Algorithmica (2001) 29: 262–306
DOI: 10.1007/s004530010061

Algorithmica
© 2001 Springer-Verlag New York Inc.

Dynamical Sources in Information Theory: Fundamental Intervals and Word Prefixes

B. Vallée[1]

Abstract. A quite general model of source that comes from dynamical systems theory is introduced. Within this model, some basic problems of algorithmic information theory contexts are analysed. The main tool is a new object, the generalized Ruelle operator, which can be viewed as a "generating" operator for fundamental intervals (associated to information sharing common prefixes). Its dominant spectral objects are linked with important parameters of the source, such as the entropy, and play a central rôle in all the results.

Key Words. Information theory, Dynamical systems, Transfer operator, Sources, Entropy, Fundamental intervals.

1. Introduction. In information theory contexts, data items are (infinite) words that are produced by a common mechanism, called a source. Real-life sources are often complex objects. We introduce here a general framework of sources related to dynamical systems theory which goes beyond the cases of memoryless and Markov sources. This model can describe non-Markovian processes, where the dependency on past history is unbounded, and as such, they attain a high level of generality.

A probabilistic dynamical source is defined by two objects: a symbolic mechanism and a density. The mechanism is related to symbolic dynamics and associates an infinite word $M(x)$ to a real number x of the $[0, 1]$ interval. It can be viewed as a generalization of numeration systems, the binary expansion of a real x, or the continued fraction expansion of the real x being well-known instances. Once the mechanism has been fixed, the density f on the $[0, 1]$ interval can vary. This then induces different probabilistic behaviours for source words: for instance, the distribution of the binary expansions of reals depends on the distribution of the reals themselves. The dependence on the initial input distribution has already been considered by Devroye [8] when he studies digital trees associated to the simplest source, the Bernoulli source. We treat here a more general model defined from a general dynamical source *and* a general initial input distribution, which we call a *probabilistic dynamic source.*

In probabilistic dynamic contexts, an important tool is the Ruelle *transfer operator*: it is classically used as a "generating operator" since it can easily generate objects that are essential in the analysis. Here, we are interested in problems which come from computational information theory. Most of the problems in this area deal with *prefixes* of words. All the source words which begin with the same prefix "come from" an interval of $[0, 1]$ that is called a fundamental interval. However, the classical Ruelle operator cannot generate both ends of these intervals at the same time, so that it is not adequate in information theory contexts. We thus devise a new tool, the generalized Ruelle operator,

[1] GREYC, Université de Caen, F-14032 Caen, France. Brigitte.Vallee@info.unicaen.fr.

Received September 15, 1998; revised March 16, 1999. Communicated by H. Prodinger and W. Szpankowski. Online publication October 6, 2000.

that acts on functions of two variables and relates to an earlier generalization of [39]. Like in its classical version, the generalized Ruelle operator depends on a complex parameter s. The operator is now a "generating" operator for fundamental intervals; more precisely, it makes it possible to express the *Dirichlet series of fundamental intervals* that plays a central rôle in all the analyses of the paper.

Furthermore, positivity properties of the (generalized) Ruelle operator (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 intervenes everywhere in the analysis of the source. First, the main intrinsic parameters of the source, like the *entropy* $h(\mathcal{S})$ and the *coincidence probability* $c(\mathcal{S})$, are proven to be independent of the initial density f on the unit interval; they depend only on the mechanism of the source and they satisfy

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

More generally, the dominant eigenvalue function intervenes in all the results of the paper.

(i) The number $B(x)$ of finite prefixes whose occurrence probability is at least equal to x (for $x \to 0$) is analysed. It satisfies, for "most" of the sources,

$$B(x) \simeq \frac{-1}{\lambda'(1)} \frac{1}{x} \qquad \text{for} \quad x \to 0.$$

(ii) The distribution of the prefixes of length k is shown to follow asymptotically a log-normal law, for "most" of the sources. This proves a strong "equipartition property", in the flavour of the Shannon–MacMillan–Breimann Theorem. The dominant behaviours of the mean and the variance involve the first two derivatives of the function $\log \lambda(s)$ at $s = 1$.

(iii) The coincidence between two source words, i.e., the length of their longest common prefix, is shown to follow a geometric law asymptotically. The ratio of the geometric law depends on the drawing of the words. If the two words are independently drawn, then the ratio equals $\lambda(2)$. More generally, the ratio equals $\lambda(2 + r)$ where $r > -1$ is a parameter which is linked to the drawing of the words.

REMARK. In the first two results, we use the informal term "most of the sources". The exceptions correspond to very particular cases that will be precisely characterized.

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. Other common instances of sources studied in the literature are only that of stationary ergodic sources, which possess good "mixing" properties, see for instance the work of Szpankowski et al. [36], [37]. To the best of our knowledge, the only instance of an explicit source with unbounded history dependency considered in the literature is that of continued fraction representations [13]. Our probabilistic dynamical source model encompasses all the usual models previously studied so that our approach both unifies previous analyses in a powerful way and extends previous results.

Plan of the paper. Section 2 describes the general framework of dynamical sources, and defines fundamental intervals, as well as two basic parameters of the source, the entropy and the coincidence probability. Section 3 is devoted to the description of the main questions. It is shown that all of them involve the Dirichlet series of fundamental measures. In Section 4 we introduce the generalized Ruelle operator, and show how it generates the fundamental measures. In Section 5 we then transfer the properties of the operator on the Dirichlet series of fundamental intervals, and we relate dominant spectral objects of the Ruelle operator to basic parameters of the source, like the entropy and the coincidence probability. In Section 6 we study more precisely some important properties of the dominant eigenvalue. We obtain classification results that will help to characterize the exceptions. With the results obtained in Sections 5 and 6, we can return to our questions and solve them in Sections 7–9. We conclude with an example in Section 10.

We focus here on the application of this model to problems which involve words, and more precisely word prefixes. A companion paper (a joint work with Julien Clément and Philippe Flajolet) deals with further analyses on an important class of trees used in information theory: the digital trees (or tries). The main tools that we introduce here, as the generalized Ruelle operator, or the Dirichlet series of fundamental intervals, also play a central rôle there. Thus, the companion paper will make great use of Sections 2 and 4–6 of this paper.

2. Probabilistic Dynamical Sources. Here, we describe the general framework of probabilistic dynamical sources. First, we introduce *symbolic* dynamical sources, with two different possible mechanisms—basic, Markovian. These mechanisms are associated to *dynamical systems* defined from expanding analytical maps of the interval. Then, when endowing the unit interval with some (analytical) density, we define the concept of probabilistic dynamical sources. Finally, we present the notion of fundamental intervals, fundamental measures, and two basic parameters of the source, the entropy and the coincidence probability.

2.1. *Basic Symbolic Dynamical Sources.* In information theory contexts, a source is a mechanism which produces infinite words written on an alphabet $\mathcal{M}$. We are first interested here in sources that are associated to *basic dynamical systems*, where the mechanism is the same at each step.

DEFINITION 2.1 (Basic Symbolic Dynamical Source). A basic dynamical source is defined by four elements:

(a) An alphabet $\mathcal{M}$ included in $\mathbf{N}$, 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., $\bar{\mathcal{I}} = \bigcup_{m\in\mathcal{M}} \bar{\mathcal{I}}_m$.
(c) A mapping σ which is constant and equal to m on each $\mathcal{I}_m$.
(d) A 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 let $\mathcal{H}$ be the set $\mathcal{H} := \{h_m,\ m \in \mathcal{M}\}$. There exists a common complex neighbourhood $\mathcal{V}$ of $\bar{\mathcal{I}}$ on which the set $\mathcal{H}$ satisfies

the following:

(d1) The mappings h_m extend to holomorphic maps on $\mathcal{V}$, mapping $\mathcal{V}$ strictly inside $\mathcal{V}$ (i.e., $h_m(\bar{\mathcal{V}}) \subset \mathcal{V}$).

(d2) The mappings $|h'_m|$ extend to holomorphic maps $\widetilde{h}_m$ on $\mathcal{V}$ and there exists $\delta_m < 1$ for which $0 < |\widetilde{h}_m(z)| \le \delta_m$ for $z \in \mathcal{V}$.

(d3) There exists $\gamma < 1$ for which the series $\sum_{m \in \mathcal{M}} \delta_m^s$ converges on $\Re(s) > \gamma$.

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 of a dynamical system that the image $T(\bar{\mathcal{I}}_m)$ is a union of some elements $\bar{\mathcal{I}}_j$ of the partition. Conditions (d1) and (d2) express that the inverse branches h_m are contractions, or that T is expansive. Condition (d3) always holds for a finite alphabet (with $\gamma = -\infty$); it is only useful for infinite alphabets.

It is sufficient that an iterate of T satisfies conditions (d). For instance, the mapping T associated to a continued fraction source does not fulfil conditions (d1) and (d2) but the iterate T^2 does.

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

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

where the kth component $M_k(x)$ of $M(x)$ is equal to $\sigma(T^{k-1}x)$, while the kth prefix $P_k(x)$ of $M(x)$ is equal to

$$P_k(x) := (M_1(x), M_2(x), \ldots, M_k(x)). \tag{2}$$

The number of branches of T equals the cardinality of the alphabet, and the alphabet is used for coding the distinct branches of T, or the distinct inverse branches of T which are denoted h_m. Here, h_m is a bijection from $\mathcal{I}$ to $\mathcal{I}_m$, which coincides with the inverse of the restriction of T to $\mathcal{I}_m$.

Memoryless sources. All the 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 Bernoulli source associated to a probability system $P = (p_m)_{m\in\mathcal{M}}$ (finite or denumerable) is the source where all the components M_k are independent and follow a Bernoulli law of parameters $(p_m)_{m\in\mathcal{M}}$. The topological partition of $\mathcal{I}$ is then defined by

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

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

$$T(x) = \{bx\}, \qquad \sigma(x) = [bx], \tag{3}$$

where $[u]$ is the integer part of u and $\{u\} = u \bmod 1 = u - [u]$ 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 Bernoulli sources (i.e., Bernoulli sources where all p_j's are equal).

Continued fraction expansion. This general framework may also create quite different sources, with memory. It is sufficient to use a mapping T with at least one non-affine branch. In a sense, it is the derivative $T'(x)$ that keeps memory of previous history. The *continued fraction* transformation is an example of this situation. The alphabet is $\mathbf{N}$, 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 function $T(x) := (1/x) - m$,

$$T_{\mathrm{CF}}(x) = \left\{\frac{1}{x}\right\}, \qquad \sigma_{\mathrm{CF}}(x) = \left[\frac{1}{x}\right], \tag{4}$$

where $[u]$ is the integer part of u and $\{u\} = u \bmod 1 = u - [u]$ is the fractional part of u. This transformation gives rise to the continued fraction expansion of x.

The inverse branches are all the linear fractional transformations (LFTs) h_m defined by $h_m(x) := 1/(x+m)$. The first branch h_1 does not satisfy (d1) and (d2), but the set of the LFTs $h_m \circ h_n$ satisfies conditions (d).

2.2. *Markov Symbolic Dynamical Sources.* Until now, the shift T used at each stage is always defined in the same way. Very often, the modelling of more realistic sources leads to the use of a shift that depends on the last emitted symbol. This gives rise to so-called Markov sources.

DEFINITION 2.2 (Markov Symbolic Dynamical Sources). Let $\mathcal{M}$ be a finite alphabet of cardinality r, and let $\mathcal{S} = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_i, \ldots, \mathcal{S}_r)$ be a set of $r+1$ different basic dynamical systems, all defined on the same alphabet $\mathcal{M}$. A Markov dynamical source is then defined as follows: the basic dynamical system $\mathcal{S}_0$ is used to begin with, and the dynamical system $\mathcal{S}_j$ is chosen when the previously emitted symbol is j.

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

$$M(x) := (M_1(x), \ldots, M_k(x), \ldots), \tag{5}$$

together with the sequence of the iterates of the real x,

$$(T^{\langle 1\rangle}(x), T^{\langle 2\rangle}(x), \ldots, T^{\langle k\rangle}(x), \ldots), \tag{6}$$

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

$$\text{if} \quad M_k(x) = j, \qquad \text{then} \quad T^{\langle k+1\rangle}(x) := T_j(T^{\langle k\rangle}(x)) \quad \text{and} \quad M_{k+1}(x) := \sigma_j(T^{\langle k\rangle}(x)). \tag{7}$$

Like previously in the case of a basic source (2), the kth prefix $P_k(x)$ of $M(x)$ is equal to

$$P_k(x) := (M_1(x), M_2(x), \ldots, M_k(x)). \tag{8}$$

Each shift T_j is associated to a topological partition $(\mathcal{I}_{i|j})$ $(1 \le i \le r)$ of the unit interval $\mathcal{I}$, and satisfies hypotheses (d) of Definition 2.1. 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 required 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 $\bar{\mathcal{I}}_{i|j}$ that extends to a holomorphic map on $\mathcal{V}$, mapping $\mathcal{V}$ strictly inside $\mathcal{V}$ (i.e., $h_{i|j}(\bar{\mathcal{V}}) \subset \mathcal{V}$ for $1 \le i \le r$, $0 \le j \le r$).

The usual model of Markov chains of order 1 is then relative to the case when all the $\mathcal{S}_j$'s are Bernoulli systems. More precisely, if the system $\mathcal{S}_j$ is a Bernoulli system of parameters $\Pi_j := (p_{i|j})_{i \le r}$, the transition matrix Π of the Markov chain is the square matrix $r \times r$,

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

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 (general) dynamical system (see Figure 1). 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 $\mathcal{S}$ associated to quasi-partition $\mathcal{I}_{i,j}$ of $]0, r+1[$ and to branches $T_{i,j}$ is a dynamical system (that is no longer basic).

One can use both interpretations of a Markovian source, but, here, we prefer stay in the unit interval and we adopt the first formalism which is closer to the intuition behind Markov chains. This is the point of view that has been adopted by Ruelle himself [29].

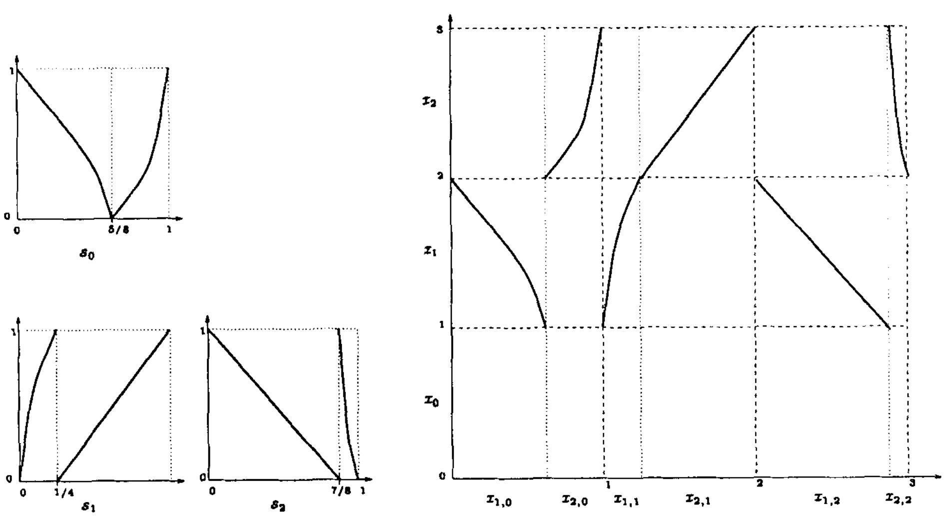

Fig. 1. Relation between a Markov source (defined by three basic sources $\mathcal{S}_0$, $\mathcal{S}_1$, $\mathcal{S}_2$) and the corresponding general dynamical system.

2.3. *Fundamental Intervals and Prefixes.* We consider now the kth iterate of the shift. In the case of a basic source, this is the true kth iterate of T. In the case of a Markovian source, this is the shift $T^{(k)}$ defined in (7). Each branch (or each inverse branch) of the kth 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 $|h|$. An inverse branch h of depth k is then associated in a unique way to a finite word $W(h) = (m_1, m_2, \dots, m_k)$ of length k which keeps the memory of the choices. In the basic case, each inverse branch of depth k associated to $(m_1, m_2, \dots, m_k)$ is of the form

$$h = h_{m_1} \circ h_{m_2} \circ \cdots \circ h_{m_k}, \tag{9}$$

where h_i denotes the ith branch of T. In the Markov case, it is of the form

$$h = h_{m_1|0} \circ h_{m_2|m_1} \circ \cdots \circ h_{m_k|m_{k-1}}. \tag{10}$$

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 begins and finishes 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 begin and finish with symbol i. In the same vein, $\mathcal{C}_k$, $\mathcal{C}_k[i]$ denote the same objects of fixed depth k.

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

DEFINITION 2.3 (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 $P_k(x)$ of length k is exactly the finite word associated to h.

2.4. *Probabilistic Dynamical Sources.* In what follows, we are interested in *probabilistic dynamical sources*, where the words are emitted by the mechanism of the source, but also with a prescribed distribution which depends on a density on the unit interval.

DEFINITION 2.4 (Probabilistic Dynamical Sources). Let $\mathcal{S}$ be a dynamical source (basic or Markovian) and let f be a real analytic density on interval $\mathcal{I}$ that extends to an analytical function on $\mathcal{V}$. Let F be the associated distribution function. The pair $(\mathcal{S}, F)$ is called a probabilistic dynamical source. The set $\mathcal{M}^\infty$ of the words produced by the dynamical probabilistic source $(\mathcal{S}, F)$ 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 $\mathcal{I}_h$,

$$u_h := |F(h(0)) - F(h(1))|, \tag{11}$$

associated to an inverse branch h defined in (9) or (10), plays an important rôle, since it equals the probability that a source word begins with the prefix of $\mathcal{M}^*$ relative to h. It is called the *fundamental measure* relative to h.

More generally, when studying two source words $M(x)$ and $M(y)$ independently drawn from the same probabilistic dynamical source $(\mathcal{S}, f)$, the square $\mathcal{I} \times \mathcal{I}$ is endowed with continuous density $g(x, y) = f(x)f(y)$, and μ denotes the associated measure on the square $\mathcal{I} \times \mathcal{I}$. In this context, the measure $\mu(C_h)$ of the fundamental square $C_h := \mathcal{I}_h \times \mathcal{I}_h$ is equal to u_h^2 and it plays an important rôle, since it represents the probability that two source words both begin with the same prefix relative to h.

2.5. *Dirichlet Series of Fundamental Intervals. Entropy, Coincidence Probability.* The *entropy* $h(\mathcal{S}, F)$ relative to probabilistic dynamical source $(\mathcal{S}, F)$ is defined as the limit, if it exists, of a quantity that involves the fundamental measures u_h,

$$h(\mathcal{S}, F) := \lim_{k\to\infty} \frac{-1}{k} \sum_{|h|=k} u_h \log u_h. \tag{12}$$

In the same vein, the probability that two independent words have the same prefix of length k equals $\sum_{|h|=k} u_h^2$. Since this quantity generally appears to decrease exponentially with k, it is natural to define the *coincidence probability* $c(\mathcal{S}, F)$ as the following limit, if it exists,

$$c(\mathcal{S}, F) := \lim_{k\to\infty} \left(\sum_{|h|=k} u_h^2\right)^{1/k}. \tag{13}$$

More generally, for an integer $b \geq 2$, the *b-coincidence probability* $c_b(\mathcal{S}, F)$, defined as

$$c_b(\mathcal{S}, F) := \lim_{k\to\infty} \left(\sum_{|h|=k} u_h^b\right)^{1/k}, \tag{14}$$

is related to the probability that b independent words have the same prefix of length k.

The previous three definitions involve the *series of fundamental measures of depth k*,

$$\Lambda_k(F, s) := \sum_{|h|=k} u_h^s = \sum_{|h|=k} |F(h(0)) - F(h(1))|^s, \tag{15}$$

since

$$h(\mathcal{S}, F) := \lim_{k\to\infty} \frac{-1}{k} \frac{d}{ds} \Lambda_k(F, s)|_{s=1}, \tag{16}$$

$$c(\mathcal{S}, F) := \lim_{k\to\infty} [\Lambda_k(F, 2)]^{1/k}, \qquad c_b(\mathcal{S}, F) := \lim_{k\to\infty} [\Lambda_k(F, b)]^{1/k}. \tag{17}$$

We show in what follow that the quantities $\Lambda_k(F, s)$ defined in (15) asymptotically behave as the kth power of a certain function $\lambda(s)$ that is well defined and analytic near the real axis. The three objects defined in (12)–(14) depend only on the mechanism $\mathcal{S}$. They are independent of the distribution F and can only be expressed with the function $\lambda(s)$ as

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

They play an important rôle in all our analyses.

2.6. *Comparing with Other Models of Sources.* First, our model encompasses the most usual sources, like memoryless sources, or Markov chains, but also most of the main numeration systems such as those associated to base b or to continued fraction expansion. Since the model is not only symbolic, but also probabilistic, it is possible to examine properties of the expansions (for instance dyadic expansions or continued fraction expansions) of numbers that are not uniformly distributed on the unit interval.

More generally, the sources studied in information theory are essentially ergodic and stationary. We show in what follows that our sources are all ergodic, but not necessarily stationary. Moreover, the proposed framework encompasses essentially all the instances of sources that satisfy some "easy" sufficient conditions of ergodicity (for instance, unicity of the invariant measure and mixing).

3. Distribution of Prefixes Source Words

3.1. *Three Questions about Prefixes.* Generally, in information theory contexts, we are interested in three basic questions about the probability that given source words appear. All these questions concern prefixes of words and can be easily translated into questions about fundamental measures.

(i) Evaluate the number $B(\rho)$ of finite prefixes whose probability is at least equal to ρ (for $\rho \to 0$). The number $B(\rho)$ is alternatively defined as the number of fundamental measures at least equal to ρ.

(ii) Describe the distribution of the prefixes of the same fixed length k. Alternatively, describe the distribution of the fundamental measures of depth k. More precisely, we let $\ell_k(x) := u_h$ when x belongs to the fundamental interval $h(\mathcal{I})$ of depth k. Since the fundamental intervals of depth k form a quasi-partition of $\mathcal{I}$, the random variable ℓ_k is almost everywhere defined on $\mathcal{I}$ and we examine characteristics of its distribution when x is distributed over the interval $\mathcal{I}$ according density f.

(iii) Describe the coincidence between two source words, i.e., the length of their longest common prefix. For two words $M(x)$ and $M(y)$, one defines $C(x, y)$ to be the length of the longest common prefix of $M(x)$ and $M(y)$,

$$C(x, y) = \mathrm{Max}\{k \in \mathbf{N};\ P_k(x) = P_k(y)\}.$$

Then C is a random variable almost everywhere defined on the square $\mathcal{I} \times \mathcal{I}$ and we examine characteristics of its distribution when (x, y) is distributed over the square $\mathcal{I} \times \mathcal{I}$ according density $g(x, y)$.

3.2. *The Main Results.* Our main results all involve the dominant eigenvalue $\lambda(s)$ of the Ruelle operator; they are the following: For all the dynamical sources, but the exceptional ones:

(i) The number $B(\rho)$ of fundamental intervals whose measure is at least equal to ρ satisfies

$$B(\rho) \simeq \frac{1}{\lambda'(1)\rho} \qquad \text{for} \quad \rho \to 0.$$

This result is true provided that the function $s \to \lambda(s)$ is not periodic (Theorem 2).

(ii) The variable $\log \ell_k$ follows asymptotically a normal law (for $k \to \infty$), provided that $\log \lambda(s)$ is strictly concave; the mean is asymptotically equivalent to $k\lambda'(1)$; the variance is asymptotically equivalent to $k[\lambda''(1) - \lambda'(1)^2]$ (Theorem 1).

(iii) The variable C follows asymptotically a geometric law. The ratio of the geometric law depends only on the behaviour of the initial density g near the diagonal $x = y$ of the unit square. If g has valuation r near the diagonal, i.e.,

$$g(x, y) = |x - y|^r \ell(x, y) \tag{18}$$

with ℓ defined in the square and strictly positive in the unit square, and $r > -1$, then the ratio is of the form $\lambda(2 + r)$ (Theorem 3).

Furthermore, we give a precise characterization of exceptional dynamical sources (Proposition 11). There are two kinds of exceptional sources, both related to exceptional properties of the dominant eigenvalue $\lambda(s)$. The first one is relative to the case when the dominant eigenvalue $\lambda(s)$ is periodic and intervenes as an exception for the first result. The second one is relative to the case when the dominant eigenvalue $\lambda(s)$ is log-affine (i.e., $\log \lambda(s)$ affine) and intervenes as an exception for the second result. We prove that exceptional sources are quite similar to the two simpler sources, memoryless sources or Markov chains, where all branches are affine and we conjecture that exceptional sources can only belong to one of these simpler models: a "complex" source cannot be exceptional.

3.3. *Asymptotic Normality of the Variable* $\log \ell_k$. We describe now how the quantity $\Lambda_k(F, s)$ plays a central rôle in this question. The random variable ℓ_k is a step function that is constant on any fundamental interval $h(\mathcal{I})$ of depth k (and equal to u_h). For studying the distribution of the random variable $\log \ell_k$, we use its moment generating function,

$$M_k(s) := \mathrm{E}\,[\exp(s \, \log \ell_k)] = \mathrm{E}\,[\ell_k^s],$$

which satisfies

$$M_k(s) = \sum_{|h|=k} u_h^s u_h = \sum_{|h|=k} u_h^{1+s} = \Lambda_k(F, 1 + s). \tag{19}$$

We will prove that $\Lambda_k(F, 1 + s)$ behaves nearly like the kth power of a fixed analytic function $\lambda(1 + s)$. The central limit theorem of probability theory asserts that exact large powers induce Gaussian laws in the asymptotic limit. Here, we use an extension of the central limit theorem to "quasi-powers" which has been developed in a general setting by Hwang [16]. This extension is valid provided that the function $\log \lambda(s)$ is strictly concave at $s = 1$. We show that this is the case for most of the sources, but not for those that behave like an unbiaised Bernoulli source.

3.4. *Longest Common Prefix of Two Words.* For two words $M(x)$ and $M(y)$, one defines $C(x, y)$ to be the length of their longest common prefix,

$$C(x, y) = \mathrm{Max}\{k \in \mathbf{N};\ P_k(x) = P_k(y)\}.$$

Then C is is a random variable almost everywhere defined on the square $\mathcal{I} \times \mathcal{I}$. The event $C(x, y) \geq k$ is formed with all the points (x, y) such that prefixes $P_k(x)$ and $P_k(y)$ are the same. Then x and y belong to the same fundamental interval of depth k, and

$$[C \geq k] = \bigcup_{|h|=k} \mathcal{I}_h \times \mathcal{I}_h.$$

In the case when the two words are independently drawn from the same probabilistic dynamical source $(\mathcal{S}, f)$, the measure on the square is defined as the product of the two measures associated to the initial distribution F, so that

$$\Pr[C \geq k] = \sum_{|h|=k} u_h^2 = \Lambda_k(F, 2).$$

Again, the Dirichlet series of fundamental measures appear, now at $s = 2$; we shall prove that more general probabilistic settings involve the Dirichlet series $\Lambda_k(F, s)$ for other real values of s that are related to the choice of words distribution.

3.5. *Asymptotic Behaviour of the Number of the Most Probable Prefixes.* We recall that we wish to describe the asymptotic behaviour of function B defined as

$$B(x) = \sum_{h; u_h \geq x} 1$$

when x tends to 0. Here, the central rôle is played by the *Dirichlet series of all fundamental measures* (of any depth)

$$\Lambda(F, s) := \sum_h u_h^s = \sum_h |F(h(0)) - F(h(1))|^s = \sum_{k \geq 0} \Lambda_k(F, s). \tag{20}$$

This function $\Lambda(F, s)$ intervenes because of the relations

$$\Lambda(F, s) = s \int_0^\infty B(x)\, x^{s-1}\, dx, \tag{21}$$

$$\Lambda(F, s) = s \int_0^\infty A(y) e^{-sy}\, dy \qquad \text{with} \quad A(y) = B(e^{-y}), \tag{22}$$

that show that $\Lambda(F, s)$ has two integral forms that are related to function B: the first integral form (21) defines $\Lambda(F, s)$ to be the Mellin transform of function B, and the second one (22) defines $\Lambda(F, s)$ to be the Laplace transform of function A. So, it is a well-known fact that the location of poles of $\Lambda(F, s)$ gives some knowledge on the asymptotics of B near 0, and we shall prove the following facts:

The function $s \to \Lambda(F, s)$ is analytic in the plane $\Re(s) > 1$, and it has a simple pole at $s = 1$. Near the line $\Re(s) = 1$, there are only two possible cases:

(a) *The periodic case: $\Lambda(F, s)$ has other poles on the line, and they are regularly distributed on the line. There is a strip $\sigma < \Re(s) < 1$ that is free of poles.*

(b) *The aperiodic case*: $\Lambda(F, s)$ *has no other poles on the line, but there is possibly an accumulation of poles on the left of the line.*

So, we proceed as follows:

(a) In the periodic case we can use the first integral form and Mellin analysis [12], because of the pole-free region. For reasons which will be explained later (essentially convergence problems), it seems that this Mellin analysis cannot be conducted on the function B itself which is too highly discontinuous in this case. We instead consider the integral D of B, defined as $D(x) := \int_0^x B(y)\,dy$.

(b) In the aperiodic case it is not always possible to locate precisely the singularities of $\Lambda(F, s)$ on the left of the line $\Re(s) = 1$. An alternative method uses the second integral form and Tauberian theorems [7], [38].

4. Introduction of Generalized Ruelle Operators. Here, we define the generalized Ruelle operator, and show how it generates the fundamental intervals, as well as the Dirichet series of fundamental measures.

4.1. *Density Transformers.* There is a direct relationship between the dynamics of source $\mathcal{S}$, the answers to the main three problems, 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], [30], [29]. In its simplest form, the transfer operator associated to a basic dynamical system is the "density transformer",

$$\mathcal{G}[f](x) := \sum_{i\in\mathcal{M}} |h_i'(x)|\; f \circ h_i(x). \tag{23}$$

If f is the initial density on $\mathcal{I}$, then the density on $\mathcal{I}$ after one iteration of the process, i.e., the density function of the iterate $T(x)$, is precisely $\mathcal{G}[f](x)$. The component operator given by the ith term is denoted by $\mathcal{G}_{[i]}$; it is defined by

$$\mathcal{G}_{[i]}[f](x) := |h_i'(x)|\; f \circ h_i(x), \tag{24}$$

so that

$$\mathcal{G} = \sum_{i\in\mathcal{M}} \mathcal{G}_{[i]}. \tag{25}$$

In the same way, one can define a "density transformer" associated to a Markov dynamical system. There are now r different densities $(f_1, f_2, \dots, f_r)$ 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 $\mathcal{S}_0$, one has

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

More generally, the sequence of "conditional densities" $(f_1, f_2, \dots, f_r)$ at one iteration, and the sequence of "conditional densities" $(g_1, g_2, \dots, g_r)$ at the following iteration, are related by an operator matrix $\mathcal{G}$ that is built from the density transformers $\mathcal{G}_j$ associated to each dynamical system $\mathcal{S}_j$. The density transformer $\mathcal{G}_j$ associated to $\mathcal{S}_j$ 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 $\mathcal{G}_{[i|j]}$,

$$\mathcal{G}_{[i|j]}[f](x) := |h'_{i|j}(x)| f \circ h_{i|j}(x), \tag{26}$$

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 now consider the $(r \times r)$-matrix $\mathcal{G}$ whose general coefficient is $\mathcal{G}_{[i|j]}$,

$$\mathcal{G} = (\mathcal{G}_{[i|j]})_{1 \le i,j \le r} \tag{27}$$

(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" $(f_1, f_2, \ldots, f_r)$ at one iteration, into the sequence of "conditional densities" $(g_1, g_2, \ldots, g_r)$ at the following iteration:

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

4.2. *Classical Ruelle Operators.* In fact, in each case (basic dynamical system or Markovian one) it proves highly useful to work with more general operators, called the Ruelle operators. Each component operator in (24) or (26) depends now on a complex parameter s and is defined with the analytic extension $\widetilde{h}$ of $|h'|$. The new component operators are respectively denoted by $\mathcal{G}_{s,[i]}$ or $\mathcal{G}_{s,[i|j]}$:

$$\mathcal{G}_{s,[i]}[f](z) := \widetilde{h}_i^s(z) f \circ h_i(z), \tag{28}$$

$$\mathcal{G}_{s,[i|j]}[f](z) := \widetilde{h}_{i|j}^s(z) f \circ h_{i|j}(z). \tag{29}$$

As in (25) or in (27), the Ruelle operators are now respectively defined by

$$\mathcal{G}_s = \sum_{i=1}^{r} \mathcal{G}_{s,[i]}, \qquad \mathcal{G}_s := (\mathcal{G}_{s,[i|j]})_{1 \le i,j \le r}, \tag{30}$$

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

4.3. *Generalized Ruelle Operators.* We introduce here a new tool: the generalized operators of Ruelle that involve secants of inverse branches,

$$H(u, v) := \left| \frac{h(u) - h(v)}{u - v} \right|, \tag{31}$$

instead of tangents $|h'(z)|$ of inverse branches. Each component operator in (28) or (29) is now defined with the analytic extension $\widetilde{H}$ of the secant H relative to branch h. The new component operators are respectively denoted by $\mathbf{G}_{s,[i]}$ or $\mathbf{G}_{s,[i|j]}$. They act on functions F of two (complex) variables in the following way:

$$(32) \qquad \mathbf{G}_{s,[i]}[F](u,v) := \widetilde{H}_i^s(u,v)F(h_i(u),h_i(u))$$

or

$$(33) \qquad \mathbf{G}_{s,[i|j]}[F](u,v) := \widetilde{H}_{i|j}^s(u,v)F(h_{i|j}(u),h_{i|j}(u)).$$

As previously in (30), the generalized Ruelle operators themselves are respectively defined by

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

in the basic case and in the Markovian case.

The generalized Ruelle operator "extends" the (usual) Ruelle operator as follows: if f denotes the diagonal mapping of F that is defined by $f(u) := F(u,u)$, one obtains on the diagonal $u = v$ the relations

$$(35) \qquad \widetilde{H}(u,u) = \widetilde{h}(u), \qquad \mathbf{G}_s[F](u,u) = \mathcal{G}_s[f](u).$$

4.4. *Pseudo-Powers of Ruelle Operators.* We now show how the operators $\mathcal{G}_s$ or $\mathbf{G}_s$ generate all the branches h of any depth. We first consider the case of a basic dynamical system. By the chain rule, the kth iterates of $\mathcal{G}_s$ and $\mathbf{G}_s$ involve inverse branches h of depth k,

$$(36) \quad \mathcal{G}_s^k[f](z) = \sum_{|h|=k} \widetilde{h}(z)^s f \circ h(z), \qquad \mathbf{G}_s^k[F](u,v) = \sum_{|h|=k} \widetilde{H}(u,v)^s F(h(u),h(v)),$$

where the functions $\widetilde{h}$ and $\widetilde{H}$ are the extensions of the derivative and the secant of branches h, and the sum now ranges over all branches of depth k.

In the Markovian case, the coefficient (i,j) of the kth iterate of matrix $\mathcal{G}_s$ or $\mathbf{G}_s$ involves all the branches h relative to a word $(m_1, \ldots m_k)$ which begins with j $(m_1 = j)$ and finishes with i $(m_k = i)$. We wish to generate all the inverse branches of depth k (i.e., all the inverse branches of $T^{(k)}$), and, for this purpose, we define pseudo-powers $\mathcal{G}_s^{\langle k\rangle}$, $\mathbf{G}_s^{\langle k\rangle}$ of the Ruelle operators. We first let $\mathcal{G}_s^{\langle 0\rangle} := I$, $\mathbf{G}_s^{\langle 0\rangle} := I$. Then we consider the operators $\mathcal{M}_s$ and $\mathbf{M}_s$ relative to the initial dynamical system $\mathcal{S}_0$:

$$(37) \qquad \mathcal{M}_s[f] := \begin{pmatrix} \mathcal{G}_{s,[1|0]} \\ \mathcal{G}_{s,[2|0]} \\ \vdots \\ \mathcal{G}_{s,[i|0]} \\ \vdots \\ \mathcal{G}_{s,[r|0]} \end{pmatrix} [f], \qquad \mathbf{M}_s[F] := \begin{pmatrix} \mathbf{G}_{s,[1|0]} \\ \mathbf{G}_{s,[2|0]} \\ \vdots \\ \mathbf{G}_{s,[i|0]} \\ \vdots \\ \mathbf{G}_{s,[r|0]} \end{pmatrix} [F].$$

If U denotes the unit rth dimensional vector, i.e., ${}^tU = (1, 1, \ldots, 1)$ (r times), then ${}^tU\mathcal{M}_s$ denotes the Ruelle operator associated to $\mathcal{S}_0$ and ${}^tU\mathbf{M}_s$ its generalized version. We let

$$\mathcal{G}_s^{\langle 1\rangle} := {}^tU\mathcal{M}_s, \qquad \mathbf{G}_s^{\langle 1\rangle} := {}^tU\mathbf{M}_s.$$

For $k \geq 2$, the pseudo-powers are defined by

$$\mathcal{G}_s^{\langle k\rangle} =: {}^tU\mathcal{G}_s^{k-1}\mathcal{M}_s, \qquad \mathbf{G}_s^{\langle k\rangle} =: {}^tU\mathbf{G}_s^{k-1}\mathbf{M}_s. \tag{38}$$

Now, the pseudo-iterate of order k generates all the inverse branches of depth k and we obtain the analog of (36):

$$\begin{aligned} \mathcal{G}_s^{\langle k\rangle}[f](z) &= \sum_{|h|=k} \widetilde{h}(z)^s f(h(z)), \\ \mathbf{G}_s^{\langle k\rangle}[F](u, v) &= \sum_{|h|=k} \widetilde{H}(u, v)^s F(h(u), h(v)). \end{aligned} \tag{39}$$

In order to unify our notations, we extend the notion of pseudo-powers to the case of a basic dynamical system, and we let, in this case,

$$\mathcal{G}_s^{\langle k\rangle} := \mathcal{G}_s^k, \qquad \mathbf{G}_s^{\langle k\rangle} := \mathbf{G}_s^k. \tag{40}$$

4.5. *The Dirichlet Series of Fundamental Measures.* Symbolically, in both cases, the kth pseudo-powers of the operators represent k iterations of shift T. Then the series of fundamental intervals of depth k defined in (15) can be easily expressed in terms of the pseudo-powers of the operator $\mathbf{G}_s$,

$$\Lambda_k(F, s) := \sum_{|h|=k} |F(h(0)) - F(h(1))|^s = \mathbf{G}_s^{\langle k\rangle}[L^s](0, 1), \tag{41}$$

where L is defined as the analytic extension of the secant of the distribution F,

$$L(x, y) = \left| \frac{F(x) - F(y)}{x - y} \right|. \tag{42}$$

Note that the diagonal application of L is exactly the extension of the density $f := F'$.

In the same vein, the quasi-inverses $(I - \mathcal{G}_s)^{\langle -1\rangle}$, $(I - \mathbf{G}_s)^{\langle -1\rangle}$, respectively being the formal sum of all the pseudo-powers of the operators, then represent *all* the possible iterations. An important case is the series of all the fundamental measures (20),

$$\begin{aligned} \Lambda(F, s) &:= \sum_{h} |F(h(0)) - F(h(1))|^s \\ &= \sum_{k\geq 0} \mathbf{G}_s^{\langle k\rangle}[L^s](0, 1) = (I - \mathbf{G}_s)^{\langle -1\rangle}[L^s](0, 1). \end{aligned} \tag{43}$$

In both cases, the series $\Lambda_k(F, s)$ of fundamental measures of depth k is also expressible in terms of true powers $\mathbf{G}_s^k$, in the basic case,

$$\Lambda_k(F, s) = \mathbf{G}_s^k[L^s](0, 1), \tag{44}$$

or in the Markovian case, for $k \geq 1$,

(45) $$\Lambda_k(F, s) = \mathbf{G}_s^{(k)}[L^s](0, 1) = {}^t U \mathbf{G}_s^{k-1} \mathbf{M}_s[L^s](0, 1).$$

Asymptotic properties of such powers are thus needed and they are closely related to *spectral properties*, especially dominant and subdominant eigenvalues of the transfer operator $\mathbf{G}_s$.

In the same vein, the series $\Lambda(F, s)$ of fundamental intervals of all depths is expressible in terms of the true quasi-inverse $(1 - \mathbf{G}_s)^{-1}$, in the basic case,

(46) $$\Lambda(F, s) = (I - \mathbf{G}_s)^{-1}[L^s](0, 1),$$

or, in the Markovian case,

(47) $$\Lambda(F, s) := \sum_k \Lambda_k(F, s) = 1 + {}^t U (I - \mathbf{G}_s)^{-1} \mathbf{M}_s[L^s](0, 1).$$

Asymptotic analysis of coefficients of this series is dependent on the location of its poles. Such poles arise from values of s where $(I - \mathbf{G}_s)^{-1}$ is singular, that is, values s for which 1 is an eigenvalue of $\mathbf{G}_s$. In this way, the poles also relate to the spectral properties of the transfer operator.

In the quite particular case when the two following conditions are fulfilled, (i) all the branches of T are linear fractional transformations, (ii) the initial distribution is uniform $F(x) := x$, then the series of fundamental intervals can be solely expressed with the usual Ruelle operator [13]. However, in the general case it appears that the generalized Ruelle operator must be introduced. We now show that the generalized operator shares its main spectral properties with the classical operator which it extends.

5. First Spectral Properties of Generalized Ruelle Operators. We now state the main properties of the generalized Ruelle operator: we prove it to be nuclear, we describe its spectrum, and we exhibit strong positivity properties (for real values of parameter s) that entail the existence of dominant (positive) spectral objects. We then transfer all these properties to the Dirichlet series of fundamental intervals, and we relate the dominant spectral objects of the operator to the basic parameters of the source, the entropy and the coincidence probability.

5.1. *Nuclearity, Trace Formula, and Fredholm Determinant.* We first recall the notion of nuclearity introduced by Grothendieck [14], [15]. Let B be a Banach space and $B^\star$ its dual space. An operator $\mathcal{L}$: $B \to B$ is nuclear of order 0 if it admits a representation

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

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

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

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, together with the analogue of the characteristic polynomial known as the Fredholm determinant,

$$F(\mathcal{L}, u) := \det(I - u\mathcal{L}) := \prod_{i \in I}(1 - \lambda_i u), \tag{49}$$

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

$$\det(I - u\mathcal{L}) = \exp[\operatorname{Tr}\log(I - u\mathcal{L})] = \exp\left[-\sum_{k=1}^{\infty}\frac{u^k}{k}\operatorname{Tr}\mathcal{L}^k\right]. \tag{50}$$

In this way, it is easy to deal with spectral properties of nuclear operators of order 0.

5.2. *Composition Operators.* Each component operator $\mathcal{G}_{s,h}$ is known as a composition operator, defined by $\mathcal{G}_{s,h}[f] := \widetilde{h}^s f \circ h$. We recall that each branch h satisfies "contracting" properties (d1) and (d2) of Definition 2.1: there exists a suitable neighbourhood $\mathcal{V}$ of $\bar{\mathcal{I}}$ such that the following holds: h and $|h'|$ extend to analytics map on $\mathcal{V}$; h maps the closure $\bar{\mathcal{V}}$ of the disk $\mathcal{V}$ inside $\mathcal{V}$, and 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 $\bar{\mathcal{V}}$. Endowed with the sup-norm,

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

$A_\infty(\mathcal{V})$ is a Banach space. Such operators are studied in an extensive way by several authors (Schwartz [31], Shapiro and Taylor [35], Shapiro [34]), and their results are well summarized in [33]. They prove the following:

The operator $\mathcal{G}_{s,h}$: $A_\infty(\mathcal{V}) \to A_\infty(\mathcal{V})$ is compact; it is moreover nuclear of order 0. Its spectrum consists of a geometric progression which involves the value $\alpha(h) := \widetilde{h}(h^\star)$ of $\widetilde{h}$ at the unique fixed point $h^\star$ of the branch h inside $\mathcal{V}$,

$$\operatorname{Sp}\mathcal{G}_{s,h} := \{\mu_n := \alpha(h)^s\,[\varepsilon(h)\alpha(h)]^n,\ n \in \mathbf{N}\} \tag{51}$$

(here $\varepsilon(h)$ denotes the sign of the function h' on $\mathcal{I}$). In particular, the trace of $\mathcal{G}_{s,h}$ is well defined and satisfies

$$\operatorname{Tr}\mathcal{G}_{s,h} = \frac{\alpha(h)^s}{1 - \varepsilon(h)\alpha(h)}. \tag{52}$$

Moreover, the eigenfunction ψ_n relative to eigenvalue μ_n has all its derivatives of order $j < n$ that are zero at the fixed point $h^\star$ and its derivative of order n is non-zero at $h^\star$. The eigenfunction ψ_0 is non-zero on $\mathcal{V}$.

Since each branch h satisfies hypotheses (d1) and (d2) of Definition 2.1, the generalized component operator $\mathbf{G}_{s,h}$ acts on the space $B_\infty(\mathcal{V})$ formed with all functions F

that are holomorphic in the domain $\mathcal{V} \times \mathcal{V}$ and are continuous on the closure $\bar{\mathcal{V}} \times \bar{\mathcal{V}}$. Endowed with the sup-norm,

$$\|F\| = \sup\{|F(u,v)|;\ (u,v) \in \mathcal{V} \times \mathcal{V}\},$$

$B_\infty(\mathcal{V})$ is a Banach space. A theorem due to Mayer [22] can be used in order to determine the spectrum of $\mathbf{G}_{s,h}$. Mayer's theorem shows that, for any inverse branch h, the spectrum of $\mathbf{G}_{s,h}$ has the same elements as the spectrum of $\mathcal{G}_{s,h}$. However, the eigenvalue μ_n appears in the spectrum of $\mathbf{G}_{s,h}$ with an algebraic multiplicity exactly equal to $n+1$. We introduce the signed operator $\widetilde{\mathcal{G}}_{s,h}$,

$$\text{(53)} \qquad \widetilde{\mathcal{G}}_{s,h} = \varepsilon(h)\mathcal{G}_{s,h}, \qquad \text{with} \quad \varepsilon(h) := \text{Sign}(h') \text{ on } \mathcal{I}.$$

Thus, the spectrum of $\mathbf{G}_{s,h}$ can be alternatively defined as a union of spectra,

$$\text{(54)} \qquad \text{Sp}\,\mathbf{G}_{s,h} = \left(\bigcup_{\ell \text{ even}} \text{Sp}\,\mathcal{G}_{s+\ell,h}\right) \cup \left(\bigcup_{\ell \text{ odd}} \text{Sp}\,\widetilde{\mathcal{G}}_{s+\ell,h}\right),$$

where the union is taken in the sense of multi-sets, and this entails a trace formula for $\mathbf{G}_{s,h}$,

$$\text{(55)} \quad \text{Tr}\,\mathbf{G}_{s,h} = \frac{\alpha(h)^s}{[1-\varepsilon(h)\alpha(h)]^2} = \sum_{\ell\geq 0} \varepsilon(h)^\ell \text{Tr}\,\mathcal{G}_{s+\ell,h} = \sum_{\ell \text{ even}} \text{Tr}\,\mathcal{G}_{s+\ell,h} + \sum_{\ell \text{ odd}} \text{Tr}\,\widetilde{\mathcal{G}}_{s+\ell,h}.$$

5.3. *Functional Spaces and Spectra of Transfer Operators.* One first makes precise the functional spaces to which the $\mathbf{G}_s$ operators are applied. In the Markovian case, we restrict ourselves to finite alphabets. However, in the basic case, we can consider infinite (denumerable) alphabets. In this case, the fact that one can choose the same open set $\mathcal{V}$ for all branches h, and the convergence condition (d3), entail "good" properties for the Ruelle operator $\mathbf{G}_s$ when s belongs to the plane $\Re(s) > \gamma$. We denote by $\mathcal{J}$ the intersection of $\mathcal{V}$ with the real axis. The secant mapping $\widetilde{H}(u,v)$ defined in (31) has a strictly positive real part on $\mathcal{V} \times \mathcal{V}$, and the operator $\mathbf{G}_s$ is well defined for any complex s in the plane $\Re(s) > \gamma$.

In the basic case, the $\mathcal{G}_s$ operators (resp. the $\mathbf{G}_s$ operators) are then taken to act on the space $A_\infty(\mathcal{V})$ (resp. $B_\infty(\mathcal{V})$) defined previously. In the Markovian case, the $\mathcal{G}_s$ operators (resp. the $\mathbf{G}_s$ operators) are taken to act on the space $A_\infty(\mathcal{V})^r$ (resp. $B_\infty(\mathcal{V})^r$). Since the component operators $\mathcal{G}_{s,h}$, $\mathbf{G}_{s,h}$ are nuclear of order 0, the operators $\mathcal{G}_s$, $\mathbf{G}_s$ are nuclear of order 0.

The *signed* operator $\widetilde{\mathcal{G}}_s$ is now defined from the signed component operators $\widetilde{\mathcal{G}}_{s,h}$ defined in (53); in the two respective cases,

$$\text{(56)} \qquad \widetilde{\mathcal{G}}_s := \sum_{|h|=1} \widetilde{\mathcal{G}}_{s,h}, \qquad \widetilde{\mathcal{G}}_s := (\widetilde{\mathcal{G}}_{s,h}),$$

and the multiplicativity of ε entails similar equalities for the powers of $\widetilde{\mathcal{G}}_s$. Then trace formulae for $\mathbf{G}_s$ involve both families $\mathcal{G}_{s+\ell}$, $\widetilde{\mathcal{G}}_{s+\ell}$. In the Markovian case, the trace of the kth iterate of $\mathcal{G}_s$ (resp. $\mathbf{G}_s$) equals the sum of the trace of the diagonal elements of the matrix $\mathcal{G}_s^k$ (resp. $\mathbf{G}_s^k$); such diagonal elements only involve inverse branches h that are cyclic: these are branches whose associated word begins and ends with the same symbol. Finally, in both cases, the trace formulae involve the set $\mathcal{C}_k$; it 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); one obtains

$$\text{(57)} \qquad \operatorname{Tr} \mathcal{G}_s^k = \sum_{h \in \mathcal{C}_k} \frac{\alpha(h)^s}{1 - \varepsilon(h)\alpha(h)},$$

$$\operatorname{Tr} \mathbf{G}_s^k = \sum_{h \in \mathcal{C}_k} \frac{\alpha(h)^s}{[1 - \varepsilon(h)\alpha(h)]^2} = \sum_{\ell \text{ even}} \operatorname{Tr} \mathcal{G}_{s+\ell}^k + \sum_{\ell \text{ odd}} \operatorname{Tr} \widetilde{\mathcal{G}}_{s+\ell}^k.$$

PROPOSITION 1 (Spectrum). *For $\Re(s) > \gamma$, the operators $\mathcal{G}_s$, $\widetilde{\mathcal{G}}_s$, $\mathbf{G}_s$ are bounded, and compact, even more nuclear. Their spectra are discrete with only an accumulation point at 0. Moreover, the spectrum of $\mathbf{G}_s$ is determined by the spectra of $\mathcal{G}_s$ and $\widetilde{\mathcal{G}}_s$,*

$$\text{(58)} \qquad \operatorname{Sp} \mathbf{G}_s = \left(\bigcup_{\ell \text{ even}} \operatorname{Sp} \mathcal{G}_{s+\ell} \right) \cup \left(\bigcup_{\ell \text{ odd}} \operatorname{Sp} \widetilde{\mathcal{G}}_{s+\ell} \right),$$

where the union is taken in the sense of multi-sets. The Fredholm determinant $\mathcal{F}(s, u)$ of $\mathcal{G}_s$ and the Fredholm determinant $\mathbf{F}(s, u)$ of $\mathbf{G}_s$ are expressible with the quantities $\alpha(h) = \widetilde{h}(h^\star)$,

$$\text{(59)} \qquad \mathcal{F}(s, u) := \det(I - u\mathcal{G}_s) = \exp\left[-\sum_{k=1}^{\infty} \frac{u^k}{k} \sum_{h \in \mathcal{C}_k} \frac{\alpha(h)^s}{1 - \varepsilon(h)\alpha(h)} \right],$$

$$\text{(60)} \qquad \mathbf{F}(s, u) := \det(I - u\mathbf{G}_s) = \exp\left[-\sum_{k=1}^{\infty} \frac{u^k}{k} \sum_{h \in \mathcal{C}_k} \frac{\alpha(h)^s}{[1 - \varepsilon(h)\alpha(h)]^2} \right],$$

where the set $\mathcal{C}_k$ contains all the inverse branches of depth k (in the basic case) or only the cyclic inverse branches of depth k (in the Markov case).

5.4. *Dominant Spectral Properties for Real s.* When $s = \sigma$ is real, the operators $\mathcal{G}_s$, $\mathbf{G}_s$ satisfy strong positivity properties related to the Perron–Frobenius theory [19].

PROPOSITION 2 (Dominant Eigenvalue). *For real $s > \gamma$, the operators $\mathcal{G}_s$ and $\mathbf{G}_s$ have a unique dominant eigenvalue (of largest modulus). It is positive and has multiplicity 1.*

PROOF. We follow the lines of Mayer's work [22] that we adapt in our context. Mayer himself uses a result due to Krasnoselskii [19].

A subset K of a real Banach space B is called a proper cone if (i) $\rho K \subset K$ for $\rho > 0$ and (ii) $K \cap -K = \{0\}$. A proper cone is called reproducing if $B = K - K$, i.e., every element g of B is a difference of two elements of K. A linear operator $\mathcal{L}$: $B \to B$ is positive with respect to K if $\mathcal{L}K \subset K$. A positive operator $\mathcal{L}$: $B \to B$ is u_0-positive, for some some u_0 in the interior K^* of K, if there exist, for every non-zero $f \in K$, an integer p and strictly positive reals α, β for which

$$\text{(61)} \qquad \alpha u_0 \leq \mathcal{L}^p[f] \leq \beta u_0,$$

where the order is defined with respect to K. Here is the result that we shall use.

POSITIVITY THEOREM [19]. *Any compact u_0-positive operator $\mathcal{L}$: $B \to B$ satisfies a Perron–Frobenius property: it has a unique eigenvector in K^* and the relative eigenvalue is simple, positive, and in absolute value strictly larger than the other eigenvalues of $\mathcal{L}$.*

We first apply Krasnoselskii's result to the operator $\mathcal{G}_s$ in the basic case. For real s, $\mathcal{G}_s$ acts on the real Banach space $A_{\infty\mathbf{R}}(\mathcal{V})$ formed with elements f of $A_\infty(\mathcal{V})$ which are real on the real segment $\mathcal{J}$. We denote by A_+ the subset of $A_{\infty\mathbf{R}}(\mathcal{V})$ formed with elements f which are positive on the real segment $\mathcal{J}$. For real s, $\mathcal{G}_s$ acts on A_+, and A_+ is a cone, proper and reproducing. The interior of the cone, denoted by A_+^*, is formed with elements f of $A_\infty(\mathcal{V})$ which are strictly positive on the real segment $\mathcal{J}$. We define the function u_0 to be equal to the constant function 1, and we show now that the operator $\mathcal{G}_s$ is u_0-positive with respect to the cone A_+: The upper bound of (61) is clear. For the lower bound, consider an element $f \in A_+$ and suppose that, for each integer p, there exists x in $\mathcal{J}$ for which $\mathcal{G}_s^p[f](x) = 0$. Then f is zero at each point $h(x)$ associated to an inverse branch h of depth p. Since f is analytic, then f is zero.

Then we apply Krasnoselskii's theorem: since $\mathcal{G}_s$: $A_{\infty\mathbf{R}}(\mathcal{V}) \to A_{\infty\mathbf{R}}(\mathcal{V})$ is a compact u_0-positive operator with respect to the proper and reproducing cone A_+, the restriction of $\mathcal{G}_s$ to the real Banach space $A_{\infty\mathbf{R}}(\mathcal{V})$ has a unique positive dominant eigenvalue $\lambda(s)$ strictly positive. One can choose the dominant eigenvector ψ_s in the cone A_+^*, which means that ψ_s is strictly positive on $\mathcal{J}$. Moreover, a direct calculus using the nuclearity (and the trace formula) shows that the spectra of the two operators, the operator $\mathcal{G}_s$: $A_\infty(\mathcal{V}) \to A_\infty(\mathcal{V})$ and its restriction to $A_{\infty\mathbf{R}}(\mathcal{V})$ are the same. Finally, the operator $\mathcal{G}_s : A_\infty(\mathcal{V}) \to A_\infty(\mathcal{V})$ has itself dominant spectral properties.

This ends the proof for the operator $\mathcal{G}_s$ in the basic case. This proof can be easily generalized to the other cases: the Markov case and/or the case of the operator $\mathbf{G}_s$. The real Banach spaces are then respectively $A_{\infty\mathbf{R}}(\mathcal{V})^r$, $B_{\infty\mathbf{R}}(\mathcal{V})$, $B_{\infty\mathbf{R}}(\mathcal{V})^r$, where $B_{\infty\mathbf{R}}(\mathcal{V})$ is the subspace of functions F whose restriction to $\mathcal{J} \times \mathcal{J}$ is real. The associated cones are $(A_+)^r$, B_+, $(B_+)^r$, where B^+ is the set formed with the zero function together with functions whose restriction to $\mathcal{J} \times \mathcal{J}$ is positive and not identically zero. □

We have shown the existence of dominant eigenvalues, $\lambda(s)$ for $\mathcal{G}_s$, $\lambda_1(s)$ for $\mathbf{G}_s$. We now prove the equality $\lambda(s) = \lambda_1(s)$. Since the spectrum of the operators is discrete, this makes it possible to separate the dominant eigenvalues from the remainder of the spectra; there is a "spectral gap", and $\mathcal{G}_s$, $\mathbf{G}_s$ decompose as [20]

$$\mathcal{G}_s = \lambda(s)\mathcal{P}_s + \mathcal{N}_s, \qquad \mathbf{G}_s = \lambda_1(s)\mathbf{P}_s + \mathbf{N}_s.$$

Here, $\mathcal{P}_s$, $\mathbf{P}_s$ are the projections over the dominant eigenspace, and $\mathcal{N}_s$, $\mathbf{N}_s$ are relative to the remainder of the spectrum, so that their spectral radius is strictly smaller than the dominant eigenvalue. More generally $\mathcal{G}_s^k$, $\mathbf{G}_s^k$ decomposes as

$$\mathcal{G}_s^k = \lambda(s)^k \mathcal{P}_s + \mathcal{N}_s^k, \qquad \mathbf{G}_s^k = \lambda_1(s)^k \mathbf{P}_s + \mathbf{N}_s^k. \tag{62}$$

The previous relations, together with positive properties of the dominant eigensubspace projections, entail the equalities

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

Since $\mathbf{G}_s$ extends $\mathcal{G}_s$ in the sense of (35), one deduces the equality $\lambda_1(s) = \lambda(s)$. The dominant projections $\mathcal{P}_s$, $\mathbf{P}_s$ can be written as

$$\mathcal{P}_s[f](u) = e_s[f]\psi_s(u), \qquad \mathbf{P}_s[F](u, v) = E_s[F]\Psi_s(u, v),$$

where ψ_s, Ψ_s are the dominant eigenfunctions, and e_s, E_s some linear forms. From (35), one deduces the equalities

$$\Psi_s(u, u) = \psi_s(u), \qquad E_s[F] = e_s[f] \quad \text{if } f \text{ is the diagonal of } F.$$

So, we have shown:

PROPOSITION 3 (Dominant Spectral Objects). *Dominant spectral objects of* $\mathbf{G}_s$ *and* $\mathcal{G}_s$ *are closely linked. Both operators have the same dominant eigenvalue* $\lambda(s)$, *and the other dominant spectral objects (dominant eigenvectors and dominant projectors) are related*:

$$\Psi_s(u, u) = \psi_s(u), \qquad E_s[F] = e_s[f] \quad \textit{if } f \textit{ is the diagonal of } F.$$

5.5. *Quasi-Power Property for* $\Lambda_k(F, s)$. By the classical theory of analytic perturbation [17], 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 $s \to \lambda(s)$, $s \to \Psi_s$, $s \to E_s$ define analytic functions in a neighbourhood of any point where $\lambda(s)$ is well defined. Then the decompositions (62) extend to a neighbourhood of the real axis. Since the Dirichlet series of fundamental measures of depth k are expressed in terms of the kth iterate of $\mathbf{G}_s$ as in (41) or in (44), these decompositions can be applied to these Dirichlet series that behave in fact as a kth power of the dominant eigenvalue $\lambda(s)$, with an exponential remainder term.

PROPOSITION 4 (Quasi-Power Property). *Let* σ *be real. Denote by* $\lambda(s)$ *the dominant eigenvalue (defined in a neighbourhood of* σ*) and by* $\mu(\sigma)$ *a subdominant eigenvalue of the operator* $\mathbf{G}_\sigma$. *For any distribution* F *associated to a density* $f \in A_\infty(\mathcal{V})$ *strictly positive on* $\mathcal{J}$, *and any constant* ρ *satisfying* $|\mu(\sigma)|/\lambda(\sigma) < \rho < 1$, *there exist a neighbourhood* $\mathcal{W}$ *of* σ *and a function* u_F *strictly positive on* $\mathcal{W}$, *for which one has, for any* $k \geq 1$ *and any* s *in* $\mathcal{W}$,

$$\Lambda_k(F, s) = \lambda(s)^k u_F(s)[1 + O_F(\rho^k)]. \tag{63}$$

5.6. *Special Values of the Spectral Objects*; *Entropy and Coincidence Probabilities.* For $s = 1$, the Ruelle operator is a density transformer, and this property entails explicit values of some spectral objects.

PROPOSITION 5 (Special Values). *The dominant eigenvalue at* $s = 1$ *equals* 1, *the dominant eigenvector satisfies* $\Psi_1(0, 1) = 1$, *and the dominant projector* E_1 *satisfies*

$$E_1[F] = \int_0^1 F(x, x)\, dx. \tag{64}$$

PROOF. Since fundamental intervals of depth k form a quasi-partition of unit interval $\mathcal{I}$, there results the equality $\Lambda_k(F, 1) = 1$ 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 (35), and, when coming back to the relation $\Lambda_k(F, 1) = 1$, the equality $\Psi_1(0, 1) = 1$. □

We recall that entropy and coincidence probabilities, defined in (12)–(14), admit expressions (16) and (17) in terms of a Dirichlet series of fixed depth. Then the asymptotic behaviour described in (63) provides expressions of entropy and coincidence probability that involve spectral objects for $s = 1$, $s = 2$, or more generally $s = b$.

PROPOSITION 6 (Entropy and Coincidence Probabilities). *The entropy of the source is equal to the opposite of the derivative of $s \to \lambda(s)$ at $s = 1$, while the coincidence probability is equal to $\lambda(2)$. More generally, the b-coincidence probability is equal to $\lambda(b)$.*

5.7. *Two Particular Cases: Bernoulli Sources or Markov Chains.* We recall that Bernoulli sources are memoryless sources on an alphabet (finite or infinite) $\mathcal{M}$ where symbol m arises with probability p_m. The standard Ruelle operator associated to the system is

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

The dominant eigenfunction is the same for each component operator $\mathcal{G}_{s,h}$ and equals the constant function, for all values of s, $\Re(s) > \sigma$; 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

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

so that the Fredholm determinant is

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

The eigenvector relative to the ℓth eigenvalue $\lambda_\ell(s)$ is a polynomial of degree ℓ. For symmetric Bernoulli sources, it is independent of s. For the symmetric Bernoulli source with two symbols, the family of eigenfunctions coincides exactly with Bernoulli polynomials [5], defined by

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

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 $p^s_{i|j}$ 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

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

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

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

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.

The two particular previous cases are characterized by the fact that all the branches are affine: we call these sources ABS (Affine Branches Sources). Then, for such sources, all the component operators share the same dominant eigenfunction which is the constant function. In what follows, we are led to study instances of sources where the components of the Ruelle operator have strongly correlated dominant spectral objects. More precisely, we give the following definition.

DEFINITION 5.1. A source $\mathcal{S}$ is said to be similar to a source with affine branches (SABS in shorthand) if:

(i) In the basic case, there exist a function $\nu \in \mathcal{A}_\infty(\mathcal{V})$, strictly positive on $\mathcal{J}$ and non-zero on $\mathcal{V}$, and a system of positive numbers $(p_h, h \in \mathcal{H})$ for which, for each inverse branch h, and all s, one has

$$\mathcal{G}_{s,h}[\nu^s] = p_h^s\,\nu^s.$$

(ii) In the Markovian case of an alphabet of cardinality r, there exist r functions $\nu_1, \nu_2, \ldots, \nu_r$ in $\mathcal{A}_\infty(\mathcal{V})$, strictly positive on $\mathcal{J}$, and non-zero on $\mathcal{V}$, and a matrix $Q = (q_{[\ell|j]})$, $1 \leq \ell, j \leq r$, such that one has

$$\mathcal{G}_{s,[\ell|j]}[\nu_j^s] = q^s_{[\ell|j]}\,\nu_\ell^s$$

for each inverse branch $h_{[\ell|j]}$.

The following result shows that these sources actually behave as sources whose branches are all affine, since their Ruelle operator has exactly the same spectrum as the Ruelle operator associated to a source with affine branches.

PROPOSITION 7. *Let $\mathcal{S}$ be a source similar to a source with affine branches. Then the Ruelle operator $\mathcal{G}_s$ has exactly the same spectrum as the Ruelle operator associated to* *a source with affine branches. More precisely:*

(i) *In the basic case, the system of reals p_h is a system of probabilities, and the Ruelle operator has the same spectrum as the Bernoulli source of probabilities p_h.*
(ii) *In the Markovian case, there exist r positive reals $(b_1, b_2, \ldots, b_r)$ such that the matrix with coefficients $p_{[\ell|j]} := (b_\ell/b_j)q_{[\ell|j]}$ is the transition matrix of a Markov chain, and the Ruelle operator has the same spectrum as the Markov chain with transition matrix $p_{[\ell|j]}$.*

PROOF. In the basic case, the positivity of ν on $[0, 1]$ proves that $\alpha(h) = \prod_{i=1}^{j} p_{h_i}$ for all h of the form $h = h_1 \circ h_2 \circ \cdots \circ h_j$. Then the trace formula shows that the Fredholm determinant $\mathcal{F}(s, u)$ is the same as the Fredholm determinant associated to the Bernoulli source of probabilities p_h. In particular, the dominant eigenvalue $\lambda(s)$ of $\mathcal{G}_s$ equals $\sum_{h\in\mathcal{H}} p_h^s$ and satisfies $\lambda(s) = 1$. Then p_h is a system of probabilities.

In the Markovian case, the positivity of ν_ℓ on $[0, 1]$ proves that

$$\alpha(h) = \prod_{i=1}^{k} q_{[j_i|j_{i-1}]} \qquad \text{for all} \quad h \in C_k[\ell] \quad \text{of the form} \quad h = \prod_{i=1}^{k} h_{j_i|j_{i-1}}$$
$$\text{with} \quad j_0 = j_k = \ell.$$

Denote by Q_s the matrix of general coefficient $q_{[\ell|j]}^s$. Then the trace formula shows that the Fredholm determinant $\mathcal{F}(s, u)$ satisfies

$$\mathcal{F}(s, u) = \prod_{\ell\geq 0} \det(I - uQ_{s+\ell}).$$

Since the dominant eigenvalue $\lambda(s)$ of $\mathcal{G}_s$ satisfies $\lambda(s) = 1$, the matrix Q has an eigenvalue equal to 1, and there exist r positive reals $(b_1, b_2, \ldots b_r)$ such that

$$\sum_{\ell=1}^{r} b_\ell\, q_{[\ell|j]} = b_j.$$

Then $p_{[\ell|j]} := (b_\ell/b_j)q_{[\ell|j]}$ is the coefficient of the transition matrix of a Markov chain which has the same spectrum as the Ruelle operator. □

Such sources actually have properties that are very similar to sources with affine branches, with respect of both eigenvalues or eigenfunctions. Moreover, the fact that all component operators share the same (dominant) eigenfunction seems to be a quite strong constraint which entails in both cases a multiplicative property for the quantities α, and a very special form for the spectrum. The author does not know other sources where all component operators share the same (dominant) eigenfunction. Since the eigenfunction of $\mathcal{G}_{s,h}$ is closely related to branch h, it seems unlikely that all component operators may have the same eigenfunction, unless these branches h are all affine. So, we state the following conjecture:

CONJECTURE 1. *The sources that are similar to sources with affine branches can only be sources with affine branches.*

5.8. *Comparing with Other Models: Stationary, Ergodicity, and Strong Mixing.* In the framework of our dynamical sources, the dominant eigenvector ψ_1, with normalizing condition $\int_0^1 \psi_1(x)\,dx = 1$, is the density of the unique invariant probability under shift T. Then classical results prove that the dynamical source is ergodic. Note that it is not stationary, unless the initial density is exactly equal to ψ_1. Strong mixing properties are easily deduced from the quasi-power property.

6. Further Spectral Properties of Generalized Ruelle Operators. Because of the quasi-power property (Proposition 4), the dominant eigenvalue function $s \to \lambda(s)$ plays a central rôle in our analyses. Here, we establish some important properties of this function. More generally, we study the spectral radius $\mathbf{R}(s)$ of the operator $\mathbf{G}_s$, and the poles of the quasi-inverse $(I - \mathbf{G}_s)^{-1}$. We obtain two results of classification which characterize possible exceptions for our further results.

6.1. *Maximum Properties on the Half-Planes.* At a point s in the half-plane $\Re(s) \geq \sigma$, we compare the spectral radius $\mathbf{R}(s)$ of $\mathbf{G}_s$ and the spectral radius $\mathbf{R}(\sigma) = \lambda(\sigma)$ of $\mathbf{G}_\sigma$.

PROPOSITION 8 (Maximum Properties).

(i) *The function $s \to \lambda(s)$ is strictly decreasing along the real axis $s > \gamma$.*
(ii) *On each vertical line $\Re(s) = \sigma$, the inequality $\mathbf{R}(s) \leq \lambda(\sigma)$ holds.*
(iii) *If the equality $\mathbf{R}(s) = \lambda(\sigma)$ holds for $s = \sigma + it$, $t \neq 0$, then $\mathbf{G}_s$ has an eigenvalue $\lambda = e^{ia}\lambda(\sigma)$ that belongs to the spectrum of $\mathcal{G}_s$ for some real a.*
(iv) *In the half-plane $\Re(s) > \sigma$, the strict inequality $\mathbf{R}(s) < \lambda(\sigma)$ holds.*

PROOF. (i) From relation (63) of Proposition 4, the dominant eigenvalue $\lambda(s)$ is alternatively defined by

$$\lambda(s) = \lim_k \Lambda_k(\mathrm{Id}, s)^{1/k}.$$

From properties (d2) and (d3) of dynamical sources, there exists $\delta < 1$ for which $|h'(x)| \leq \delta$ for any inverse branch of depth 1, and any x in the unit interval. One deduces the inequalities $|h(0) - h(1)| \leq \delta^k$, valid for all inverse branch of depth k, and then

$$\lambda(s+u) \leq \delta^u\, \lambda(s),$$

so that the function $\lambda(s)$ strictly decreases along the real axis.

We now consider vertical lines, and we prove (ii). The description of the spectrum given in Proposition 1 shows that this spectral radius depends on both of the spectral radii $R(s+\ell)$ and $\widetilde{R}(s+\ell)$ of operators $\mathcal{G}_s$ and $\widetilde{\mathcal{G}}_s$. We begin to study the spectral radii of operators $\widetilde{\mathcal{G}}_s$ and $\mathcal{G}_s$ on vertical lines, and we prove the following fact:

On the line $\Re(s) = \sigma$, the spectral radius $\widetilde{R}(s)$ of the operator $\widetilde{\mathcal{G}}_s$ and the spectral radius $R(s)$ of the operator $\mathcal{G}_s$ both satisfy

$$R(s) \leq \lambda(\sigma), \qquad \widetilde{R}(s) \leq \lambda(\sigma). \tag{65}$$

We consider first the operator $\mathcal{G}_s$ in the basic case. Let λ be an eigenvalue of $\mathcal{G}_s$ and let f denote an eigenvector relative to λ. In the same way, the vector f_σ denotes a dominant eigenvector relative to $\lambda(\sigma)$. This function is strictly positive on the segment $\mathcal{J}$, non-zero on $\mathcal{V}$ and normalized by the condition $f_\sigma(0) = 1$. Moreover, one can suppose that the function μ

$$\mu(x) := \frac{f(x)}{f_\sigma(x)} \tag{66}$$

is of modulus at most 1 on $[0, 1]$ and attains modulus 1 at point x_0. One always has

$$|\lambda\, f(x_0)| = |\mathcal{G}_s[f](x_0)| = \left|\sum_{|h|=1} \widetilde{h}(x_0)^s f \circ h(x_0)\right| \leq \sum_{|h|=1} \widetilde{h}(x_0)^\sigma |f \circ h(x_0)| \tag{67}$$

$$\leq \sum_{|h|=1} \widetilde{h}(x_0)^\sigma f_\sigma \circ h(x_0) = \lambda(\sigma) f_\sigma(x_0), \tag{68}$$

and the definition of x_0 proves the inequality $|\lambda| \leq \lambda(\sigma)$.

For the operator $\mathcal{G}_s$ in the Markovian case, we consider the same objects: λ is an eigenvalue of $\mathcal{G}_s$ and $f = (f_1, f_2, \dots f_r)$ denotes an eigenvector relative to λ. In the same way, the vector $f_\sigma = (f_{\sigma,1}, f_{\sigma,2}, \dots, f_{\sigma,r})$ denotes a dominant eigenvector relative to $\lambda(\sigma)$. This function has all its components strictly positive on the segment $\mathcal{J}$, non-zero on $\mathcal{V}$. Moreover, one can suppose that all the functions μ_i,

$$\mu_i(x) := \frac{f_i(x)}{f_{\sigma,i}(x)}, \tag{69}$$

are of modulus at most 1 on $[0, 1]$, and one function μ_ℓ attains modulus 1 at point x_0. One always has

$$|\lambda\, f_\ell(x_0)| = \left|\sum_j \widetilde{h}_{\ell|j}(x_0)^s f_j \circ h_{\ell|j}(x_0)\right| \leq \sum_j \widetilde{h}_{\ell|j}(x_0)^\sigma |f_j \circ h_{\ell|j}(x_0)| \tag{70}$$

$$\leq \sum_j \widetilde{h}_{\ell|j}(x_0)^\sigma f_{\sigma,j} \circ h_{\ell|j}(x_0) = \lambda(\sigma) f_{\sigma,\ell}(x_0), \tag{71}$$

and the definition of point x_0 and index ℓ proves the inequality $|\lambda| \leq \lambda(\sigma)$.

The property (65) is proven for the operator $\mathcal{G}_s$; it is clear that it can be easily adapted to the operator $\widetilde{\mathcal{G}}_s$. We return now to operator $\mathbf{G}_s$ with the spectrum formula (58) together with the strict decreasing of λ along the real axis, and we easily prove (iii) and (iv). □

6.2. *Singularities of the Quasi-Inverse* $(I - \mathbf{G}_s)^{-1}$. We have explained in Section 3 why it is necessary to locate the poles of the series $\Lambda(F, s)$ precisely. We recall that $\lambda(1) = 1$ (Proposition 5). Then, from Proposition 8, the operator $I - \mathbf{G}_s$ is invertible in the plane $\Re(s) > 1$. Thus, the series $\Lambda(F, s)$ is analytic there and it has a simple pole at $s = 1$. We focus on what may take place near the line $\Re(s) = 1$ and we consider so-called *particular points*: they are points $s = 1 + it$, with $t \neq 0$ for which the spectrum of $\mathbf{G}_s$ contains an eigenvalue equal to 1. The following result, which extends results of [10], [28], and [40], gives a characterization of particular points and describes the only two possible types of behaviour.

PROPOSITION 9 (Periodicity and Aperiodicity). *The operator* $\mathbf{G}_s$ *may only behave in two different ways on the line* $\Re(s) = 1$:

(i) *The aperiodic case. There are no particular points, and the operator* $I - \mathbf{G}_s$ *is invertible in the punctured plane* $\Re(s) \geq 1, s \neq 1$.
(ii) *The periodic case. There are particular points, and they are regularly spaced on the line. They form a sequence of the form* $s_k := 1 + kit, k \in \mathbf{Z}$, *for some* $t > 0$. *The operator* $(I - \mathbf{G}_s)^{-1}$ *has simple poles at these points, and there is a strip on the left of the line* $\Re(s) = 1$ *that is free of poles. In this case, the source is similar to a source with affine branches (SABS) and the Fredholm determinant* $\mathcal{F}(s, u)$ *of* $\mathcal{G}_s$ *is periodic of period* it, *i.e.,* $\mathcal{F}(s + it, u) = \mathcal{F}(s, u)$.

We shall see in what follows that these two cases may occur, mainly for simple sources, as Bernoulli sources or Markov chains. However, we conjecture that "correlated" sources whose branches are not all affine will always be aperiodic.

PROOF. Assertion (iii) of Proposition 8 shows that it is sufficient to work with the $\mathcal{G}_s$ operator. We prove the previous statement in the two main cases: the basic case and the Markov case.

Basic case. We keep the notations of Section 6.1. We consider the point x_0 where the function μ attains its maximum. From the equalities $\lambda = \lambda(\sigma)$ and $|\mu(x_0)| = 1$, we deduce that the sequence of inequalities (67), (68) becomes a sequence of equalities at x_0,

$$\lambda f(x_0) = \left|\sum_{|h|=1} \widetilde{h}(x_0)^{\sigma+it} f \circ h(x_0)\right| = \sum_{|h|=1} \widetilde{h}(x_0)^{\sigma} |f \circ h(x_0)| \tag{72}$$

$$= \sum_{|h|=1} \widetilde{h}(x_0)^{\sigma} f_\sigma \circ h(x_0) = \lambda(\sigma) f_\sigma(x_0). \tag{73}$$

For any h of depth 1, the equality

$$|f \circ h(x_0)| = f_\sigma \circ h(x_0)$$

holds, and an inductive argument proves that the function μ defined in (66) satisfies $|\mu \circ h(x_0)| = 1$ for any inverse branch h of any depth. Since any real x in $[0, 1]$ is the limit of a sequence $h(x_0)$, one gets

$$|\mu(x)| = 1 \qquad \text{for any } x \text{ in } [0, 1].$$

Then the sequence of equalities (72), (73), now valid for any x in $[0, 1]$, entails the relation

$$\sum_{|h|=1} \widetilde{h}(x)^{\sigma} |f \circ h(x)| = \left|\sum_{|h|=1} \widetilde{h}(x)^{\sigma+it} f \circ h(x)\right|.$$

The sequence $a_h(x) := \widetilde{h}(x)^{\sigma+it} f \circ h(x)$ satisfies the equality $|\sum a_h(x)| = \sum |a_h(x)|$. Then there exists $\theta(x)$ (of modulus 1) such that $a_h(x) = \theta(x)|a_h(x)|$ for any h of depth 1.

When returning to our problem, we note that $\theta(x) = \mu(x)$, and we deduce the equality

$$\widetilde{h}(x)^{it}\mu \circ h(x) = \mu(x) \qquad \text{for any } h \text{ of depth 1,} \tag{74}$$

and, more generally,

$$\widetilde{h}(x)^{it}\mu \circ h(x) = \mu(x) \qquad \text{for any } h \text{ of any depth.} \tag{75}$$

The equality extends (by analytic continuation) to $\mathcal{V}$. Since μ is of modulus 1 on [0, 1], Section 5.2 and (75) show that μ is a dominant eigenfunction of all the component operators $\mathcal{G}_{it,h}$, so that we deduce the relations

$$\alpha(h)^{it} = 1 \qquad \text{for any } h \text{ of any depth,}$$

which involve the quantity α defined in Section 5.2. Then the Fredholm determinant satisfies, from (60), the relation

$$\mathcal{F}(s + it, u) = \mathcal{F}(s, u).$$

Moreover, the function μ is non-zero on $\mathcal{V}$ and it can be written as $\mu := \exp(itL)$ with some analytic function L. Since μ is of modulus 1 on [0, 1], the function L is real on [0, 1], so that the function $\nu := \exp L$ is strictly positive on [0, 1], and ν^s is a dominant eigenfunction of all the component operators $\mathcal{G}_{s,h}$. The source is thus similar to a source with affine branches.

Markov case. Again, we keep the notations of Section 6.1. The equality $|\lambda| = \lambda(\sigma)$ transforms the sequence of inequalities (70), (71) into a sequence of equalities:

$$\lambda\, f_\ell(x_0) = \left|\sum_j \widetilde{h}_{\ell|j}(x_0)^s f_j \circ h_{\ell|j}(x_0)\right| = \sum_j \widetilde{h}_{\ell|j}(x_0)^\sigma |f_j \circ h_{\ell|j}(x_0)| \tag{76}$$

$$= \sum_j \widetilde{h}_{\ell|j}(x_0)^\sigma f_{\sigma,j} \circ h_{\ell|j}(x_0) = \lambda(\sigma) f_{\sigma,\ell}(x_0). \tag{77}$$

In particular, for any symbol j, we deduce the equality

$$|f_j \circ h_{\ell|j}(x_0)| = f_{\sigma,j} \circ h_{\ell|j}(x_0).$$

Then μ_j, defined in (69), has modulus 1 at the point $x_j := h_{\ell|j}(x_0)$; when writing the sequence of equalities similar to (76), (77) but due to the relation $|\mu_j(x_j)| = 1$, we obtain

$$|f_\ell \circ h_{j|\ell}(x_j)| = f_{\sigma,\ell} \circ h_{j|\ell}(x_j),$$

and, finally, an inductive argument proves that all the μ_j's have modulus 1 on [0, 1]. Then the sequence of equalities (76), (77), now valid for any x in $\mathcal{J}$ and any symbol ℓ, proves the relation

$$\widetilde{h}_{\ell|j}(x)^{it}\mu_j \circ h_{\ell|j}(x) = \mu_\ell(x) \qquad \text{for any symbol } \ell, j.$$

The equality extends (by analytic continuation) to $\mathcal{V}$. In particular,

$$\widetilde{h}(x)^{it}\mu_j \circ h(x) = \mu_j(x) \qquad \text{for any } \quad h \in \mathcal{C}[j]. \tag{78}$$

Since μ_j is not zero on [0, 1], Section 5.2 and (78) prove that μ_j is a dominant eigenvector of all the component operators $\mathcal{G}_{it,h}$, for any $h \in \mathcal{C}[j]$. In particular, μ_j is non-zero on $\mathcal{V}$. Then we deduce the relations

$$\alpha(h)^{it} = 1 \qquad \text{for} \quad h \in \mathcal{C},$$

which involve the quantity α defined in Section 5.2. Then the Fredholm determinant satisfies, from relation (60), the equality

$$\mathcal{F}(s + it, u) = \mathcal{F}(s, u).$$

Moreover, the functions μ_j are non-zero on $\mathcal{V}$ and can be written as $\mu_j = \exp(itL_j)$ with some analytic functions L_j. Since μ_j's are of modulus 1 on [0, 1], the functions L_j are real on [0, 1], so that the functions $\nu_j := \exp L_j$ are strictly positive on [0, 1]. Then the functions ν_j^s satisfy, for any $h = h_{[\ell|j]}$,

$$(79) \quad \widetilde{h}(x)^s \nu_j^s \circ h(x) = q_{[\ell|j]}^s \, \nu_\ell^s(x), \qquad \text{with positive } q_{[\ell|j]} \text{ such that} \quad q_{[\ell|j]}^{it} = 1,$$

and the source is similar to a source with affine branches. This concludes the proof for the Markov case. □

6.3. *Log-Concavity of the Dominant Eigenvalue.* This property intervenes mainly in the study of the variance of the random variable $\log \ell_k$. It will play an important rôle in the height of tries in a companion paper. Again, as in Section 6.2, the function $\lambda(s)$ may have only two different behaviours with respect to concavity.

PROPOSITION 10 (Log-Concavity). *For real $s > \gamma$, the function $\log \lambda(s)$ is always concave. There are only two different possible behaviours*:

(i) *The function $\log \lambda(s)$ is strictly concave.*
(ii) *The function $\log \lambda(s)$ is affine. Then the alphabet $\mathcal{M}$ is finite, of cardinality r, and the source is similar to a symmetric Bernoulli source.*

In both cases, and for any positive reals s, t such that $s > t > \gamma$, one has $\lambda(s)^t < \lambda(t)^s$.

PROOF. We shall prove the following inequality:

$$(80) \qquad \lambda(\delta + \beta) \leq \lambda(2\delta)^{1/2} \, \lambda(2\beta)^{1/2},$$

for any real pair (δ, β). If equality holds for $\delta \neq \beta$, then the function $\log \lambda(s)$ is affine, and we prove that the operator has exactly the same spectrum as the Bernoulli operator.

Basic case. The function

$$(81) \qquad \psi(x) := f_{\delta+\beta}(x)(f_{2\delta}(x))^{-1/2}(f_{2\beta}(x))^{-1/2},$$

defined with the dominant eigenvectors f_σ of $\mathcal{G}_\sigma$, can be normalized by the condition

$$\operatorname{Sup}\{\psi(x);\ x \in [0, 1]\} = 1.$$

We denote by x_0 a point where $\psi(x_0) = 1$. One always has

$$\begin{aligned}
(82) \quad \lambda(\delta+\beta) f_{\delta+\beta}(x_0) &= \sum_{|h|=1} \widetilde{h}(x_0)^{\delta+\beta} f_{\delta+\beta} \circ h(x_0) \\
&\leq \sum_{|h|=1} \widetilde{h}(x_0)^{\delta} \, (f_{2\delta} \circ h(x_0))^{1/2} \, h(x_0)^{\beta} \, \left(f_{2\beta} \circ h(x_0)\right)^{1/2} \\
(83) \quad &\leq \left(\sum_{|h|=1} \widetilde{h}(x_0)^{2\delta} f_{2\delta} \circ h(x_0)\right)^{1/2} \left(\sum_{|h|=1} \widetilde{h}(x_0)^{2\beta} f_{2\beta} \circ h(x_0)\right)^{1/2} \\
&= (\lambda(2\delta) f_{2\delta}(x_0))^{1/2} \left(\lambda(2\beta) f_{2\beta}(x_0)\right)^{1/2}.
\end{aligned}$$

(Inequality (83) is due to the Cauchy–Schwarz property.) The definition of x_0 now proves inequality (80). If now equality holds in (80), then (82) becomes an equality, and then the function ψ satisfies $\psi(h(x_0)) = 1$ for any inverse branch h of depth 1. Now, an inductive argument proves that the function ψ satisfies $\psi(h(x_0)) = 1$ for any inverse branch h of any depth. Since any real x in $[0, 1]$ is the limit of a sequence $h(x_0)$, one gets

$$\psi(x) = 1 \qquad \text{for any } x \text{ in } [0, 1].$$

Then the sequence of inequalities (82), (83) are now equalities for any x in $[0, 1]$ and the Cauchy–Schwarz inequality becomes an equality. Thus, there exists $\psi^{\langle k\rangle}(x)$ such that, for every h of depth k,

$$(84) \qquad \frac{\widetilde{h}(x)^{2\delta} f_{2\delta} \circ h(x)}{\widetilde{h}(x)^{2\beta} f_{2\beta} \circ h(x)} = \psi^{\langle k\rangle}(x).$$

Let $\gamma := 2(\delta - \beta)$ (γ is supposed to be non-zero) and $\psi_0(x) := f_{2\delta}(x)/f_{2\beta}(x)$. Then the first member of (84) involves the component operator $\mathcal{G}_{\gamma,h}$, and

$$(85) \qquad \psi^{\langle k\rangle} = \mathcal{G}_{\gamma,h}[\psi_0].$$

Consider the particular case when h is of the form h_0^k, with h_0 of depth ρ. Then (85) can be written as $\psi^{\langle k\rho\rangle}(x) = \mathcal{G}^k_{\gamma,h_0}[\psi_0](x)$. Since ψ_0 is strictly positive on $\mathcal{J}$, the sequence $\psi^{\langle k\rangle}$ has a quasi-power property and satisfies

$$\lim_{k\to\infty} (\psi^{\langle k\rho\rangle}(x))^{1/k} = \lim_{k\to\infty} \left(\mathcal{G}^k_{\gamma,h_0}[\psi_0](x)\right)^{1/k} = \alpha(h_0)^{\gamma}.$$

We deduce that all the $\alpha(h)$ relative to any branch h of depth ρ are equal, and equal to α^{ρ}, for some constant α. Returning to (84), we then deduce that

$$(86) \qquad \theta(x) := \lim_{k\to\infty} \frac{\psi^{\langle \ell\rangle}(x)}{\alpha^{\gamma\ell}}$$

exists, is strictly positive on $[0, 1]$, and is a common eigenfunction for all the component operators $\mathcal{G}_{\gamma,h}$. Then we let $\nu := \theta^{1/\gamma}$, and ν^s is a common eigenfunction for all the component operators $\mathcal{G}_{s,h}$. The source is similar to a Bernoulli source with all probabilities equal to α. Then the alphabet is finite, of cardinality r, and the source is similar to a symmetric Bernoulli source.

Markovian case. The functions

$$\psi_\ell(x) := f_{\delta+\beta,\ell}(x)(f_{2\delta,\ell}(x))^{-1/2}(f_{2\beta,\ell}(x))^{-1/2},$$

defined with the components $f_{\sigma,\ell}$ of dominant eigenvectors f_σ of $\mathcal{G}_\sigma$, can be normalized by the condition

$$\mathrm{Sup}\{\psi_\ell(x);\ x \in [0,1];\ \ell \in [1..r]\} = 1.$$

We consider a point x_0 and a symbol ℓ for which $\psi_\ell(x_0) = 1$. One always has

$$
\begin{aligned}
(87)\quad & \lambda(\delta+\beta)\, f_{\delta+\beta,\ell}(x_0) \\
&= \sum_j \widetilde{h}_{\ell|j}(x_0)^{\delta+\beta} f_{\delta+\beta,j} \circ h_{\ell|j}(x_0) \\
&\le \sum_j \widetilde{h}_{\ell|j}(x_0)^{\delta} \left(f_{2\delta,j} \circ h_{\ell|j}(x_0)\right)^{1/2} h_{\ell|j}(x_0)^{\beta} \left(f_{2\beta,j} \circ h_{\ell|j}(x_0)\right)^{1/2} \\
(88)\quad &\le \left(\sum_j \widetilde{h}_{\ell|j}(x_0)^{2\delta} f_{2\delta,j} \circ h_{\ell|j}(x_0)\right)^{1/2} \left(\sum_j \widetilde{h}_{\ell|j}(x_0)^{2\beta} f_{2\beta,j} \circ h_{\ell|j}(x_0)\right)^{1/2} \\
&= \left(\lambda(2\delta) f_{2\delta,\ell}(x_0)\right)^{1/2} \left(\lambda(2\beta) f_{2\beta,\ell}(x_0)\right)^{1/2}.
\end{aligned}
$$

(Inequality (88) is due to the Cauchy–Schwarz property.) The definition of x_0 implies inequality (80). If now the equality holds in (80), we have a sequence of equalities in (87) and (88), and an inductive argument proves that all the ψ_j are equal to 1 on $[0, 1]$. Then the sequence of equalities in (87) and (88) is now valid for any x in $[0, 1]$ and the Cauchy–Schwarz inequality becomes an equality. Thus, there exists $\phi_\ell^{\langle k\rangle}(x)$ such that, for every h of depth k which begins with symbol j and ends with symbol ℓ,

$$\frac{\widetilde{h}(x)^{2\delta} f_{2\delta,j} \circ h(x)}{\widetilde{h}(x)^{2\beta} f_{2\beta,j} \circ h(x)} = \phi_\ell^{\langle k\rangle}(x). \tag{89}$$

Let $\gamma := 2(\delta - \beta)$ (γ is supposed to be non-zero) and $\phi_j(x) := f_{2\delta,j}(x)/f_{2\beta,j}(x)$. Then the first member of (89) involves the component operator $\mathcal{G}_{\gamma,h}$, and

$$\phi_\ell^{\langle k\rangle}(x) = \mathcal{G}_{\gamma,h}[\phi_\ell]. \tag{90}$$

Consider the particular case when h is of the form h_0^k, with $h_0 \in \mathcal{C}_\rho[\ell]$. Then the first member of (89) can be written as $\phi_\ell^{\langle k\rho\rangle}(x) = \mathcal{G}^k_{\gamma,h_0}[\phi_\ell](x)$. Since ϕ_ℓ is strictly positive on $[0, 1]$, the sequence $\phi_\ell^{\langle k\rho\rangle}$ satisfies a quasi-power property so that

$$\lim_{k\to\infty} (\phi_\ell^{\langle k\rho\rangle}(x))^{1/k} = \alpha(h_0)^\gamma,$$

and then all the $\alpha(h)$ relative to elements $h \in \mathcal{C}_\rho[\ell]$ are equal to $\alpha(\ell)^\rho$. Moreover, by another application of the quasi-power property, the limit

$$\theta_\ell(x) := \lim_{k\to\infty} \frac{\phi_\ell^{\langle k\rangle}(x)}{\alpha(\ell)^{\gamma k}} \tag{91}$$

defines a common eigenfunction for all the component operators $\mathcal{G}_{\gamma,h}$ relative to branches $h \in \mathcal{C}[\ell]$. Furthermore, relation (89) proves that the functions $\widetilde{h}_{\ell|j}(x)^{\gamma}\theta_j \circ h_{\ell|j}(x)$ and $\theta_\ell(x)$ are proportional, i.e.,

$$\widetilde{h}_{\ell|j}(x)^{\gamma}\theta_j \circ h_{\ell|j}(x) = q_{\ell|j}^{\gamma}\theta_\ell(x).$$

Since the functions θ_j are strictly positive on [0, 1], and non-zero on $\mathcal{V}$, we let $\nu_j := \theta_j^{1/\gamma}$, and the functions ν_j^s satisfy

$$\widetilde{h}_{\ell|j}(x)^{s}\nu_j^s \circ h_{\ell|j}(x) = q_{\ell|j}^{s}\nu_\ell^s(x).$$

Now, the equalities

$$q_{\ell|j}q_{j|\ell} = \alpha(\ell)^2 = \alpha(j)^2$$

prove that all the $\alpha(\ell)$ are equal (to α), and the trace formula proves that the Ruelle operator has the same spectrum as a Bernoulli symmetric operator $\mathcal{B}_{r,s}$. This ends the proof of the two first assertions.

We now prove the third assertion. If $\log\lambda(s)$ is affine, it is of the form $\log\lambda(s) = (1-s)\log r$ with an integer $r > 1$. Then

$$t\log\lambda(s) = t(1-s)\log r < s(1-t)\log r = s\log\lambda(t).$$

In the other case, $\log\lambda(s)$ is strictly concave. Consider an integer k such that s belongs to the interval $[t, kt]$, then the strict concavity of $\log\lambda(s)$ proves the inequality:

$$\lambda(s) < \lambda(t)^{\alpha}\lambda(kt)^{\beta}, \qquad \text{with} \quad \alpha+\beta = 1, \qquad \frac{s}{t} = \alpha + k\beta.$$

On the other hand, for any integer k,

$$\lambda(kt) = \lim_{\ell\to\infty}\left[\sum_{|h|=\ell} u_h^{kt}\right]^{1/\ell} \leq \lim_{\ell\to\infty}\left[\sum_{|h|=\ell} u_h^{t}\right]^{k/\ell} = \lambda(t)^k.$$

Finally,

$$\lambda(s) < \lambda(t)^{\alpha+\beta k} = \lambda(t)^{s/t}.$$

This ends the proof of Proposition 10. □

6.4. *Exceptional Cases.* The proofs of Propositions 9 and 10 are quite similar. When studying periodicity, we use triangular inequality, whereas the study about log-concavity uses Cauchy–Schwartz inequality. In our proofs, we have exhibited strong properties of eigenfunctions and eigenvalues. In exceptional cases, all the sources are similar to sources with affine branches, and the systems of the α's are highly correlated.

So, we obtain characterizations of exceptional cases from both points of view, log-affinity and periodicity.

PROPOSITION 11. *The following two conditions are equivalent*:

(a) *The dominant eigenvalue $\lambda(s)$ of the operator $\mathcal{G}_s$ is log-affine.*
(b) *The alphabet $\mathcal{M}$ is finite, of cardinality r and the source is similar to a symmetric Bernoulli source.*

Let $t > 0$ be a real number and denote by a the real number $a = \exp(-2\pi/t)$. The following two conditions are equivalent:

(c) *The dominant eigenvalue $\lambda(s)$ of the operator $\mathcal{G}_s$ is periodic of period it.*
(d) *The source is similar to a source with affine branches. In the basic case, all the quantities α's are positive powers of the real a. In the Markovian case, all the quantities $q_{[\ell|j]}$ are positive powers of the real a.*

PROOF. It is clear from Propositions 7, 9, and 10. □

We now describe the exceptional sources with affine branches:

COROLLARY. *The only Bernoulli sources that are log-affine are exactly symmetric Bernoulli sources. In particular, their alphabet is finite. A Bernoulli source is periodic if and only if there exists a real number $a < 1$ such that all the probabilities p_m belong to the semi-group $\langle a \rangle$ generated by a.*

The only cases of Markov chains that are log-affine correspond to degenerate cases, where each Bernoulli source $\mathcal{S}_j$ is symmetric. A Markov chain is periodic if and only if there exist a real number $a < 1$ and a sequence of real numbers $(b_1, b_2, \ldots, b_r)$ such that all the quantities $p_{i|j}(b_i/b_j)$ belong to the semi-group $\langle a \rangle$ generated by a.

Here are some examples of periodic Bernoulli sources:

$$\left(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}\right), \qquad (p, p, p^2) \quad \text{with} \quad p = \tfrac{1}{2}(\sqrt{2} - 1),$$
$$(p_m)_{m \geq 1} \quad \text{with} \quad p_m = \left(\tfrac{1}{2}\right)^m.$$

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

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

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 $\langle a \rangle$ generated by a and all the numbers p_m/b belong to this cyclic group $\langle a \rangle$. An instance of this situation is $(\frac{2}{5}, \frac{2}{5}, \frac{1}{5})$. As Pollicott [28] and Fayolle et al. [11] remark, there is an accumulation of s for which $\lambda(s) = 1$ on the left of the line $\Re(s) = 1$. Our Tauberian argument directly shows that the total contribution of these poles gives a term $o(1/x)$ in the asymptotic expansion of $B(x)$.

Non-degenerate instances of transition matrices of periodic Markov chains are

$$\Pi := \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \quad \text{or}$$

$$\Pi := \begin{pmatrix} \dfrac{1}{c^3} & \dfrac{c}{c^3 - 1} \\ \dfrac{c^3 - 1}{c^3} & \dfrac{1}{c^2} \end{pmatrix} \qquad \text{with} \quad c > 1 \quad \text{root of} \quad c^5 - 2c^3 - c^2 + 1 = 0.$$

We recall that our Conjecture 1 states that any source similar to a source with affine branches has actually all its branches affine. Thus, we restate it in a weaker form:

CONJECTURE 2. *The exceptional sources can only be Bernoulli sources or Markov chains described in the previous corollary.*

Now, in the following three sections, we return to the three main problems, and solve them using the tools and the results previously described.

7. The Strong Equipartition Property of the Prefixes of Fixed Length. Our approach has been first described in Section 3.3: our study of the distribution of the random variable $\log \ell_k$ uses its moment generating function,

$$M_k(s) := \mathrm{E}\,[\exp(s\ \log \ell_k)] = \mathrm{E}\,[\ell_k^s],$$

which satisfies

$$M_k(s) = \sum_{|h|=k} u_h^s\, u_h = \sum_{|h|=k} u_h^{1+s} = \Lambda_k(F, 1+s).$$

Now, the quasi-power property of Section 5.6 proves that $M_k(s)$ behaves nearly like a "large power" of the fixed function λ. More precisely, there exists a sufficiently small complex neighbourhood of $s = 1$ where

$$M_k(s) = \exp(k \log \lambda(1+s) + V(s)) \cdot (1 + O(\rho^k)). \tag{92}$$

Here,

$$V(s) = \log(E_{1+s}[L^s]\, \Psi_{1+s}(0, 1))$$

is analytic near $s = 1$, and ρ is any number satisfying $|\mu(1)| < \rho < 1$, where $\mu(1)$ is a subdominant eigenvalue of $\mathbf{G}_1$.

The central limit theorem of probability theory asserts that large powers—in the "pure" case $V = \alpha = 0$ at least—induce Gaussian laws in the asymptotic limit. There are two differences here: one is the analytic factor $e^{V(s)}$; the other corresponds to the error term $O(\rho^k)$ which is negligible in the scale of the problem. The extension of the central limit theorem to "quasi-powers" of the form (92) has been developed in a general setting by Hwang [16]. Hwang's technology is based on the Berry–Esseen inequality that relates the L_∞ distance between distribution functions to a distance between characteristic functions.

THEOREM (Hwang's Quasipower Theorem). *Let Z_k be a sequence of random variables whose moment generating functions admit the asymptotic estimate*

$$M_k(s) := \mathrm{E}\,[\exp(s\, Z_k)] = \exp(kU(s) + V(s)) \left(1 + O\left(\frac{1}{W_k}\right)\right), \qquad W_k \to \infty,$$

the error term being uniform for s in a disk $|s| \le s_0$ for some $s_0 > 0$. Assume that $U(s)$ and $V(s)$ are analytic for $|s| \le s_0$ and $U(s)$ satisfies the "second moment condition"

$U''(0) \neq 0$. *Then the distribution of Z_k is asymptotically Gaussian*:

$$\Pr\left[\frac{Z_k - kU'(0)}{\sqrt{kU''(0)}} < t\right] = \Phi(t) + O\left(\frac{1}{S_k}\right) \qquad \text{where} \quad \Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-w^2/2}\,dw$$

uniformly for all x in $\mathbf{R}$, as k tends to ∞, with $S_k = \min(\sqrt{k}, W_k)$.

Under these strong analyticity conditions, the mean and variance of Z_k are obtained by differentiation of the asymptotic form of the moment (or the cumulant) generating functions:

$$\mathrm{E}[Z_k] = kU'(0) + V'(0) + O\left(\frac{1}{W_k}\right), \qquad \mathrm{Var}[Z_k] = kU''(0) + V''(0) + O\left(\frac{1}{W_k}\right).$$

The theorem is applicable to the relation (92), with $U(s) = \log\lambda(1+s)$ that is an analytic function near $s = 0$ and the second moment condition $U''(0) \neq 0$ holds provided that the function $\log\lambda(s)$ is strictly concave. Thus, we can state:

THEOREM 1. *Let $(\mathcal{S}, F)$ be a probabilistic dynamical source. If $\mathcal{S}$ is not log-affine, then the distribution of the random variable $\log \ell_k(x)$ is asymptotically Gaussian,*

$$\Pr\left[\frac{\log \ell_k(x) - Ak}{\sqrt{Bk}} < t\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-w^2/2}\,dw + O\left(\frac{1}{\sqrt{k}}\right),$$

uniformly for $x \in \mathbf{R}$, as $k \to +\infty$. The constants A and B are expressible in terms of the derivatives of $\log\lambda(s)$ at $s = 1$,

$$A = [\log\lambda(s)]'_{s=1} = \lambda'(1) \quad \text{and} \quad B = [\log\lambda(s)]''_{s=1} = \lambda''(1) - \lambda'(1)^2.$$

More precisely,

$$\mathrm{E}[\log \ell_k] = Ak + C + O(\rho^k) \quad \text{and} \quad \mathrm{Var}[\log \ell_k] = Bk + D + O(\rho^k),$$

where ρ is any real that is larger than a subdominant eigenvalue $\rho > |\mu(1)|$. The quantities C, D are constants which depend on the initial distribution F, while the main terms are independent of the initial distribution and depend only on the mechanism $\mathcal{S}$ of the source.

This result is a strong form of the Almost Equipartition Property, known as the Shannon–MacMillan–Breiman Theorem, that can be described as follows (see [4], [32], and [41] for more details):

THEOREM (Shannon–MacMillan–Breiman Theorem). *Let $\mathcal{S}$ be a stationary ergodic source with entropy $h(\mathcal{S})$ and alphabet $\mathcal{M}$. Then, for any $\varepsilon > 0$, there exists a positive integer $K_0(\varepsilon)$ such that, if $k > K_0(\varepsilon)$, the set $\mathcal{M}^k$ of prefixes of length k decomposes into two sets $\mathcal{E}_k$ and $\mathcal{T}_k$ satisfying*

(i) $\Pr[\mathcal{E}_k] < \varepsilon$,
(ii) $\exp(-k[h(\mathcal{S}) - \varepsilon]) < \Pr[\{t\}] < \exp(-k[h(\mathcal{S}) + \varepsilon])$ *for any prefix $t \in \mathcal{T}_k$.*

In other words, the set $\mathcal{M}^k$ of prefixes of length k consists of a set of low probability or atypical prefixes (namely $\mathcal{E}_k$) and a disjoint set of high probability or typical prefixes, each of which has a probability of occurrence approximatively $\exp[-kh(\mathcal{S})]$. Our Theorem 1 sharpens the result in the case of the dynamical probabilistic source. We recall that such a source is always ergodic, but not stationary in general.

8. Most Probable Prefixes. We wish to evaluate the asymptotic behaviour of $B(\rho)$, and we have explained in Section 3.5 how the properties of Dirichlet series $\Lambda(F, s)$ intervene, mainly via the location of poles near the line $\Re(s) = 1$. Now, Propositions 8 and 9 make these properties precise, and we consider the two main cases: first, the aperiodic case; then, the periodic case.

Aperiodic case. Here, we begin with the integral expression of $\Lambda(F, s)$, (22):

$$\Lambda(F, s) = s \int_0^\infty A(y)e^{-sy}\,dy \qquad \text{with} \quad A(y) = B(e^{-y}) = \sum_{u_h \ge e^{-y}} 1,$$

and we use the following Tauberian Theorem due to Delange [7], [38].

TAUBERIAN THEOREM [7]. *Let $V(s)$ be a function that admits in the half-plane $\Re(s) > \sigma > 0$ the integral representation*

$$V(s) = s \int_0^\infty A(y)e^{-sy}\,dy,$$

where A is increasing and positive. Assume that

(i) *$V(s)$ is analytic on $\Re(s) = \sigma$, $s \neq \sigma$, and*
(ii) *for some $\gamma \ge 0$,*

$$V(s) = \frac{g(s)}{(s-\sigma)^{\gamma+1}} + \ell(s),$$

where g, ℓ are analytic at σ, with $g(\sigma) \neq 0$.

Then, as $x \to \infty$,

$$A(x) = \frac{g(\sigma)}{\sigma\Gamma(\gamma+1)} e^{x\sigma} x^\gamma \,[1 + \varepsilon(x)\,].$$

The hypotheses of the Tauberian Theorem are fulfilled by $\Lambda(F, s)$ for $\sigma = 1$ and $\gamma = 0$. Validity of hypothesis (i) comes from the aperiodicity. For hypothesis (ii), we begin with the decompositions (62) of iterates of $\mathbf{G}_s$ that are valid on a neighbourhood of any real point s, for any $k \ge 0$. Then this decomposition generalizes to the pseudo-iterates of $\mathbf{G}_s$ defined in (38) and finally to the quasi-inverse, and the quasi-inverses themselves decompose as

$$(I - \mathbf{G}_s)^{-1} = \frac{\mathbf{P}_s}{1 - \lambda(s)} + (I - \mathbf{N}_s)^{-1}$$

or

$$(I - \mathbf{G}_s)^{-1} = \frac{{}^t U \mathbf{P}_s \circ \mathbf{M}_s}{1 - \lambda(s)} +^t U (I - \mathbf{N}_s)^{-1} \circ \mathbf{M}_s,$$

near $s = 1$. Here, $\mathbf{M}_s$ is defined in (37), $\mathbf{P}_s$ is the dominant projector, and $\mathbf{N}_s$ has a spectral radius strictly smaller than 1, so that $(I - \mathbf{N}_s)^{-1}$ is analytic near $s = 1$. Both decompositions involve the quantity $1/(1 - \lambda(s))$ which can be written near $s = 1$ as

$$\frac{1}{1 - \lambda(s)} = \frac{-1}{\lambda'(1)} \frac{a(s)}{s - 1},$$

where a is analytic near $s = 1$ and satisfies $a(1) = 1$. Returning to $\Lambda(F, s)$ with expression (43), and using the secant L of distribution F defined in (42), we obtain the decomposition of hypothesis (ii), with

$$g(s) := \frac{-1}{\lambda'(1)} a(s) \mathbf{P}_s [L^s](0, 1), \qquad \ell(s) := (I - \mathbf{N}_s)^{-1} [L^s](0, 1)$$

in the basic case, or

$$g(s) := \frac{-1}{\lambda'(1)} a(s)^t U \mathbf{P}_s \circ \mathbf{M}_s [L^s](0, 1), \qquad \ell(s) := {}^t U (I - \mathbf{N}_s)^{-1} \circ \mathbf{M}_s [L^s](0, 1)$$

in the Markovian case.

Finally, we can apply the Tauberian Theorem. At $s = 1$, Proposition 6 gives special values of the spectral objects, from which one deduces that $g(1) = -1/\lambda'(1)$. Then

$$A(y) = \frac{-1}{\lambda'(1)} e^y [1 + \varepsilon(y)] \qquad \text{as} \quad x \to \infty$$

and we conclude that

$$B(x) = \frac{-1}{\lambda'(1)x} + o\left(\frac{1}{x}\right) \qquad \text{as} \quad x \to 0.$$

This ends the aperiodic case.

Periodic case. In this case the function B may be highly discontinuous. For instance, in the case of a symmetric Bernoulli source relative to an alphabet of cardinality r, the function B is a step function that satisfies

$$B(x) = \frac{r^{k+1} - 1}{r - 1} \qquad \text{for} \quad x \in \left] \frac{1}{r^{k+1}}, \frac{1}{r^k} \right].$$

Then one cannot expect a result of the same type as previously. On the other hand, it seems that the analysis cannot be conducted in this case on the function B itself which is too highly discontinuous. We instead consider the integral D of B,

$$D(x) := \int_0^x B(y)\, dy, \tag{93}$$

and we begin with the integral expression of $\Lambda(F, s)$:

$$\Lambda(F, s+1) = -s(s+1) \int_0^{\infty} D(x)\, x^{s-1}\, dx.$$

Here, we use Mellin analysis, and we denote by $g^*(s)$ the Mellin transform of a function $g(x)$, defined by

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

so that the Mellin transform $D^*(s)$ of integral D of B,

(95) $$D^*(s) = \frac{1}{s(s+1)}\, \Lambda(F, s+1),$$

is closely linked to $\Lambda(F, s)$.

We use the following theorem [12] that explains how the behaviour of Λ near its singularities can give the asymptotic expansion of D.

MELLIN INVERSION THEOREM. *Suppose that D has a Mellin transform $D^*(s)$ having a non-empty fundamental strip $\langle \alpha, \beta \rangle$. Assume that $D^*(s)$ admits a meromorphic continuation to the strip $\langle \gamma, \beta \rangle$ for some $\gamma < \alpha$, and is analytic on the line $\Re(s) = \gamma$. Assume that there exists a real number $\nu \in (\alpha, \beta)$ such that*

(96) $$D^*(s) = O(|s|^{-r}) \qquad \text{with} \quad r > 1$$

for infinitely many values of $\Im(s)$ when $|s| \to \infty$ in the strip $\langle \gamma, \nu \rangle$. If $D^*(s)$ admits the singular expansion for $s \in \langle \gamma, \alpha \rangle$,

$$\sum_{(\xi,k) \in A} d_{\xi,k}\, \frac{1}{(s-\xi)^k},$$

then an asymptotic expansion of $D(x)$ at 0 is

$$D(x) = \sum_{(\xi,k)\in A} d_{\xi,k} \left(\frac{(-1)^{k-1}}{(k-1)!} x^{-\xi} (\log x)^{k-1} \right) + O(x^{-\gamma}).$$

We prove first that function D eventually fulfils the hypotheses of the theorem. The expression of $D^*(s)$ in (95) together with the properties of Λ shows that $D^*(s)$ has a fundamental strip $\langle 0, +\infty \rangle$. The periodicity assumption entails the existence of a region for Λ that is free of poles. Thus, there exists a real γ $(-1 < \gamma < 0)$ for which $D^*(s)$ admits a meromorphic continuation to the strip $\langle \gamma, +\infty \rangle$ and is analytic on the line $\Re(s) = \gamma$. The periodicity also entails that $\Lambda(F, s)$ is bounded on vertical lines of the strip $\langle 0, +\infty \rangle$, so that $D^*(s)$ satisfies (96) with $r = 2$. Moreover, in this case the poles of $\Lambda(F, s+1)$ in the strip $\langle \gamma, +\infty \rangle$ are all of the form $s_k := ikt_0$ for some $t_0 \neq 0$, with $k \in \mathbf{Z}$. Then $D^*(s)$ has a double pole at $s = 0$ and simple poles at points s_k for $k \neq 0$.

Finally, the periodicity assumption shows that the residues of $\Lambda(F, s)$ at all the points ξ_k equal $-1/\lambda'(1)$, and the singular expression of $D^*(s)$,

$$D^*(s) \asymp \frac{-1}{\lambda'(1)} \frac{1}{s^2} + \frac{C_F}{s} + \frac{-1}{\lambda'(1)} \sum_{\substack{k \in \mathbf{Z} \\ k \neq 0}} \frac{1}{s_k(1+s_k)} \frac{1}{(s-s_k)},$$

leads to the asymptotic expansion of $D(x)$ at 0:

$$D(x) = \frac{1}{\lambda'(1)} \log x + C_F + \frac{1}{\lambda'(1)} P(t_0 \log x) + O(x^{-\gamma}),$$

where P is a periodic function. The only term that may depend on the distribution F is the constant term C_F.

The Mellin transform $B^*(s)$ of $B(x)$ is itself closely linked with $\Lambda(F, s)$,

$$B^*(s) = \frac{1}{s} \Lambda(F, s), \tag{97}$$

but we cannot apply the previous theorem to B^*, since it only satisfies (96) with $r = 1$. If one tries to apply this theorem, we obtain an asymptotic expression for B which is not convergent. However, the derivative of the above expression of $D(x)$ can give some information about $B(x)$. The function $P(t)$ is a continuous periodic function with a very explicit Fourier series of the form

$$P(t) = \sum_{k \in \mathbf{Z} \setminus \{0\}} \frac{e^{itk}}{ik(1+ik)}.$$

Furthermore, the derivative of $P(t)$ exists and (after combining k and $-k$) is given by

$$P'(t) = 2 \sum_{k>0} \frac{\cos(kt)}{1+k^2} + 2 \sum_{k>0} \frac{k \sin(kt)}{1+k^2},$$

an expression which is indeed bounded: the first part is clearly bounded (and continuous). Furthermore, the second sum is bounded, too; we just have to consider the case $0 < t \leq \pi/2$, and we split the sum. Firstly,

$$\sum_{k \leq 1/t} \frac{k \sin(kt)}{1+k^2} \ll \sum_{k \leq 1/t} \frac{k^2 t}{1+k^2} \ll \frac{1}{t} t = 1.$$

Secondly, by partial summation

$$\sum_{k > 1/t} \frac{k \sin(kt)}{1+k^2} \ll \sum_{k > 1/t} \frac{1}{k^2} \frac{1}{t} \ll t \frac{1}{t} = 1.$$

Since the function $P(t)$ has a bounded derivative, the derivative (with respect to x) of the term $P(t_0 \log x)$ is of the form $\Theta(x)$. Then the above expansion of $D(x)$ combined with the monoticity of $B(x)$ directly implies (by elementary means) that $B(x) = \Theta(x)$.

Finally, the following theorem describes in both cases (periodic or aperiodic) the asymptotic behaviour of the number of most probable prefixes.

THEOREM 2. *Let $(\mathcal{S}, F)$ be a probabilistic dynamical source. Let $B(x)$ be the number of prefixes whose probability is at least equal to x.*

(a) *In the case when the source $\mathcal{S}$ is aperiodic, $B(x)$ has the following asymptotic behaviour:*

$$B(x) = \frac{-1}{\lambda'(1)x} + o\left(\frac{1}{x}\right) \qquad as \quad x \to 0.$$

(b) *In the case when the source is $\mathcal{S}$ periodic, there exists a strip $1 - \gamma < \Re(s) < 1$ where the operator $(I - \mathcal{G}_s)^{-1}$ has no poles. Then the integral D of function B defined by $D(x) := \int_0^x B(y)\,dy$ admits the asymptotic expansion at $x = 0$,*

$$D(x) = \frac{1}{\lambda'(1)} \log x + C_F + \frac{1}{\lambda'(1)} P(t_0 \log x) + O(x^{-\gamma}),$$

where P is a periodic function. The only term that depends a priori on the distribution F is the constant term C_F. Moreover, there exist two strictly positive constants A and C (that may depend on the distribution F) such that

$$\frac{A}{x} \leq B(x) \leq \frac{C}{x}.$$

9. Coincidence of Prefixes. In our approach described in Section 3.4, we consider a point (x, y) in the unit square $\mathcal{Q} = [0, 1]^2$, and we analyse the random variable C that represents the coincidence between the two words $M(x)$ and $M(y)$. Then the event $[C \geq k]$ is formed of all pairs (x, y) whose associated words $M(x)$ and $M(y)$ coincide till depth k. In that case, x and y belong to the same fundamental interval of depth k, so that (x, y) lies in a "fundamental square" $\mathcal{C}_h = \mathcal{I}_h \times \mathcal{I}_h$. Thus, one has

$$(98) \qquad [C \geq k] = \bigcup_{|h|=k} \mathcal{I}_h \times \mathcal{I}_h = \bigcup_{|h|=k} \mathcal{C}_h.$$

We consider three cases in the following: first, the case when the two words are drawn independently, second, the case when the density on the unit square $\mathcal{Q} = [0, 1]^2$ is proportional to $|x - y|^r$, and finally the general case of densities of valuation r.

If the two words are drawn independently from the same probabilistic source, the probability of the event $[C \geq k]$ involves all the fundamental measures u_h associated to inverse branches h of depth k,

$$(99) \qquad \rho_k := \Pr[C \geq k] = \sum_{|h|=k} \mu(\mathcal{C}_h) = \sum_{|h|=k} (u_h)^2 = \Lambda_k(F, 2).$$

Now, the quasi-power property (Proposition 4) applies, and, as a consequence, the ρ_k decrease geometrically:

$$(100) \qquad \rho_k = A_F \cdot \lambda(2)^k (1 + O(\gamma^k)),$$

for some real $A > 0$ that depends on the initial distribution F, and some constant γ with $0 < \gamma < 1$ (one may take any $\gamma > |\mu(2)|/\lambda(2)$, where $\mu(2)$ is a subdominant eigenvalue of $\mathbf{G}_2$). The expectation of C satisfies

$$\text{(101)} \qquad \mathrm{E}[C] = \sum_{k \geq 0} \rho_k = \Lambda(F, 2).$$

The analysis generalizes directly to densities $g(x, y)$ on the unit square that are proportional to $|x - y|^r$ for some real parameter $r > -1$. The case $r = 0$ is the uniform model; the cases $-1 < r < 0$ correspond to giving a heavier weight to similar words. We explain now why statements similar to (99) then hold true but with the operator $\mathbf{G}_{2+r}$ replacing $\mathbf{G}_2$. In this case the measure of a square "built" on the diagonal satisfies

$$\mu([a, b] \times [a, b]) = |b - a|^{r+2},$$

so that the probability of the event $[C \geq k]$ is now

$$\rho_k = \sum_{|h|=k} |h(0) - h(1)|^{2+r} = \Lambda_k(\mathrm{Id}, 2 + r).$$

This analysis further generalizes to more general non-uniform densities over the unit square, so-called of valuation r. This notion is defined in (18). For studying the coincidence in this quite general model, we use another generalization of Ruelle operators,

$$\widehat{\mathbf{G}}_{s,t}[F](u, v) := \sum_{|h|=1} \widetilde{h}(u)^{s/2}\, \widetilde{h}(v)^{s/2}\, \widetilde{H}(u, v)^t\, F(h(u), h(v)),$$

which is well adapted to these more general densities. This another generalized operator also extends the classical operator $\mathcal{G}_{s+t}$, in the sense of (35). It is also a generating operator since its iterate of order k involves all the branches of depth k,

$$\widehat{\mathbf{G}}^k_{s,t}[F](u, v) := \sum_{|h|=k} \widetilde{h}(u)^{s/2}\, \widetilde{h}(v)^{s/2}\, \widetilde{H}(u, v)^t\, F(h(u), h(v)).$$

Moreover, it shares its main spectral properties with the classical operator $\mathcal{G}_{s+t}$ which it extends. In particular, for s, t real, it has dominant spectral objects, and its iterate of order k behaves (for large k) as a true kth power of the dominant eigenvalue $\lambda(s + t)$ of $\mathcal{G}_{s+t}$. More precisely, a quasi-power property holds for $k \to \infty$ and

$$\widehat{\mathbf{G}}^k_{s,t}[F](u, v) \sim \lambda(s + t)^k\, \widehat{\Psi}_{s,t}(u, v) \widehat{\mathbf{E}}_{s,t}[F],$$

where $\widehat{\mathbf{E}}_{s,t}$, $\widehat{\Psi}_{s,t}$ represent dominant spectral objects of the operator $\widehat{\mathbf{G}}_{s,t}$.

With respect to densities of valuation r, defined in (18), the measure of the fundamental square $\mathcal{C}_h := \mathcal{I}_h \times \mathcal{I}_h$ satisfies

$$\mu(\mathcal{C}_h) := \iint_{\mathcal{C}_h} d\mu(x, y) = \iint_{\mathcal{C}_h} |x - y|^r \ell(x, y)\, dx\, dy.$$

Then, when using the change of variables defined by $x = h(u)$, $y = h(v)$, we obtain

$$\mu(\mathcal{C}_h) = \iint_{\mathcal{C}} \widetilde{h}(u)\, \widetilde{h}(v)\, \widetilde{H}(u, v)^r |u - v|^r\, \ell(h(u), h(v))\, du\, dv,$$

so that the probability ρ_k of the event $[C \geq k]$ is now expressible in terms of the kth iterate of the operator $\widehat{\mathbf{G}}_{2,r}$:

$$\rho_k = \iint_C |u - v|^r \, \widehat{\mathbf{G}}^k_{2,r}[\ell](u, v) \, du \, dv.$$

Since ℓ is strictly positive, a quasi-power property for $\widehat{\mathbf{G}}^k_{2,r}[\ell]$ holds, and then

$$\rho_k \sim \lambda(2 + r)^k \, \widehat{\mathbf{E}}_{2,r}[\ell] \iint_C |u - v|^r \, \widehat{\Psi}_{2,r}(u, v) \, du \, dv,$$

where $\widehat{\mathbf{E}}_{2,r}$, $\widehat{\Psi}_{2,r}$ represent dominant spectral objects of the operator $\widehat{\mathbf{G}}_{2,r}$.

THEOREM 3. *When two words are drawn from the same source with respect to density of valuation $r > -1$, the length of their longest common prefix has a probability distribution that satisfies*

$$\rho_k := \Pr[C \geq k] = A_{\ell,r} \cdot \lambda(2 + r)^k (1 + O(\gamma^k)).$$

Here, γ is some constant γ with $0 < \gamma < 1$, $\lambda(2 + r)$ is the dominant eigenvalue of the Ruelle operator $\mathcal{G}_s$ associated to the source for $s = 2 + r$, and the constant A depends on the initial distribution.

NOTE. The random variable C has another interesting algorithmic meaning. When comparing two real numbers of $\mathcal{I}$, one can use their expansions in the same numeration system. One runs in parallel two "lazy" versions of the same numeration process, one on each number, and execution is halted as soon as a discrepancy of expansions is detected. The variable C is now exactly the number of iterations of this comparing algorithm.

10. An Important Example: The Continued Fraction Source. Some results about this source have already been discussed in [13], but only in the case when the initial density is uniform. Here, we explain how our main results (Theorems 1–3) can apply to this case. The operator of Ruelle is then called the Ruelle–Mayer operator and is defined by

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

for complex s satisfying $\Re(s) > \frac{1}{2}$.

The entropy of the source is linked to the so-called Lévy constant which plays a central rôle in the analysis of the Euclidean algorithm, and the coincidence probability is a constant that intervenes in two-dimensional generalizations of the Euclidean algorithm [6], [13]. Here, one has

$$\lambda'(1) = -\frac{\pi^2}{6 \log 2}, \qquad \lambda(2) \approx 0.1994.$$

The dominant eigenfunction is $1/(1 + x)$.

The source is aperiodic, and the poles of $(I - \mathbf{G}_s)^{-1}$ are well known because they intervene in several deep mathematical questions. They include all the non-trivial 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 operator and they lie on the line $\Re(s) = \frac{1}{2}$ (see [9]). These last values do not however occur as poles of the Dirichlet series $\Lambda(\text{Id}, s)$. In the half-plane $\Re(s) > 0$, $\Lambda(\text{Id}, s)$ can be represented as

$$\Lambda(\text{Id}, s) = 2\frac{\zeta(2s-1)}{\zeta(2s)}\frac{2^{1-s}-1}{1-s} + \frac{R(s)}{\zeta(2s)},$$

where $R(s)$ is analytic in $\Re(s) > 0$. So, when the initial density is uniform, the asymptotic expansion of $B(x)$ given in Theorem 2 solely involves the non-trivial zeros of the Riemann zeta function. For an arbitrary density, the same situation occurs, because the eigenvectors f of the hyperbolic Laplace operators satisfy $\int_0^1 f(x)\,dx = 0$ [24]. Now, if the Riemann hypothesis is true, all the non-trivial zeros of the Riemann zeta function are on the line $\Re(s) = \frac{1}{2}$. Then $\Lambda(F, s)$ has no poles in the strip $\frac{1}{4} < \Re(s) < 1$. With Mellin analysis, and some tools of Prime Number Theory, Theorem 2 should imply (under the validity of Riemann hypothesis)

$$B(x) = \frac{6\log 2}{\pi^2}\frac{1}{x} + O\left(\frac{1}{x^{1/4+\varepsilon}}\right) \qquad \text{for any} \quad \varepsilon > 0.$$

In the case of the continued fraction source, the length of the fundamental intervals is closely related to the continuants. So, Theorem 1 has another interpretation in terms of continuants: the asymptotical log-normality of continuants is well known with respect to uniform density (see [27], [25], and [26]). The author has extended the result to non-uniform densities [39], and this result has been obtained with the same methods as here.

The HAKMEM memo [3] first described the comparison algorithm with continued fraction expansions. Later, the same algorithm has been used by Knuth in the Metafont system [18], and, more recently, Avnaim et al. [1] have proposed it for computing (in single-precision) the sign of (2×2)-determinants with integer entries. The comparison algorithm with continued fraction expansions has been previously studied in [40], and Theorem 3 can be seen as a far-reaching generalization of [40].

Acknowledgements. Many thanks to Julien Clément, Philippe Flajolet, Mireille Régnier, and Wojciech Szpankowski for valuable discussions.

References

[1] Avnaim, F., Boissonnat, J.-D., Devillers, O., Preparata, F., and Yvinec, M. Evaluation of a new method to compute signs of determinants. In *Proceedings of the Eleventh Annual ACM Symposium on Computational Geometry*, 1995, pp. C16–C17. Full paper in *Algorithmica*, **17** (1997), 111–132.

[2] Bedford, T., Keane, M., and Series, C., eds. *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*. Oxford University Press, 1991.

[3] Beeler, M., Gosper, R. W., and Schroeppel, R. HAKMEM. Memorandum 239, Artificial Intelligence Laboratory, M.I.T., Feb. 1972.

[4] Billingsley, P. *Ergodic Theory and Information.* Wiley, New York, 1965.

[5] Bogomolny, E. B., and Carioli, M. Quantum maps from transfer operators. *Physica D* **67** (1993), 88–112.

[6] 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.

[7] Delange, H. Généralisation du Théorème d'Ikehara. *Annales Scientifiques de l'Ecole Normale Supériore* **71** (1954), 213–242.

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

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

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

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

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

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

[14] Grothendieck, A. Produits tensoriels topologiques et espaces nucléaires. *Memoirs of the American Mathematical Society* **16** (1955).

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

[16] Hwang, H.-K. Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques. Ph.D. thesis, École Polytechnique, Dec. 1994.

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

[18] Knuth, D.E. *Computers and Typesetting, MF: the program,* Volume D. Addison-Wesley, Reading, MA, 1986.

[19] Krasnoselskii, M. *Positive Solutions of Operator Equations.* Noordhoff, Groningen, 1964.

[20] Lorch, E. R. *Spectral Theory.* Oxford University Press, New York, 1962.

[21] Mayer, D. H. On a ζ function related to the continued fraction transformation. *Bulletin de la Société Mathématique de France* **104** (1976), 195–203.

[22] Mayer, D. H. Spectral properties of certain composition operators arising in statistical mechanics, *Communications in Mathematical Physics* **68** (1979), 1–8.

[23] Mayer, D. H. Continued fractions and related transformations. In *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces,* T. Bedford, M. Keane, and C. Series, eds. Oxford University Press, Oxford, 1991, pp. 175–222.

[24] Mayer, D. H. Private communication.

[25] Mischyavichyus, G. A. Estimate of the remainder in the limit theorem for the denominators of continued fractions. *Litovskiĭ Matematičeskiĭ Sbornik* **21**(3) (1987), 63–74.

[26] Morita, T. Local limit theorem and distribution of periodic orbits of Lasota–Yorke transformations with infinite Markov partitions. *Journal of the Mathematical Society of Japan* **46**(2) (1994), 309–343. Corrections, op. cit. **47**(1) (1997), 191–192.

[27] Philipp, W. Ein zentraler Grenzwertsatz mit Anwendungen auf die Zahlentheorie. *Zeitschrift für Wahrscheinlichkeitstheorie* **8** (1967), 195–203.

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

[29] Ruelle, D. *Thermodynamic Formalism.* Addison-Wesley, Reading, MA, 1978.

[30] Ruelle, D. *Dynamical Zeta Functions for Piecewise Monotone Maps of the Interval.* Volume 4 of CRM Monograph Series. American Mathematical Society, Providence, RI, 1994.

[31] Schwartz, H. Composition operators in $\mathcal{H}^p$. Ph.D. Thesis, University of Toledo.

[32] Shannon, C. A mathematical theory of communication. *Bell Systems Technical Journal* **27** (1948), 379–423, 623–656.

[33] Shapiro, J. *Composition Operators and Classical Function Theory.* Universitext: Tracts in Mathematics. Springer-Verlag, New York, 1993.

[34] Shapiro, J. Compact composition operators on spaces of boundary regular holomorphic functions. *Proceedings of the AMS* **100** (1997), 49–57.

[35] Shapiro, J., and Taylor, P.D. Compact, nuclear, and Hilbert–Schmidt composition operators on $\mathcal{H}^2$. *Indiana University Mathematics Journal* **23** (1973), 471–496.

[36] Szpankowski, W., and Frieze, A. Greedy algorithms for the shortest common superstring that are asymptotically optimal. Preprint, 1998

[37] Szpankowski, W., and Luczak, T. A suboptimal lossy date compression based on appoximate pattern matching. *IEEE Transactions on Information Theory* **43**(5) (1997), 1439–1451.

[38] Tenenbaum, G. *Introduction à la théorie analytique des nombres*, vol. 13. Institut Élie Cartan, Nancy, 1990.

[39] Vallée, B. Opérateurs de Ruelle–Mayer généralisés et analyse des algorithmes d'Euclide et de Gauss. *Acta Arithmetica* **81**(2) (1997), 101–144.

[40] Vallée, B. Algorithms for computing signs of 2×2 determinants: dynamics and average-case algorithms, *Proceedings of the 8th Annual European Symposium on Algorithms, ESA '97*, pp. 486–499. LNCS 1284. Springer-Verlag, Berlin, 1997.

[41] Welsh, D. *Codes and Cryptography.* Oxford Science Publications, Clarendon Press, Oxford, 1989.