.

Theorem 1 (Poisson expectations of additive parameters)**.** *Let* (S, F) *be a probabilistic dynamical source and* $\mathcal{P}_z$ *the Poisson model of rate* z. *Then expectations in the model* $(\mathcal{P}_z, \mathsf{S}, F)$ *of the toll parameters relative respectively to the size of a trie, path length of an array-trie trie, path length of an ordered-list trie, path length of a bst-trie are*

$$\begin{aligned} \widehat{S}(z) &= \sum_{\mathbf{w} \in \mathcal{M}^*} \left[1 - (1 + z u_{\mathbf{w}}) e^{-z u_{\mathbf{w}}}\right], \\ \widehat{P}_A(z) &= \sum_{\mathbf{w} \in \mathcal{M}^*} z u_{\mathbf{w}} \left[1 - e^{-z u_{\mathbf{w}}}\right] \\ \widehat{P}_L(z) &= \sum_{\mathbf{w} \in \mathcal{M}^*} \sum_{i \in \mathcal{M}} z\, U_{\mathbf{w} \cdot [>i]} (1 - e^{-z u_{\mathbf{w} \cdot i}}) \\ \widehat{P}_B(z) &= 2 \sum_{\mathbf{w} \in \mathcal{M}^*} \sum_{\substack{(i,j) \in \mathcal{M}^2 \\ i < j}} \frac{u_{\mathbf{w} \cdot i} u_{\mathbf{w} \cdot j}}{U_{\mathbf{w} \cdot [j,i]}^{\,2}} \left[e^{-z U_{\mathbf{w} \cdot [j,i]}} - 1 + z U_{\mathbf{w} \cdot [j,i]}\right], \end{aligned}$$

where $U_{\mathbf{w}\cdot[j,i]} = \sum_{k=j}^{i} u_{\mathbf{w}\cdot k}$, *and* $U_{\mathbf{w}\cdot[>j]} = \sum_{k>j} u_{\mathbf{w}\cdot k}$.

We can now return to the Bernoulli model thanks to the principles of "algebraic depoissonization" summarized in (12) and (13) that induce a formal dictionary from Poisson to Bernoulli expectations:

$$e^{-az} \mapsto (1-a)^n, \quad ze^{-az} \mapsto n(1-a)^{n-1}.$$

Therefore the expectations of the four additive parameters in the Bernoulli model $(\mathcal{B}_n, \mathsf{S}, F)$ are also expressible purely in terms of fundamental measures.

Theorem 2 (Bernoulli expectations of additive parameters)**.** *Let* $(\mathcal{B}_n, \mathsf{S}, F)$ *be the Bernoulli model relative to a fixed number* n *of words independently drawn from a probabilistic dynamical source* (S, F)*. Then the expectations for the size of a trie, path length of an array-trie, path length of an ordered-list trie, path length of a bst trie are*

$$\begin{aligned}
S(n) &= \sum_{\mathbf{w}\in\mathcal{M}^*} \left[1-(1+(n-1)\,u_{\mathbf{w}})\,(1-u_{\mathbf{w}})^{n-1}\right] \\
P_A(n) &= \sum_{\mathbf{w}\in\mathcal{M}^*} n u_{\mathbf{w}}[1-(1-u_{\mathbf{w}})^{n-1}] \\
P_L(n) &= \sum_{\mathbf{w}\in\mathcal{M}^*} \sum_{i\in\mathcal{M}} nU_{\mathbf{w}\cdot[>i]}(1-(1-u_{\mathbf{w}\cdot i})^{n-1}) \\
P_B(n) &= 2\sum_{\mathbf{w}\in\mathcal{M}^*} \sum_{\substack{(i,j)\in\mathcal{M}^2\\ i<j}} \frac{u_{\mathbf{w}\cdot i}u_{\mathbf{w}\cdot j}}{{U_{\mathbf{w}\cdot[j,i]}}^2}\left[(1-U_{\mathbf{w}\cdot[j,i]})^n-1+nU_{\mathbf{w}\cdot[j,i]}\right],
\end{aligned}$$

where $U_{\mathbf{w}\cdot[i,j]} = \sum_{k=i}^{j} u_{\mathbf{w}\cdot k}$ *and* $U_{\mathbf{w}\cdot[>j]} = \sum_{k>j} u_{\mathbf{w}\cdot k}$ *as before.*

3.4. **Height of a trie.** Consider now a random trie (or equivalently, its array-trie implementation) that is produced by a probabilistic source (S, F) in the Poisson model $(\mathcal{P}_z, \mathsf{S}, F)$. Such a trie has height at most k provided that no fundamental interval of depth k contains more than one word. The probability of this event, is given by the independence property of the Poisson model, from which the mean value results.

Theorem 3 (Height under the Poisson model)**.** *The distribution of trie height under the Poisson model* $(\mathcal{P}_z, \mathsf{S}, F)$ *is given by*

$$\pi_k(z) := \Pr[h \le k] = \prod_{|\mathbf{w}|=k} (1+zu_{\mathbf{w}})e^{-zu_{\mathbf{w}}} = e^{-z} \prod_{|\mathbf{w}|=k} (1+zu_{\mathbf{w}}) \tag{22}$$

and the expectation of height is

$$\widehat{H}(z) = \mathrm{E}[h; \mathcal{P}_z, \mathsf{S}, F] = \sum_{k=0}^{\infty} [1-\pi_k(z)]\,. \tag{23}$$

The corresponding quantities under the Bernoulli model are obtained next by means of the "algebraic depoissonization" principle (12), (13).

Theorem 4 (Height under the Bernoulli model)**.** *Let* $\pi_{k,n}$ *denote the probability that a trie built on an* n*–tuple of random items has height at most* k *under the*

Bernoulli model $(\mathcal{B}_n, \mathsf{S}, F)$. The exponential generating function of the $\pi_{k,n}$ satisfies

$$\Pi_k(z) := \sum_n \pi_{k,n} \frac{z^n}{n!} = \prod_{|\mathbf{w}|=k} (1 + z u_{\mathbf{w}}),$$

and the corresponding expectation is

$$\mathrm{E}[h; \mathcal{B}_n, \mathsf{S}, F] = n![z^n] \sum_{k=0}^{\infty} \left(e^z - \prod_{|\mathbf{w}|=k} (1 + z u_{\mathbf{w}}) \right).$$

3.5. **Dirichlet series associated to trie parameters.** The expectations in the Poisson model and in the Bernoulli model belong to the paradigm of harmonic sums that are general sums of the form

$$F(z) = \sum_{k \in K} \lambda_k f(\mu_k z).$$

The asymptotic analysis of such sums is classically done by Mellin transform techniques (see Section 6.1) whose application depends on the location and nature of poles of an associated Dirichlet series

$$\Lambda(s) := \sum_{k \in K} \lambda_k {\mu_k}^s.$$

The Mellin approach thus leads to considering a variety of Dirichlet series of fundamental intervals. For instance, in the Poisson model, three kinds of Dirichlet series are involved for array-tries, list-tries, and bst-tries:

$$\Lambda^{\langle A \rangle}(F, s) = \sum_{\mathbf{w} \in \mathcal{M}^*} u_{\mathbf{w}}^s, \qquad \Lambda^{\langle L \rangle}(F, s) = \sum_{\mathbf{w} \in \mathcal{M}^*} \sum_{i \in \mathcal{M}} U_{\mathbf{w}\cdot[>i]} \ {u_{\mathbf{w}\cdot i}}^{s-1}, \tag{24}$$

$$\Lambda^{\langle B \rangle}(F, s) = 2 \sum_{\mathbf{w} \in \mathcal{M}^*} \sum_{\substack{(i,j) \in \mathcal{M}^2 \\ i<j}} u_{\mathbf{w}\cdot i} \ u_{\mathbf{w}\cdot j} \ {U_{\mathbf{w}\cdot[i,j]}}^{s-2}. \tag{25}$$

Similar but modified versions arise in the Bernoulli model.

In contrast to additive parameters, the quantity expressing the probability distribution of height is not a harmonic sum, however its logarithm is,

$$\log \pi_k(z) = -\sum_{|\mathbf{w}|=k} z u_{\mathbf{w}} \log(1 + z u_{\mathbf{w}}),$$

and it is associated to the Dirichlet series

$$\Lambda_k^{\langle A \rangle}(F, s) = \sum_{\mathbf{w} \in \mathcal{M}^k} {u_{\mathbf{w}}}^s \tag{26}$$

corresponding to branches of depth k.

At this stage, we need a way to study Dirichlet series of fundamental measures in order to estimate asymptotically average values of parameters. This necessitates locating the poles of the Dirichlet series defined in Equations (24) and (25) as well as determining the behaviour of the family of Dirichlet series (when k varies) defined in (26). We introduce next several generalizations of Ruelle operators that play the rôle of generating operators for fundamental intervals and open an avenue to the analytic study of such Dirichlet series.

4. Generalized Ruelle operators: analytic properties

This section starts with the easy notion of density transformers (Section 4.1) that lift into transfer operators upon introducing an extra complex parameter s (Section 4.2). We then design a class of generalized Ruelle operators, the multi-secant operators (Section 4.3), where derivatives present in transfer operators are replaced by secants *and* multivariate functions are acted on. The operators based on multi-secants generate a variety of Dirichet series of fundamental measures, including the ones arising from the analysis of trie parameters (Sections 4.4, 4.5).

4.1. **Density transformers.** There is a direct relationship between the dynamics of source S, the answers to the main problems of Section 3.5, and spectral properties of an operator closely related to the way the shift T transforms probability distributions. The basic ingredient, well-developed in dynamical systems theory, is the class of *transfer operators* [2, 49]. In its simplest form, the transfer operator associated to a basic dynamical system is the "density transformer",

$$\mathcal{G}[f](x) := \sum_{i\in\mathcal{M}} |h_i'(x)|\ f\circ h_i(x). \tag{27}$$

The term comes from the following obvious fact: if X is a random variable whose distribution has density f, then the density of $T(X)$ is $\mathcal{G}[f]$. In other words, the operator $\mathcal{G}$ describes one step of the source process. The component operator given by the i-th term is denoted by $\mathcal{G}_{[i]}$; it is defined by

$$\mathcal{G}_{[i]}[f](x) := |h_i'(x)|\ f\circ h_i(x), \tag{28}$$

so that

$$\mathcal{G} = \sum_{i\in\mathcal{M}} \mathcal{G}_{[i]}. \tag{29}$$

In the same way, one can define a "density transformer" associated to a Markov dynamical system. There are now r different densities $(f_1, f_2, \ldots, f_r)$ that correspond to "conditional densities": $f_j(x)$ is the density at the point x when the last emitted symbol equals j. One begins with density f, and, after one iteration of the shift associated to the initial system S_0, one has

$$f_j(x) = |h_{j|0}'(x)|\ f\circ h_{j|0}(x).$$

More generally, the sequence of conditional densities $(f_1, f_2, \ldots, f_r)$ at one iteration stage, and the sequence of conditional densities $(g_1, g_2, \ldots, g_r)$ at the next iteration stage are related by a matrix of operators $\mathcal{G}$ that is built from the density transformers $\mathcal{G}_j$ associated to each dynamical system S_j. The density transformer $\mathcal{G}_j$ associated to S_j acts on f_j

$$\mathcal{G}_j[f_j](x) := \sum_{i\in\mathcal{M}} |h_{i|j}'(x)|\ f_j\circ h_{i|j}(x).$$

Each term of the previous sum defines an operator which will be denoted by $\mathcal{G}_{[i|j]}$,

$$\mathcal{G}_{[i|j]}[f](x) := |h_{i|j}'(x)|\ f\circ h_{i|j}(x), \tag{30}$$

and each term $\mathcal{G}_{[i|j]}[f_j]$ represents the "part" of the new density g_i that "comes from" the density f_j. We consider now the $r\times r$–matrix $\mathcal{G}$ whose general coefficient is

$\mathcal{G}_{[i|j]}$

$$\mathcal{G} = \left(\mathcal{G}_{[i|j]}\right), \tag{31}$$

(i is the index for lines, and j the index for columns). This matrix $\mathcal{G}$ is itself the density transformer, since it transforms the sequence of conditional densities $(f_1, f_2, \dots, f_r)$ at one iteration stage, into the sequence of conditional densities $(g_1, g_2, \dots, g_r)$ at the following iteration stage:

$$\begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_r \end{pmatrix} = \left(\mathcal{G}_{[i|j]}\right) \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_r \end{pmatrix}.$$

4.2. **Classical Ruelle operators.** In each case, of a basic dynamical system or of a Markovian one, it proves highly useful to work with more general operators, called the Ruelle operators. Each component operator in (28) or (30) depends now on a complex parameter s and is defined through the analytic extension $\widetilde{h}$ of $|h'|$. The new component operators are respectively denoted by $\mathcal{G}_{s,[i]}$ or $\mathcal{G}_{s,[i|j]}$

$$\mathcal{G}_{s,[i]}[f](z) := \widetilde{h}_i(z)^s \ f \circ h_i(z). \tag{32}$$

$$\mathcal{G}_{s,[i|j]}[f](z) := \widetilde{h}_{i|j}(z)^s \ f \circ h_{i|j}(z). \tag{33}$$

As in (29) or in (31), the Ruelle operators are now respectively defined by

$$\mathcal{G}_s = \sum_{i=1}^{r} \mathcal{G}_{s,[i]}, \qquad \mathcal{G}_s := \left(\mathcal{G}_{s,[i|j]}\right) \tag{34}$$

in the basic case, or in the Markovian case. The dynamics of the process is a priori described by $s = 1$ (i.e., $\mathcal{G} = \mathcal{G}_1$), but many other properties appear to be dependent upon complex values of s other than 1.

4.3. **Generalized Ruelle operators.** In [59, 61], Vallée has introduced a new tool, the generalized Ruelle operator, that involves *secants* of inverse branches instead of *tangents* $|h'(z)|$ of inverse branches. Here, we design a further generalization that involves *multi-secants*, so that the "hyper-generalized" operator, $\mathfrak{G}_s$, acts on a space of functions of m complex variables; it is then said to be an extension of degree m. In our applications, we shall need various extensions of degrees $m = 1, 2, 3$ or 4.

We first define two mappings that extend an inverse branch h into a mapping defined on $\mathbb{C}^m$. The first one that only depends on m is

$$V_m[h](x_1, \dots, x_m) = (h(x_1), \dots, h(x_m)).$$

The second one, denoted by $H_s[h]$, must satisfy the following three properties.

(i) H_s extends the tangent mapping and its restriction to the diagonal coincides with the tangent mapping.

$$H_s[h](x, x, \dots, x) = \widetilde{h}^s(x) \quad \text{and} \quad H_s[h](x) = \widetilde{h}^s(x) \text{ for the special case } m = 1.$$

(ii) H_s satisfies multiplicative properties in the style of the chain-rule,

$$H_s[h \circ g](x_1, x_2, \dots, x_m) = H_s[h](g(x_1), g(x_2), \dots, g(x_m)) \ H_s[g](x_1, x_2, \dots, x_m).$$

(*iii*) Finally, when the parameter s and the points x_i are all real, H_s possesses a positivity property

$$H_s[h](x_1, x_2, \dots, x_m) > 0 \text{ for real } s \text{ and real } x_i.$$

Particularly important realizations of H_s are the multi-secant mappings defined from any collection of affine functions $s \mapsto d_{i,j}(s)$

$$(35)\quad H_s[h](x_1, \cdots, x_m) := \prod_{1\le i,j\le m} \left|\frac{h(x_i) - h(x_j)}{x_i - x_j}\right|^{d_{i,j}(s)} \quad \text{with} \sum_{1\le i,j\le m} d_{i,j}(s) = s.$$

The previous properties $(i), (ii), (iii)$ are then clearly all satisfied.

Each component operator in (32) or (33) is now defined with the analytic extension $\widetilde{H}_s[h]$ of the multi-secant $H_s[h]$ relative to branch h. The new component operators are denoted by $\mathfrak{G}_{s,[i]}$ or $\mathfrak{G}_{s,[i|j]}$ respectively. They act on functions F of m (complex) variables as follows

$$\mathfrak{G}_{s,[i]}[F] := \widetilde{H}_s[h_i]\ F \circ V_m[h_i], \qquad \mathfrak{G}_{s,[i|j]}[F] := \widetilde{H}_s[h_{i|j}]\ F \circ V_m[h_{i|j}].$$

In a way analogous to (34), the generalized Ruelle operators are defined by

$$\mathfrak{G}_s := \sum_{i=1}^{r} \mathfrak{G}_{s,[i]}, \qquad \mathfrak{G}_s := \left(\mathfrak{G}_{s,[i|j]}\right)$$

in the basic case and in the Markovian case, respectively.

The generalized Ruelle operator constitutes an extension of the classical Ruelle operator in the following sense: If f is the diagonal of F, i.e., $f(u) := F(u, \cdots, u)$ (m times), the following relation holds on the diagonal $x_1 = \cdots = x_m = u$

$$(36)\qquad \mathfrak{G}_s[F](u, \cdots, u) = \mathcal{G}_s[f](u).$$

The case of dimension $m = 1$ corresponds exactly to the classical Ruelle operator $\mathcal{G}_s$.

We next show how the operators $\mathfrak{G}_s$ generate all the branches h of any depth, first in the case of a basic dynamical system (Section 4.4), then in the Markovian case (Section 4.5).

4.4. **The Dirichlet series: basic case.** By the multiplicative property (ii), the k-th iterate $\mathfrak{G}_s$ involves all the inverse branches h of depth k,

$$(37)\qquad \mathfrak{G}_s^k[F] = \sum_{|h|=k} \widetilde{H}_s[h]\ F \circ V_m,$$

where the function $\widetilde{H}_s[h]$ is the analytic extension of the multi-secant $H_s[h]$, and the sum now ranges over all inverse branches of depth k, or equivalently over all prefixes $\mathbf{w}$ of length k.

In the same vein, the quasi-inverse $(I - \mathfrak{G}_s)^{-1}$, being the formal sum of all the powers of the operator, then represents *all* the possible iterations, and is consequently expressed as a sum that ranges over all inverse branches

$$(I - \mathfrak{G}_s)^{-1}[F] = \sum_h \widetilde{H}_s[h]\ F \circ V_m[h].$$

From Section 3.5 and Equations (24), (25), (26), the Dirichlet series arising from the analysis of tries are

$$\Lambda^{\langle A\rangle}(F,s) = \sum_{\mathbf{w}\in\mathcal{M}^*} u_{\mathbf{w}}{}^s, \qquad \Lambda_k^{\langle A\rangle}(F,s) = \sum_{\mathbf{w}\in\mathcal{M}^k} u_{\mathbf{w}}{}^s$$

$$\Lambda^{\langle B\rangle}(F,s) = 2\sum_{\mathbf{w}\in\mathcal{M}^*}\sum_{\substack{(i,j)\in\mathcal{M}^2\\ i<j}} u_{\mathbf{w}\cdot i}\ u_{\mathbf{w}\cdot j}\ U_{h\cdot[i,j]}{}^{s-2},$$

$$\Lambda^{\langle L\rangle}(F,s) = \sum_{\mathbf{w}\in\mathcal{M}^*}\sum_{i\in\mathcal{M}} U_{\mathbf{w}\cdot[>i]}\ u_{\mathbf{w}\cdot i}{}^{s-1}.$$

The first two series correspond to the Dirichlet series of fundamental measures defined in Section 1 and they already play a central rôle in a large number of analyses involving prefixes of words; see [61]. They can be generated by the operator relative to the case $m = 2$ and the "true" secant

$$H_s^{\langle A\rangle}[h](x_1,x_2) = \left|\frac{h(x_1)-h(x_2)}{x_1-x_2}\right|^s, \tag{38}$$

applied to the function $L_s^{\langle A\rangle} := H_s^{\langle A\rangle}[F]$ which is the secant of the initial distribution F. More precisely,

$$\Lambda_k^{\langle A\rangle}(F,s) \ := \ \sum_{|h|=k} |F(h(0))-F(h(1))|^s = \left(\mathfrak{G}_s^{\langle A\rangle}\right)^k [L_s^{\langle A\rangle}](0,1), \tag{39}$$

$$\Lambda^{\langle A\rangle}(F,s) \ := \ \sum_{h} |F(h(0))-F(h(1))|^s = (I-\mathfrak{G}_s^{\langle A\rangle})^{-1}[L_s^{\langle A\rangle}](0,1), \tag{40}$$

with

$$\mathfrak{G}_s^{\langle A\rangle}[F] := \sum_{|h|=1} H_s^{\langle A\rangle}[h]\ F\circ V_2.$$

The other two series, $\Lambda^{\langle L\rangle}, \Lambda^{\langle B\rangle}$, arise from hybrid tries where order between symbols matters. There is an order on the alphabet; there is also the natural order on the topological partition $\{\mathcal{I}_m\}$ induce by the order on the $(0,1)$ interval. We assume that these orders are either identical or the reverse of each other. The measure $U_{\mathbf{w}\cdot[i,j]}$ is then the measure of an interval. When the inverse branch h is relative to the prefix $\mathbf{w}$, the two end points are to be chosen amongst the four points $h\circ h_i(0), h\circ h_i(1), h\circ h_j(0), h\circ h_j(1)$. This choice depends on the signs $\epsilon(h), \epsilon(h_i), \epsilon(h_j)$ of the derivatives of the inverse branches h, h_i, h_j. The measure $U_{\mathbf{w}\cdot[>i]}$ is also the measure of an interval. The two endpoints are to be chosen amongst the four points $h\circ h_i(0), h\circ h_i(1), h(0), h(1)$. This choice depends on signs $\epsilon(h), \epsilon(h_i)$.

Thus, each term of the Dirichlet series relative to hybrid tries involves either three points (in the case of list-tries) or four points (in the case of bst-tries). Accordingly, the corresponding Ruelle operators act on functions of three variables ($m = 3$) in the case of the list-trie and on on functions of four variables ($m = 4$) in the case of the bst-trie. We denote them respectively by $\mathfrak{G}_s^{\langle L\rangle}$ and $\mathfrak{G}_s^{\langle B\rangle}$. They are defined through the multi-secants $H_s^{\langle B\rangle}$ and $H_s^{\langle L\rangle}$

$$H_s^{\langle B\rangle}[h](x_1,x_2,x_3,x_4) = \left|\frac{h(x_1)-h(x_2)}{x_1-x_2}\right| \cdot \left|\frac{h(x_3)-h(x_4)}{x_3-x_4}\right| \cdot \left|\frac{h(x_1)-h(x_4)}{x_1-x_4}\right|^{s-2}$$

$$H_s^{\langle L\rangle}[h](x_1,x_2,x_3) = \left|\frac{h(x_2)-h(x_3)}{x_2-x_3}\right| \cdot \left|\frac{h(x_1)-h(x_2)}{x_1-x_2}\right|^{s-1}, \tag{41}$$

so that

$$\mathfrak{G}_s^{\langle B\rangle}[F] = \sum_{|h|=1} H_s^{\langle B\rangle}[h]\ F\circ V_4[h], \qquad \mathfrak{G}_s^{\langle L\rangle}[F] = \sum_{|h|=1} H_s^{\langle L\rangle}[h]\ F\circ V_3[h]. \tag{42}$$

The particular functions used to generate the Dirichlet series are the secants relative to initial distribution F,

$$L_s^{\langle B\rangle} = H_s^{\langle B\rangle}[F], \qquad L_s^{\langle L\rangle} = H_s^{\langle L\rangle}[F].$$

As opposed to $\Lambda^{\langle A\rangle}(F,s)$, the series $\Lambda^{\langle B\rangle}(F,s)$ and $\Lambda^{\langle L\rangle}(F,s)$ involve sums of subdivisions of fundamental intervals. The fundamental canonical measures $u_i^\star$ of depth one and relative to a uniform density will play an important rôle in the expression of both Dirichlet series. They are defined by

$$u_i^\star := |h_i(0)-h_i(1)|,$$

and, in a way consistent with previous notations, we set $U^\star_{[i,j]} = \sum_{k=i}^{j} u_k^\star$ and $U^\star_{[>i]} = \sum_{k>i} u_k^\star$.

In the homoclinal case with all branches being increasing (i.e. $\epsilon(h)=1$), one has

$$\Lambda^{\langle L\rangle}(F,s) = \sum_i U^\star_{[>i]} u_i^\star \left(I-\mathfrak{G}_s^{\langle L\rangle}\right)^{-1}\left[L_s^{\langle L\rangle}\right](h_i(0),h_i(1),1), \tag{43}$$

$$\Lambda^{\langle B\rangle}(F,s) = 2\sum_{i<j} u_i^\star\ u_j^\star\ {U^\star_{[i,j]}}^{s-2} \left(I-\mathfrak{G}_s^{\langle B\rangle}\right)^{-1} \left[L_s^{\langle B\rangle}\right](h_i(0),h_i(1),h_j(0),h_j(1)). \tag{44}$$

Heteroclinal case. In the heteroclinal case, where some branches increase and other decrease, there are eight possible expressions for the measure $U_{\mathbf{w}\cdot[i,j]}$, depending on the values of $\epsilon(h),\epsilon(h_i),\epsilon(h_j)$, and four possible expressions for the measure $U_{\mathbf{w}\cdot[>i]}$, depending on the values of $\epsilon(h)$ and $\epsilon(h_i)$. So, we split the symbols i of the alphabet $\mathcal{M}$ into two subsets, depending of the possible values of $\epsilon(h_i)$, and we split the set of all possible branches h (relative to all possible prefixes $\mathbf{w}$) into two subsets, corresponding to increasing branches and decreasing branches; we wish to generate separately these two subsets.

For this last splitting, we introduce the signed operator $\widetilde{\mathfrak{G}}_s$ whose component operators are defined from the sign $\epsilon(h)$ of the derivative h',

$$\widetilde{\mathfrak{G}}_{s,[i]} := \epsilon(h_i)\ \mathfrak{G}_{s,[i]}. \tag{45}$$

The multiplicativity of ϵ entails that

$$\widetilde{\mathfrak{G}}_s^k = \sum_{|h|=k} \epsilon(h)\ \mathfrak{G}_{s,h} \tag{46}$$

so that the two relations

$$(47)\qquad \begin{aligned} 2\sum_{\epsilon(h)=1} \mathfrak{G}_{s,h} &= (I-\mathfrak{G}_s)^{-1}[F] + (I-\widetilde{\mathfrak{G}}_s)^{-1}[F], \\ 2\sum_{\epsilon(h)=-1} \mathfrak{G}_{s,h} &= (I-\mathfrak{G}_s)^{-1}[F] - (I-\widetilde{\mathfrak{G}}_s)^{-1}[F] \end{aligned}$$

show that the previous expressions (43) (44) can be extended to the heteroclinal case, with expressions that now involve the two quasi–inverses $(I-\mathfrak{G}_s)^{-1}$ and $(I-\widetilde{\mathfrak{G}}_s)^{-1}$.

In particular, the "part" of each Dirichlet series that involves the quasi-inverse $(I-\mathfrak{G}_s)^{-1}$ can be made explicit. We associate to $\epsilon(h_i)\in\{-1,1\}$ the quantity $\beta(i) := (1-\epsilon(h_i))/2 \in \{0,1\}$. In the case of bst-tries, we consider 4–tuples and operate with the group $\mathcal{B}$ of order 4 generated by the two transpositions τ and ρ acting on the first two components and on the last two components, respectively. We denote by $\mathcal{B}(i,j)$ the subset of $\mathcal{B}$ that contains the two elements $\tau^{\beta(i)}\circ\rho^{\beta(j)}$ and $\tau^{1-\beta(i)}\circ\rho^{1-\beta(j)}$. In the case of list-tries, we consider 3–tuples of the form (a,b,c) with $c\in\{0,1\}$ and operate within the group $\mathcal{L}$ of order 4 generated by the two transpositions τ and μ which respectively act on the first two components, and exchanges the two possible values of the last one. We denote by $\mathcal{L}(i)$ the subset of $\mathcal{L}$ that contains the two elements $\tau^{\beta(i)}\circ\mu$ and $\tau^{1-\beta(i)}$. Then, the "part" of each Dirichlet series which involves the quasi-inverse $(I-\mathfrak{G}_s)^{-1}$ equals respectively

$$(48)\qquad \frac{1}{2}\sum_i U^\star_{[>i]}\; {u_i^\star}^{s-1} \sum_{r\in\mathcal{L}(i)} \left(I-\mathfrak{G}_s^{\langle L\rangle}\right)^{-1}\left[L_s^{\langle L\rangle}\right]\; r(h_i(0),h_i(1),1),$$

in the list-trie case, and

$$(49)\quad \sum_{i<j} u_i^\star\; u_j^\star\; {U^\star_{[i,j]}}^{s-2} \sum_{r\in\mathcal{B}(i,j)} \left(I-\mathfrak{G}_s^{\langle B\rangle}\right)^{-1}\left[L_s^{\langle B\rangle}\right] r(h_i(0),h_i(1),h_j(0),h_j(1))$$

in the bst-trie case.

4.5. **The Dirichlet series: Markovian case.** In the Markovian case, the coefficient (i,j) of the k–th iterate of matrix $\mathfrak{G}_s$ involves all the branches h relative to a word $\mathbf{w}=(m_1,\dots m_k)$ which starts with j $(m_1=j)$ and ends with i $(m_k=i)$.

For the Dirichlet series relative to array-tries, we wish to generate all the inverse branches of depth k. We first consider the operator $\mathfrak{M}_s$ relative to the initial dynamical system S_0

$$\mathfrak{M}_s[F] := \begin{pmatrix} \mathfrak{G}_{s,[1|0]} \\ \mathfrak{G}_{s,[2|0]} \\ \vdots \\ \mathfrak{G}_{s,[i|0]} \\ \vdots \\ \mathfrak{G}_{s,[r|0]} \end{pmatrix} [F].$$

If $\mathbf{e}$ denotes the unit r-th dimensional vector, i.e., ${}^t\mathbf{e}=(1,1,\dots 1)$ (r times), then ${}^t\mathbf{e}\;\mathfrak{M}_s$ denotes the generalized Ruelle operator associated to S_0. We let

$$\mathfrak{G}_s := {}^t\mathbf{e}\;\mathfrak{M}_s.$$

For $k \geq 2$, the operators ${}^t\mathbf{e}\ \mathfrak{G}_s^{k-1}\ \mathfrak{M}_s$ generate all the inverse branches of depth k and we obtain the analog of (37)

$$\begin{aligned} {}^t\mathbf{e}\ \mathfrak{G}_s^{k-1}\ \mathfrak{M}_s\ [F] &= \sum_{|h|=k} \widetilde{H}_s[h]\ F \circ V_m[h], \\ \left[I + {}^t\mathbf{e}\,(I - \mathfrak{G}_s)^{-1}\,\mathfrak{M}_s\right][F] &= \sum_h \widetilde{H}_s[h]\ F \circ V_m[h]. \end{aligned}$$

We then deduce the formulae for the Dirichlet series relative to the array-tries

$$\Lambda^{\langle A\rangle}(F,s) = \left[I + {}^t\mathbf{e}\left(I - \mathfrak{G}_s^{\langle A\rangle}\right)^{-1}\mathfrak{M}_s\right][L_s^{\langle A\rangle}](0,1), \tag{50}$$

$$\Lambda_k^{\langle A\rangle}(F,s) = {}^t\mathbf{e}\left(\mathfrak{G}_s^{\langle A\rangle}\right)^{k-1}\mathfrak{M}_s\ [L_s^{\langle A\rangle}](0,1) \tag{51}$$

In the same vein, if $\mathbf{e}_\ell$ is the ℓ-th vector of the canonical basis, the operator

$${}^t\mathbf{e}_\ell\ \mathfrak{G}_s^{k-1}\ \mathfrak{M}_s$$

generates all the inverse branches of depth k relative to prefixes $\mathbf{w}$ that end with symbol ℓ. In the case when all the branches are increasing, the expressions of the Dirichlet series relative to list-tries or bst-tries involve the canonical fundamental measure extended to the Markovian case by

$$u_{i|\ell}^\star := |h_{i|\ell}(0) - h_{i|\ell}(1)|,$$

and, consistently with previous notations, we set $U_{[i,j]|\ell}^\star = \sum_{k=i}^j u_{k|\ell}^\star$ and $U_{[>i]|\ell}^\star = \sum_{k>i} u_{k|\ell}^\star$.

$$\begin{aligned} \Lambda^{\langle B\rangle}(F,s) = 2\sum_{\substack{\ell,i,j\\ i>j}} u_{i|\ell}^\star u_{j|\ell}^\star {U_{[i,j]|\ell}^\star}^{s-2} &\left[I + {}^t\mathbf{e}_\ell\left(I - \mathfrak{G}_s^{\langle B\rangle}\right)^{-1}\mathfrak{M}_s\right] \\ &[L_s^{\langle B\rangle}](h_{i|\ell}(0), h_{i|\ell}(1), h_{j|\ell}(0), h_{j|\ell}(1)), \end{aligned} \tag{52}$$

$$\begin{aligned} \Lambda^{\langle L\rangle}(F,s) = \sum_{\ell,i} U_{[>i]|\ell}^\star u_{i|\ell}^\star &\left[I + {}^t\mathbf{e}_\ell\left(I - \mathfrak{G}_s^{\langle L\rangle}\right)^{-1}\mathfrak{M}_s\right] \\ &[L_s^{\langle L\rangle}](h_{i|\ell}(0), h_{i|\ell}(1), 1). \end{aligned} \tag{53}$$

Heteroclinal case. In the general case when the branches may be increasing or decreasing, we consider the signed operators $\widetilde{\mathcal{G}}_s$ or $\widetilde{\mathfrak{G}}_s$ whose component operators are defined from the sign $\epsilon(h)$ of the derivative h',

$$\widetilde{\mathcal{G}}_{s,[i|j]} := \epsilon(h_{i|j})\ \mathcal{G}_{s,[i|j]}, \qquad \widetilde{\mathfrak{G}}_{s,[i|j]} := \epsilon(h_{i|j})\ \mathfrak{G}_{s,[i|j]}.$$

The multiplicativity of ϵ entails formulae as in (46) and (47). Then, the operators $\widetilde{\mathfrak{G}}_s$ intervene, via their quasi–inverses, in the expressions of the Dirichlet series relative to list-tries or bst-tries, together with the quasi -inverses of the operators $\mathfrak{G}_s$.

4.6. **Asymptotic analysis.** Asymptotic analysis of coefficients of these Dirichlet series is dependent on spectral properties of the (generalized) Ruelle operator that generate them. In the study of the height, the asymptotic behaviour of $\Lambda_k^{\langle A\rangle}$ (as k tends to infinity) is related to dominant spectral properties of the Ruelle operator $\mathfrak{G}_s^{\langle A\rangle}$. For size or path length, the Dirichlet series involve quasi-inverses $(I-\mathfrak{G}_s)^{-1}$ and $(I-\widetilde{\mathfrak{G}}_s)^{-1}$ of various Ruelle operators $\mathfrak{G}_s$ and signed Ruelle operators $\widetilde{\mathfrak{G}}_s$. Asymptotic analysis of coefficients of these Dirichlet series is dependent on the location of its poles. Such poles arise from values of s where $(I-\mathfrak{G}_s)^{-1}$ or $(I-\widetilde{\mathfrak{G}}_s)^{-1}$ is singular, that is, values for which 1 is an eigenvalue of $\mathfrak{G}_s$ or $\widetilde{\mathfrak{G}}_s$. In this way, the poles of Dirichlet series also relate to the spectral properties of the transfer operators.

5. Generalized Ruelle operators: analytic properties

The generalized multi-secant operators belong to the category of nuclear operators (defined in Section 5.1) which, in particular, have a discrete spectrum. Such infinite dimensional operators in many ways "behave like" finite matrices, with well-defined notions of trace and determinant attached to them. First, we introduce composition operators that are the basic building blocks of transfer operators (Section 5.2) and have an explicitly characterized spectrum (Section 5.3). Transfer operators, including the multi-secant variety, then have spectra that are determined from those of composition operators by means of trace formulae (Section 5.4). On the other hand, the generalized operators possess strong positivity properties of the Perron-Frobenius type for real values of parameter s. Such properties entail the existence of dominant (positive) spectral objects (Section 5.5) that can be eventually related to the basic parameters of the source, the entropy and the coincidence probability (Section 5.8). A related "quasi-power property" gives access to the behaviour of operators under iteration (Sections 5.6). Other properties can then be transferred to the Dirichlet series of fundamental intervals of interest for the analysis of tries (Section 5.7, 5.8, 5.9).

5.1. **Nuclearity, trace formula, and Fredholm determinant.** We first recall the notion of nuclearity introduced by Grothendieck [23, 24]. Let B be a Banach space and $B^\star$ its dual space. An operator $\mathcal{M}: B \to B$ is nuclear of order 0 if it admits a representation

$$\mathcal{M}[f] = \sum_{i\in I} \mu_i \; e_i^\star(f) \; e_i \quad \text{for all } f \in B,$$

with $e_i \in B$, $e_i^\star \in B^\star$ such that $||e_i|| = ||e_i^\star|| = 1$ and the μ_i are p-summable for all $p > 0$ (i.e., $\sum |\mu_i|^p < +\infty$). Most of matrix algebra extends to such operators; in particular, one can define the trace,

$$\operatorname{Tr} \mathcal{L} = \sum_{i\in I} \mu_i \; e_i^\star(e_i), \quad \text{also equal to} \quad \operatorname{Tr} \mathcal{L} = \sum_{i\in I} \lambda_i,$$

where the λ_i's are the eigenvalues of $\mathcal{L}$, counted with their algebraic multiplicities. The traces of the iterates of $\mathcal{L}$ are also well-defined, as is the analogue of the characteristic polynomial known as the Fredholm determinant,

$$F(\mathcal{L}, u) := \det(I - u\mathcal{L}) := \prod_{i\in I} (1 - \lambda_i u),$$

where the λ_i's are the eigenvalues of $\mathcal{L}$, (again counted with their algebraic multiplicities). There exists an important relation between the Fredholm determinant and traces of iterates,

$$\det(I - u\mathcal{L}) = \exp\left[\operatorname{Tr}\log(I - u\mathcal{L})\right] = \exp\left[-\sum_{k=1}^{\infty}\frac{u^k}{k}\operatorname{Tr}\,\mathcal{L}^k\right] \tag{54}$$

that extends the familiar Jacobi formula of matrix theory: $\log\circ\det = \operatorname{Tr}\circ\log$. These properties give access to spectral properties of nuclear operators of order 0.

5.2. **Composition operators.** Each component operator $\mathcal{G}_{s,h}$ defined by

$$\mathcal{G}_{s,h}[f] := \widetilde{h}^s\ f\circ h$$

is known as a composition operator. We recall that each branch h satisfies contraction properties (d_1) and (d_2) of Definition 1: there exists a suitable neighbourhood $\mathcal{V}$ of $\overline{\mathcal{I}}$ such that h and $|h'|$ extend to analytics map on $\mathcal{V}$; h maps the closure $\overline{\mathcal{V}}$ of $\mathcal{V}$ inside $\mathcal{V}$; there exists $\delta < 1$ for which $0 < |\widetilde{h}(z)| \le \delta$ for all $z \in \mathcal{V}$.

Then, the operator $\mathcal{G}_{s,h}$ acts on the space $A_\infty(\mathcal{V})$ formed with all functions f that are holomorphic in the domain $\mathcal{V}$ and are continuous on the closure $\overline{\mathcal{V}}$. Endowed with the sup-norm,

$$||f|| = \sup\ \{|f(u)|;\ u \in \overline{\mathcal{V}}\},$$

$A_\infty(\mathcal{V})$ is a Banach space. Such operators have been studied in an extensive way by several authors (Schwartz [50], Shapiro and Taylor [53], Shapiro [52]).

Since each branch h satisfies assumptions (d_1) and (d_2) of Definition 1, the generalized component operator $\mathfrak{G}_{s,h}$ acts on the space $\mathfrak{A}_\infty(\mathcal{V})$ formed with all functions F that are holomorphic in the domain $\mathcal{V}^m$ and are continuous on the closure $\overline{\mathcal{V}}^m$. Endowed with the sup-norm,

$$||F|| = \sup\ \{|F(u_1, u_2, \ldots u_m)|;\ (u_1, u_2, \ldots u_m) \in \overline{\mathcal{V}}^m\},$$

$\mathfrak{A}_\infty(\mathcal{V})$ is a Banach space.

5.3. **Spectra of composition operators.** All the extended composition operators $\mathfrak{G}_{s,h}$ relative to the same branch h have the same eigenvalues. However, the multiplicities of these eigenvalues depend on the degree m of the extension.

Proposition 2. *Let $\alpha(h) = \widetilde{h}(\underline{h})$ denote the value of $\widetilde{h}$ at the fixed point $\underline{h}$ of h. Then the spectrum of the operator $\mathfrak{G}_{s,h}$ is formed with the eigenvalues $\mu_\ell, (\ell \ge 0)$,*

$$\mu_\ell := \alpha(h)^s[\epsilon(h)\alpha(h)]^\ell. \tag{55}$$

Each eigenvalue μ_ℓ appears in $\operatorname{Sp}\mathfrak{G}_{s,h}$ with a multiplicity $\binom{\ell+m-1}{m-1}$. Consequently, a trace formula for $\mathfrak{G}_{s,h}$ holds

$$\operatorname{Tr}\mathfrak{G}_{s,h} = \frac{\alpha(h)^s}{(1-\epsilon(h)\alpha(h))^m}.$$

Proof. The proof uses a theorem due to Mayer [42]. Here, $\mathsf{Multi}\,[A]$ denotes the set of multisets built over a multiset A. More precisely, for a multiset $A = \{a_1, a_2, a_3, \ldots, a_r\}$, (with possible equality between symbols a_i), one has

$$\mathsf{Multi}\,[A] = \prod_{1\le i\le r} a_i^*.$$

Theorem (Mayer). *Let Ω be an open domain of $\mathbb{C}^m$, and let $B_\infty(\Omega)$ be the set of functions that are holomorphic on Ω and continuous on $\overline{\Omega}$. Let $\mathcal{R}_\phi$ be defined on $\mathcal{B}_\infty(\Omega)$ by $\mathcal{R}_\phi[f] := \phi\ f \circ \psi$, where ϕ is in $B_\infty(\Omega)$ and ψ strictly maps Ω on itself. The spectrum of $\mathcal{R}_\phi$ is*

$$\mathrm{Sp}(\mathcal{R}_\phi) = \phi(\underline{z})\ .\ \mathsf{Multi}\ [\mathrm{Sp}(\psi'(\underline{z})]$$

where $\underline{z}$ is the unique fixed point of ψ inside Ω and ψ' is the differential of ψ.

Since the components operators $\mathfrak{G}_{s,h}$ fulfill the conditions of the theorem, their spectra are then precisely determined. Note that, for purposes of trie analysis, all the cases $m = 1, 2, 3, 4$ are of interest.

In the case $m = 1$ when the operator $\mathfrak{G}_{s,h}$ coincides with the usual Ruelle operator $\mathcal{G}_{s,h}$, the previous theorem applies with $\phi = \widetilde{h}^s, \psi = h$, and $\mathrm{Sp}(\psi'(\underline{z}) = \{h'(\underline{h})\} = \{\epsilon(h)\alpha(h)\}$, so that

$$\mathrm{Sp}\,\mathcal{G}_{s,h} = \{\alpha(h)^s(\epsilon(h)\alpha(h))^\ell \mid \ell \in \mathbb{N}\}$$

is exactly the set formed with elements μ_ℓ defined in (55).

In the general case when the operator $\mathfrak{G}_{s,h}$ acts on m-variables spaces, Mayer's theorem applies with

$$\phi = \widetilde{H}_s[h], \qquad \psi = V_m[h],$$

so that $\psi'(x_1, \ldots, x_m)$ is the diagonal matrix with coefficients $h'(x_1), h'(x_2), \ldots h'(x_m)$. The fixed point of $V_m[h]$ is the point $(\underline{h}, \underline{h}, \ldots, \underline{h})$, so that

$$\mathrm{Sp}\,V_m'[h]\ (\underline{h}, \underline{h}, \ldots, \underline{h}) = \{h'(\underline{h})\}^{[m]},$$

where the symbol $A^{[m]}$ denotes the multiset obtained by repeating m times each element of the (multi)set A. Then, the multiplicity of the eigenvalue μ_ℓ in spectra of $\mathfrak{G}_{s,h}$ equals the number of words of size ℓ in the language $a_1{}^*a_2{}^* \cdots a_m{}^*$, that is

$$[z^\ell]\ \frac{1}{(1-z)^m} = \binom{\ell+m-1}{m-1}.$$

Accordingly, these relations entail trace formulae involving the quantity $\alpha(h) = \widetilde{h}(\underline{h})$

$$\mathrm{Tr}\,\mathfrak{G}_{s,h} = \frac{\alpha(h)^s}{(1-\epsilon(h)\alpha(h))^m}.$$

□

Thus, the spectra of $\mathcal{G}_{s,h}$ and $\mathfrak{G}_{s,h}$ contain the same elements. We now compare them more precisely, via their trace formulae. With the notation $p_m(\ell) = \binom{\ell+m-2}{m-2}$, the generalized trace can be expressed in terms of the trace of the classical operator,

$$\mathrm{Tr}\,\mathfrak{G}_{s,h} = \frac{\mathrm{Tr}\,\mathcal{G}_{s,h}}{(1-\epsilon(h)\alpha(h))^{m-1}} = \sum_{\ell \geq 0} p_m(\ell)\ (\epsilon(h)\alpha(h))^\ell\,\mathrm{Tr}\,\mathcal{G}_{s,h}.$$

Consider again the signed operators $\widetilde{\mathcal{G}}_{s,h} = \epsilon(h)\ \mathcal{G}_{s,h}$ and $\widetilde{\mathfrak{G}}_{s,h} = \epsilon(h)\ \mathfrak{G}_{s,h}$ already introduced in (45). Then the trace of the generalized operators $\mathfrak{G}_{s,h}$, $\widetilde{\mathfrak{G}}_{s,h}$ are expressed with the traces of operators $\mathcal{G}_{s+\ell,h}$, $\widetilde{\mathcal{G}}_{s+\ell,h}$,

$$\operatorname{Tr}\mathfrak{G}_{s,h} = \sum_{\substack{\ell \text{ even} \\ \ell\geq 0}} p_m(\ell)\operatorname{Tr}\mathcal{G}_{s+\ell,h} + \sum_{\substack{\ell \text{ odd} \\ \ell\geq 0}} p_m(\ell)\operatorname{Tr}\widetilde{\mathcal{G}}_{s+\ell,h} \tag{56}$$

$$\operatorname{Tr}\widetilde{\mathfrak{G}}_{s,h} = \sum_{\substack{\ell \text{ even} \\ \ell\geq 0}} p_m(\ell)\operatorname{Tr}\widetilde{\mathcal{G}}_{s+\ell,h} + \sum_{\substack{\ell \text{ odd} \\ \ell\geq 0}} p_m(\ell)\operatorname{Tr}\mathcal{G}_{s+\ell,h} \tag{57}$$

5.4. **Spectra of transfer operators.** We first make precise the functional space to which the $\mathfrak{G}_s$ operator is applied. Here, the operator is determined by an integer m and a multi-secant $\widetilde{H}_s$. In the Markovian case, we restrict ourselves to finite alphabets, but allow infinite (denumerable) alphabets in the basic case. In this situation, the possibility of choosing the same open set $\mathcal{V}$ for all branches h in conjunction with the convergence condition (d_3) entails "nice" properties for the Ruelle operator $\mathfrak{G}_s$ when s belongs to the half-plane $\mathrm{Re}(s) > \gamma$. We denote by $\mathcal{J}$ the intersection of $\mathcal{V}$ with the real axis. The secant mapping $\widetilde{H}_s[h]$ defined in (35) has a strictly positive real part on $\mathcal{V}^m$, and the operator $\mathfrak{G}_s$ is well defined for any complex s in the half-plane $\mathrm{Re}(s) > \gamma$.

In the basic case, the $\mathfrak{G}_s$ operators are then taken to act on the space $\mathfrak{A}_\infty(\mathcal{V})$ defined as the set of functions that are holomorphic in the domain $\mathcal{V}^m$ and are continuous on the closure $\overline{\mathcal{V}}^m$, endowed with the sup-norm. In the Markovian case, the $\mathfrak{G}_s$ operators are taken to act on the space $\mathfrak{A}_\infty(\mathcal{V})^r$. Both functional spaces are Banach spaces. Since the component operators $\mathfrak{G}_{s,h}$ are nuclear of order 0, the operators $\mathfrak{G}_s$ are nuclear of order 0. In particular they are bounded and compact. So, their spectra are discrete.

The *signed* operators $\widetilde{\mathfrak{G}}_s$ are now defined from the signed component operators $\widetilde{\mathfrak{G}}_{s,h} = \epsilon(h)\ \mathfrak{G}_{s,h}$,

$$\widetilde{\mathfrak{G}}_s := \sum_{i\in\mathcal{M}} \widetilde{\mathfrak{G}}_{s,[i]} \qquad \widetilde{\mathfrak{G}}_s := (\widetilde{\mathfrak{G}}_{s,[i|j]})$$

The multiplicativity of ϵ entails equalities for the powers of $\widetilde{\mathfrak{G}}_s$ similar to (45) and (46).

The following proposition relates the spectrum of the generalized operator to the spectra of the classical ones.

Proposition 3. *The spectra of the generalized Ruelle operators $\mathfrak{G}_s$, $\widetilde{\mathfrak{G}}_s$ are related to the spectra of the classical Ruelle operators $\mathcal{G}_s$, $\widetilde{\mathcal{G}}_s$*

$$\operatorname{Sp}\mathfrak{G}_s = \left(\bigcup_{\substack{\ell \text{ even} \\ \ell\geq 0}} (\operatorname{Sp}\mathcal{G}_{s+\ell})^{[p_m(\ell)]}\right) \bigcup \left(\bigcup_{\substack{\ell \text{ odd} \\ \ell\geq 0}} \left(\operatorname{Sp}\widetilde{\mathcal{G}}_{s+\ell}\right)^{[p_m(\ell)]}\right).$$

$$\operatorname{Sp}\widetilde{\mathfrak{G}}_s = \left(\bigcup_{\substack{\ell \text{ even} \\ \ell\geq 0}} \left(\operatorname{Sp}\widetilde{\mathcal{G}}_{s+\ell}\right)^{[p_m(\ell)]}\right) \bigcup \left(\bigcup_{\substack{\ell \text{ odd} \\ \ell\geq 0}} (\operatorname{Sp}\mathcal{G}_{s+\ell})^{[p_m(\ell)]}\right),$$

where $p_m(\ell) = \binom{l+m+2}{m-2}$. Here, the union is taken in the sense of multisets and the notation $A^{[p]}$ denotes the multi-set obtained by repeating p times each element of the (multi)-set A.

Proof. In the Markovian case, the trace of the k-th iterate of $\mathcal{G}_s$ (resp. $\mathfrak{G}_s$) equals the sum of the trace of the diagonal elements of the matrix $\mathcal{G}_s^k$ (resp. $\mathfrak{G}_s^k$); such diagonal elements only involve inverse branches h that are cyclic: these are branches whose associated word starts and ends with the same symbol. Finally, in both cases, the trace formulae involve the set $\mathcal{C}_k$ which is the set of all the inverse branches of depth k (in the basic case) or the set of the cyclic inverse branches of depth k (in the Markov case),

$$\operatorname{Tr} \mathcal{G}_s^k = \sum_{h \in \mathcal{C}_k} \operatorname{Tr} \mathcal{G}_{s,h}, \qquad \operatorname{Tr} \mathfrak{G}_s^k = \sum_{h \in \mathcal{C}_k} \operatorname{Tr} \mathfrak{G}_{s,h}.$$

Then, relations (56) and (57) extend to powers of transfer operators,

$$\operatorname{Tr} \mathfrak{G}_s^k \quad = \quad \sum_{\substack{\ell \text{ even} \\ \ell \geq 0}} p_m(\ell) \operatorname{Tr} \mathcal{G}_{s+\ell}^k + \sum_{\substack{\ell \text{ odd} \\ \ell \geq 0}} p_m(\ell) \operatorname{Tr} \widetilde{\mathcal{G}}_{s+\ell}^k.$$

Now, via the trace formulae (54), the Fredholm determinant $\mathfrak{F}(s,u) := \det(I - u\mathfrak{G}_s)$ can be expressed in terms of the traces of the powers of the operators, and we obtain

$$\log \det(I - u\mathfrak{G}_s) = \sum_{\substack{\ell \text{ even} \\ \ell \geq 0}} p_m(\ell) \log \det(I - u\mathcal{G}_{s+\ell}) + \sum_{\substack{\ell \text{ odd} \\ \ell \geq 0}} p_m(\ell) \log \det(I - u\widetilde{\mathcal{G}}_{s+\ell}).$$

The Fredholm determinant of the generalized Ruelle operator $\mathfrak{G}_s$ acting on a functional space with m variables then satisfies

$$\det(I - u\, \mathfrak{G}_s) = \prod_{\substack{\ell \text{ even} \\ \ell \geq 0}} \left[\det(I - u\, \mathcal{G}_{s+\ell})\right]^{p_m(\ell)} \prod_{\substack{\ell \text{ odd} \\ \ell \geq 0}} \left[\det(I - u\, \widetilde{\mathcal{G}}_{s+p(\ell)})\right]^{p_m(\ell)}.$$

Finally one obtains the relation between spectra of $\mathfrak{G}_s$ and spectra of $\mathcal{G}_s$ and $\widetilde{\mathcal{G}}_s$. We proceed in the same way for the signed operator $\widetilde{\mathfrak{G}}_s$. ☐

5.5. **Dominant spectral properties for real s.** When $s = \sigma > \gamma$ is real, the operators $\mathcal{G}_s, \mathfrak{G}_s$ satisfy strong positivity properties related to the Perron-Frobenius theory [35]. The proof given in [41] can be easily extended to this generalized framework, using the positivity property (iii) of $\widetilde{H}_s$ in Section 4.3. We obtain:

> *For real $s > \gamma$, each operator $\mathfrak{G}_s$ has a unique dominant eigenvalue (i.e., an eigenvalue of largest modulus) $\lambda(s)$ that is is positive and has multiplicity 1.*

A priori, this dominant eigenvalue is dependent on the chosen extension. We prove in this section that all the extensions of $\mathcal{G}_s$ share the same dominant eigenvalue. Since each operator is compact, its spectrum is discrete and there is a "spectral gap" between the dominant eigenvalue and the remainder of the spectrum. This makes it possible to decompose $\mathfrak{G}_s$ as

$$\mathfrak{G}_s = \lambda(s)\mathfrak{P}_s + \mathfrak{N}_s.$$

Here, $\mathfrak{P}_s$ is the projection onto the dominant eigenspace, and $\mathfrak{N}_s$ is relative to the remainder of the spectrum, so that its spectral radius is strictly smaller than the dominant eigenvalue.

More generally $\mathfrak{G}_s^k$ decomposes as

$$\mathfrak{G}_s^k = \lambda(s)^k\,\mathfrak{P}_s + \mathfrak{N}_s^k. \tag{58}$$

This is true in particular for the classical Ruelle operator $\mathcal{G}_s$ whose dominant eigenvalue is denoted by $\lambda_1(s)$. The previous relations, together with positivity properties of the dominant eigensubspace projections, entail the equalities

$$\lambda(s) = \lim_{k\to\infty}\ \mathfrak{G}_s^k\,[1](0,\dots,0) \qquad \lambda_1(s) = \lim_{k\to\infty}\ \mathcal{G}_s^k\,[1](0).$$

Relation (36), expressing that $\mathfrak{G}_s$ is an extension of $\mathcal{G}_s$, entails the equality $\lambda(s) = \lambda_1(s)$; thus all the operators $\mathfrak{G}_s$ have the same dominant eigenvalue $\lambda(s)$.

The projection $\mathfrak{P}_s$ relative to the dominant subspace can be written as

$$\mathfrak{P}_s[G](x_1,\dots,x_m) = E_s[G]\ \Psi_s(x_1,\dots,x_m),$$

where Ψ_s is the dominant eigenfunction, and E_s a linear form. The extension formula (36) enables us to relate these dominant objects to the dominant objects relative to the classical Ruelle operator $\mathcal{G}_s$, namely, its dominant eigenfunction ψ_s and its dominant projector e_s,

$$\Psi_s(u,u,\dots,u) = \psi_s(u), \qquad E_s[G] = e_s[g] \text{ if } g \text{ is the diagonal of } G.$$

We have thus proved:

Proposition 4. *The multi-secant operators $\mathfrak{G}_s$ all have dominant spectral properties. They share the same dominant eigenvalue $\lambda(s)$, and the other dominant spectral objects (the dominant eigenvectors Ψ_s and the dominant projectors E_s) are extensions of the corresponding dominant spectral objects of the classical operator $\mathcal{G}_s$ $(m=1)$*

$$\Psi_s(u,\dots,u) = \psi_s(u), \qquad E_s[G] = e_s[g] \textit{ if } g \textit{ is the diagonal of } G.$$

5.6. **Quasi-Power Property.** By the classical theory of analytic perturbation [30], for s in a sufficiently small neighbourhood of any point σ of the real axis, unicity of the dominant eigenvalue is preserved, so that the mappings $s \mapsto \lambda(s)$, $s \mapsto \Psi_s$, $s \mapsto E_s$ define analytic functions in a neighbourhood of any point where $\lambda(s)$ is well defined. Then, by extending (58) in a neighbourhood of the real axis, one deduces that the k-th iterate of $\mathfrak{G}_s$ as in (39) behaves as a k-th power of the dominant eigenvalue $\lambda(s)$:

$$\mathfrak{G}_s[F](x_1,\dots,x_m) \sim \rho\cdot\lambda(\sigma)^k, \tag{59}$$

for some $\rho > 0$, under the assumption that F is positive on $[0,1]$ and the x_j satisfy $x_j \geq 0$. The Dirichlet series $\Lambda_k(F,s)$ associated to the array-trie is a special case that we state in a detailed form for future reference in the course of the proofs of Theorems 7 and 8 relative to trie height.

Proposition 5 (Quasi-power Property). *Let $\sigma > \gamma$ be real. For any distribution F associated to a density $f \in A_\infty(\mathcal{V})$ strictly positive on $\mathcal{J}$, there exists a positive constant ρ such that*

$$\rho = \lim_{k\to\infty} \frac{\Lambda_k^{\langle A\rangle}(F,\sigma)}{\lambda(\sigma)^k}. \tag{60}$$

Let $\mu(\sigma)$ be the modulus of a subdominant eigenvalue of $\mathfrak{G}_\sigma$, and ν any constant with $\nu > \mu(\sigma)$. Then, there exist three strictly positive constants α, β, δ, such that, for any $k \geq 1$, one has

$$\alpha\, \lambda(\sigma)^k \leq \Lambda_k^{\langle A \rangle}(F, \sigma) \leq \beta\, \lambda(\sigma)^k, \qquad |\Lambda_k^{\langle A \rangle}(F, \sigma) - \rho\, \lambda(\sigma)^k| \leq \delta \nu^k. \tag{61}$$

5.7. **Maximum properties of the dominant eigenvalue.** Because of the dominant spectral property, the dominant eigenvalue function $s \mapsto \lambda(s)$ plays a central rôle in our analyses. Here, we establish some important properties of this function. More generally, we study the spectral radius $\mathfrak{R}(s)$ of the operator $\mathfrak{G}_s$. At a point s in the half–plane $\mathrm{Re}(s) \geq \sigma$, we compare the spectral radii $\widetilde{R}(s)$ of $\widetilde{\mathcal{G}}_s$, $\mathfrak{R}(s)$ of $\mathfrak{G}_s$, $\widetilde{\mathfrak{R}}(s)$ of $\widetilde{\mathfrak{G}}_s$ and the spectral radius $\mathfrak{R}(\sigma) = \lambda(\sigma)$ of $\mathfrak{G}_\sigma$.

Proposition 6 (Maximum properties of the dominant eigenvalue). *Let γ be the constant referred to in the condition (d_3) of Definition 1. The following properties hold:*

- (i) *The function $s \mapsto \lambda(s)$ is strictly decreasing along the real axis $s > \gamma$.*
- (ii) *For any positive reals s, t such that $s > t$, one has $\lambda(s)^t < \lambda(t)^s$.*
- (iii) *On each vertical line* $\mathrm{Re}(s) = \sigma$, *the inequalities $\widetilde{R}(s) \leq \lambda(\sigma)$, $\mathfrak{R}(s) \leq \lambda(\sigma)$, $\widetilde{\mathfrak{R}}(s) \leq \lambda(\sigma)$ hold.*
- (iv) *In the half-plane* $\mathrm{Re}(s) > \sigma$, *the strict inequalities $\mathfrak{R}(s) < \lambda(\sigma)$, $\widetilde{\mathfrak{R}}(s) < \lambda(\sigma)$ hold.*
- (v) *If the equality $\mathfrak{R}(s) = \lambda(\sigma)$ holds for $s = \sigma + it, t \neq 0$, then $\mathfrak{G}_s$ has an eigenvalue $\lambda = e^{ia}\lambda(\sigma)$ that belongs to the spectrum of $\mathcal{G}_s$.*
- (v') *If the equality $\widetilde{\mathfrak{R}}(s) = \lambda(\sigma)$ holds for $s = \sigma + it, t \neq 0$, then $\widetilde{\mathfrak{G}}_s$ has an eigenvalue $\lambda = e^{ia}\lambda(\sigma)$ that belongs to the spectrum of $\widetilde{\mathcal{G}}_s$.*
- (vi) *Provided that all the ϵ are not equal, the strict inequality $\widetilde{\mathfrak{R}}(\sigma) < \lambda(\sigma)$ holds.*

Proof. The first two properties are proven in [61]. For properties $(iii), (iv), (v)$, we use the Spectrum formula of Proposition 3, together with the strict decrease of λ along the real axis and the inequality $\widetilde{R}(s) \leq \lambda(s)$, in a way similar to the proofs in [61]. Finally, Property (vi) is established by means of the converse of the triangular inequality, used in the same way as in the proof of Proposition 9 in [61]. ☐

5.8. **Special values.** For $s = 1$, the classical Ruelle operator $\mathcal{G}_s$ is well-defined (since, by (d_4) of Definition 1, one has $\gamma < 1$) and $\mathcal{G}_1$ is a density transformer. This property entails explicit evaluations of some of the spectral objects at $s = 1$.

Proposition 7 (Special values). *The dominant eigenvalue at $s = 1$ equals 1, the dominant projector e_1 of $\mathcal{G}_s$ at $s = 1$ satisfies $e_1[f] = \int_0^1 f(x)\, dx$. Let F be the distribution relative to density f on $[0, 1]$. Then, the residue at $s = 1$ of the quasi-inverse $(I - \mathfrak{G}_s)^{-1}[H_s[F]]$ is independent of F and only involves the dominant eigenfunction Ψ_s at $s = 1$ under the form*

$$(I - \mathfrak{G}_s)^{-1}[H_s[F]] \asymp \frac{-1}{\lambda'(1)(s-1)} \Psi_1 \qquad (s = 1).$$

Proof. Since fundamental intervals of depth k form a quasi-partition of the unit interval $\mathcal{I}$, there results the equality $\Lambda_k^{\langle A \rangle}(F, 1) = 1$ for any distribution function

F, and thus $\lambda(1) = 1$. Moreover, the operator $\mathcal{G}_1$ is a density transformer: for $f(x) > 0$ when x is real,

$$\int_0^1 \mathcal{G}_1^k[f](t)\,dt = \int_0^1 f(t)\,dt = e_1[f] \int_0^1 \psi_1(t)\,dt + O(\rho^k),$$

from which $e_1[f]$ is obtained, provided that ψ_1 is defined as a density function with the normalization condition $\int_0^1 \psi_1(t)\,dt = 1$. One deduces the expression of projector E_1 by the extension property given in Proposition 4. Then, since the diagonal mapping relative to $H_1[F]$ is exactly $F' = f$, one obtains $E_1[H_1[F]] = e_1[f] = 1$. □

We recall that entropy and coincidence probabilities are defined by (6), (7) in terms of limits involving Dirichlet series of fixed depth. Then, the Quasi-power property (60) provides expressions of entropy and coincidence probability that are seen to involve spectral objects for $s = 1$ and $s = 2$.

Proposition 8 (Entropy and coincidence probabilities). *The entropy of the source is equal to the opposite of the derivative of $s \mapsto \lambda(s)$ at $s = 1$, while the coincidence probability is equal to $\lambda(2)$.*

5.9. **Singularities of the quasi–inverses $(I - \mathfrak{G}_s)^{-1}$, $(I - \widetilde{\mathfrak{G}}_s)^{-1}$.** As explained in Section 3, it is necessary to locate precisely the poles of the various Dirichlet series $\Lambda(F, s)$. We recall that $\lambda(1) = 1$. Then, from Property (iv) of Proposition 6 the operators $I - \mathfrak{G}_s$, $I - \widetilde{\mathfrak{G}}_s$ are invertible in the half-plane $\mathrm{Re}(s) > 1$. Thus, the operator $(I - \mathfrak{G}_s)^{-1}$, is analytic there and has a simple pole at $s = 1$. Assertion (vi) of Proposition 6 implies that this is also the case at $s = 1$, except when all the ϵ are equal to 1 (i.e. the case $\widetilde{\mathfrak{G}}_s = \mathfrak{G}_s$). We focus on what may take place near the line $\mathrm{Re}(s) = 1, s \neq 1$ and we consider so-called *particular points*: they are points $s = 1 + it$, with $t \neq 0$ for which the spectrum of $\mathfrak{G}_s$ or $\widetilde{\mathfrak{G}}_s$ contains an eigenvalue equal to 1. Assertions (v) and (v') of Proposition 6 prove that, at these particular points, the spectrum of classical operators $\mathcal{G}_s$ or $\widetilde{\mathcal{G}}_s$ contains an eigenvalue equal to 1. The following result, which extends results of [14], [44], [59], gives a characterization of particular points and describes the two types of behaviour that may be encountered.

Proposition 9 (Singularities of quasi–inverses). *The operators $\mathfrak{G}_s$, $\widetilde{\mathfrak{G}}_s$ may only behave in two different ways on the line* $\mathrm{Re}(s) = 1$*:*

- (i) Aperiodic case. *There are no particular points, and the operators $I - \mathfrak{G}_s$, $I - \widetilde{\mathfrak{G}}_s$ are invertible in the punctured half-plane* $\mathrm{Re}(s) > 1, s \neq 1$.
- (ii) Periodic case. *There are particular points for the operator $\mathfrak{G}_s$. Then they are regularly distributed on the line, and of the form $1 + kit, k \in \mathbb{Z}$. The operator $I - \widetilde{\mathfrak{G}}_s$ may have possible particular points of the form $1 + i(2k+1)t/2, k \in \mathbb{Z}$. The operators $(I - \mathfrak{G}_s)^{-1}$, $(I - \widetilde{\mathfrak{G}}_s)^{-1}$ have a simple pole at each of their particular points, and there is a strip on the left of the line* $\mathrm{Re}(s) = 1$ *that is free of poles.*

Proof. The proof follows the same lines as in Proposition 9 of [61]. The existence of particular points for $\mathfrak{G}_s$ implies the equalities $\alpha(h)^{it} = 1$ for all inverse branch h. Then, there are two possible cases, depending on whether all the equalities

$\alpha(h)^{it/2} = \epsilon(h)$ hold or not. In the first case, the operator $\widetilde{\mathfrak{G}}_s$ has particular points at $s = 1 + i(2k+1)t/2$ for $k \in \mathbb{Z}$. In the second case, the operator $\widetilde{\mathfrak{G}}_s$ has no particular points. □

6. Asymptotic analysis of size and path length

We can now return to the analysis of trie parameters starting with the additive parameters of size and the various path lengths. We first introduce the main asymptotic tool, the Mellin transform, that relates poles of Dirichlet series and asymptotic forms of harmonic sums, including those expressing additive trie parameters. Then, we obtain our main results in the Poisson model. A further stage of "Dirichlet depoissonization" provides the corresponding asymptotic estimates in the Bernoulli model.

6.1. **Mellin transform.** The Mellin transform is the method of choice for deriving asymptotic expansions whenever harmonic sums are involved. We briefly review here the properties needed, following the survey article [18]. The Mellin transform is the integral transform defined by

$$g^*(s) := \int_0^\infty g(x)\; x^{s-1}\, dx.$$

It is defined on a strip $\mathrm{Re}(s) \in (\alpha, \beta)$ called the fundamental strip that is denoted by $\langle \alpha, \beta \rangle$. For instance, the Mellin transform of e^{-x} is the Euler gamma function $\Gamma(s)$ with fundamental strip $\langle 0, +\infty \rangle$. A harmonic sum is any sum of the form

$$G(x) = \sum_{k \in K} \lambda_k g(\mu_k x), \tag{62}$$

where the coefficients $\{\lambda_k\}$ and $\{\mu_k\}$ are called amplitudes and frequencies, and the function g is called the base function. The analysis of tries reduces essentially to asymptotic estimations of particular harmonic sums.

There are two basic principles of Mellin analysis that are used in this paper and that we now list, referring to [18] for detailed conditions.

(M1) *Harmonic sum property.* The transform of the general harmonic sum defined by (62) satisfies

$$G^*(s) = \Xi(s) \cdot g^*(s), \qquad \text{with } \Xi(s) = \sum_{k \in K} \lambda_k \mu_k^{-s}, \tag{63}$$

in the intersection of the strips of absolute convergence of $\Xi(s)$ and $g^*(s)$.

(M2) *Mapping property.* The Mellin transform maps the individual terms in the asymptotic expansion of a function $G(x)$ (here, typically, a harmonic sum arising from (M1)) to the singularities of the transform $G^*(s)$. The correspondence fares both ways and is given by the following rule: Assume that $G^*(s)$ defined in some strip $\langle \alpha, \beta \rangle$ admits a meromorphic continuation to an extended strip $\langle \alpha, \gamma \rangle$ with $\gamma \geq \beta$, is analytic on the line $\mathrm{Re}(s) = \gamma$, and satisfies $G^*(s) = O(|s|^{-r})$ for some $r > 1$ as $\mathrm{Im}(s) \to \pm\infty$ in $\alpha \leq \mathrm{Re}(s) \leq \gamma$. Then each singular term in a local expansion of $G^*(s)$ at a pole in the extended strip provides a corresponding term in the asymptotic expansion of $G(x)$ at

$+\infty$, according to the dictionary:

$$\left(\frac{d}{(s-\xi)^k}\right) \quad \Longrightarrow \quad \left(\frac{(-1)^k d}{(k-1)!} x^{-\xi} (\log x)^{k-1}\right), \tag{64}$$

with an error term that is

$$O(x^{-\gamma}). \tag{65}$$

The combination of (M1) and (M2) then yields the basic principle of asymptotic analysis by Mellin transforms:

> *The Mellin transform factorizes a harmonic sum $G(x)$, in such a way that the asymptotic estimation of the sum reduces to analysing separately two collections of singularities: those of the basic transform $g^*(s)$ and those of the associated Dirichlet series $\Xi(s)$.*

Mellin analysis of harmonic sums is often conducted in situations where: (i) the transform g^* of the base function is of exponential decay (this is systematically the case when there appears the exponential function); (ii) the Dirichlet series $\Xi(s)$ is of at most polynomial growth as $|\mathrm{Im}(s)| \to \infty$. In this case, the transform $G^*(s)$ is of fast decay at $\pm i\infty$ and the application of (64) is legitimate.

It is apparent from (64) that complex poles induce periodic fluctuations. As detailed in [18], the framework extends to any function that has infinitely many poles in a finite strip, provided it remains of controlled growth $O(|s|^{-r})$ with $r > 1$ on a set of parallels to the real axis that escape to $\pm i\infty$. We refer to the term of asymptotics *"in the weak sense"* for such conditions that are only imposed on a vertical ladder. In that case, there is an infinite superposition of periodic elements that may or may not be collectively periodic: this fact depends on whether poles are regularly spaced vertically, or not.

When poles accumulate from the right near the boundary line $\mathrm{Re}(s) = \gamma$ of the strip $\langle \alpha, \gamma \rangle$, the general paradigm of (64) and (65) needs to be mildly amended. In effect, it is not immediately clear that the induced fluctuation must contribute globally a term that remains $o(x^{-\gamma})$. The precise result that follows is logically needed as the situation arises in the context of dynamical sources.

Proposition 10 (Improved Mellin asymptotics). *Assume that the Mellin transform $G^*(s)$ of $G(x)$ defined in $\langle \alpha, \beta \rangle$ admits a meromorphic continuation to the strip $\langle \alpha, \gamma \rangle$ with $\gamma \geq \beta$, satisfies $O(|s|^{-r})$ with $r > 1$ in the weak asymptotic sense in $\alpha \leq \mathrm{Re}(s) \leq \gamma$, and is meromorphic on the line $\mathrm{Re}(s) = \gamma$ with only finitely many poles on that line. Then one has the implication*

$$G^*(s) \asymp \sum_{(\xi,k)} \frac{d_{\xi,k}}{(s-\xi)^k} \implies G(x) = \sum_{(\xi,k)} \frac{(-1)^k d_{\xi,k}}{(k-1)!} x^{-\xi} (\log x)^{k-1} + o(x^{-\gamma}), \tag{66}$$

as $x \to +\infty$, where ξ ranges over all poles such that $\alpha \leq \mathrm{Re}(\xi) \leq \gamma$.

The singular expansion of a function in a domain is denoted by '$\asymp$'. It is to be understood in the sense of [18], as the formal sum of the local singular expansions of $\Omega(s)$ at each singularity of the domain.

Proof. By possibly subtracting elementary functions to $G(x)$, we may assume without loss of generality that $G(x)$ is analytic on $\mathrm{Re}(s) = \gamma$. (Suitable combinations of exponentials and monomials in x will do, as they as their transforms are of fast decay at $\pm i\infty$.) The classical proof of (64) in (M2) appeals to integration on a large rectangle with vertical sides $\mathrm{Re}(s)=\alpha$ and $\mathrm{Re}(s)=\gamma$; see [18]. We are then reduced to proving that, for the right boundary,

$$\int_{\gamma-i\infty}^{\gamma+i\infty} G^*(s)x^{-s}\,ds = o(x^{-\gamma}).$$

Set $s=\gamma+iu$ and $x=e^y$. Then, the equivalent estimate

$$x^{-\gamma}\int_{-\infty}^{+\infty} G^*(\gamma+iu)e^{-iuy}\,du = o(x^{-\gamma})$$

is granted by the Riemann-Lebesgue Lemma: Whenever $h(u)$ is in $\mathcal{L}^1(-\infty,+\infty)$, then one has

$$\int_{-\infty}^{+\infty} h(u)e^{iuy}\,du = o(1),$$

as $y\to\pm\infty$. □

6.2. **Size and path-length in the Poisson model: finite alphabets.** The operator $I-\mathfrak{G}_s$ is invertible in the plane $\mathrm{Re}(s)>1$, so that the series $\Lambda(F,s)$ is analytic there and it has a simple pole at $s=1$. Two types of situations need to be distinguished depending on the periodicity of the source.

(a) In the periodic case, $\Lambda(F,s)$ has poles on $\mathrm{Re}(s)=1$ that are regularly spaced on the line $\mathrm{Re}(s)=1$ and contribute periodic terms in the asymptotic expansion. Moreover, there is a vertical strip $\langle 1-\delta,1\rangle$ (with $0<\delta<1$) free of poles that provides good bounds for the error terms.

(b) In the aperiodic case, there are no other poles on the line $\mathrm{Re}(s)=1$; however, there may be an accumulation of poles on the left of the line $\mathrm{Re}(s)=1$. Proposition 10 then allows us to quantify directly the contribution of these poles.

Theorem 5 (Additive parameters in the Poisson model; finite alphabet)**.** *Let $(\mathcal{P}_z,\mathsf{S},F)$ be a Poisson model of rate z relative to a source S with finite alphabet and initial distribution F.*

(i) *In the case when the source is aperiodic, the expectations of size and path length are*

$$\widehat{S}(z)=\frac{1}{h(\mathsf{S})}z+o(z),\qquad \widehat{P}_A(z)=\frac{1}{h(\mathsf{S})}z\log z + C_A(F,\mathsf{S})z+o(z).$$

The expectations of path length of list-tries and bst-tries are of the form (where N is B or L for bst's and lists)

$$\widehat{P}_N(z)=\frac{K_N(\mathsf{S})}{h(\mathsf{S})}z\log z + C_N(F,\mathsf{S})z+o(z).$$

(ii) *In the case when the source is periodic, the expectations of size and path length are*

$$\begin{aligned}\widehat{S}(z) &= \frac{1}{h(\mathsf{S})}z\,[1+Q_A(\log z)]+o(z^{1-\delta})\\ \widehat{P}_A(z) &= \frac{1}{h(\mathsf{S})}z\log z + z\,[C_A(F,\mathsf{S})+Q_A(\log z)]+o(z^{1-\delta}).\end{aligned}$$

The expectations of path length of list-tries and bst-tries are of the form (where N is B or L)

$$\widehat{P}_N(z) = \frac{K_N(\mathsf{S})}{h(\mathsf{S})} z \log z + z\left[C_N(F,\mathsf{S}) + Q_N(\log z)\right] + o(z^{1-\delta}).$$

The function $Q_N(u)$ depends on the source S and is of very small amplitude; δ is a positive constant, satisfying $0 < \delta < 1$, that is determined by the width of the pole-free region of $\Lambda(F,s)$ left of the line $\mathrm{Re}(s) = 1$). The constants $K_N(\mathsf{S})$ do not depend on the initial distribution whereas the constants $C_N(F,\mathsf{S})$ may depend on it.

Proof. The expressions given in Theorem 1 are harmonic sums that involve the exponential function. The Mellin transform of the exponential is the gamma function $\Gamma(s)$ that has known singularities at the nonpositive integers. With our conventions, the Dirichlet series of these harmonic sums are of the form $\Xi(s) = \Lambda(-s)$ for $\Lambda(s)$ a Dirichlet series of fundamental intervals of sorts.

Precisely, the expressions of Theorem 1 have the following Mellin transforms tabulated along with their fundamental strip:

$$\begin{array}{rll} \text{Size:} & \widehat{S}(z) \mapsto -\Lambda^{\langle A\rangle}(F,-s)(s+1)\Gamma(s) & s \in \langle -2,-1\rangle \\ \text{Path length:} & \widehat{P}_A(z) \mapsto -\Lambda^{\langle A\rangle}(F,-s)\Gamma(s+1) & s \in \langle -2,-1\rangle \\ \text{Path length in list-tries:} & \widehat{P}_L(z) \mapsto -\Lambda^{\langle L\rangle}(F,-s)\Gamma(s+1) & s \in \langle -2,-1\rangle \\ \text{Path length in bst-tries:} & \widehat{P}_B(z) \mapsto \Lambda^{\langle B\rangle}(F,-s)\Gamma(s) & s \in \langle -2,-1\rangle\,. \end{array}$$

All these transforms extend to meromorphic functions in a larger strip to the right of the line $\mathrm{Re}(s) = -1$ in a way suitable for Mellin analysis.

From properties of $\mathfrak{G}_s$, the singular expansion of $\Lambda^{\langle A\rangle}(F,s)$ at $s = 1$ reads,

$$\Lambda^{\langle A\rangle}(F,s) \asymp \frac{-1}{\lambda'(1)}\,\frac{1}{s-1} + C_A(F,\mathsf{S}) \qquad (s=1).$$

In the case of hybrid tries (list-tries or bst-tries), expressions (43), (44) for the basic homoclinal case, (48), (49) for the basic heteroclinal case, (52), (53) for the Markovian case, together with special values given in Proposition 7, imply, for the singular expansions of $\Lambda^{\langle N\rangle}(F,s)$ at $s = 1$, the form

$$\Lambda^{\langle N\rangle}(F,s) \asymp \frac{-1}{\lambda'(1)}\,\frac{K_N(\mathsf{S})}{s-1} + C_N(F,\mathsf{S}) \qquad (s=1).$$

The constants $K_N(\mathsf{S})$ can be made explicit: they only depend on the "uniform fundamental measures" of depth one, denoted by $u_i^\star$ and on the values of the dominant eigenfunction Ψ_1 at the boundaries of fundamental intervals of depth one. For instance, in the case when the dynamical source is basic with all its branches increasing, one has

$$K_L(\mathsf{S}) = \sum_i U^\star_{[>i]}\;\Psi_1^{\langle L\rangle}(h_i(0),h_i(1),1),$$

$$K_B(\mathsf{S}) = 2\sum_{i<j} \frac{u_i^\star\; u_j^\star}{U^\star_{[i,j]}}\;\;\Psi_1^{\langle B\rangle}(h_i(0),h_i(1),h_j(0),h_j(1)).$$

In the basic case, but with not all branches being increasing, the assertion (vi) of Proposition 6 shows that the poles of $(I - \widetilde{\mathfrak{G}}_s)^{-1}$ do not intervene in the dominant term, and the previous expressions can be generalized thanks to (48) and (49),

$$K_L(\mathsf{S}) = \frac{1}{2}\sum_i U^\star_{[>i]} \sum_{r\in\mathcal{L}(i)} \Psi_1^{\langle L\rangle} r(h_i(0), h_i(1), 1),$$

$$K_B(\mathsf{S}) = \sum_{i<j} \frac{u_i^\star\ u_j^\star}{U^\star_{[i,j]}} \sum_{r\in\mathcal{B}(i,j)} \Psi_1^{\langle B\rangle} r(h_i(0), h_i(1), h_j(0), h_j(1))$$

In the Markovian case, the formulae are similar and involve the components of the vector eigenfunction Ψ_1 (see Equations (52) and (53)).

The constants $C_N(F, \mathsf{S})$ are determined through residue calculations and also involve dominant spectral objects of $\mathfrak{G}_s$ and $\widetilde{\mathfrak{G}}_s$ at $s = 1$.

Due to the fast decrease of the function $\Gamma(s)$ towards $\pm i\infty$, the conditions of Mellin analysis apply. The proof is then completed by means of Equation (43) and (44) combined with basic residue calculations. □

We remark here that the fluctuations for size can have an asymptotically dominant order, while the fluctuations for path length are always subdominant.

6.3. **Size and path-length in the Poisson model: infinite alphabets.** In the case of an infinite alphabet, the series of "uniform fundamental measures" of depth one may involve other singularities, so that the average path lengths of hybrid tries are not always of order $z \log z$.

We consider three Dirichlet series

$$L(s) := \sum_{i<j} u_j^\star\, u_i^{\star s-1}, \qquad B(s) := \sum_{i<j} \frac{u_j^\star\, u_i^\star}{(u_i^\star + \ldots u_j^\star)^{2-s}}, \qquad A(s) = \sum_i u_i^{\star s}.$$

The last series is exactly the Dirichlet series $\Lambda_1^{\langle A\rangle}(1, s)$. By hypothesis (d_3) of Definition 1, there exists $\gamma < 1$ such that this series is convergent for $\mathrm{Re}(s) > \gamma$. Since the dominant spectral objects Ψ_s and E_s are strictly positive for real s, the dominant singularity of $\Lambda^{\langle L\rangle}(F, s)$ is located at $s = \max(1, s_L)$ where s_L is the dominant singularity of $L(s)$. In the same vein, the dominant singularity of $\Lambda^{\langle B\rangle}(F, s)$ is located at $s = \max(1, s_B)$ where s_B is the dominant singularity of $B(s)$.

This discussion enables us to state:

Proposition 11 (Additive parameters in the Poisson model; infinite alphabet)**.** *Assume that the canonical fundamental measures $u_i^\star$ are decreasing and the dominant singularity s_A is strictly less than 1. Then the average path length of the bst-trie is necessarily of the form*

$$\widehat{P}_B(z) = O(z^{1+\epsilon}) \qquad \text{for all } \epsilon > 0.$$

For any $1 < \delta < 2$, there exist sources such that the average path length of the list-trie is of the form

$$\widehat{P}_L(z) = \Theta(z^\delta).$$

Proof. The proof is based on relations between the dominant singularities s_L and s_B:

(*i*) *There holds* $s_B \le s_L \le 2$. In effect, the inequality $u_i^\star \le (u_i^\star + \dots u_j^\star)$ implies that $B(s) \le L(s)$ for s real, $s < 2$.

(*ii*) *If the sequence* $u_j^\star$ *is decreasing, then* $s_B \le 1$. Since $(u_j^\star)$ is decreasing, one has $u_i^\star + \dots + u_j^\star \ge (j-i+1)\, u_j^\star$, so that, for $1 < s < 2$,

$$B(s) \le \sum_i u_i^\star \sum_{j>i} \frac{{u_j^\star}^{s-1}}{(j-i+1)^{2-s}} \le \left(\sum_i u_i^\star\right) \left(\sum_j \frac{{u_j^\star}^{s-1}}{j^{2-s}}\right).$$

The Hölder inequality with $\alpha = r/(s-1)$ for some $r \in]\max(\gamma, s-1), 1[$ and $\beta := r/(r+1-s)$ entails the convergence of the second series.

(*iii*) *If* $u_i^\star = \Theta(i^{-\alpha})$ *with* $1 < \alpha < 2$, *then* $s_L = 2/\alpha$ *and* $s_B < 1$. In this case, the series $B(1)$ is convergent with a general term that is $\Theta(i^{-\alpha})$, while $L(s)$ has its general term that is $\Theta(i^{-\rho})$ with $\rho = \alpha(s-1) + \alpha - 1$.

□

Such a discrepancy between the orders of growth of path length in array tries, list-tries, and bst-tries occurs for instance for the generalized Zipf laws,

$$p_k^{(\alpha)} = \zeta(\alpha)^{-1} k^{-\alpha}, \qquad \alpha > 1,$$

($\zeta(s)$ is the Riemann zeta function) that have "soft tails"; see also the case of the continued fraction source in Section 8.

6.4. **Dirichlet depoissonization.** Mellin transforms are ideally suited to the analysis of harmonic sums, of which the Poisson sum relative to path length in array tries,

$$F(x) = \sum_h u_h \left(1 - e^{-x u_h}\right), \tag{67}$$

is typical. Sums very much like (67) involving an exponential occur in the analysis of Poisson models; see Theorem 1. On the other hand, by algebraic depoissonization (12), (13), the Bernoulli model leads to sums exemplified by the counterpart of (67),

$$G(x) = \sum_h u_h \left(1 - (1-u_h)^x\right); \tag{68}$$

see Theorem 2. The problem of analysing asymptotically (68) is a bit trickier than that of (67).

The approach developed here is based on approximating directly the singularities of a Dirichlet series (the one associated to the Bernoulli model) by the singularities of another simpler series (the one arising from the Poisson model), this in the perspective of Mellin analysis. The approach, called *Dirichlet depoissonization*, is briefly introduced and put to use for the analysis of multidimensional search, but without much justification, in [19]. It will be immediately realized that the heart of the matter lies in the following result.

Proposition 12 (Dirichlet depoissonization). *Let* $u_h \to 0$ *with* u_h *positive. Define the two Dirichlet series*

$$\Omega(s) = \sum_h u_h^s, \qquad \widetilde{\Omega}(s) = \sum_h \left(\log \frac{1}{1-u_h}\right)^s.$$

Assume that $\Omega(s)$ *has a nonempty domain of convergence, is meromorphically continuable to the whole of* $\mathbb{C}$*, and is in any finite strip of polynomial growth in the weak asymptotic sense. Then the same properties hold for* $\widetilde{\Omega}(s)$*. In addition, the singularities of* $\Omega, \widetilde{\Omega}$ *are related by*

$$\mathrm{Sing}(\widetilde{\Omega}) = \{s-k \mid s \in \mathrm{Sing}(\Omega) \text{ and } k \in \mathbb{N}\},$$

and the singular expansions of $\Omega(s)$ *and* $\widetilde{\Omega}(s)$ *are related by*

$$\widetilde{\Omega}(s) \asymp \Omega(s) + c_1(s)\Omega(s+1) + c_2(s)\Omega(s+2) + \cdots, \tag{69}$$

where

$$c_j(s) = [x^j]\exp\left[s\log(\frac{1}{x}\log\frac{1}{1-x})\right].$$

(The term "weak" refers again to the fact that growth need only be controlled on certain parallels to the real axis tending to $\pm i\infty$.)

Proof. What we want to do is justify the chain of formal transformations,

$$\begin{aligned}
\widetilde{\Omega}(s) &\approx \sum {u_h}^s \exp\left(s\log\left(1+\frac{1}{2}u_h+\frac{1}{3}{u_h}^2+\cdots\right)\right) \\
&\approx \sum {u_h}^s \exp\left(s\left(\frac{12}{u_h}+\frac{5}{24}{u_h}^2+\cdots\right)\right) \\
&\approx \sum {u_h}^s \left(1+\frac{1}{2}su_h+(\frac{5}{24}s+\frac{18}{s^2}){u_h}^2+(\frac{1}{8}s+\frac{5}{48}s^2+\frac{1}{48}s^3){u_h}^3+\cdots\right) \\
&\approx \Omega(s)+c_1(s)\Omega(s+1)+c_2(s)\Omega(s+2)+\cdots.
\end{aligned}$$

Let σ_0 denotes the abscissa of convergence of $\Omega(s)$. The indices of the u_h have been dropped and it is assumed implicitly that sums range over the whole set of possible $\{u_h\}$. We also fix an (arbitrary) integer m that will control the order of expansions.

Analytic continuation. In a first pass, we consider the problem of transforming $\Omega(s)$ in a way that ensures analytic continuation, but without worrying about uniformity with respect to s. We assume that $\mathrm{Re}(s) > \sigma_0$. First, since $u_h \to 0$, Taylor expansions produce

$$\log\left(\frac{1}{u}\log(1-u)^{-1}\right) = D(u) + O(u^{m+1}) \qquad (u \to 0),$$

for some computable polynomial D of degree m. Thus,

$$\widetilde{\Omega}(s) = \sum u^s \exp(sD(u)) \cdot \exp(O(u^{m+1})).$$

Now,

$$\exp(O(u^{m+1})) = 1 + O(u^{m+1}) \qquad (u \to 0).$$

In particular, the difference

$$\Delta(s) := \widetilde{\Omega}(s) - \sum u^s \exp(sD(u)) \tag{70}$$

has its general term that decreases like $O(u^{\sigma+m+1})$, where $\sigma = \mathrm{Re}(s)$. It is in particular analytic in $\mathrm{Re}(s) \geq \sigma_0 - m$.

Next, we turn to

$$\Omega_1(s) = \sum u^s \exp(sD(u)),$$

for which one has

$$\exp(sD(u)) = 1 + c_1(s)u + c_2(s)u^2 + \cdots + c_m(s)u^m + O(u^{m+1}),$$

where the c_j are obtained by Equation (12). Indeed, for any fixed s , the quantity $sD(u)$ lies in a neighbourhood of 0 for which the expansion of $\exp(v) = 1 + v + \cdots + O(v^{m+1})$ is applicable. Thus, the difference

$$\Delta_1(s) = \Omega_1(s) - (\Omega(s) + c_1(s)\Omega(s+1) + \cdots + c_m(s)\Omega(s+m)) \tag{71}$$

has its general term that decreases like $u^{\sigma+m+1}$ so that it is analytic in $\text{Re}(s) \geq \sigma_0 - m$. Equations (70) and (71) imply that

$$\widetilde{\Omega}(s) = \Delta(s) + \Delta_1(s) + (\Omega(s) + c_1(s)\Omega(s+1) + \cdots + c_m(s)\Omega(s+m))$$

originally defined for $\text{Re}(s) > \sigma_0$ is meromorphic $\text{Re}(s) \geq \sigma_0 - m$. Since m was fixed arbitrarily at the beginning, meromorphicity of $\widetilde{\Omega}(s)$ in the whole of C results. In addition, the singularities of $\widetilde{\Omega}$ are those of Ω shifted by $0, 1, 2, \ldots$, and the singular expansion (69) is justified.

Growth. There remains to examine the growth of $\widetilde{\Omega}(s)$ as $\text{Im}(s) \to \pm\infty$. This is achieved by modifying the previous proof, taking uniformity with respect to s into account. We may assume that $\text{Re}(s) < \sigma_0 + 1$, since it is analytic continuation to the left that matters. The basic idea is that

$$\exp(sO(u^{m+1})) - 1 = \begin{cases} O(1) = O(|s|u^{m+1}) & \text{if } |s|u^{m+1} > 1 \\ O(|s|u^{m+1}) & \text{if } |s|u^{m+1} \leq 1. \end{cases} \tag{72}$$

The first line is justified by the fact that u is real and $\text{Re}(s)$ is bounded from above, so that we have a bound of $O(1)$ that is replaced by the less precise estimate $O(|s|u^{m+1})$; the second line follows from the standard expansion of $e^x - 1$ near the origin. Consequently, the function $\Delta(s)$ defined in Eq. (70) is $O(|s|)$ for $\text{Re}(s) \geq \sigma_0 - m$.

In a manner similar to (72), separating the two cases $|s|u > 1$ and $|s|u \leq 1$ yields a uniform version of (70):

$$\exp(sD(u)) = 1 + c_1(s)u + c_2(s)u^2 + \cdots + c_m(s)u^m + O(|s|^{m+1}u^{m+1}),$$

There results that $\Delta_1(s)$ is, for $\text{Re}(s) \geq \sigma_0 - m$, of growth at most $O(|s|^{m+1})$. Therefore, $\widetilde{\Omega}(s)$ is a sum of quantities of the form $c_j(s)\Lambda(s+j)$ and of an analytic quantity $\Delta(s) + \Delta_1(s)$ that is $O(|s|^{m+1})$. It is thus of weak polynomial growth in $\text{Re}(s) \geq \sigma_0 - m$. Since m is chosen arbitrarily, the result follows. □

The argument adapts to the more general pair of Dirichlet series

$$\Omega(s) = \sum_h \lambda_h u_h^s, \qquad \widetilde{\Omega}(s) = \sum_h \lambda_h \left(\log \frac{1}{1-u_h}\right)^s,$$

with $\lambda_h > 0$, and the property stated in Proposition 12 still holds in this case.

6.5. **Asymptotics in the Bernoulli model.** The principle of Dirichlet depoissonization shows that the only poles introduced by the method are at distance at least one from the dominant poles arising from the Poisson model. Consequently, the estimates of Theorem 5 remain valid in that case.

Theorem 6 (Asymptotics in the Bernoulli model, finite alphabet)**.** *Under the Bernoulli model of parameter n and for a source with a finite alphabet, the $O(n)$ character of the expected trie size and the $O(n \log n)$ character of the expected path lengths hold: the estimates given in Theorem 5 remain valid provided the Poisson parameter z is replaced by the Bernoulli parameter n.*

Proposition 11 is similarly susceptible to extensions in the Bernoulli model.

7. Asymptotic analysis of height

Finally, we consider the last parameter, the height. We first perform an asymptotic reduction to a harmonic sum from which the asymptotic form of the expected height in the Poisson model derives painlessly. A further step of saddle point depoissonization provides the final result under the Bernoulli model.

7.1. **Double exponential approximation.** The expectation of height in a Poisson model relative to a rate z, a source S, and an initial distribution F, is expressed by the infinite sum (see Theorem 3):

$$\text{(73)} \qquad \mathrm{E}[h; \mathcal{P}_z, \mathsf{S}, F] = \sum_{k=0}^{\infty} [1 - \pi_k(z)], \quad \text{with } \pi_k(z) = \prod_{|h|=k} (1 + z u_h) e^{-z u_h}$$

Such a series does not fall directly in the orbit of Mellin transform techniques but it can be well approximated by a harmonic sum. Indeed, the Quasi-Power Property of Proposition 5 provides the existence of a constant ρ defined by

$$\text{(74)} \qquad \rho := \frac{1}{2} \lim_{k \to \infty} \frac{\Lambda_k^{\langle A \rangle}(F, 2)}{\lambda(2)^k}.$$

The following result gives first an asymptotic approximation to the distribution π_k stated under a strong form of $\mathcal{L}^1$-convergence.

Proposition 13. *The distribution of trie height under the Poisson model admits a double exponential approximation: the two sequences*

$$\pi_k(z) = \prod_{|h|=k} (1 + z u_h) e^{-z u_h}, \qquad \widehat{\pi}_k(z) := \exp[-\rho z^2 \lambda(2)^k],$$

satisfy

$$\sum_{k \geq 0} [\pi_k(z) - \widehat{\pi}_k(z)] = o(1).$$

Proof. We use the fact that $\log \pi_k(z)$ is a harmonic sum relative to $\Lambda_k^{\langle A \rangle}(F, s)$ and proceed in two steps.

First step. First, we compare $\pi_k(z)$ with $\widetilde{\pi}_k(z)$ where

$$\widetilde{\pi}_k(z) = \prod_{|h|=k} \exp[-\frac{z^2 {u_h}^2}{2}] = \exp(-\frac{z^2}{2} \sum_{|h|=k} {u_h}^2);$$

Choose an arbitrary number d in the interval

$$\left]\frac{3\log z}{|\log\lambda(3)|},\ \frac{2\log z}{|\log\lambda(2)|}\right[$$

and set

$$\kappa(z) := \lfloor d\log z\rfloor\,.$$

The existence of these quantities is granted by the strict inequality $\log\lambda(2)/\log\lambda(3) < 2/3$ of assertion (ii) of Proposition 6. Then, there exist positive numbers ϵ, ϵ' and $\rho \in [0,1[$ such that the following three properties hold:

$$(C_1):\ z^2{\lambda_2}^{\kappa(z)} \geq z^{\epsilon} \quad (C_2):\ z^3{\lambda_3}^{\kappa(z)} \leq z^{-\epsilon'}, \quad (C_3):\ (\forall h,\ |h| \geq \kappa(z))\ zu_h \leq \rho.$$

The sum in (73) is split into two parts (left tail; central domain and right tail) according to the integer $\kappa(z)$.

(i) *Left tail.* We first consider the part relative to indices $k \leq \kappa(z)$. The inequality (true for $x \geq 0$)

$$\exp(-\frac{x^2}{2}) \leq (1+x)e^{-x} \leq \exp(-\frac{x^2}{2(1+x)})$$

implies that

$$\widetilde{\pi}_k(z) - \pi_k(z) \leq \exp(-\frac{z^2}{2}\sum_{|h|=k}{u_h}^2) - \exp(-\frac{z^2}{2}\sum_{|h|=k}\frac{{u_h}^2}{1+zu_h}). \tag{75}$$

This introduces the sequence

$$B_k = \frac{1}{2}\sum_{|h|=k}\frac{{u_h}^2}{1+zu_h}.$$

For any positive sequence $(a_i)_{i\in I}$ of sum s, the inequality

$$\sum_{i\in I}\frac{{a_i}^2}{1+a_i} \leq \frac{s^2}{1+s}$$

holds, and this implies that B_k is decreasing. This fact, in conjunction with the Mean Value Theorem, entails the inequality

$$\widetilde{\pi}_k(z) - \pi_k(z) \leq z^3\exp[-z^2\,B_{\kappa(z)}].$$

Now, Property (C_3) of index $\kappa(z)$, the Quasi-Power Property, and finally Property (C_1) of index $\kappa(z)$ imply that

$$\sum_{k\leq\kappa(z)}\widetilde{\pi}_k(z) - \pi_k(z) \leq \kappa(z)\,z^3\,\exp[-dz^{\epsilon}] = o(1) \tag{76}$$

(ii) *Central domain and right tail.* The second part of the sum relative to $k > \kappa(z)$ is also $o(1)$. Indeed, the inequality

$$\pi_k(z) - \widetilde{\pi}_k(z) \leq \log\pi_k(z) - \log\widetilde{\pi}_k(z).$$

and the relations

$$\log\pi_k(z) \leq \sum_{|h|=k}\left[-\frac{(zu_h)^2}{2} + \frac{(zu_h)^3}{3}\right], \qquad \log\widetilde{\pi}_k(z) = -\sum_{|h|=k}\frac{(zu_h)^2}{2},$$

entail that

$$\pi_k(z) - \widetilde{\pi}_k(z) \le z^3 \sum_{|h|=k} {u_h}^3.$$

Once again, the Quasi-Power Property at $s = 3$ in conjunction with Property (C_2) of index $\kappa(z)$ yields the following inequalities

$$\sum_{k \ge \kappa(z)} [\pi_k(z) - \widetilde{\pi}_k(z)] \le d_0 z^3 \sum_{k \ge \kappa(z)} {\lambda_3}^k = d_0 z^3 \frac{{\lambda_3}^{\kappa(z)}}{1-\lambda_3} \le z^{-\epsilon'} = o(1). \tag{77}$$

Second step. The distribution $\widetilde{\pi}_k(z)$ under the Poisson model is thus well approximated by $\widehat{\pi}_k(z) = \exp[-\rho z^2 \lambda(2)^k]$. The Quasi-Power Property once again provides two constants d_1 and d_2 such that the following two inequalities hold

$$\left| \frac{1}{2} \sum_{|h|=k} {u_h}^2 - \rho \lambda(2)^k \right| \le d_1\, \nu^k, \qquad \sum_{|h|=k} {u_h}^2 \ge 2 d_2\, \lambda(2)^k,$$

where ν lies strictly between the dominant eigenvalue $\lambda(2)$ and the modulus $\mu(2)$ of a subdominant eigenvalue of the operator $\mathfrak{G}_2$. Hence, the inequalities

$$\left| \sum_{k \ge 0} \widetilde{\pi}_k(z) - \widehat{\pi}_k(z) \right| \le \sum_{k \ge 0} |\widetilde{\pi}_k(z) - \widehat{\pi}_k(z)| \le d_1 z^2 \sum_{k \ge 0} \nu^k \exp[-d_2\, z^2\, \lambda(2)^k], \tag{78}$$

entail, by an elementary argument or by Mellin transforms, that

$$\sum_{k \ge 0} \nu^k \exp[-d_2 z^2 \lambda(2)^k] = o(1). \tag{79}$$

Finally, Equations (76), (77) and (79) imply the result. □

7.2. **Distribution of height under a Poisson model.** First, note that Equations (76), (77) and (79) provide the asymptotic distribution of height $\widehat{h}_z$ in the model $(\mathcal{P}_z, \mathsf{S}, F)$, since they entail

$$\lim_{z \to \infty} \sup_{k \ge 0} \left| \Pr[\widehat{h}_z < k] - \exp[-\rho z^2 \lambda(2)^k] \right| = 0.$$

On the other hand, the harmonic sum that approximates the average height in a Poisson model

$$D(z) = \sum_{k=0}^{\infty} \left(1 - \exp[-\rho z^2 \lambda(2)^k] \right), \tag{80}$$

has Mellin transform

$$D^*(s) = -\frac{1}{2}\, \rho^{-s/2} \frac{\Gamma(s/2)}{1 - \lambda(2)^{-s/2}}.$$

This is a textbook example of a Mellin analysis. The fundamental strip is $\langle -2, 0 \rangle$, with the singular expansion at $s = 0$ being

$$D^*(s) \asymp -\frac{2}{\log \lambda(2)} \frac{1}{s^2} + \left[\frac{\gamma - \log \rho}{\log \lambda(2)} - \frac{1}{2} \right] \frac{1}{s} \qquad (s = 0).$$

There are also regularly spaced poles on the line $\mathrm{Re}(s) = 0$ that entail periodic fluctuations. We can state:

Theorem 7 (Asymptotic height under the Poisson model)**.** *In the Poisson model of rate z relative to a source* S *with initial distribution F, the average height is*

$$\widehat{H}(z) = \frac{2}{|\log \lambda(2)|}\log z + Q_F(\log z) - \left[\frac{\gamma - \log\rho}{\log\lambda(2)} - \frac{1}{2}\right] + o(1),$$

where ρ is a positive constant depending on the source (S, F) and $Q_F(u)$ is a periodic function of very small amplitude.
Moreover, the asymptotic distribution of height is of double exponential type,

$$\lim_{z\to\infty}\sup_{k\geq 0}\left|\Pr\{\widehat{h}_z < k\} - \exp\left[-\rho z^2\lambda(2)^k\right]\right| = 0.$$

7.3. **Height under the Bernoulli model.** We first recall the notations of Theorem 4. The quantity $\pi_{k,n}$ denotes the probability that a trie built on an n–tuple of random items has height at most k under the Bernoulli model. The exponential generating function $\Pi_k(z)$ of the $\pi_{k,n}$ has value

$$\Pi_k(z) := \prod_{|h|=k}(1 + zu_h)\,.$$

In the case of a memoryless source, the analysis can be conducted from there using the saddle point method [16, 20] and it even becomes completely elementary in the case of unbiased binary tries; see [39]. We propose to follow precisely the same approach in order to analyse height in the case of a general dynamical source. The reader is also referred to Jacquet and Szpankowski's interesting paper [29] for a general framework of saddle-point depoissonization (under the name of "analytic depoissonization").

The formal idea of the analysis is the following: Start with the Cauchy integral formula

$$\pi_{k,n} = \frac{n!}{2i\pi}\int_\gamma \Pi_k(z)\,\frac{dz}{z^{n+1}}, \tag{81}$$

where γ is a simple closed contour encircling positively the origin. Since one has, for *any* $x \in (0, +\infty)$ and an arbitrary fixed m:

$$\log(1+x) = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{m-1}\frac{x^{m-1}}{m-1} + O(x^m),$$

the "exp-log" transformation applied to the integrand gives

(82)

$$\begin{aligned}\Pi_k(z) &= \exp\sum\log(1+zu_h)\\ &= \exp\left(z\sum u_h - \frac{z^2}{2}\sum {u_h}^2 + \frac{z^3}{3}\sum {u_h}^3 - \cdots + O(z^m\sum {u_h}^m)\right),\end{aligned}$$

where all the sums are taken over $|h| = k$. Retaining only the first two terms in (82) leads us to expect, for suitable values of k at least, the validity of the approximation

$$\pi_{k,n} \approx \frac{n!}{2i\pi}\int_\gamma \exp\left(-\frac{z^2}{2}\sum {u_h}^2\right)\,\frac{e^z\,dz}{z^{n+1}}. \tag{83}$$

This may now be viewed as a perturbation of the Cauchy coefficient integral applied to e^z. But it is well-known that the latter integral can be estimated by the saddle point method [10] which consists in the following steps:

(i) integrate along the circle $|z| = n$ (an approximate saddle point of the integrand);
(ii) observe that the contribution is concentrated in a small sector $ne^{i\theta}$ with $|\theta| \leq \theta_0$ and $\theta_0 = n^{-1/2+\alpha}$ for any α such that $0 < \alpha < \frac{1}{6}$ (this conditions ensures that one has simultaneously $n{\theta_0}^2 \to +\infty$ while $n{\theta_0}^3 \to 0$);
(iii) finally reduce the integral asymptotically to a complete Gaussian integral that can be evaluated.

In the case of e^z, this is of course one of the standard ways of deriving Stirling's formula for $1/n!$. If we regard the factor of e^z/z^{n+1} as almost constant on the tiny fraction of the integration contour that matters, we are then led to the approximations

$$\begin{aligned} \pi_{k,n} &\approx \frac{n!}{2i\pi} \exp\left(-\frac{n^2}{2}\sum {u_h}^2\right) \int_{\gamma_0} \frac{e^z\,dz}{z^{n+1}} \\ &\approx \exp\left(-\frac{n^2}{2}\sum u_h^2\right) &(84)\\ &\approx \exp\left(-\rho n^2 \lambda(2)^k\right), &(85)\end{aligned}$$

in view of the saddle point integral of e^z/z^{n+1}, of Stirling's formula for (84), and of basic results relative to fundamental intervals for (85). The estimate (85) is nothing but the double-exponential approximation of the distribution that was already encountered under the Poisson model. It involves the constant ρ already defined in (74).

Theorem 8 (Asymptotic height under the Bernoulli model)**.** *Under the Bernoulli model of parameter n, the logarithmic character of the expected trie height and the uniform double exponential approximation hold: the estimates given in Theorem 6 remain valid provided the Poisson parameter z is replaced by the Bernoulli parameter n.*

Proof. We now explain how to make the preceding strategy fully rigorous. We consider a small constant α with $0 < \alpha < \frac{1}{6}$, and we set

$$\theta_0 = n^{-1/2+\alpha}, \quad \gamma = \{ne^{i\theta} \mid |\theta| \leq \pi\}, \quad \gamma_0 = \{ne^{i\theta} \mid |\theta| \leq \theta_0\}.$$

The argument needs to distinguish two domains of variation of k with a boundary $\kappa_1(n)$ that is

$$\kappa_1(n) = \lfloor d \log n \rfloor,$$

for a constant d chosen to satisfy the following constraints that are refinements of constraints already encountered in the study of the Poisson model.

(C_1) $d < 2/|\log \lambda(2)|$, which ensures that for $k = \kappa_1(n)$, the term $n^2\lambda(2)^k$ is of exact order $\Theta(n^{\epsilon_0})$ with $\epsilon_0 > 0$.
(C_2) $d > 2/|\log \nu|$, where ν is strictly enclosed between the dominant eigenvalue $\lambda(2)$ and the modulus ν of a subdominant eigenvalue of the operator $\mathfrak{G}_2$. This makes the approximation $n^2 \sum u_h^2 \sim 2\rho n^2 \lambda(2)^k$ accurate to $O(n^{-\epsilon_1})$ absolute error as soon as $k \geq \kappa_1(n)$ (see Eq. (59).
(C_3) $d > 3/|\log \lambda(3)|$, so that for $k \geq \kappa_1(n)$, the cubic terms $n^3\lambda(3)^k$ are $O(n^{-\epsilon_2})$ (with $\epsilon_2 > 0$) and can be neglected;

(C_4) $d > (1+2\alpha)/|\log\lambda(2)|$, which ensures that the terms $n^2\lambda(2)^k{\theta_0}^2$ are $O(n^{-\epsilon_3})$ (with $\epsilon_3 > 0$) for $k \ge \kappa_1(n)$. This is a technical requirement needed to validate a saddle point perturbation estimate.

Observe that (C_1) and (C_3) are compatible given the log-concavity properties of $\lambda(\sigma)$ while the compatibility of (C_1) on the one hand, (C_2) and (C_4) on the other hand is automatically granted. Consequently, the constant d can be chosen anywhere in the interval

$$\left]\max\left(\frac{3}{|\log\lambda(3)|}, \frac{2}{|\log\nu|}, \frac{1+2\alpha}{|\log\lambda(2)|}\right), \frac{2}{|\log\lambda(2)|}\right[.$$

We also need repeatedly estimates of the form $|e^{h(z)}|$ for $z = ne^{i\theta}$ and $h(z) = z - cz^2$ with $c = \lambda(2)^k$ small. One has

$$g_n(\theta) := \log\left|e^{h(ne^{i\theta})}\right| = \mathrm{Re}(h(ne^{i\theta})) = n\cos\theta - cn^2\cos(2\theta).$$

An elementary study shows that for n large enough and $cn = o(1)$ (this is granted for $c = \lambda(2)^k$ and $k \ge \kappa_1(n)$ by (C_4)), $g_n(\theta)$ attains uniquely its maximum on $(-\pi, +\pi)$ at $\theta = 0$ and is unimodal on $(-\frac{\pi}{2}, +\frac{\pi}{2})$.

(i) *Left tail.* At $k = \kappa_1(n)$ and on the circle $z = ne^{i\theta}$, we have, with condition (C_1),

$$\log|\Pi_k(z)| \le n\cos\theta - c_0 n^{\epsilon_0},$$

for some $c_0 > 0$ and $\epsilon_0 > 0$, since the quadratic term dominates and is of exact order $\Theta(n^{\epsilon_0})$. Thus, trivial bounds applied to the Cauchy integral entail

$$\pi_{\kappa_1(n),n} \le \frac{n!e^n}{n^n}\cdot e^{-c_0 n^{\epsilon_0}}, \tag{86}$$

which corresponds to the left tail of the distribution, $k \le \kappa_1(n)$, being *globally* exponentially small.

(ii) *Central domain and right tail.* For $k \ge \kappa_1(n)$, we need to check that the saddle point method applies. On the part $\gamma\setminus\gamma_0$ of the contour, we have for n large enough

$$|\Pi_k(z)| \le \left|e^{n\cos\theta_0}\exp\left(-\rho n^2\lambda(2)^k + O(n^2\nu^k) + O(n^3\lambda(3)^k)\right)\right|.$$

But conditions (C_2) and (C_3) imply that the contribution of the integral in the noncentral region $\gamma\setminus\gamma_0$ satisfies

(87)
$$\left|\frac{n!}{2i\pi}\int_{\gamma\setminus\gamma_0}\Pi_k(z)\,\frac{dz}{z^{n+1}}\right| \le \frac{n!}{n^n}\sup_{\gamma\setminus\gamma_0}|\Pi_k(z)| = O\left(\exp[-n{\theta_0}^2/4]\exp[-\rho n^2\lambda(2)^k]\right).$$

(We used a replacement of ${\theta_0}^2/2$ by ${\theta_0}^2/4$ in order to absorb terms of the form $\sqrt{n}$ that come from Stirling Formula). From (87), the contribution on $\gamma\setminus\gamma_0$ is thus exponentially small.

For similar reasons, the quantity $\Pi_k(z)$ in the central region γ_0 satisfies, for some $\alpha_3 > 0$,

$$\begin{aligned} \Pi_k(z) &= \exp\left(z - \rho z^2 \lambda(2)^k + O(n^2\nu^k) + O(n^3\lambda(3)^k)\right) \\ &= e^z \exp[-\rho n^2 \lambda(2)^k] \left(1 + O(\frac{1}{n^{\alpha_3}})\right). \end{aligned} \tag{88}$$

This results from conditions $(C_2), (C_3)$, and condition (C_4) ensures that the variation of $z^2\lambda(2)^k$ on γ_0 is bounded.

Since the Cauchy coefficient integral applied to e^z yields $1/n!$, given the bound (87) on $\gamma \setminus \gamma_0$ and the uniform approximation (88) on γ_0, we conclude that, for $k \geq \kappa_1(n)$,

$$\pi_{k,n} = \exp[-\rho n^2 \lambda(2)^k] \left(1 + O(\frac{1}{n^{\alpha_3}})\right). \tag{89}$$

Equations (86) and (89) thus fully characterize the whole domain $k \geq 1$. They establish an approximation of the very same type as in the the Poisson model. The proof is now completed. □

As a methodological aside, we observe that the proof given above essentially amounts to a lifting of the estimates established for large *real* values of the Poisson parameter z to the case where z assumes values on (parts of) a large *complex* circle $z = ne^{i\theta}$. This fact is in agreement with the general principles of saddle-point depoissonization [16, 29]. It is also to be noted, following [16], that the double exponential limit formula for the distribution of height is not a limit probability distribution in the strictest classical sense: setting

$$\kappa_0(n) := \left\lfloor \frac{\log(\rho n^2)}{|\log \lambda(2)|} \right\rfloor$$

and leaving aside error terms, we have

$$\pi_{\kappa_0(n)+j,n} \approx e^{-\beta(n)\lambda(2)^j} \qquad \text{where} \qquad \beta(n) = \left(\frac{1}{\lambda(2)}\right)^{\{\log(\rho n^2)/|\log\lambda(2)|\}}, \tag{90}$$

with $\{u\}$ the fractional part function. In other words, the distribution is a discrete sampling of the double exponential function, as expressed by (89). Equivalently, there is a family (90) of distributions (governed by the bounded parameter $\beta(n)$) that change gradually and periodically as n increases through powers of $1/\lambda(2)$. Such periodicity phenomena expressed by discrete samplings of an extreme value distribution are by now fairly common in tries; see for instance [33] for a detailed discussion.

8. Applications

We briefly develop here three main applications to Bernoulli sources, Markov chains, and the continued fraction source. Statements here are specializations of Theorems 1 to 8 of the paper; they are systematically given under the Bernoulli model of index n and simplified forms are often stated so as to emphasize better the shape of results. We then conclude with a few open problems.

8.1. **Memoryless sources.** Memoryless sources are sources on a finite or infinite alphabet $\mathcal{M}$, where symbol m occurs independently with probability p_m. The standard Ruelle operator associated to the system is

$$\mathcal{G}_s[f](z) := \sum_{m\in\mathcal{M}} p_m^s\, f(q_m + p_m z), \qquad \text{with} \;\; q_m := \sum_{i<m} p_i.$$

The dominant eigenfunction is the same for each component operator $\mathcal{G}_{s,h}$ and it equals the constant function, for all values of $s, \mathrm{Re}(s) > \sigma$; then, the constant function is also the dominant eigenfunction of $\mathcal{G}_s$ itself, for all values of s, and $\lambda(s) = \sum_{m\in\mathcal{M}} p_m^s$ is the dominant eigenvalue. The dominant projector $e_s[f]$ is the integral $\int_0^1 f(t)dt$. More generally, the spectrum of $\mathcal{G}_s$ is

$$\mathrm{Sp}\,\mathcal{G}_s = \{\lambda_\ell(s) := \sum_{m\in\mathcal{M}} p_m^{s+\ell},\ \ell \geq 0\},$$

so that the Fredholm determinant is

$$\mathcal{F}(s,u) = \prod_{\ell\geq 0}\, [1 - u \sum_{m\in\mathcal{M}} {p_m}^{\ell+s}].$$

The eigenvector relative to the ℓ-th eigenvalue $\lambda_\ell(s)$ is a polynomial of degree ℓ. For symmetric memoryless sources, this ℓth eigenvalue is independent of s and in the case of two symbols, the family of eigenfunctions coincides exactly with Bernoulli polynomials [4], defined by

$$B_\ell(z) := \ell!\ [t^\ell]\, \frac{te^{zt}}{e^t - 1}.$$

The results obtained here are to be compared to the classical results on digital trees built on a finite alphabet $\mathcal{M} = \{1, 2, \dots, r\}$ where symbol i has probability p_i. We state[3]

Corollary 1. *Consider a finite memoryless source* S *with finite alphabet of cardinality* r *and probabilities* $\{p_i\}_{i=1}^r$. *The entropy and the coincidence probabilities are*

$$h(\mathsf{S}) = -\sum_i p_i \log p_i, \qquad c(\mathsf{S}) = \sum_i p_i^2.$$

(*i*) *Trie size and standard path length have expectations of the form*

$$S(n) \approx \frac{1}{h(\mathsf{S})} n, \qquad P(n) \sim \frac{n\log n}{h(\mathsf{S})}.$$

(*ii*) *Trie height has asymptotically a double exponential distribution with mean*

$$H(n) \sim \frac{2}{|\log c(\mathsf{S})|}\, \log n.$$

(*iii*) *The path lengths of hybrid tries satisfy*

$$P_L(n) = \frac{K_L(\mathsf{S})}{h(\mathsf{S})} n\log n, \qquad P_B(n) = \frac{K_L(\mathsf{S})}{h(\mathsf{S})} n\log n,$$

[3] We use the notation '$\approx$' for 'approximately equal', (as opposed to '$\sim$' reserved for the more precise 'asymptotically equivalent'), i.e., up to possible fluctuations induced by nonreal poles.

where

$$K_L(\mathsf{S}) = \sum_{i=1}^{r} (i-1)p_i, \qquad K_B(\mathsf{S}) = 2\sum_{i<j} \frac{p_i p_j}{p_i + \cdots + p_j}.$$

The statements relative to path length and height extend the analyses of [43, 54, 55] to the case of a general analytic density. The analysis of hybrid tries was already given in [7]. Devroye [11] considered the effect on the unbiased memoryless source ($p_1 = p_2 = \frac{1}{2}$) of a density that need not be analytic. Devroye showed that either $H(n) \sim 2\log n$ or $H(n) = +\infty$ depending on whether $\int f^2$ (with f the density) exist or not, and that $P(n) \sim n \log n$ as soon as f is square integrable.

Quite often, authors do not take into account the possible periodic fluctuations that may occur in the main asymptotic estimate of size. Here are some examples of periodic memoryless sources:

$(1/2, 1/4, 1/4)$, $\qquad (p, p, p^2)$ with $p = (1/2)(\sqrt{2}-1)$, $\qquad (p_m)_{m\geq 1}$ with $p_m = (1/2)^m$.

The case when the Fredholm determinant is pseudo–periodic, i.e.,

$$\mathcal{F}(s+it, u) = \mathcal{F}(s, e^{ia}u) \qquad \text{with } a \neq 2k\pi$$

is also interesting. A Bernoulli source is pseudo–periodic if and only if there exist two real numbers a and b such that b does not belong to the cyclic group $< a >$ generated by a and all the numbers p_m/b belong to this cyclic group $< a >$. An instance of this situation is $(2/5, 2/5, 1/5)$. As noted by Pollicott [44] and Fayolle et al. [15], there is an accumulation of s for which $\lambda(s) = 1$ on the left of the line $\mathrm{Re}(s) = 1$. Our improved Mellin argument of Prop. 12 shows directly that the total contribution of these poles contributes a term to the asymptotic expansion of size that remains $o(n)$, and a term $o(n \log n)$ in the expansion of path length.

8.2. **Markov chains.** We consider now the particular case of Markov chains. Here, the alphabet $\mathcal{M}$ is finite, of cardinality r, and the matrix Π_s whose general term is ${p_{i|j}}^s$ plays a central rôle. For $s = 1$, it is the transition matrix of the Markov chain. The spectrum of the matrix operator $\mathcal{G}_s$ is exactly the union of the spectra of the matrices $\Pi_{s+\ell}$, for all integers $\ell \geq 0$, so that

$$\mathrm{Sp}\,\mathcal{G}_s = \bigcup_{\ell\geq 0} \mathrm{Sp}\ \Pi_{s+\ell} \qquad \mathcal{F}(s,u) = \prod_{\ell \geq 0} \det(I - u\,\Pi_{s+\ell}).$$

If the eigenvalues of matrix Π_s are denoted by $\lambda^{(i)}(s)$ for $1 \leq i \leq r$, then

$$\mathrm{Sp}\,\mathcal{G}_s = \{\lambda^{(i)}(s+\ell)\,|1 \leq i \leq r,\ \ell \geq 0\},$$

and the eigenvector relative to eigenvalue $\lambda^{(i)}(s+\ell)$ has all its components that are polynomials of degree ℓ. Finally, the dominant eigenvalue of the operator $\mathcal{G}_s$ is exactly the dominant eigenvalue of the matrix Π_s, and the associated eigenfunction has all its components that are constants. Finally, the k-th component of the eigenfunction Ψ_1 is nothing more than the vector of stationary probabilities of the Markov chain.

Corollary 2. *Consider a Markov chain* S *with transition probabilities* $\{p_{i|j}\}_{1\leq i,j\leq r}$. *The entropy and the coincidence probabilities are*

$$h(\mathsf{S}) = -\sum_k \pi_k \sum_j p_{j|k} \log p_{j|k}, \qquad c(\mathsf{S}) = \lambda(2)$$

where π_k the k-th component of the vector of stationary probabilities of the Markov chain, and $\lambda(2)$ is the dominant eigenvalue of the termwise squared matrix $\left(p_{i|j}{}^2\right)$.

(*i*) *Trie size and standard path length have expectations of the form*

$$S(n) \approx \frac{1}{h(\mathsf{S})} n, \qquad P(n) \sim \frac{1}{h(\mathsf{S})} n \;\log n.$$

(*ii*) *Trie height has asymptotically a double exponential distribution with mean*

$$H(n) \sim \frac{2}{|\log \lambda(2)|} \log n.$$

(*iii*) *The expected values of path length in hybrid tries satisfy*

$$P_L(n) \sim \frac{K_L(\mathsf{S})}{h(\mathsf{S})} n \log n, \qquad P_B(n) \sim \frac{K_B(\mathsf{S})}{h(\mathsf{S})} n \log n,$$

where,

$$K_L(\mathsf{S}) = \sum_k \pi_k \sum_i (i-1) p_{i\,|\,k}, \quad K_B(\mathsf{S}) = 2 \sum_k \pi_k \sum_{i<j} \frac{p_{i|k} p_{j|k}}{p_{i|k} + \cdots + p_{j|k}}.$$

As mentioned in the introduction, this result supplements reference [26].

8.3. **The continued fraction source.** Preliminary results about the continued fraction source have been already reported in [21], but were limited to an initial density that is uniform. The Ruelle operator in this case is called the Ruelle–Mayer operator,

$$\mathcal{G}_s[f](z) := \sum_{m \geq 1} \frac{1}{(m+z)^{2s}} \; f(\frac{1}{m+z}),$$

and convergence is granted for complex s satisfying $\mathrm{Re}(s) > 1/2$.

The entropy of the source is related to what is known as Lévy's constant that plays a key rôle in the metric theory of continued fractions and the analysis of the Euclidean algorithm. The coincidence probability $\lambda(2)$ relative to the Ruelle Mayer–operator is a constant, sometimes known as "Vallée's constant", that intervenes in two–dimensional generalizations of the Euclidean algorithm [9, 21, 58]. (For these aspects, see also [21] and references therein.) The dominant eigenfunction of $\mathcal{G}_1$, known as Gauss's measure, is proportional to $1/(1+x)$.

Corollary 3. *Consider the continued fraction source* S *and any initial density that is analytic. The entropy and the coincidence probability are*

$$h(\mathsf{S}) = -\lambda'(1) = \frac{\pi^2}{6 \log 2} \qquad c(\mathsf{S}) = \lambda(2) \doteq 0.19945\,88183\,43767.$$

(*i*) *The expected values of size and path length satisfy*

$$S(n) = \frac{6 \log 2}{\pi^2} n + o(n), \qquad P_A(n) = \frac{6 \log 2}{\pi^2} \; n \log n + O(n).$$

(*ii*) *Height obeys a double exponential law with mean*

$$H(n) = \frac{2}{|\log \lambda(2)|} \log n + O(1).$$

(iii) The expected values of path length in hybrid tries satisfy

$$P_L(n) \sim \frac{3\log 2}{2\pi^2}\, n \log^2 n$$

$$P_B(n) \sim Cn\log n \ \textit{with}\ C := \frac{3}{\pi^2}\left[1 + 4\sum_{k\geq 2}\frac{1}{k^2-1}\log\frac{k+1}{2}\right].$$

The example of the continued fraction source is interesting, on many counts. First, as stated in (iii) above, the average path lengths $P_L(n)$ of the list-trie and the average path length $P_B(n)$ of the bst-trie turn out not to be of the same asymptotic order. when n tends to ∞. (Not too unexpectedly, the continued fraction source resembles a generalized Zipf source of parameter $\alpha = 2$; see Section 6.3.) Second, periodicity related issues are particularly fascinating.

The continued fraction source is aperiodic, and the poles of $(I - \mathcal{G}_s)^{-1}$ intervene in deep mathematical problems: They include all the nontrivial zeros of the Riemann zeta function. The other values s for which $\mathcal{G}_s$ has eigenvalue 1 are related to the eigenvalues of the hyperbolic Laplace operators and they lie on the line $\mathrm{Re}(s) = 1/2$ (see Efrat's paper [12]). However, these last values do not occur as poles of the Dirichlet series $\Lambda^{\langle A\rangle}(Id, s)$. In effect, in the half-plane $\mathrm{Re}(s) > 0$, the Dirichlet series $\Lambda(\mathrm{Id}, s)$ can be represented as (see Prop. 2, 3 of [21])

$$\Lambda(s) \equiv \Lambda^{\langle A\rangle}(Id, s) = 2\frac{\zeta(2s-1)}{\zeta(2s)}\,\frac{2^{1-s}-1}{1-s} + \frac{R(s)}{\zeta(2s)}$$

where $R(s)$ is analytic in $\mathrm{Re}(s) > 0$. So, when the initial density is uniform, the asymptotic expansions of $S(n)$ or $P_A(n)$ given in Theorem 6 solely involve the nontrivial zeros of the Riemann zeta function. For an arbitrary density, the same situation occurs, because the eigenvectors f of the hyperbolic Laplace operators satisfy $\int_0^1 f(x)dx = 0$, as pointed out to us by [40]. Thus, the nature of second order asymptotics of additive parameters must be related to the Riemann hypothesis.

Corollary 4. *The fluctuations in the mean value of path length and size of standard tries under the continued fraction source are related to the Riemann hypothesis. For instance, for size, one has*

$$S(n) = \frac{6\log 2}{\pi^2} n + \Phi(n) + O(1),$$

where $\Phi(n)$ is a sum indexed by the set Z of zeros of the Riemann zeta function in the critical strip $0 < \mathrm{Re}(s) < 1$,

$$\Phi(n) = \sum_{\chi\in Z} \mathrm{Res}\left(\Lambda(-s)\Gamma(s+1)n^{-s}\right)_{s=\chi/2}. \tag{91}$$

In particular, under the Riemann hypothesis (R.H.), we have $\Phi(n) = O(n^{1/4+\epsilon})$ for any $\epsilon > 0$. In addition, unconditionally with respect to R.H., the following two implications hold.

(i) If $\sigma_0 = \sup\{\mathrm{Re}(\xi),\ \xi \in Z\}$, then, for any $\epsilon > 0$, one has

$$\lim_{n\to\infty} \frac{\Phi(n)}{n^{\sigma_0/2+\epsilon}} = 0.$$

(*ii*) *Conversely, if* σ_1 *is the supremum of all* σ *such that* $\Phi(n)/n^{\sigma/2}$ *is unbounded, then* $\zeta(s)$ *has at least one zero in* $\mathrm{Re}(s) \geq \sigma_1$.

Proof. Equation (91) follows, at least formally, from a residue computation along a large rectangle $(-11/10 \pm iT, 1/10 \pm iT)$. Analytically, (91) is then justified by the fact that $\zeta(s)$ grows somewhat large along horizontal lines that traverse the critical strip at high altitudes, thereby avoiding zeros in the critical strip. Accordingly, $1/\zeta(s)$ is of weak polynomial growth in a strip that contains the critical strip. This last fact is granted by known properties of the Riemann zeta function; we omit details as they are entirely similar to the discussion of a formula of Ramanujan in Titchmarsh's book [56]; see Sections 9.7 and 9.8.

Once (91) is granted, the other implications follow easily. □

Numerical aspects of fluctuations. We offer here a succinct discussion of the fluctuating function $\Phi(n)$. We know, from the preceding discussion, that the function $\Lambda(s)$ is of weak polynomial growth. Let $\Phi^{[T_0]}(n)$ be the truncated sum of (91) restricted to zeros ξ of the zeta function such that $|\mathrm{Im}(\xi)| \leq T_0$. The difference, $|\Phi(n) - \Phi^{[T_0]}(n)|$ is the product of a polynomial in T_0 and of a quantity that is approximately $e^{-\pi T_0/2} n^{1/2}$, given the fast exponential decay of the Gamma function at $\pm i\infty$. In this discussion, we use $\doteq$ in the informal sense of 'roughly equal'.

Take for instance $T_0 = 5 \times 10^8$ corresponding to the first 1.5×10^9 zeros of the zeta function, as computed by van de Lune, te Riele and Winter in 1986. These zeros all satisfy the Riemann hypothesis. Forgetting about ancillary polynomial factors, an upper bound on the truncation error, when the terms corresponding to the high zeros of Z are neglected, is then at worst

$$|\Phi(n) - \Phi^{[T_0]}(n)| \leq E_1(n) \quad \text{where} \quad E_1(n) \doteq \left(10^{-6\times 10^9} \cdot n\right)^{1/2}.$$

This is thus totally negligible for all values of n less than huge threshold of about $10^{6\times 10^9}$.

For all practical purposes here the Riemann hypothesis is thus to be regarded as "true". We next focus attention on the first few terms in (91). Two ingredients intervene: the fast decay of the Gamma function and the fact that the first nonreal zeros of the zeta function have a fairly large imaginary part, being at $\xi_1, \xi_2 = \frac{1}{2} \pm 14.134i$. Assuming for the sake of illustration that the residue of $\Lambda(-s)$ at ξ_1, ξ_2 is of modulus at most 50, the corresponding contribution to $\Phi(n)$ —the one that should be numerically dominating anyway— is now bounded from above by $(8 \times 10^{-7}) \cdot n^{1/4}$. The next zeros are farther away where the Gamma function is even smaller, so that we may reasonably expect

$$|\Phi(n)| \leq 10^{-6} \cdot n^{1/4} \quad \text{for} \quad n \leq 10^{6\times 10^9}.$$

This quantity will become 1 only for n about 10^{24}.

In summary, the Riemann hypothesis, whether true or false, does not really affect the physical nature of the fluctuations. In addition, $\Phi(n)$, even though it oscillates from $-\infty$ to $+\infty$, is still going not to be detectable until n is quite large ($n > 10^{24}$). Fluctuations will be in particular completely offset by the *constant term* in the expansion of $S(n)$. This fact is also interesting since it provides a naturally occurring instance of the "Dumont phenomenon" (first discovered by Dumont in the 1980's and proved rigorously by Delange). The surprising phenomenon is that, empirically for values of x very close to 1, one has apparently the numerical convergence

$$\sum_{n=1}^{\infty} \mu(n) x^n \xrightarrow{?} \frac{1}{\zeta(0)} = -2, \qquad (x \to 1^-),$$

($\mu(n)$ is the Moebius function), although, in reality, the left hand side eventually oscillates unboundedly as $n \to +\infty$.

8.4. **Extensions and open problems.** For reasons related to database applications [13, 36], one often considers the variant of tries called bucket tries or b-tries, where recursion is halted as soon as subsets of words with cardinality $\leq b$ are encountered. (Standard tries correspond to $b = 1$ and subtrees of size $\leq b$ are then stored in "buckets" or "pages".). The framework developed in this paper applies almost verbatim, so that we only indicate briefly the corresponding conclusions.

Corollary 5. *For bucket tries with parameter b, height obeys a double exponential law*

$$\lim_{n\to\infty} \sup_{k\geq 0} \left| \Pr\{h_n \leq k\} - \exp[-\rho_b\, \lambda(b+1)^k\, n^{b+1}] \right| = 0,$$

for some $\rho_b > 0$ and its expected value satisfies

$$\mathrm{E}[h_n] \sim \frac{b+1}{|\log \lambda(b+1)|} \log n.$$

Size, defined as the number of internal nodes of the tree, has mean

$$S(n) \approx \frac{1}{b\, h(\mathsf{S})}\, n.$$

Various types of path lengths can be defined and similarly analysed: on average, a random branch in the tree still has depth $\log n / h(\mathsf{S})$ and it is only an $O(1)$ term that modifies the corresponding estimate of standard tries. It is to be noted that the parameters governing the distribution of height change radically, while the estimate for size shows that pages tend to be used to a fraction $h(\mathsf{S})$ of their capacity since on average a b-trie behaves in a way quantitatively similar to a perfect packing into pages of capacity $b \cdot h(\mathsf{S})$.

Preliminary investigations suggest that our theory may also applicable to suffix trees, despite difficulties due to correlations inherently present in that structure; see [27] for a treatment of classical models.

Globally, we may regard standard trie height as "well known", since its moments and its distribution are well characterized. We leave as an open problem of interest the following question.

(P_1) *Analysis of the height of hybrid tries* (list tries and bst-tries), where height is defined as the length of the maximal chain of pointers connecting the root to any external node. Perhaps probabilistic methods might be of use, since the problem amounts to determining the "balance" between the few long branches of the abstract trie and the many large collections of long branches inside structured nodes.

For additive parameters under general source models, little distributional information is available at the moment, even for standard tries. Here is a set of three open problems.

(P_2) *Concentration of distribution* almost certainly holds for size and path length. This is equivalent to saying that the standard deviation is of order strictly smaller than the mean. What one needs to establish is the exact cancellation between the main asymptotic orders of the squared mean and of the moment of order 2. This study should be interesting since, already for the standard trie and the uniform Bernoulli model, Prodinger and his collaborators have discovered surprising connections between such variance estimations and modular form identities [31, 32].

(P_3) *A limit distribution of the Gaussian type for size and path length* is likely to hold, at least in many subcases. This conjecture is suggested by hard analytical results of Jacquet, Régnier, and Szpankowski [25, 28] under classical source models; see also Mahmoud's book [39] for an account of some aspects. However, our feeling at the moment is that such limit laws might be dependent on deep conjectures on the eigenvalues $\lambda_j(s)$ and on perturbations of the equation $\lambda_j(s) = 1$ in the form of $\lambda_j(s) = e^{it}$ with t small. (In the continued fraction case, the Riemann hypothesis lies immediately below the surface.)

(P_4) *The limit distribution of the search cost of a random item in standard and hybrid tries* is most probably Gaussian in many cases. The property is known for biased memoryless models, although it does not hold for the unbiased model. This question enables us to conclude on an optimistic note with a problem likely to be somewhat tractable, since it reduces to an analysis of the mean number of nodes at each level in a random tree.

Acknowledgements. This work was supported in part by the Long Term Research Project ALCOM-IT (# 20244) of the European Union.

References

[1] ALLEN, B., AND MUNRO, I. Self-organizing binary search trees. *Journal of the ACM 25*, 4 (Oct. 1978), 526–535.

[2] BEDFORD, T., KEANE, M., AND SERIES, C. *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces.* Oxford University Press, 1991.

[3] BENTLEY, J., AND SEDGEWICK, R. Fast algorithms for sorting and searching strings. In *Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* (January 1997), SIAM Press, pp. 360–369. New Orleans.

[4] BOGOMOLNY, E. B., AND CARIOLI, M. Quantum maps from transfer operators. *Physica D 67* (1993), 88–112.

[5] BURGE, W. H. An analysis of binary search trees formed from sequences of nondistinct keys. *JACM 23*, 3 (July 1976), 451–454.

[6] CLAMPETT, H. A. Randomized binary searching with tree structures. *Communications of the ACM 7*, 3 (Mar. 1964), 163–165.

[7] CLÉMENT, J., FLAJOLET, P., AND VALLÉE, B. The analysis of hybrid trie structures. In *Proceedings of the Ninth Annual ACM–SIAM Symposium on Discrete Algorithms* (Philadelphia, 1998), SIAM Press, pp. 531–539.

[8] CLÉMENT, J., AND ZIMMERMANN, P. An efficient spell-checker based on ternary search tries (provisional title). Research report, Institut National de Recherche en Informatique et en Automatique, 1999. In preparation.

[9] DAUDÉ, H., FLAJOLET, P., AND VALLÉE, B. An average-case analysis of the Gaussian algorithm for lattice reduction. *Combinatorics, Probability and Computing 6* (1997), 397–433.

[10] DE BRUIJN, N. G. *Asymptotic Methods in Analysis.* Dover, 1981. A reprint of the third North Holland edition, 1970 (first edition, 1958).

[11] DEVROYE, L. A probabilistic analysis of the height of tries and of the complexity of triesort. *Acta Informatica 21* (1984), 229–237.

[12] EFRAT, I. Dynamics of the continued fraction map and the spectral theory of $SL(2, Z)$. *Inventiones Mathematicae 114* (1993), 207–218.

[13] Fagin, R., Nievergelt, J., Pippenger, N., and Strong, R. Extendible hashing: A fast access method for dynamic files. *A.C.M. Transactions on Database Systems 4* (1979), 315–344.

[14] Faivre, C. Distribution of Lévy's constants for quadratic numbers. *Acta Arithmetica 1*, 61 (1992), 13–34.

[15] Fayolle, G., Flajolet, P., and Hofri, M. On a functional equation arising in the analysis of a protocol for a multi-accessbroadcast channel. *Adv. Appl. Prob.*, 18 (1986), 441–472.

[16] Flajolet, P. On the performance evaluation of extendible hashing and trie searching. *Acta Inf. 20* (1983), 345–369.

[17] Flajolet, P., Gardy, D., and Thimonier, L. Birthday paradox, coupon collectors, caching algorithms, and self–organizing search. *Discrete Applied Mathematics 39* (1992), 207–229.

[18] Flajolet, P., Gourdon, X., and Dumas, P. Mellin transforms and asymptotics : Harmonic sums. *Theoretical Computer Science 144*, 1–2 (June 1995), 3–58.

[19] Flajolet, P., and Puech, C. Partial match retrieval of multidimensional data. *Journal of the ACM 33*, 2 (1986), 371–407.

[20] Flajolet, P., and Steyaert, J.-M. A branching process arising in dynamic hashing, trie searching and polynomial factorization. In *Automata, Languages and Programming* (1982), M. Nielsen and E. M. Schmidt, Eds., vol. 140 of *Lecture Notes in Computer Science*, Springer Verlag, pp. 239–251. Proceedings of 9th ICALP Colloquium, Aarhus, Denmark, July 1982.

[21] Flajolet, P., and Vallée, B. Continued fraction algorithms, functional operators and structure constants. *Theoretical Computer Science*, 194 (1998), 1–34.

[22] Gonnet, G. H., and Baeza-Yates, R. *Handbook of Algorithms and Data Structures: in Pascal and C*, second ed. Addison–Wesley, 1991.

[23] Grothendieck, A. *Produits tensoriels topologiques et espaces nucléaires.* No. 16 in Memoirs of the American Mathematical Society. A.M.S., Providence, 1955.

[24] Grothendieck, A. La théorie de Fredholm. *Bulletin de la Société Mathématique de France 84* (1956), 319–384.

[25] Jacquet, P., and Régnier, M. Trie partitioning process: Limiting distributions. In *CAAP'86* (1986), P. Franchi-Zanetacchi, Ed., vol. 214 of *Lecture Notes in Computer Science*, pp. 196–210. Proceedings of the 11th Colloquium on Trees in Algebra and Programming, Nice France, March 1986.

[26] Jacquet, P., and Szpankowski, W. Analysis of digital tries with Markovian dependency. *IEEE Transactions on Information Theory 37*, 5 (1991), 1470–1475.

[27] Jacquet, P., and Szpankowski, W. Autocorrelation on words and its applications: analysis of suffix trees by string-ruler approach. *Journal of Combinatorial Theory. Series A 66*, 2 (1994), 237–269.

[28] 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.

[29] Jacquet, P., and Szpankowski, W. Analytical de-Poissonization and its applications. *Theoretical Computer Science 1-2*, 201 (1998), 1–62.

[30] Kato, T. *Perturbation Theory for Linear Operators.* Springer-Verlag, 1980.

[31] Kirschenhofer, P., and Prodinger, H. Eine Anwendung der Theorie der Modulfunktionen in der Informatik. *Sitzungsberichte der Österreichischen Akademie der Wissenschaften 197* (1988), 339–366.

[32] Kirschenhofer, P., and Prodinger, H. On some applications of formulæ of Ramanujan in the analysis of algorithms. *Mathematika 38* (1991), 14–33.

[33] Kirschenhofer, P., Prodinger, H., and Szpankowski, W. Analysis of a splitting process arising in probabilistic counting and other related algorithms. *Random Structures & Algorithms 9* (1996), 379–402.

[34] Knuth, D. E. *The Art of Computer Programming*, third ed., vol. vol. 3: Sorting and Searching. Addison-Wesley, 1998.

[35] Krasnoselskii, M. *Positive solutions of operator equations.* P. Noordhoff, Groningen, 1964.

[36] Larson, P. A. Dynamic hashing. *BIT 18* (1978), 184–201.

[37] Lind, D., and Marcus, B. *An introduction to symbolic dynamics and coding.* Cambridge University Press, Cambridge, 1995.

[38] Lothaire, M. *Combinatorics on Words*, vol. 17 of *Encyclopedia of Mathematics and its Applications.* Addison–Wesley, 1983.

[39] Mahmoud, H. *Evolution of Random Search Trees.* John Wiley, New York, 1992.

[40] Mayer, D. H. Private communication.

[41] Mayer, D. H. Spectral properties of certain composition operators arising in statistical mechanics. *Commun. Math. Phys. 68* (1979), 1–8.

[42] Mayer, D. H. On composition operators on Banach spaces of holomorphic functions. *journal of functional analysis*, 35 (1980), 191–206.

[43] Pittel, B. Paths in a random digital tree: limiting distributions. *Advances in Applied Probability 18*, 1 (1986), 139–155.
[44] Pollicott, M. A complex Ruelle–Perron–Frobenius theorem and two counterexamples. *Ergod. Th. and Dynam. Syst.*, 4 (1984).
[45] Prodinger, H. Combinatorics of geometrically distributed random variables: Left-to-right maxima. *Discrete Mathematics 153* (1996), 253–270.
[46] Régnier, M. On the average height of trees in in digital search and dynamic hashing. *Information Processing Letters 13* (1982), 64–66.
[47] Rivest, R. L. Partial match retrieval algorithms. *SIAM Journal on Computing 5* (1976), 19–50.
[48] Ruelle, D. *Thermodynamic formalism.* Addison Wesley, 1978.
[49] Ruelle, D. *Dynamical Zeta Functions for Piecewise Monotone Maps of the Interval*, vol. 4 of *CRM Monograph Series.* American Mathematical Society, Providence, 1994.
[50] Schwartz, H. *Composition operators in $\mathcal{H}^p$.* PhD thesis, Univ. of Toledo, 1969.
[51] Sedgewick, R. *Algorithms in C: Fundamentals, Data Structures, Sorting, Searching*, third ed. Addison–Wesley, Reading, Mass., 1988.
[52] Shapiro, J. Compact composition operators on spaces of boundary regular holomorphic functions. In *Proceedings of the AMS* (1997), vol. 100, pp. 49–57.
[53] Shapiro, J., and Taylor, P. Compact, nuclear, and Hilbert–Schmidt composition operators on $\mathcal{H}^2$. *Indiana Univ. Math. J.*, 23 (1973), 471–496.
[54] Szpankowski, W. Some results on V-ary asymmetric tries. *Journal of Algorithms 9* (1988), 224–244.
[55] Szpankowski, W. On the height of digital trees and related problems. *Algorithmica 6*, 2 (1991), 256–277.
[56] Titchmarsh, E. C., and Heath-Brown, D. R. *The Theory of the Riemann Zeta-function*, second ed. Oxford Science Publications, 1986.
[57] Trabb Pardo, L. Set representation and set intersection. Tech. rep., Stanford University, 1978.
[58] Vallée, B. Algorithms for computing signs of 2×2 determinants: dynamics and average–case analysis. In *Algorithms—ESA'97* (1997), G. W. R. Burkhard, Ed., no. 1136 in Lecture Notes in Computer Science, pp. 486–499. Proceedings of the Fifth European Symposium on Algorithms, Graz, September 1997.
[59] Vallée, B. Opérateurs de Ruelle-Mayer généralisés et analyse des algorithmes d'Euclide et de Gauss. *Acta Arithmetica 2*, 81 (1997), 101–144.
[60] Vallée, B. Dynamics of the binary euclidean algorithm: Functional analysis and operators. *Algorithmica 22*, 4 (1998), 660–685.
[61] Vallée, B. Dynamical sources in information theory: Fundamental intervals and words prefixes. *Technical report of the University of Caen, Les cahiers du GREYC*, 2 (1999). Submitted.
[62] Vallée, B., and Lemée, C. Average-case analyses of three algorithms for computing the jacobi symbol. Rapport de Recherche de l'Université de Caen, Les Cahiers du GREYC # 1, 1999. Submitted.
[63] Yao, A. C.-C. A note on the analysis of extendible hashing. *Information Processing Letters 11* (1980), 84–86.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
(France)
http://www.inria.fr
ISSN 0249-6399