Dynamical systems and average-case analysis of general tries

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

June 3, 1997

Abstract

Digital trees or tries are a general purpose flexible data structure that implements dictionaries built on words. This paper is devoted to the average-case analysis of tries in a unified model that encompasses the models of independent symbols, Markov chains, and more generally all sources associated to a dynamical system. The three major parameters, number of nodes, path length, and height, are analysed precisely. The results can all be stated in terms of two intrinsic characteristics of the source: the entropy and the probability of letter coincidence. These characteristics relate in a natural way to spectral properties of a transfer operator of the Ruelle type. Our results extend and improve earlier ones obtained when the source produces independent symbols, or is derived from a Markov chain.

1 Introduction.

Digital trees or tries are a versatile data structure that implements "dictionary" operations on sets of words (namely, insert, delete and query), as well as set-theoretic operations like set union or set intersection. It can also be used for text searching, interval search and partial match retrieval; see [13, 29]. In a recent paper, Bentley and Sedgewick even show that a carefully design implementation can be more efficient than hashing while offering considerably wider functionality [2].

The cost of these operations is mainly measured by three parameters of the tree structure: height, number of internal nodes and external path length. This paper is devoted to the average case analysis of these three basic parameters. Here, data items are (infinite) words that are produced by a common mechanism, called a source in information theory contexts.

Tries have been analysed to some extent under two models: a memoryless source, where symbols in words are each emitted independently of the previous ones, and occasionally, Markov chains, where the probability of emitting a symbol depends solely on a bounded part of the past history. While methods used earlier are rather diverse (from probabilsitic to analytic), all the estimates appear to involve two intrinsic parameters of the souce: the entropy and the probability of symbol coincidence. The first parameter intervenes in the expectation of size and path length, the second one in the expectation of height. For memoryless sources, generating functions are classically used, in conjunction with Mellin transforms [9, 14, 20, 22, 23, 26, 31]. When the source has memory, classical generating function techniques no longer apply directly. For instance, under the Markov chain model, Jacquet et Szpankowski [15] resort to a notion of alignment in conjunction with a suitable inclusion-exclusion principle.

As opposed to hashing, data on which tries are built often arise from more complex sources. We work here inside a quite general framework of sources, related to dynamical systems theory, which goes beyond the cases of memoryless and Markov sources. A dynamical source can describe nonmarkovian processes, where the dependency on the past history is not bounded, and as such, they attain a high level of generality. The mechanism associates a word to a real number of the $[0, 1]$ interval. When the mechanism is fixed, the initial distribution on the interval $[0, 1]$ can change and this induces different behaviours for trie parameters.

The dependence on the initial input distribution has already been considered by Devroye in the case of the simplest souce, the Bernoulli source. We treat here the case of a general dynamic source *and* a general initial input distribution, which we call a *probabilistic dynamic source.*

To the best of our knowledge, the only instance of a source with unbounded history dependency considered in the literature is that of continued fraction representations [11]. One uses there quite different methods that are based on the underlying dynamical system and the main tool is the Ruelle *transfer operator* that extends ordinary or matrix generating functions. The Ruelle operator may then be considered as a "generating operator" as it generates objects that are essential in the analysis of tries, namely, *fundamental intervals.*

Here, we show that we can adapt these methods to the general framework of dynamical sources. Our main tool is a Ruelle operator linked to the dynamical system associated to the source. However, contrary to what happens with the continued fraction source, the classical Ruelle operator is not adequate when the initial input density varies. We thus devise a new tool, the generalized Ruelle operator, that acts on functions of two variables and relates to an earlier generalization of [32]. The operator makes it possible to express the Dirichlet series of fundamental intervals. The Mellin transform is then no longer the major tool since poles of the Dirichlet series cannot always be precisely located. However, the Ruelle operator satisfies strong positivity properties so that Tauberian methods can be used. Furthermore, positivity properties entail the existence of dominant spectral objects that are closely linked to intrinsic parameters of the source, like *entropy* and *coincidence probability.* This approach unifies previous analyses in a powerful way while widely extending their range to the case of a probabilistic dynamic source.

The following results are established:

- The height of a trie has expectation $\sim B \log n$ and its limit probability distribution is of the doubly exponential type with sharp tail properties (Theorem 2).
- The average search time is $\sim A \log n$ (Theorem 3).
- The average size of the trie is, up to possible small fluctuations, well approximated by a quantity that is $\approx An$. (Theorem 4)

There, the constants A and B are well-characterized quantities that are expressible in terms of entropy and coincidence probability. They can be also expressed in terms of spectral objects of transfer operators and are computable for each particular model. They are independent of the density used to draw the words produced by the source. This last result in particular generalizes and improves the results obtained by Devroye [6].

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 trie data structure. It is shown that the expectations of the three parameters studied (height, size, path length) can be expressed as harmonic sums that involve the measures of fundamental intervals. Then, the Dirichlet series of fundamental intervals is introduced, via Mellin transforms and Tauberian theory. In Section 4, we make explicit the basic properties of the generalized Ruelle operator, and show how it generates the fundamental intervals. 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 5, we describe how the singular behaviour of the Dirichlet series of fundamental intervals yields the asymptotic form of trie parameters expressed in terms of of entropy and coincidence probability. We conclude with some examples in Section 6.

2 Probabilistic dynamical sources.

Here, we describe the general framework of dynamical sources, and define the notion of the fundamental intervals, and two basic parameters of the source, the entropy and the coincidence probability.

In the context of information theory, a source is a mechanism which produces infinite words written on an alphabet $\mathcal{M}$. We are mainly interested here in sources that are associated to dynamical systems. Such a system is defined with four elements:

(i) an alphabet $\mathcal{M}$ included in N. (ii) a quasi–partition of $\mathcal{I} :=]0,1[$ with intervals $\mathcal{I}_m, m \in \mathcal{M}$, (iii) a mapping σ which is constant and equal to m on each $\mathcal{I}_m$ (iv) a mapping T which satisfies two properties: the restriction of T to $\mathcal{I}_m$ is a real analytic bijection from $\mathcal{I}_m$ to $\mathcal{I}$; the mapping T is expansive, i.e. there exists $\delta > 1$ for which $|T'(x)| \geq \delta > 1$ on $\mathcal{I}$.

The words emitted by the source are then produced as follows: The mapping $T : \mathcal{I} \to \mathcal{I}$ (almost everywhere defined) is used for iterating the process, as a shift mapping; the mapping $\sigma : I \to \mathcal{M}$ 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

$$M(x) := (\sigma(x), \sigma(Tx), \sigma(T^2x), \ldots \sigma(T^kx) \ldots). \tag{1}$$

The k-th component of $M(x)$ is equal to $M_k(x) := \sigma(T^{k-1}x)$.

The number of branches of T equals the cardinality of the alphabet, and the alphabet is used for coding the distinct branches of T, 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 such a framework. A source is said memoryless when the random variables M_k are independent and follow the same law. The Bernouilli 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 Bernouilli law of parameters $(p_m)_{m \in \mathcal{M}}$.

The quasi–partition of $\mathcal{I}$ is then defined dy $\mathcal{I}_m :=]q_m, q_{m+1}[$, where q_m is the sum of the terms p_j with index j less than m; 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

$$T(x) = \{bx\}, \qquad \sigma(x) = [bx], \tag{2}$$

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 Bernouilli sources.

Continued fraction expansion. The general framework may also create quite different sources, with memory. It is sufficient to use a mapping T whose branches are not all affine. The *continued fraction* transformation is an example of this situation. The alphabet is N, the quasi–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 homography $T(x) := (1/x) - m$,

$$T_{CF}(x) = \{\frac{1}{x}\}, \qquad \sigma_{CF}(x) = [\frac{1}{x}], \tag{3}$$

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

Markov sources. Until now, the shift used at each stage is always defined in the same way. Very often, the modeling of more realistic sources leads to the use of a shift that depends on the last emitted symbol. If r denotes the cardinality of the alphabet (finite or denumerable), one considers a set of $r+1$ different systems $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_i, \ldots, \mathcal{S}_r$ all defined on the same alphabet. The system $\mathcal{S}_0$ is used to begin with, and the system $\mathcal{S}_j$ is chosen when the previously emitted symbol is j. One thus associates to a real x of $\mathcal{I}$ an infinite word $M(x)$ on alphabet $\mathcal{M}$

$$M(x) := (\sigma(x), \sigma(Tx), \sigma(T^2x), \ldots \sigma(T^kx) \ldots),$$

with a formula analogous to (1), where the mappings T et σ are now defined by

$$\sigma(x) := \sigma_0(x) \quad T(x) := T_j(x) \text{ if } \sigma(x) = j; \quad \sigma(Tx) := \sigma_j(T_j x) \text{ if } \sigma(x) = j.$$

The model of Markov chains of order 1 is then relative to the case when all the $\mathcal{S}_j$'s are Bernouilli systems. More precisely, if the system $\mathcal{S}_j$ is a Bernouilli system of parameters $P_j := (p_{ij})_{j \leq r}$, the transition matrix P of the Markov chain is the square matrix $r \times r$

$$P := (p_{ij}) \quad 1 \leq i, j \leq r,$$

and the initial probability system is the vector P_0.

Fundamental intervals, entropy, coincidence probability. Each inverse branch of T^k

$$h_{m_1} \circ h_{m_2} \circ \cdots \circ h_{m_k}$$

is called an inverse branch of depth k. It is associated to a word of $\mathcal{M}^k$. For an alphabet of cardinality r, there are r^k branches of depth k. The depth of h is denoted by $|h|$. The *fundamental interval* relative to the inverse branch h is $\mathcal{I}_h := h(\mathcal{I})$. 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 of length k is exactly the word associated to h.

In the sequel, we are interested in *probabilistic dynamical sources*, where the words are emitted with a prescribed distribution: The interval $\mathcal{I}$ is endowed with continuous density f, F denotes the associated distribution function, and the set $M(\mathcal{I})$ of the words produced by the dynamical source $\mathcal{S}$ is endowed with the probability induced by M. In this context, the measure u_h of the fundamental interval $\mathcal{I}_h$

$$u_h := |F(h(0)) - F(h(1))| \tag{4}$$

plays an important rôle, since it equals the probability that a source word begins with the prefix of $\mathcal{M}^*$ relative to h.

The *entropy* $h(\mathcal{S}, F)$ and the *coincidence probability* $c(\mathcal{S}, F)$ relative to source $\mathcal{S}$ are defined as the limits, when they exist, of two quantities that involve the measures u_h of fundamental intervals,

$$h(\mathcal{S}, F) := \lim_{k \to \infty} \frac{-1}{k} \sum_{|h|=k} u_h \log u_h \quad c(\mathcal{S}, F) := \lim_{k \to \infty} \left(\sum_{|h|=k} u_h^2 \right)^{1/k}. \tag{5}$$

We show in the sequel that these limits exist and are independent of the distribution F. They play an important rôle in the average–case analysis of the tries.

3 Tries associated to a general source.

We consider the problem of comparing n infinite words independently produced by the same dynamical source when proceeding by elementary comparisons between their digits. The underlying structure is a tree, called a *trie* [17, 22]. We first describe the trie data structure, then we show that the expectations of the three studied parameters (height, size, path length) can be expressed as harmonic sums that involve the measures of fundamental intervals. Finally, the Dirichlet series of fundamental intervals is introduced, via Mellin transform and Tauberian theorems.

For any word produced by the dynamical source $\mathcal{S}$

$$M(x) := (\sigma(x), \sigma(Tx), \sigma(T^2 x), \ldots \sigma(T^k x) \ldots),$$

we define the head and tail functions

$$\text{head}(M(x)) = \sigma(x), \qquad \text{tail}(M(x)) = T(x).$$

Any finite set of infinite words produced by the same source can be written as $M(X)$, and one associates to X a *trie*, Trie(X), defined by the following recursive rules:

(R_1) If $X = \{x\}$ has cardinality equal to 1, then Trie(X) consists of a single *leaf node* that contains $M(x)$.

(R_2) If X has cardinality at least 2, then Trie(X) is an *internal node* represented generically by 'o' to which are attached ℓ subtrees, where $\ell = \text{card}\{\sigma(x)), \ x \in X\}$ is the number of different head symbols in $M(X)$. Let $b_1 < b_2 < \cdots < b_\ell$ be these different head symbols; Trie(X) is defined by

$$\text{Trie}(X) = \langle o, \text{Trie}(X_1), \text{Trie}(X_2), \ldots, \text{Trie}(X_\ell)\rangle,$$

where

$$X_j = \{T(x) \ \big| \ \sigma(x) = b_j, \ x \in X\} = \text{Trie}(X \cap \mathcal{I}_{b_j})$$

is the trie which collects all the suffixes of words that begin with b_j.

Such a tree structure underlies classical radix sorting methods. It can be built by following the recursive rules R_1, R_2, and its nodes are closely linked to fundamental intervals.

The *level* of a node in a trie is the number of edges that connect it to the root. The *height* of the trie is the maximum level of any leaf, the *path length* of the trie is the sum of the levels of all leaves and the *size* of the tree is the number of its internal nodes.

The height of Trie(X) is then a measure of distance between the two closest elements of X. The path length equals the total number of symbols that need to be examined in order to distinguish all elements of X, and the size of the tree plus the cardinality of X is the number of fundamental intervals necessary to isolate all elements of X.

The purpose of average–case analysis of structures is to characterize the mean value of their parameters under a well-defined probabilistic model that describes the initial distribution of its inputs. Here, we adopt the following quite general model : we work with a finite set of infinite words independently produced by the same dynamical source. This set is of the form $M(X)$, has cardinality n and is obtained by n independent drawings $x_1, x_2, \ldots, x_n$ in the interval $\mathcal{I}$, when the interval is endowed with density f (F denotes the associated distribution).

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

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

In this model, N is narrowly concentrated near its mean n with a high probability so that the rate n plays a rôle much similar to size. This model is called the Poisson model of rate n and distribution F and is denoted by $\mathcal{P}(n, F)$. (It is possible to go back to the model where n is fixed by analytic "depoissonization" techniques, and first order asymptotic results appear not to be affected by the Poisson assumption.) The interest of the Poisson model is that there is complete independence on what happens in disjoint subintervals of $\mathcal{I}$. In particular, the number of elements that fall into any interval of measure u is itself distributed as a Poisson variable of rate nu. Such a property holds notably for the fundamental intervals associated to the source $\mathcal{S}$, whose measure u_h is given by (4). This strong independence property gives access to the expectation of basic parameters.

Consider a random trie produced by a source $\mathcal{S}$ that is endowed with the Poisson model $\mathcal{P}(n, F)$. Such a trie has height $\le k$ provided no fundamental interval of depth k contains more than 1 element. The probability of this event, is given by the independence property of the Poisson model,

$$\pi_k(n) = \prod_{|h|=k} (1 + nu_h)e^{-nu_h}, \tag{6}$$

which yields the expected height

$$D(n) = \sum_{k=0}^{\infty} [1 - \pi_k(n)]\,. \tag{7}$$

For size, we observe that an internal node in the trie corresponds to a fundamental interval occupied by at least two numbers. Similarly, path length is equal to the sum over all internal nodes of the number of elements that fall into the associated fundamental interval. The expectations of size and path length result, since

$$\left[1-(1+\lambda)e^{-\lambda}\right], \qquad \lambda\left[1-e^{-\lambda}\right],$$

are respectively the probability that a Poisson variable of rate λ has value ≥ 2 and the expectation of such a variable conditioned to be at least 2. In summary:

Theorem 1 *The expectations of height, size, and path length of a trie produced by a source $\mathcal{S}$ are, under the Poisson model $\mathcal{P}(n, F)$*

$$\begin{aligned} D(n) &= \sum_k [1 - \prod_{|h|=k} (1+nu_h)e^{-nu_h}] \\ S(n) &= \sum_h \left[1-(1+nu_h)e^{-nu_h}\right] \\ P(n) &= \sum_h nu_h\left[1-e^{-nu_h}\right], \end{aligned} \tag{8}$$

where the sums and products are indexed by inverse branchs h associated to the source $\mathcal{S}$, and u_h is the measure of the fundamental interval $\mathcal{I}_h$, equal to $|F(h(0)) - F(h(1))|$.

Mellin transform. The general relations (8) of Theorem 1 show that $S(n)$ and $P(n)$ are so-called harmonic sums

$$G(n) = \sum_h g(n\,u_h), \qquad u_h > 0. \tag{9}$$

The Mellin transform of a function $g(x)$ is defined by

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

and the Mellin transform of e^{-x} is the Euler gamma function $\Gamma(s)$. The Mellin transform is known to map the asymptotic expansions of the original function $g(x)$ at 0 and $+\infty$ to the singularities of the transformed function $g^*(s)$. In addition, it transforms the harmonic sum $G(n)$ of type (9) into a factored form,

$$G^*(s) = \Lambda(F,-s)\cdot g^*(s), \qquad \Lambda(F,s) = \sum_h u_h^s.$$

Here, Mellin transforms of $S(n)$ and $P(n)$ are found to be

$$S^*(s) = -\Lambda(F,-s)(s+1)\Gamma(s), \qquad P^*(s) = -\Lambda(F,-s)s\Gamma(s). \tag{10}$$

There, the key quantity for the analysis of size and path length is the *series of fundamental intervals*,

$$\Lambda(F,s) = \sum_{|h|\geq 0} u_h^s = \sum_{|h|\geq 0} |F(h(0)) - F(h(1))|^s \tag{11}$$

considered for complex values of s. By the principles of Mellin asymptotics, the precise location of singularities of $\Lambda(F,s)$ is needed.

Tauberian methods. For a general source, it is not always possible to locate precisely the singularities of $\Lambda(F,s)$. An alternative method is based on Tauberian theorems. The Dirichlet series $\Lambda(F,s)$ admits an integral expression

$$\Lambda(F,s) = s\int_0^\infty A(x)e^{-sx}dx \quad \text{with} \quad A(x) = \sum_{u_h \geq e^{-x}} 1,$$

so that the harmonic sum in (9) also admits an integral expression

$$G(n) = \int_0^\infty g'(x)\, B(\frac{n}{x})dx,$$

that involves the function B defined as $B(x) := A(\log x)$. Here, one obtains

$$P(n) := \int_0^\infty [1 - e^{-x} + xe^{-x}]\, B(\frac{n}{x})dx, \qquad S(n) := \int_0^\infty xe^{-x}\, B(\frac{n}{x})dx, \tag{12}$$

where the function B is defined as previously. If now the properties of $\Lambda(F,s)$ permit to apply a Tauberian Theorem, it gives the main term of the asymptotic expansion of $B(y)$ which can be injected in the expressions (12). One thus obtains the main terms of the asymptotic expansion of $P(n)$ and $S(n)$.

The analysis of height. The analysis of height is different; it is based on estimates of the individual probabilities $\pi_k(n)$ of (6) followed by a Mellin analysis. First, taking logarithms, one gets

$$\log \pi_k(n) = \sum_{|h|=k} \big[-nu_h + \log(1+nu_h)\big]. \tag{13}$$

The inequality

$$|\log \pi_k(n) + \frac{n^2}{2}\sum_{|h|=k} u_h^2| \leq \frac{n^3}{2}\sum_{|h|=k} u_h^3, \tag{14}$$

shows that the *series of fundamental intervals of depth k*

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

is now involved for real values of parameter s (namely $s=2, s=3$).

4 Generalized Ruelle operators.

Here, we make explicit the basic properties of the generalized Ruelle operator, and show how it generates the fundamental intervals. We transfer the properties of the operator on the Dirichlet series of fundamental intervals, and we link the dominant spectral objects of the Ruelle operator to the basic parameters of the source, the entropy and the coincidence probability.

There is a direct relationship between the dynamics of source $\mathcal{S}$, the average-case analysis of $\mathcal{S}$–tries, and spectral properties of an operator closely related to the way T transforms probability distributions. The basic ingredient, well-developed in dynamical systems theory, is the class of *transfer operators* [1, 28]. In its simplest form, the transfer operator is the "density transformer",

$$\mathcal{G}[f](x) := \sum_{m\in\mathcal{M}} |h_m'(x)|\, f(h_m(x)) = \sum_{|h|=1} |h'(x)|\, f(h(x)).$$

If f is the initial density on $\mathcal{I}$, the density function of the iterate $T(x)$ is precisely $\mathcal{G}[f](x)$. In fact, it proves highly useful to work with the more general operator,

$$\mathcal{G}_s[f](x) := \sum_{|h|=1} \tilde{h}(z)^s f(h(x)), \tag{16}$$

where $\tilde{h}$ is the analytic extension of $|h'|$ and s is a complex parameter. The dynamics of the basic algorithm 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.

We introduce here a new tool: the generalized operators of Ruelle that involve secants of inverse branches

$$H(u,v) := \left| \frac{h(u) - h(v)}{u - v} \right| \tag{17}$$

instead of tangents $|h'(z)|$ of inverse branches; they are defined by the expression

$$G_s[F](u,v) = \sum_{|h|=1} \tilde{H}(u,v)^s \; F(h(u), h(v)), \tag{18}$$

where $\tilde{H}$ is the analytic extension of H and s is a complex parameter. There the sum is taken over all branches of depth 1. Other useful generalisations of Ruelle operators have been previously given [32].

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 relation

$$G_s[F](u,u) = \sum_{|h|=1} |h'(u)|^s F(h(u), h(u)) = \sum_{|h|=1} |h'(u)|^s f(h(u)) = \mathcal{G}_s[f](u). \tag{19}$$

By the chain rule, the kth iterate of $\mathcal{G}_s$ and G_s are then

$$\mathcal{G}_s^k[f](z) = \sum_{|h|=k} \tilde{h}(z)^s \; f(z),$$

$$G_s^k[F](u,v) = \sum_{|h|=k} \tilde{H}(u,v)^s \; F(h(u), h(v)),$$

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

Symbolically, the kth powers of the operators represent k iterations of shift T. The quasi-inverses $(I - \mathcal{G}_s)^{-1}$, $(I - G_s)^{-1}$, respectively being the formal sum of all the powers of the operators then represent *all* the possible iterations. Important instances of these schemes are the series of fundamental intervals (15), (11) which can be easily expressed in terms of the operator G_s,

$$\Lambda_k(F,s) := \sum_{|h|=k} |F(h(0)) - F(h(1))|^s = G_s^k[L^s](0,1), \tag{20}$$

$$\Lambda(F,s) := \sum_{h} |F(h(0)) - F(h(1))|^s = (I - G_s)^{-1}[L^s](0,1), \tag{21}$$

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

$$L(x,y) = \left| \frac{F(x) - F(y)}{x - y} \right|. \tag{22}$$

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

The series $\Lambda_k(F,s)$ of fundamental intervals of depth k is expressible in terms of powers G_s^k. 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 G_s. The series $\Lambda(F,s)$ of fundamental intervals of all depths is expressible in terms of the quasi-inverse $(1-G_s)^{-1}$. Asymptotic analysis of coefficients of this series is dependent on the location of its poles. Such poles arise from values of s where $(I-G_s)^{-1}$ is singular, that is, values for which 1 is an eigenvalue of 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 homographies, (ii) the initial distribution is uniform $F(x) := x$, then the series of fundamental intervals can be solely expressed with the usual Ruelle operator. 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.

Functional spaces and spectra of transfer operators. In order to develop such spectral properties of G_s, one first needs to make precise the functional spaces to which the G_s operators are applied. Furthermore, we wish to avoid convergence issues, and only consider the case of a finite alphabet. Most of the results can be adapted to the infinite case, as we will show for the continued fraction source. In this finite case, there exists an open domain $\mathcal{V} \subset C$ that contains $\bar{\mathcal{I}}$ and is strictly mapped inside itself by all the inverse branches h of depth 1, and hence by all the branches of arbitrary depth. We denote by $\mathcal{J}$ the intersection of $\mathcal{V}$ with the real axis. The secant mapping $\tilde{H}(u,v)$ used in (18) has a strictly positive real part on $\mathcal{V}\times\mathcal{V}$, and the operator G_s is well defined for any complex s. The G_s operators are then taken to act 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 cloture $\bar{\mathcal{V}}$. 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.

Analytically, the operator G_s acts on an infinite dimensional function space. However, the starting point of the functional analysis side shows that G_s behaves very much like a finite dimensional object.

> **Spectrum.** (P_0)*The operator G_s is bounded, and compact. Its spectrum is discrete with only an accumulation point at 0 and is determined by the spectrum of $\mathcal{G}_s$*
>
> $$Sp\ G_s = \bigcup_{\ell\geq 0}(-1)^\ell\ Sp\ \mathcal{G}_{s+\ell},$$
>
> *where the algebraic multiplicity of λ in G_s equals the sum of the algebraic multiplicities of $(-1)^\ell\lambda$ in each of the operators $\mathcal{G}_{s+\ell}$.*

Dominant spectral properties for real s. When $s=\sigma$ is real, the operator G_s satisfies a strong positivity property related to the Perron-Frobenius theory [19]. For *real* s, the operator G_s has a unique dominant eigenvalue (of largest modulus) $\lambda(s)$. It is positive, has multiplicity 1. In particular, this makes it possible to separate $\lambda(s)$ and a subdominant eigenvalue $\mu(s)$ (one of second largest modulus): $\lambda(s) > |\mu(s)|$. Then, G_s decomposes as

$$G_s = \lambda(s)P_s + N_s,$$

where P_s is the projection over the eigenspace determined by the eigenvector Ψ_s and N_s has a spectral radius strictly smaller than $\lambda(s)$, and more generally G_s^k decomposes as

$$G_s^k = \lambda(s)P_s^k + N_s^k.$$

If F is a function of $B_\infty(\mathcal{V})$, there results, for any $k\geq 1$ and any $(u,v)\in\mathcal{V}\times\mathcal{V}$,

$$G_s^k[F](u,v) = \lambda(s)^k E_s[F]\Psi_s(u,v) + N_s^k[F](u,v),$$

for some continuous linear form $E_s[F]$, the spectral radius of N_s^k being $|\mu(s)|^k$. The dominant eigenfunction $\Psi_s(u,v)$ is strictly positive on $\mathcal{J} \times \mathcal{J}$. Also, if F is strictly positive on $\mathcal{J} \times \mathcal{J}$, $E_s[F]$ is strictly positive, so that one can factor out the first term.

Positivity (P_1) *For any function F of $B_\infty(\mathcal{V})$ positive on $\mathcal{I}$, there exists $M > 1$ for which one has*

$$G_s^k[F](u,v) = \lambda(s)^k E_s[F] \Psi_s(u,v)\ [1 + O(M^{-k})], \tag{23}$$

for any $k \geq 1$.

Such a formula with an exponentially small relative error is the one that is most useful in applications. Dominant spectral objects of G_s and $\mathcal{G}_s$ are closely linked. Both operators have the same dominant eigenvalue, and from (19), one deduces

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

Special properties at $s = 1$ and $s = 2$. The dominant eigenvalue at $s = 1$ equals 1 , the dominant eigenvector satisfies $\Psi_1(0,1) = 1$, and the dominant projector E_1 satisfies

$$E_1[F] = \int_0^1 F(x,x)\,dx, \tag{24}$$

The asymptotic behaviour described in (23) provides expressions of entropy and coincidence probability that involve spectral objects for $s = 1$ and $s = 2$.

Special cases (P_2) *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 $\lambda(2)$.*

Properties of the dominant eigenvalue. By the classical theory of analytic perturbation [16], for s in a sufficiently small neighbourhood of any point σ of the real axis, unicity of the dominant eigenvalue is preserved, so that

Perturbation (P_3) *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.*

The dominant eigenvalue $\lambda(s)$ is alternatively defined by

$$\lambda(s) = \lim_k \left(\sum_{|h|=k} |h(0) - h(1)|^s \right)^{1/k},$$

since the sum on the right is $G_s^k[1](0,1)$ to which (23) applies provided s is real. One deduces

$$\lambda(s+u) \leq (\frac{1}{\delta})^u\ \lambda(s),$$

so that $\lambda(s)$ strictly decreases along the real axis. Furthermore, on the line $\Re(s) = \sigma$, from the triangular inequality, any eigenvalue μ of G_s satisfies $|\mu| \leq \lambda(\sigma)$. In summary:

Maximum property(P_4) *In the half-plane $\Re(s) > \sigma$, any eigenvalue μ of G_s satisfies $|\mu| < \lambda(\sigma)$. On the line $\Re(s) = \sigma$, any eigenvalue μ of G_s satisfies $|\mu| \leq \lambda(\sigma)$.*

Singularities of the quasi–inverse $(I-G_s)^{-1}$. It is necessary to investigate the nature of the series $\Lambda(F,s)$, and especially to locate its poles. From property (P_4), the operator $I-G_s$ is invertible in the plane $\Re(s)>1$. Thus, the series $\Lambda(F,s)$ is analytic there and, from property (P_3), it has a simple pole at $s=1$. We focus on what may take place on the line $\Re(s)=1$ and we consider so-called $e^{i\alpha}$*–particular points*: they are points $1+it$, with $t\neq 0$ for which the spectrum of G_{1+it} contains an eigenvalue equal to $e^{i\alpha}$. The following result, which extends results of [8], [25], [33], gives a characterisation of particular points and describe the three possible types of behaviour.

> **Particular points** (P_5) *The point $1+it$ is $e^{i\alpha}$–particular if and only if all the elementary components of $\mathcal{G}_{1+it}$ admit $e^{i\alpha}$ as a dominant eigenvalue with the same common eigenvector. Moreover, at such a particular point, the operator G_s has a unique dominant eigenvalue equal to $e^{i\alpha}$.*
>
> *The operator G_s may behave in three different ways on the line $\Re(s)=1$:*
>
> (i) *Aperiodic case. There are not particular points, and the operator $I-G_s$ is invertible in the punctured plane $\Re(s)>1, s\neq 1$.*
>
> (ii) *Quasi–periodic case. There are particular points, but not 1–particular points. The operator $(I-G_s)^{-1}$ has an analytic extension to the punctured plane $\Re(s)>1, s\neq 1$.*
>
> (iii) *Periodic case. There are 1– particular points, and they are regularly distributed on the line. The operator $I-G_s$ has poles at these points, but there is a strip in the left on the line $\Re(s)=1$ that is free of poles.*

The terminology comes from the properties of the dominant eigenvalue. We shall see in the sequel that these three cases may occur, mainly for simple sources, as Bernouilli sources or Markov sources. However, a "correlated" source whose branches are not all affine will be aperiodic in the most of the cases.

5 Average–case analysis of general tries.

We now state the main results of the paper. First, we analyse the height, and obtain results that involve the coincidence probability, equal to $\lambda(2)$. Then, we consider the two other parameters (size, path length) whose analysis involve the quasi–inverse. Two alternative methods are used, depending on the aperiodicity of the source, Mellin transform, or Tauberian theorems. The results involve the entropy of the source, equal to the derivative of $s\rightarrow\lambda(s)$ at $s=1$.

Analysis of height. We begin with the formula

$$\pi_k(n)=\prod_{|h|=k}(1+nu_h)e^{-nu_h},$$

and the inequality

$$|\log\pi_k(n)+\frac{n^2}{2}\sum_{|h|=k}u_h^2|\leq\frac{n^3}{2}\sum_{|h|=k}u_h^3, \tag{25}$$

that introduce the *series of fundamental intervals of depth k*

$$\Lambda_k(F,s):=\sum_{|h|=k}u_h^s=\sum_{|h|=k}|F(h(0))-F(h(1))|^s=G_s^k[L^s](0,1),$$

where L is defined in (22) and s is real. For real s, Property (P_1) shows the existence of two constants C_s and $M_s>1$ such that one has

$$\sum_{|h|=k}u_h^s=C_s\,\lambda(s)^k+O(\frac{1}{M_s})^k. \tag{26}$$

Thus, Eq. (25), (26) imply

$$\log \pi_k(n) = -\frac{C_2 n^2}{2}\lambda(2)^k + O(n^2(\frac{1}{M_2})^k + n^3\lambda(3)^k), \tag{27}$$

where the constants $\lambda(2), \lambda(3)$ are the dominant eigenvalues of G_2 and G_3. One thus obtains

$$\lim_{n\to\infty} |\, \pi_k(n) - \exp[-\frac{C_2 n^2}{2}\lambda(2)^k] \,| = 0,$$

where the constant C_2 depends on F and is equal to $e_2[f^2]\,\Psi_2(0,1)$.
For the expected height, Eq. (27) and Eq. (7) leads us to introduce the quantities,

$$\widetilde{\pi}_k(n) = \exp[-\frac{c_2}{2}n^2\lambda(2)^k], \qquad \widetilde{D}(n) = \sum_{k=0}^{\infty}\left[1 - \widetilde{\pi}_k(n)\right], \tag{28}$$

and it can be proved (by splitting the sums) that

$$D(n) = \widetilde{D}(n) + o(1).$$

The Mellin analysis is then similar to the case of digital trees built on binary expansions. The transform of the approximation $\widetilde{D}(n)$ defined in (28) is, for $-2 < \Re(s) < 0$,

$$\widetilde{D}^*(s) = -\frac{1}{2}\left(\frac{c_2}{2}\right)^{-s/2}\frac{\Gamma(s/2)}{1-\lambda(2)^{-s}}.$$

The double pole at $s = 0$ induces a logarithmic term while the imaginary poles at $s = 2ik\pi/\log\lambda(2)$ contribute a periodic term.

Theorem 2 *The expected height of a trie produced by a source $\mathcal{S}$ is, under the Poisson model $\mathcal{P}(n, F)$,*

$$D(n) = \frac{2}{|\log c(\mathcal{S})|}\log n + P_F(n) + \gamma + A_F + o(1),$$

with $P_F(u)$ a continuous periodic function whose frequency only depends on the source. The coefficients of P_F and the constant A_F linearly depend on the square f^2 of the density $f = F'$. Under the same model, the distribution of the height admits a limit distribution of the doubly exponential type,

$$\lim_{n\to\infty}\sup_k \left|\pi_k(n) - \exp[-\frac{D_F n^2}{2}\lambda(2)^k]\right| = 0,$$

where the constant D_F linearly depends on the square f^2 of the density f.

The fluctuations in the expected height are thus of a standard form with Fourier coefficients that are values of the gamma function at regularly spaced points of the imaginary axis.

Analysis of size and path-length. There are two main classes of methods, depending on the aperiodicity of the source. In all the cases, from (P_4) property, the operator $I - G_s$ is invertible in the plane $\Re(s) > 1$. Thus, the series $\Lambda(F, s)$

$$\Lambda(F, s) = (I - G_s)^{-1}[L^s](0, 1)$$

is analytic there and it has a simple pole at $s = 1$. The singular expansion of $\Lambda(F, s)$ at $s = 1$ reads, from Property (P_3) and special properties at $s = 1$,

$$\Lambda(F, s) = \frac{-1}{\lambda'(1)}\frac{1}{s-1} + O(1) \qquad (s \to 1), \tag{29}$$

On the other hand, $\Gamma(s)$ has a unique pole in the strip $-1 \le \Re(s) < 0$, at $s = -1$, and near $s = -1$, one has $\Gamma(s) = -1/(s+1) + O(1)$; it is then sufficient to study the poles of $\Lambda(F, s)$ in the strip $0 < \Re(s) \le 1$.

In the periodic case, $\Lambda(F, s)$ has other poles on $\Re(s) = 1$, but there exists a strip $1 - \alpha < \Re(s) < 1$ $(\alpha > 0)$ free of poles, so that one can apply principles of Mellin analysis. Moreover, the poles are regularly distributed on line $\Re(s) = 1$ and give periodic terms in the asymptotic expansion.

In the two other cases, from Property (P_5), the operator $I - G_s$ is invertible in the punctured plane $\Re(s) \ge 1, s \ne 1$ (aperiodic case), or the operator $(I - G_s)^{-1}$ has an analytic extension near each particular point (quasi–periodic case). In these two cases, the functions

$$P^*(s) = -\Lambda(-s)s\Gamma(s) \qquad S^*(s) = -\Lambda(F, -s)(s+1)\Gamma(s),$$

defined in (10) extend to analytic functions in the punctured plane $\Re(s) \ge 1, s \ne 1$. However, there may be an accumulation of poles in the left of the line $\Re(s) = 1$, and one cannot easily evaluate directly the contribution of these poles in the asymptotic expansion. We thus apply a Tauberian Theorem.

For $\Re(s) > 1$, $\Lambda(F, s)$ admits an integral expression

$$\Lambda(F, s) = s \int_0^\infty A(x) e^{-sx} dx \quad \text{with} \quad A(x) = \sum_{u_h \ge e^{-x}} 1.$$

The properties $(P_3), (P_4)$ of G_s in the punctured plane $\Re(s) \ge 1, s \ne 1$ permit to apply a Tauberian Theorem to $\Lambda(F, s)$ that gives an equivalent form for B,

$$B(y) := A(\log y) = \frac{-1}{\lambda'(1)}\, y\ [1 + \varepsilon(y)] \quad \text{with } \varepsilon(y) \to 0 \text{ for } y \to \infty.$$

By splitting the integral expressions of the expectations $P(n)$ et $S(n)$

$$P(n) := \int_0^\infty [1 - e^{-x} + xe^{-x}]\, B(\frac{n}{x}) dx, \quad S(n) := \int_0^\infty xe^{-x} B(\frac{n}{x}) dx,$$

that involve the function B, one obtains the main term in the asymptotic expansions of $P(n)$ and $S(n)$,

$$P(n) = \frac{-1}{\lambda'(1)}\, n \log n\ [1 + \varepsilon_1(n)] \quad S(n) = \frac{-1}{\lambda'(1)}\, n\ [1 + \varepsilon_2(n)],$$

with $\varepsilon_i(n) \to 0$ for $n \to \infty$. Furthermore, property (P_2) links the entropy and the quantity $\lambda'(1)$, and we state the main results of the analysis of path length and size.

Theorem 3 *The expectation of path length of a random trie produced by a dynamical source $\mathcal{S}$ satisfies, under the Poisson model $\mathcal{P}(n, F)$*

$$P(n) = \frac{1}{h(\mathcal{S})} n \log n + o(n \log n).$$

The main term does not depend on the distribution F. Moreover, in the case when the source is periodic, the remainder term is explicit,

$$P(n) = \frac{1}{h(\mathcal{S})} n \log n + nP_F(n) + A_F n + o(n).$$

Theorem 4 *The expectation of size of a random trie produced by a dynamical source $\mathcal{S}$ satisfies, under the Poisson model $\mathcal{P}(n, F)$*

(i) *in the case when the source is not periodic*

$$S(n) = \frac{1}{h(\mathcal{S})} n + o(n).$$

(ii) *in the case when the source is periodic*

$$S(n) = \frac{1}{h(\mathcal{S})} n + nP_F(n) + o(n).$$

The periodic term depends on the distribution F.

We remark here that the fluctuations for size can have an asymptotically main order, while the the fluctuations for path length always have an asymptotically smaller order.

6 Some important particular cases.

We develop now two main applications: Bernouilli sources, and the continued fraction source. It is also possible to recover easily the classical results in the case of a Markov chain.

Bernouilli sources. The results obtained here are to be compared to the classical results on digital trees built on a finite alphabet $\mathcal{M} = \{1, 2, \ldots, r\}$ where symbol i has probability p_i. In that case, the expected height is known to be asymptotic to

$$\frac{2}{\log K} \log n, \qquad \text{with} \qquad K = \sum_{i=1}^{r} p_i^2$$

the probability that two digits coincide. The expected size and path length under the Poisson model of rate n are known to be approximated by

$$\frac{1}{H} n, \qquad \frac{1}{H} n \log n \qquad \text{with} \qquad H = \sum_{i=1}^{r} p_i \log p_i$$

being the entropy function. Very often, the authors do not take into account the possible periodic fluctuations which intervene in the main term of the size. We give now an example of what happens for a Bernouilli source with three symbols, associated to probability system (p, q, r), $p + q + r = 1$. The standard Ruelle operator associated to the system is

$$\mathcal{G}_s[f](z) := p^s f(pz) + q^s f(p + qz) + r^s f(p + q + rz).$$

The constant function is the dominant eigenvector for all values of s, and $\lambda(s) = p^s + q^s + r^s$ is the dominant eigenvalue. The dominant projector $e_s[f]$ is the integral $\int_0^1 f(t)dt$. Since the constant function is also an eigenvector common to the three components of $\mathcal{G}_s$, it is easy to build Bernouilli sources of the three different types described in Property (P_5).

Periodic case: $(1/3, 1/3, 1/3)$ or $(1/2, 1/4, 1/4)$

Aperiodic case: $(1/2, 1/3, 1/6)$

Quasi–periodic case: (p, p, p^2) with $p = (1/2)(\sqrt{2} - 1)$.

In the quasi–periodic case, the function $\lambda : s \to p^s + q^s + r^s$ is quasi–periodic. As Pollicott [25] and Fayolle et al. [12] remark it , there is an accumulation of s for which $\lambda(s) = 1$ in the left of the line $\Re(s) = 1$. Our Tauberian argument directly shows that the total contribution of these poles gives a term $o(n)$ in the asymptotic expansion of the size, and a term $o(n \log n)$ in the asymptotic expansion of the path length.

Moreover, our techniques permit us to show that the expectation of size is, up to possible small fluctuations, independent of the initial density. This result sharpens the results of Devroye, and extends them to the case of any probabilistic dynamical sources.

The continued fraction source. This example is discussed in [11], in the case when the initial density is uniform. The operator of Ruelle is called in this case the Ruelle–Mayer operator and is defined by

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

for complex s satisfying $\Re(s) > 1/2$.

The entropy of the source is linked to the so–called Lévy's 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 [5], [11]. Here, one has

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

The source is aperiodic, and the poles of $(I - G_s)^{-1}$ are well–known because they intervene in deep mathematical problems. They include all the nontrivial zeros of the Riemann zeta function. The other values s for which $\mathcal{G}_s$ has eigenvalue 1 are related to the eigenvalues of the hyperbolic Laplace operators and they lie on the line $\Re(s) = 1/2$.(see Efrat's paper [7]). But these last values do not however occur as poles of the Dirichlet series $\Lambda(s, Id)$. In the half-plane $\Re(s) > 0$, $\Lambda(s, Id)$ can be represented as

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

where $R(s)$ is analytic in $\Re(s) > 0$. So, when the initial density is uniform, the asymptotic expansion of $S(n)$ and $P(n)$ solely involve the nontrivial zeros of the Riemann zeta function. For an arbitrary density, the same situation is likely to occur but we cannot prove this at the moment.

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

References

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

[2] Bentley, J, and Sedgewick, R. Fast Algorithms for Sorting and Searching Strings. In *Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, January, 1997.

[3] Billingsley, P. *Probability and Measure*, 2nd ed. John Wiley & Sons, 1986.

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

[5] Daudé, H., Flajolet, P., and Vallée, B. An average-case analysis of the Gaussian algorithm for lattice reduction. To appear in *Combinatorics, Probability and Computing*, Oct. 1995. 30 pages.

[6] Devroye, L. A probabilistic analysis of the height of tries and of the complexity of triesort. *Acta Informatica*, vol 21, pp 229-237

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

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

[9] Flajolet, P. On the performance of evaluaion of extendible hashing and trie searching. *Acta Informatica* 20, 1983, 345–369.

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

[11] Flajolet, P., and Vallée, B. Continued fraction Algorithms, Functional operators and Structure constants. Rapport de Recherche de l'Université de Caen, Les Cahiers du GREYC #12, 1996. To appear in *Theoretical Computer Science.*

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

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

[14] Jacquet, P., Régnier, M. Limit distribution for tries parameters. *Lecture Notes in Computer Science* 214, 196–210, 1986.

[15] Jacquet, P. and Szpankowski, W. Analysis of Digital tries with Markovian Dependency. *IEEE Trans. Inform. Theory* Vol. 37, 5, sept. 91, 1470–1475

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

[17] Knuth, D. E. *The Art of Computer Programming,* vol. 3: Sorting and Searching. Addison-Wesley, 1973.

[18] Knuth, D. E. *The Art of Computer Programming,* 2nd ed., vol. 2: Seminumerical Algorithms. Addison-Wesley, 1981.

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

[20] Kirschenhofer, P. and Prodinger, H. Further results on digital search trees. *Theoretical Comput. Sci.* , vol 58, 143–154, 1988.

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

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

[23] Pittel, B. Path in a random digital tree: Limiting distributions. *Adv. Appl. Prob.*, vol 18, 139–155, 1986

[24] Pittel, B. Asymptotical growth of a class of random trees. *Ann. Prob.*, vol. 13, 414–427, 1985.

[25] Pollicott, M. A complex Ruelle–Perron–Frobenius Theorem and two counterexamples. *Ergod. Th. and Dynam. Syst.* (1984) 4,pp 135-146

[26] Régnier, M., and Jacquet, P. New results in the size of tries. *IEEE Trans. Inform. Theory,* vol 35, no.1, 203–205, 1989.

[27] Régnier, M. Trie hashing analysis. Proc. 4th Int.Conf. Data Eng., Los Angeles, CA, 1988, 377–387.

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

[29] Sedgewick, R. *Algorithms.* Addison–Wesley, 1988.

[30] Szpankowski, W. On the height of digital trees and related problem. *Algorithmica* (6) 1991 256–277

[31] Szpankowski, W. Some results on V–ary asymmetric tries *J. Algorithms*, vol. 9, 224–244, 1988.

[32] Vallée, B. Opérateurs de Ruelle-Mayer généralisés et analyse des algorithmes d'Euclide et de Gauss. Rapport de Recherche de l'Université de Caen, Les Cahiers du GREYC # 4, 1995. To appear in *Acta Arithmetica.*

[33] Vallée, B. Dynamique des fractions continues a contraintes périodiques. Rapport de Recherche de l'Université de Caen, Les Cahiers du GREYC #1, 1997.