Size and Path Length of Patricia Tries: Dynamical Sources Context

Jérémie Bourdon
GREYC, Université de Caen, F-14032 Caen, France;
e-mail: bourdon@info.unicaen.fr

Received 8 February 2001; revised 4 April 2001; accepted 1 August 2001

ABSTRACT: Digital trees, such as tries, and Patricia tries are data structures routinely used in a variety of computer and communication applications including dynamic hashing, partial match retrieval, searching and sorting, conflict resolution algorithms for communication broadcast, data compression, and so forth. Here, we consider tries and Patricia tries built from n words emitted by a probabilistic dynamical source. Such sources encompass classical and many models such as memoryless sources and finite Markov chains. The probabilistic behavior of its main parameters, namely, the size and the path length, appears to be determined by some intrinsic characteristics of the source, such as Shannon entropy and entropy-like constants, that depend on the spectral properties of specific transfer operators of Ruelle type.
 Random Struct. Alg., 19, 289–315, 2001

Key Words: Average-case analysis of data structures; information theory; trie; Mellin analysis; dynamical systems; Ruelle operator; functional analysis

1. INTRODUCTION

Tries are an abstract data structure that can be superimposed on a set of words. As an abstract structure, tries are split according to the symbols encountered in words. Consider a fixed alphabet $\Sigma = \{a_1, \dots, a_r\}$ and let $X \subset \Sigma^\infty$ be a finite set

DOI 10.1002/rsa.10015

of infinite words over Σ. The trie associated with X is then defined recursively by the rule

$$\mathrm{Tr}\,(X) = \langle \mathrm{Tr}\,(\underline{T}_{[a_1]}X), \ldots, \mathrm{Tr}\,(\underline{T}_{[a_r]}X)\rangle, \quad \text{where } \underline{T}_{[\alpha]}X := \{x \mid \alpha \cdot x \in X\}.$$

In other words, $\underline{T}_{[\alpha]}X$ is the set of all words following the symbol α. The recursion ends when the associated set X contains zero or one element. The advantage of the trie is that it only maintains the minimal prefix set of symbols that is necessary to distinguish all the elements of X.

Digital trees are a standard data structure for sorting and searching [5, 7, 15, 19], data compression [1, 17, 30, 31], and pattern-matching [13]. The need of efficient storage and transmission of multimedia data [14], and applications to DNA sequencing [13] emphasizes the importance of such data structures. Patricia tries have been introduced in 1968 by Morrison [22]. This structure is a variation of tries that eliminates the waste of space caused by nodes having only one son. This is done by collapsing one-way branches into a single node. This structure finds a number of applications, notably suffix trees. Sedgewick [24] and Knuth [19] describe various techniques for implementing search and insertion using Patricia tries.

The performance of algorithms that use these structures strongly depends on the shape of the underlying trees. The number of internal nodes is proportional to the number of pointers needed to store the data structure, whereas the external path length is related to the number of comparisons during the creation of the trie. The shape itself depends on the way words are generated. In information theory context, the mechanism which produces words is called a *source*. The two simple models of sources are memoryless sources, where symbols are emitted independently, and Markov chains, where the probability of emitting a symbol solely depends on a finite number of previously generated symbols.

The main parameters of Patricia tries have been already studied for the above classical sources. The external path length of Patricia tries has been analyzed by Kirschenhofer, Prodinger, and Szpankowski in [18], Szpankowski in [27], who obtained the moments of the depth and Rais, Jacquet, and Szpankowski in [25] proved the convergence in distribution of the depth for tries and Patricia tries built on memoryless sources. Devroye [4] also has obtained results for the depth of Patricia tries under a probabilistic model on which the keys are i.i.d. random variables with a continuous density f on $[0, 1]$.

However, data on which tries are built often arise from real sources that may involve intricate dependencies between symbols. Here, we adopt the model of dynamical sources introduced by Vallée [28]. This model associates a word $M(x)$ to a real x of $[0, 1]$ and an initial density f on $[0, 1]$. The mechanism can be viewed as a limiting process of consecutive refinements of Markov chains that take into account a higher-level of dependency on the symbols at each step. Consequently, it can describe non-Markovian phenomena where the dependency on past history is *unbounded*. A high-level of generality is thus obtained by the model. This model fits the framework of mixing model as described by Szpankowski in [26].

The size and the path length of standard and hybrid tries have been studied extensively in the context of dynamical sources by Clément [2] and Clément, Flajolet, Vallée [3]. First, we recall their methods and then, we state the new results concerning the size and the path length of Patricia tries. Our probabilistic model is the so-called Bernoulli model of size n denoted by $(\mathcal{B}_n, \mathcal{S})$: it considers all possible sets X of a fixed cardinality n consisting of independent source words of infinite length. We aim to analyze the probabilistic behavior of the size and the path length of a Patricia trie $\mathrm{PaTr}(X)$ when the cardinality n of the set X becomes large.

The analysis of tries mainly involves the prefixes of the words: in a dynamical source, all source words that start with the same prefix w come from a common interval of $[0, 1]$. The probability of such a measure is denoted by p_w. These intervals are called the *fundamental intervals* and their measure is called the *fundamental probability*. In this article, we use Mellin transform and Dirichlet series of fundamental probabilities $\sum_w p_w^s$. Our basic mathematical tool is a generalization of *Ruelle transfer operator* that is used as a "generating operator" of fundamental probabilities. In previous articles, Vallée [28], Clément [2] and Clément, Flajolet, Vallée [3] have introduced successive generalizations of the Ruelle operator, mainly based on a secant (and multisecant) construction, that act on functions of two (or more) variables. Such operators depend on a complex parameter s and suitably generate fundamental intervals and simultaneously several fundamental intervals. Finally, the analysis is performed in a so-called Poisson model but basic depoissonization arguments allow a return to the Bernoulli model. Furthermore, positive properties of the Ruelle operators (for real values of parameter s) entail the existence of dominant spectral objects, in particular, the existence of the *dominant eigenvalue function* $\lambda(s)$ defined in the neighborhood of the real axis.

The analysis of Patricia tries parameters leads us to also consider conditional probabilities, and then more complicated Dirichlet series that involve both fundamental probabilities and conditional probabilities. We thus use an entire family of Ruelle operators.

In the Bernoulli model $(\mathcal{B}_n, \mathcal{S})$ relative to dynamical sources $\mathcal{S}$, the average values of the size $\widehat{S}_P(n)$ and the path length $\widehat{L}_P(n)$ of Patricia tries built over n words have the following asymptotic behavior [1]

$$\widehat{S}_P(n) \approx \frac{1}{h(\mathcal{S})}[1 - C_1(\mathcal{S})]n, \quad \widehat{L}_P(n) - \frac{1}{h(\mathcal{S})} n \log n \approx \left[\frac{\gamma - C_2(\mathcal{S})}{h(\mathcal{S})} + C(\mathcal{S}, f)\right] n.$$

Here, $h(\mathcal{S})$ denotes the entropy of the source $\mathcal{S}$, $C_1(\mathcal{S})$ and $C_2(\mathcal{S})$ are constants depending solely on the mechanism of the source while $C(\mathcal{S}, f)$ is a constant that depends on both the source and the initial density. These results are to be compared with those obtained by Clément, Flajolet, and Vallée in [3] for standard tries in the same model. In the Bernoulli model $(\mathcal{B}_n, \mathcal{S})$ relative to dynamical sources $\mathcal{S}$, the average values of the size $\widehat{S}(n)$ and the path length $\widehat{L}(n)$ of standard tries have the following asymptotic behavior

$$\widehat{S}(n) \approx \frac{1}{h(\mathcal{S})} n, \qquad \widehat{L}(n) - \frac{1}{h(\mathcal{S})} n \log n \approx \left[\frac{\gamma}{h(\mathcal{S})} + C(\mathcal{S}, f)\right] n.$$

Our results exhibit a different asymptotic behavior for Patricia tries and standard tries. They point out some correcting terms, namely C_1 and C_2. The constant C_1

[1] Here, $\approx$ is used for approximately equal, i.e., up to possible fluctuations induced by nonreal poles. The omitted periodic function is of mean zero.

appears in the main term of the asymptotic expansion for the size, while the constant C_2 appears in the second-order term of the asymptotic expansion for the external path length.

The structure of the article is as follows. Section 2 describes specifications of tries and Patricia tries and Section 3 presents the basic algebraic analysis of the additive parameters of tries and Patricia tries. Section 4 introduces the general model of sources and shows that the generalized Ruelle operators generate the adequate Dirichlet series. In Section 5, we come back to the average-case analysis, obtain precise estimates of size and path length, and conclude with examples of memoryless sources, Markov chains, and the continued fraction source.

2. TRIE STRUCTURES AND MODEL OF ANALYSIS

Here, we describe in more detail the trie structure and its compressed version, namely the Patricia trie. In particular, these structures can be built recursively. Then, we present a probabilistic model, namely the Poisson model, that is often used in the study of the expectation of shape parameters for both the structures.

This is our framework: consider an alphabet $\Sigma := \{a_1, a_2, \ldots, a_r\}$ of cardinality r (finite or denumerable) and a source $\mathcal{S}$ which could be of a quite general type[2] and produces infinite words of Σ^∞. Two main operations on infinite words w are useful: the map $\underline{\sigma}$: $\Sigma^\infty \to \Sigma$ that returns the first letter of a word and the shift function $\underline{T}$: $\Sigma^\infty \to \Sigma^\infty$ that returns the first suffix of a word (i.e., the word stripped of its first letter). Then, the function $\underline{T}_{[a]}$ is the restriction of $\underline{T}$ to the set $\underline{\sigma}^{-1}\{a\}$ of words beginning with symbol a and, for a finite prefix $w = a_1, \ldots, a_k$, $\underline{T}_{[w]}$ denotes the composition $\underline{T}_{[a_k]} \circ \underline{T}_{[a_{k-1}]} \circ \cdots \circ \underline{T}_{[a_1]}$.

We deal with the problem of comparing n infinite words independently produced by the same general. It follows that the probabilities p_w that a word begins with a prefix w will play a central rôle in the analysis.

2.1. Trie Structure

With any finite set X of infinite words produced by the same source, we associate a trie, $\mathrm{Tr}(X)$, defined by the following recursive rules:

- (R_0) if $X = \emptyset$, then $\mathrm{Tr}(X)$ is the empty tree,
- (R_1) if $X = \{x\}$ has a cardinality equal to 1, then $\mathrm{Tr}(X)$ consists of a single *leaf node* represented by $\boxed{x}$,
- (R_2) if X has a cardinality of at least 2, then $\mathrm{Tr}(X)$ is an *internal node* represented generically by $\bullet$ to which r subtrees are attached,

$$\mathrm{Tr}(X) = \langle \bullet, \mathrm{Tr}(\underline{T}_{[a_1]}X), \mathrm{Tr}(\underline{T}_{[a_2]}X), \ldots, \mathrm{Tr}(\underline{T}_{[a_r]}X)\rangle.$$

The edge attaching the subtrie $\mathrm{Tr}(\underline{T}_{[a_j]}X)$ is labeled by the symbol a_j.

[2] We describe precisely the model of source in Section 4.

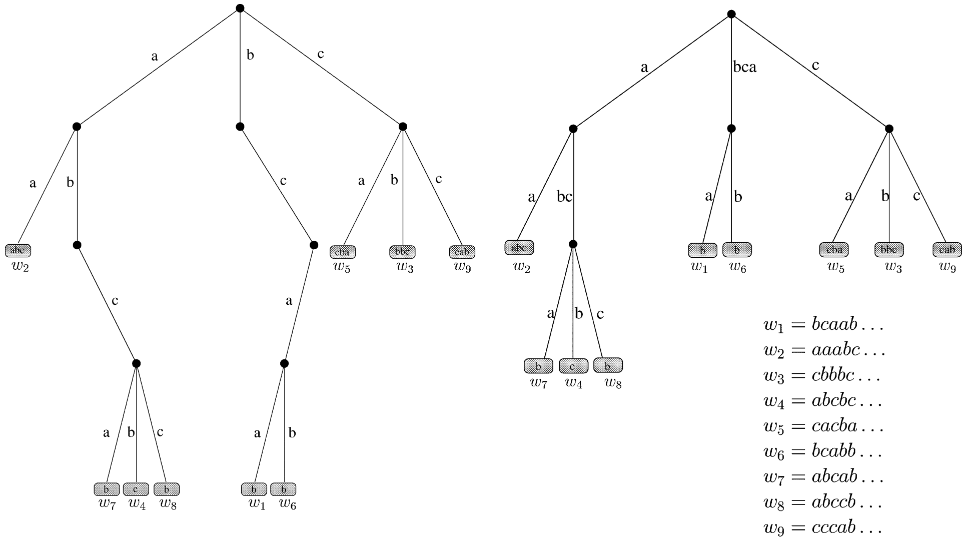


Fig. 1. An example of a ternary trie and its associated Patricia trie built on the set $\{w_1, \ldots, w_9\}$.

Such a tree structure underlies the classical radix sorting methods. It can be built by following recursive rules R_0, R_1, R_2. Its internal nodes are closely linked to prefixes of words of X. More precisely, each internal node of $\mathrm{Tr}(X)$ corresponds to a prefix w that is obtained by concatenating all the labels of the path from the root to the node. Since the node is internal, this prefix is shared by at least two words of X. Figure 1 shows an example of a trie with eight internal nodes that correspond to the prefixes $\{\varepsilon, a, b, c, ab, bc, abc, bca\}$. In the sequel, the probability p_w that an infinite source word begins with prefix w thus plays an important role.

2.2. Patricia Trie Structure

Patricia tries eliminate all internal nodes with only one son, i.e., the nodes where there exist only one distinct symbol in the set $\underline{\sigma}X$. With any finite set X of infinite words produced by the same source, we associate a Patricia trie, $\mathrm{PaTr}(X)$, defined by the following recursive rules:

(R_0) if $X = \emptyset$, then $\mathrm{PaTr}(X)$ is the empty tree,
(R_1) if $X = \{x\}$ has a cardinality equal to 1, then $\mathrm{PaTr}(X)$ consists of a single *leaf node* represented by $\boxed{x}$,
(R'_2) if X has a cardinality of at least 2, two cases must be considered depending on the number of distinct symbols contained in the multiset $\underline{\sigma}X$ that groups all the first symbols of X:
  ($R'_{2,1}$) if $\underline{\sigma}X$ contains only one symbol, then $\mathrm{PaTr}(X)$ equals $\mathrm{PaTr}(TX)$,
  ($R'_{2,2}$) otherwise, if $\underline{\sigma}X$ has at least two distinct symbols, then $\mathrm{PaTr}(X)$ is an *internal node* represented generically by $\bullet$ to which are attached r subtrees,

$$\mathrm{PaTr}(X) = \langle \bullet, \mathrm{PaTr}(\underline{T}_{[a_1]}X), \mathrm{PaTr}(\underline{T}_{[a_2]}X), \ldots, \mathrm{PaTr}(\underline{T}_{[a_r]}X) \rangle.$$

The edges of the Patricia trie are labeled by words. These words are obtained from the associated trie by concatenating all the labels of the collapsed edges.

Figure 1 shows an example of a trie and its associated Patricia trie built on a set of nine words on the alphabet $\{a, b, c\}$. The prefixes used in the trie are $\{aa, abca, abcb, abcc, bcaa, bcab, ca, cb, cc\}$. Each prefix corresponds to a leaf.

2.3. Additive Parameters

Let us consider a tree (a standard trie or a Patricia trie). The *typical depth* of a node in the tree is the number of edges that connects it to the root. The *size* of the tree is the number of its internal nodes. The *path length* of the tree is the sum of the typical depths of all (nonempty) external nodes. These parameters are additive in the sense that they can be evaluated simply by summing over all the internal nodes with a cost function at the nodes. Then the analysis of such parameters is closely linked with the recursive definition of the tree.

In the sequel, the size and the path length of the Patricia tries are considered as parameters of the standard tries with adapted cost functions: the nodes with one-way branches get a cost of zero in the analysis of Patricia tries.

2.4. Bernoulli and Poisson Models

The purpose of an average-case analysis of data structures is to characterize the mean value of their parameters under a well-defined probabilistic model that describes the initial distribution of its inputs. In the present article, we adopt the following general model: we work with a finite set X of infinite words independently produced by the same source $\mathcal{S}$. The cardinality n of the set X is usually fixed and the probabilistic model is then called the Bernoulli model of size n relative to the source $\mathcal{S}$ and denoted by $(\mathcal{B}_n, \mathcal{S})$.

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

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

In this model, N is narrowly concentrated near its mean z with high probability, so that the rate z plays a rôle much similar to the size n in the Bernoulli model. This model is called the Poisson model of rate z relative to the source $\mathcal{S}$ and is denoted by $(\mathcal{P}_z, \mathcal{S})$. Later, we will see that it is possible to go back to the Bernoulli model in which n is fixed by analytic "depoissonization" techniques (see [26]). The Poisson model is of interest because it implies complete independence of events involving what infinite words associated with a set of independent prefixes (i.e., a set that does not contain a word which is the prefix of another word of the set). In particular, if p_w is the probability that a given infinite word begins with prefix w, then the number of infinite words that begin with the prefix w is itself a Poisson variable of rate zp_w. This strong independence property gives access to the analysis of our basic parameters.

3. ALGEBRAIC ANALYSIS OF ADDITIVE PARAMETERS

In the standard trie built on the set $X := (x_1, \ldots, x_n)$, the structure of the node labeled by a prefix w is a finite string fully determined by the prefix w,

$$\underline{\sigma}\ \underline{T}_{[w]}X := (\underline{\sigma}\ \underline{T}_{[w]}x_1, \ldots, \underline{\sigma}\ \underline{T}_{[w]}x_n),$$

where the mapping $\underline{\sigma}$ and $\underline{T}$ are defined in Section 2. This finite string is called a slice.

In the previous example, the slice that corresponds to the prefix a is (a, b, b, b); it is composed of the second letters of the words $\{aaabc, \ldots, abcbc, \ldots, abcab, \ldots, abccb, \ldots\}$.

First, the root of the trie is determined by the slice $\underline{\sigma}X$ and each subtrie is relative to a shifted set $\underline{T}_{[m]}X$. Now consider an additive parameter γ on X defined recursively by the rule

$$\gamma[X] = 0 \text{ if } |X| \leq 1$$
$$\gamma[X] = \delta[\underline{\sigma}X] + \sum_{m\in\Sigma} \gamma[\underline{T}_{[m]}X] \quad \text{if } |X| \geq 2.$$

The parameter δ is sometimes called the "toll" and is defined on finite strings. The recurrence relation can be solved, leading to

$$\gamma[X] = \sum_{w\in\Sigma^*} \delta[\underline{\sigma}\ \underline{T}_{[w]}X],$$

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

Our goal is to study the parameters of Patricia tries in the Bernoulli model. This model leads to intricate and subtle analysis. To simplify it, it is convenient to study Patricia in the Poisson model which replaces the fixed parameter n by the Poisson process N (cf. [26]). First, we analyze Patricia in the Poisson model and then recover the results in the Bernoulli model. The method is called as depoissonization.

We now describe the probabilistic model induced by the Poisson model at each possible node of the trie determined by a prefix w. Recall the probability that a word starts with prefix w is the fundamental measure p_w. When w is already emitted, the probability that the next symbol emitted is m equals

$$p_{[m|w]} = \frac{p_{w\cdot m}}{p_w}.$$

Since all the words of X are independently drawn, at the internal node labeled by w, the symbols of the slice are then emitted by the memoryless source B_w relative to probabilities $\{p_{[m|w]}\}_{m\in\Sigma}$.

Moreover, if the cardinality of X is a random Poisson variable of rate z, then the length of the slice $\sigma\underline{T}_{[w]}X$ is also a random Poisson variable of rate zp_w. It follows that the expectation of parameter γ is a sum of expectations of parameter δ,

$$E[\gamma; \mathcal{P}_z, \mathcal{S}] = \sum_{w\in\Sigma^*} E[\delta; \mathcal{P}_{zp_w}, B_w].$$

3.1. Search Costs at Nodes

Here we consider the additive parameters of interest. We define their corresponding tolls and the independence property of the Poisson model gives access to the evaluation of the toll expectations.

Toll Parameters. First, the toll δ_S equals 1 provided that the slice $\underline{\sigma T_{[w]}} X$ has *at least two* symbols. The toll δ_{PS} associated with the size of the Patricia trie equals 1 provided that the slice $\underline{\sigma T_{[w]}} X$ contains *at least two different* symbols. One has

$$\delta_S(s) = \begin{cases} 1 & \text{if } |s| \geq 2, \\ 0 & \text{otherwise}, \end{cases} \qquad \delta_{PS}(s) = \begin{cases} 1 & \text{if } \#(s) \geq 2, \\ 0 & \text{otherwise}, \end{cases}$$

where $|s|$ and $\#(s)$ denote the number of symbols of s and the number of distinct symbols of s, respectively.

In the same vein, the toll δ_L for the path length of the trie and the toll δ_{PL} for the path length of the Patricia trie are simply

$$\delta_L(s) = \begin{cases} |s| & \text{if } |s| \geq 2, \\ 0 & \text{otherwise}, \end{cases} \qquad \delta_{PL}(s) = \begin{cases} |s| & \text{if } \#(s) \geq 2, \\ 0 & \text{otherwise}. \end{cases}$$

The following result is the key step of the algebraic part of the treatment of additive parameters.

Proposition 1. *Let B be a memoryless source with probabilities $\{p_i\}_{i\in\Sigma}$. Then, in the Poisson model $(\mathcal{P}_z, B)$ of parameter z relative to the source B, the expectations of the toll parameters are*

Size of tries	$E[\delta_S, \mathcal{P}_z, B] = 1 - (1+z)e^{-z}$,
Path length of tries	$E[\delta_L, \mathcal{P}_z, B] = z(1 - e^{-z})$,
Size of PaTries	$E[\delta_{PS}, \mathcal{P}_z, B] = 1 - e^{-z} - \sum_{i\in\Sigma}\left(e^{-z(1-p_i)} - e^{-z}\right)$,
Path length of PaTries	$E[\delta_{PL}, \mathcal{P}_z, B] = z\left(1 - \sum_{i\in\Sigma} p_i e^{-z(1-p_i)}\right)$

Proof. We consider an ordered alphabet $\Sigma = \{a_1, \ldots, a_r\}$. For any set $\mathcal{L} \subseteq \Sigma^*$, the exponential generating function (egf) relative to a parameter δ over $\mathcal{L}$ is defined as

$$F_\delta(z, u, x_1, \ldots, x_r) = \sum_{s\in\mathcal{L}} \frac{z^{|s|}}{|s|!} u^{\delta(s)} x_1^{|s|_1} \cdots x_r^{|s|_r},$$

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

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

$$E[\delta; \mathcal{P}_z, B] = e^{-z}\frac{\partial}{\partial u} F_\delta(z, u, p_1, \ldots, p_r)|_{u=1}. \tag{1}$$

Then, the expressions of the parameters are direct consequences of the independence property of the Poisson process expressed in the generating function framework. The egfs are defined over the set $\mathcal{L} = \Sigma^\star$ consisting of all possible finite strings. The decomposition

$$\Sigma^\star = \varepsilon + \Sigma + \sum_{k\geq 2}\Sigma^k,$$

that corresponds to the three cases of the recursive definition of tries, once translated into egfs yields

$$\begin{aligned}
&(\text{for } \delta = \delta_S) \quad 1 + z(x_1 + \cdots + x_r) + u\big(e^{z(x_1+\cdots+x_r)} - 1 - z(x_1 + \cdots + x_r)\big),\\
&(\text{for } \delta = \delta_L) \quad 1 + z(x_1 + \cdots + x_r) + e^{zu(x_1+\cdots+x_r)} - 1 - zu(x_1 + \cdots + x_r),
\end{aligned}$$

as egfs relative to the parameters δ_S and δ_L of standard tries.

For the Patricia trie parameters, we isolate the case when the slice is of the form $(a, a, \ldots, a)$. The decomposition is now

$$\Sigma^\star = \varepsilon + \Sigma + \sum_{k\geq 2}\sum_{i\in\Sigma}\{i\}^k + \sum_{k\geq 2}[\Sigma^k - \sum_{i\in\Sigma}\{i\}^k].$$

Once translated into egfs, this leads to

$$\begin{aligned}
(\text{for } \delta = \delta_{PS}) \quad & 1 + z(x_1 + \cdots + x_r) + \sum_{i\in\Sigma}(e^{zx_i} - 1 - zx_i)\\
& +u[e^{z(x_1+\cdots+x_r)} - 1 - \sum_{i\in\Sigma}(e^{zx_i} - 1)]\\
& = 1 + \sum_{i\in\Sigma}(e^{zx_i} - 1) + u\left[e^{z(x_1+\cdots+x_r)} - 1 - \sum_{i\in\Sigma}(e^{zx_i} - 1)\right],\\
(\text{for } \delta = \delta_{PL}) \quad & 1 + z(x_1 + \cdots + x_r) + \sum_{i\in\Sigma}(e^{zx_i} - 1 - zx_i)\\
& +e^{zu(x_1+\cdots+x_r)} - 1 - \sum_{i\in\Sigma}(e^{zux_i} - 1)\\
& = \sum_{i\in\Sigma}(e^{zx_i} - 1) + e^{zu(x_1+\cdots+x_r)} - \sum_{i\in\Sigma}(e^{zux_i} - 1),
\end{aligned}$$

as egfs relative to δ_{PS} and δ_{PL}. An application of (1) then gives the results. ∎

3.2. Expectations of Parameters

The expectations of the four parameters can be expressed solely with the fundamental measures.

Proposition 2. *Let $(\mathcal{P}_z, \mathcal{S}, f)$ be the Poisson model of parameter z relative to the source $\mathcal{S}$. Then the expectations of the four parameters of interest are*

Size of tries $$S(z) = \sum_{w\in\Sigma^\star} [1 - (1 + zp_w)e^{-zp_w}],$$

Path length of tries $$L(z) = \sum_{w\in\Sigma^\star} zp_w[1 - e^{-zp_w}],$$

Size of PaTries $$S_P(z) = \sum_{w\in\Sigma^\star} \left[1 - e^{-zp_w} - \sum_{i\in\Sigma}(e^{-zp_w(1-p_{[i|w]})} - e^{-zp_w})\right],$$

Path length of PaTries $$L_P(z) = \sum_{w\in\Sigma^\star} zp_w[1 - \sum_{i\in\Sigma} p_{[i|w]}e^{-zp_w(1-p_{[i|w]})}].$$

Here, $p_{[i|w]}$ denotes the conditional probability $p_{w\cdot i}/p_w$.

We can now return to the Bernoulli model using the principles of "algebraic depoissonization" described in detail by Jacquet and Szpankowski [16]. This principle is mainly based on the equalities

$$E[Y;\mathcal{P}_z] = e^{-z}\sum_{n\geq 0} E[Y;\mathcal{B}_n]\frac{z^n}{n!}, \quad \text{and thus} \quad E[Y;\mathcal{B}_n] = n!\ [z^n]e^z E[Y;\mathcal{P}_z],$$

that relate the expectations of the random variable Y under the Poisson and Bernoulli models.

Proposition 3. *Let $(\mathcal{B}_n, \mathcal{S})$ be the Bernoulli model relative to a probabilistic dynamical source $\mathcal{S}$. Then the expectations of the four parameters of interest are*

Size of tries $$\widehat{S}(n) = \sum_{w\in\Sigma^\star} [1 - (1 + (n-1)p_w)(1-p_w)^{n-1}],$$

Path length of tries $$\widehat{L}(n) = \sum_{w\in\Sigma^\star} np_w[1 - (1-p_w)^{n-1}],$$

Size of PaTries $$\widehat{S_P}(n) = \sum_{w\in\Sigma^\star} [1 - (1-p_w)^n - \sum_{i\in\Sigma}((1-p_w(1-p_{[i|w]}))^n - (1-p_w)^n)],$$

Path length of PaTries $$\widehat{L_P}(n) = \sum_{w\in\Sigma^\star} np_w\left[1 - \sum_{i\in\Sigma} p_{[i|w]} \times (1-p_w(1-p_{[i|w]}))^{n-1}\right].$$

3.3. Mellin Analysis and Dirichlet Series

In the sequel, the analysis is relative to the Poisson model. Standard depoissonization principles enable us to return to the Bernoulli model.

The expressions of average values in the Poisson model belong to the paradigm of harmonic sums (see [8]) that are general sums of the form

$$G(x) = \sum_{w\in\mathcal{U}} \lambda_w\, g(xp_w), \quad \text{for some set } \mathcal{U}. \tag{2}$$

For such sums, the Mellin transform is the appropriate tool to achieve asymptotic analysis when $x \to \infty$. For a function g defined over $]0, +\infty[$, the Mellin transform $g^*(s)$ of g is

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

provided that the integral converges. The largest open strip $\langle\alpha, \beta\rangle$ where the integral converges is called as *fundamental strip*. Since the Mellin transform of $x \mapsto \lambda g(\mu x)$ is $\lambda\mu^{-s}$ times the transform $g^*(s)$ of g, the Mellin transform of G defined in (2) is

$$G^*(s) = g^*(s)\cdot\Delta_{\mathcal{U}}(-s), \quad \text{with } \Delta_{\mathcal{U}}(s) := \sum_{w\in\mathcal{U}} \lambda_w p_w^s.$$

There is a general phenomenon which makes the Mellin transform quite useful. The poles of the Mellin transform are in direct correspondence with the terms in the asymptotic expansion of the original function at ∞ and 0. For the asymptotic evaluation of a harmonic sum $G(x)$, this principle applies provided that the Dirichlet series $\Delta_{\mathcal{U}}(s)$ and the transform $g^*(s)$ are each analytically continuable and are of proper growth. Then, the asymptotic expansion of $G(x)$ when $x \to \infty$ is closely related to the sum of residues right to the fundamental strip. For details about the methodology, we refer to [8].

For parameters of standard tries, the expressions of Proposition 2 show that the analysis involves the so-called Dirichlet series of prefix probabilities

$$\Lambda(s) := \sum_{w\in\Sigma^\star} p_w^s, \tag{3}$$

with functions $g_S(x) = 1-(1+x)e^{-x}$, and $g_L(x) = x(1-e^{-x})$ whose Mellin transforms respectively equal $-(s+1)\Gamma(s)$ and $-\Gamma(s+1)$. The Mellin transforms relative to the parameters of tries are defined on the fundamental strip $\langle -2, -1\rangle$ and equal, respectively

Size of tries $\qquad S^*(s) = -\Lambda(-s)(s+1)\Gamma(s),$

Path length of tries $\qquad L^*(s) = -\Lambda(-s)\Gamma(s+1).$

For the parameters of Patricia tries, the analysis deals with Dirichlet series whose general term involves the expression $(p_w - p_{w\cdot i})^{-s} = p_w^{-s}(1-p_{[i|w]})^{-s}$. More precisely, the Mellin transforms relative to Patricia parameters are defined on the strip

$\langle -2, -1\rangle$ and equal, respectively

Size of PaTries $\qquad S_P^*(s) = \Gamma(s)\Lambda_S(-s)$

$$\text{with } \Lambda_S(s) = -\sum_{w\in\Sigma^\star} p_w^s - \sum_{w\in\Sigma^\star} p_w^s \sum_{i\in\Sigma}[(1-p_{[i|w]})^s - 1],$$

Path length of PaTries $\qquad L_P^*(s) = -\Gamma(s+1)[\Lambda(-s)+\Lambda_L(-s)]$

$$\text{with } \Lambda_L(s) = \sum_{w\in\Sigma^\star} p_w^s \sum_{i\in\Sigma} p_{[i|w]}[(1-p_{[i|w]})^{s-1} - 1].$$

For the sequel, it proves useful to get alternative expressions of both Dirichlet series $\Lambda_S(s)$ and $\Lambda_L(s)$. Using the series expansion of $(1-x)^u$, we obtain two expressions that involve the family $\Lambda^{[m]}(s)$ for $m \geq 1$,

$$\Lambda^{[m]}(s) := \sum_{w\in\Sigma^*} p_w^s \sum_{i\in\Sigma} p_{[i|w]}^m, \tag{4}$$

under the form

$$\Lambda_S(s) = (s-1)\Lambda(s) - s\sum_{m\geq 2}\frac{(-1)^m}{m!}\left(\prod_{i=2}^{m-1}(s-i)\right)[(s-1)\Lambda^{[m]}(s)], \tag{5}$$

$$\Lambda_L(s) = -\sum_{m\geq 2}\frac{(-1)^m}{(m-1)!}\left(\prod_{i=2}^{m-1}(s-i)\right)[(s-1)\Lambda^{[m]}(s)]. \tag{6}$$

Notice that $\Lambda^{[1]}(s)$ coincides with $\Lambda(s)$ defined in (3). The sequel of the analysis is strongly dependent on the set of prefix probabilities. For standard tries, it is sufficient to study the set of probabilities p_w associated to $w \in \Sigma^*$. For Patricia tries, the set of conditional probabilities $p_{[i|w]}$ also plays a fundamental role. Here, we adopt the framework of dynamical sources developed by Vallée in [28] and used by Clément, Flajolet, and Vallée in [3] in their study of standard tries. In this case, the prefix probabilities p_w are expressed with generating operators of the Ruelle type. We generalize their method to generate, at the same time, the conditional probabilities $p_{[i|w]}$.

4. DYNAMICAL SOURCES

Dynamical sources encompass and generalize the two classical models of sources; namely, the memoryless sources and the Markovian sources. They are associated with expanding maps of the interval $[0, 1]$. We refer to [28] for more details. We first recall the definition of such sources and the main properties.

Definition 1. *A dynamical source $\mathcal{S}$ is defined by four elements*:

- **(a)** *an alphabet Σ finite or denumerable,*
- **(b)** *a topological partition of $\mathcal{I} :=]0,1[$ with disjoint open intervals $\mathcal{I}_a$, $a \in \Sigma$,*
- **(c)** *an encoding mapping σ which is constant and equal to a on each $\mathcal{I}_a$,*
- **(d)** *a shift mapping T whose restriction to $\mathcal{I}_a$ is a real analytic bijection from $\mathcal{I}_a$ to $\mathcal{I}$. Let h_a be the local inverse of T restricted to $\mathcal{I}_a$ and $\mathcal{H}$ be the set $\mathcal{H} := \{h_a, a \in \Sigma\}$. There exists a complex neighborhood $\mathcal{V}$ of $\overline{\mathcal{I}}$ on which the set $\mathcal{H}$ satisfies the following*:

(d$_1$) *the mappings h_a extend to holomorphic maps on $\mathscr{V}$, that map $\mathscr{V}$ strictly inside $\mathscr{V}$ (i.e. $h_a(\overline{\mathscr{V}}) \subset \mathscr{V}$),*
(d$_2$) *the mappings $|h'_a|$ extend to holomorphic maps $\tilde{h}_a$ on $\mathscr{V}$ and the supremum $\delta_a := \sup\{|\tilde{h}_a(z)|, z \in \mathscr{V}\}$ satisfies $\delta_a < 1$,*
(d$_3$) *there exists $\mu < 1$ for which the series $\sum_{a\in\Sigma} \delta_a^s$ converges on $\Re(s) > \mu$,*
(d$_4$) *there exists a constant K that bounds the ratio $|h''_a(x)/h'_a(x)|$ for all branch h_a and all $x \in [0, 1]$.*

Remarks. The quantity $\delta := \sup \delta_a$ satisfies $\delta < 1$ and is called the contraction ratio. The condition (d$_4$) is often referred as Rényi's condition and plays an important rôle in the study of conditional probabilities. The word $M(x)$ of Σ^∞ emitted by the source is then formed with the sequence of symbols $\sigma T^j(x)$

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

Notice that the functions σ and T that act on real numbers are related to the functions $\underline{\sigma}$ and $\underline{T}$ that act on words

$$\underline{\sigma} M(x) = \sigma x, \quad \underline{T} M(x) = M(Tx).$$

The mappings $h_w := h_{m_1} \circ h_{m_2} \circ \cdots \circ h_{m_k}$ associated with prefix words $w := m_1 \cdots m_k$ are then the inverse branches of T^k. All the infinite words that begin with the same prefix w correspond to real numbers x that belong to the same fundamental interval $\mathscr{I}_w =]h_w(0), h_w(1)[$. If the unit interval is endowed with a real analytic density f that is strictly positive, then the source is called a *Probabilistic Dynamical Source* and is denoted by $(\mathscr{S}, f)$. In the sequel, we denote by F the distribution associated to the initial density f. This distribution is called the *initial distribution*. The probability p_w that a word begins with prefix w is then the measure of this interval $\mathscr{I}_w$, i.e.,

$$p_w := |F(h_w(0)) - F(h_w(1))|.$$

The fundamental probabilities relative to the uniform density are denoted by p_w^* and are called the *fundamental canonical probabilities*.

4.1. Classical Sources

Here we show that all the classical sources are actually particular instances of dynamical sources. We explain why dynamical sources can be viewed as a limiting process of Markov chains.

Memoryless Sources. All the memoryless sources can be described inside this framework with affine branches. If $(p_a)_{a\in\Sigma}$ is the probability system, then the corresponding topological partition is defined by

$$\mathscr{I}_a :=]q_a; q_{a+1}[, \quad \text{where } q_a := \sum_{i<a} p_i,$$

and the restriction of T on $\mathscr{I}_a$ is the affine mapping with a slope $1/p_a$ on $\mathscr{I}_a$.

Markov Chains. Any Markov chain with a finite alphabet Σ can be associated to a dynamical system with branches that are piecewise affine. The partition $(\mathcal{I}_a)_{a\in\Sigma}$ of $\mathcal{I}$ is induced by the initial probabilities of the source (i.e., $|\mathcal{I}_a| = p_a$).

For Markov chains of order 1, each interval $\mathcal{I}_a$ is divided in r subintervals $(\mathcal{I}_{a.b})_{b\in\Sigma}$ of length $p_{a.b} = p_{[b|a]}p_a$. Thus, the subintervals $(\mathcal{I}_{a.b})_{b,\,a\in\Sigma}$ constitute a topological partition of $[0, 1]$. The restriction of the mapping T on $\mathcal{I}_{a.b}$ is the increasing affine branch that maps $\mathcal{I}_{a.b}$ on $\mathcal{I}_b$ with a slope

$$\frac{p_b}{p_{ab}} = \frac{p_b}{p_{[b|a]}} \cdot \frac{1}{p_a}.$$

More generally, Markov chains of order d are obtained by refining Markov chains of order $d-1$ with a similar method. Each interval $\mathcal{I}_w$ associated to a prefix w of length d is divided in r subintervals $(\mathcal{I}_{w\cdot b})_{b\in\Sigma}$ that correspond to the probability of emitting the symbol b after emission of w. The length of the subinterval $\mathcal{I}_{w\cdot b}$ is $p_{[b|w]}p_w$. The restriction of the mapping T on $\mathcal{I}_{w\cdot b}$ is the increasing affine branch that maps $\mathcal{I}_{w\cdot b}$ on $\mathcal{I}_{\underline{T}w\cdot b}$ with a slope

$$\frac{p_{\underline{T}w\cdot b}}{p_{w\cdot b}} = \frac{p_{[b|\underline{T}w]}}{p_{[b|w]}} \cdot \frac{p_{\underline{T}w}}{p_w}.$$

Then, the Markov chains of order d are relative to dynamical sources whose branches are piecewise affine with r^d pieces. Finally, a dynamical source can be considered as a limiting process of refinement of Markov chains whose order tends to ∞, provided that these successive Markov chains converge in some sense. In this limiting process, there exists an unbounded degree of dependence on symbols and piecewise affine branches become analytic. It is the variation of the derivative $T'(x)$ that keeps memory of the previous history.

Continued Fraction. The continued fraction transformation is an example of a source with unbounded memory. The alphabet is then the set of all positive integers, the topological partition is defined by $\mathcal{I}_a :=]1/(a+1), 1/a[$ and the restriction of T on $\mathcal{I}_a$ is the decreasing linear fractional transformation $Tx := (1/x) - a$. In other words,

$$Tx = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor, \quad \text{and } \sigma x = \left\lfloor \frac{1}{x} \right\rfloor.$$

The Figure 2 presents three examples of sources represented by their shift function T, namely, a memoryless source of probabilities $(\frac{1}{2}, \frac{1}{6}, \frac{1}{3})$, a Markov chain of order 1 on three symbols whose initial probabilities are $(\frac{1}{2}, \frac{1}{6}, \frac{1}{3})$, and the continued fraction expansion.

4.2. Generating Operators

We now wish to "generate" prefix probabilities p_w and conditional probabilities $p_{[i|w]}$ relative to symbol i and prefix w when the source is a dynamical source. The

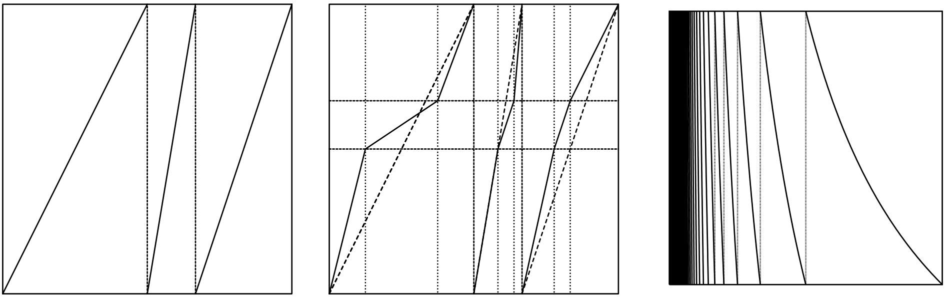


Fig. 2.

basic ingredient, well-developed in dynamical system theory, is the classical Ruelle operator

$$\mathcal{G}_s[f](x) := \sum_{a\in\Sigma} \bar{h}_a(x)^s f \circ h_a(x),$$

which depends on a parameter s and is defined through the analytic extensions $\bar{h}$ of $|h'|$. When $s = 1$, $\mathcal{G}_1[f]$ is a density transformer since if X is a random variable with density f, then the density of TX is $\mathcal{G}_1[f]$. The dynamics of the process is *a priori* described by $s = 1$, but many other properties appear to be dependent upon complex values in the neighborhood of $s = 1$.

However, the classical Ruelle operator is not directly adapted to produce the prefix probabilities p_w, since it cannot generate at the same time both ends of the fundamental intervals. In information theory contexts, Vallée [28] has introduced a new tool, the generalized Ruelle operator, that involves secants of inverse branches

$$H[h](x, y) := \left|\frac{h(x) - h(y)}{x - y}\right|$$

instead of tangents $|h'(x)|$ of inverse branches. Each symbol a produced by source $\mathcal{S}$ is "generated" by a Ruelle operator $\mathbf{G}_{s,[a]}$

$$\mathbf{G}_{s,[a]}[L](x, y) := \widetilde{H}^s[h_a](x, y)L(h_a(x), h_a(y)),$$

that involves the analytic extension $\widetilde{H}[h_a]$ of the secant $H[h_a]$ of the inverse branch h_a. These operators act on functions L of two (complex) variables. A finite word $w := m_1m_2\cdots m_k$ is generated by the composition operator $\mathbf{G}_{s,[w]} := \mathbf{G}_{s,[m_k]} \circ \cdots \circ \mathbf{G}_{s,[m_1]}$. All possible prefixes of length k are then generated by the kth power $\mathbf{G}_s^k$ of the Ruelle operator $\mathbf{G}_s$ relative to source $\mathcal{S}$,

$$\mathbf{G}_s := \sum_{a\in\Sigma} \mathbf{G}_{s,[a]},$$

and all possible prefixes of any finite length are generated by the quasi-inverse $(I - \mathbf{G}_s)^{-1}$ of $\mathbf{G}_s$.

Then, the series $\Lambda(s)$ of prefix probabilities, defined in (3), is expressed by means of the quasi-inverse of $\mathbf{G}_s$

$$\Lambda(s) = (I - \mathbf{G}_s)^{-1}[Q^s](0, 1), \tag{7}$$

where $Q := H[F]$ is the secant of the initial distribution F.

Notice that the operators $\mathbf{G}_s$ constitute an extension of the classical Ruelle operator: if ℓ is the diagonal of L, i.e., $\ell(u) := L(u, u)$, then the operators $\mathbf{G}_s$ and $\mathcal{G}_s$ satisfy

$$\mathbf{G}_s[L](u, u) := \mathcal{G}_s[\ell](u).$$

In the context of Patricia tries, we use an operator that generates at the same time the probability measures for two different depths. Clément in his Ph.D. thesis [2], and Clément, Flajolet, and Vallée [3] have introduced generalized Ruelle operators that involve products of secants instead of secants. Here, we adopt this "multisecant operator" $\mathfrak{G}_s^{[m]}$ to the study of Patricia tries parameters. It involves the "multisecant" $\mathfrak{H}_s^{[m]}[h]$ of the inverse branches h,

$$\mathfrak{H}_s^{[m]}[h](x, y, z, t) := H[h]^s(x, y)\left(\frac{H[h](z, t)}{H[h](x, y)}\right)^m$$
$$= \left|\frac{h(x) - h(y)}{x - y}\right|^{s-m}\left|\frac{h(z) - h(t)}{z - t}\right|^m$$

and acts on functions of four complex variables. It deals with the functional

$$V[h](x, y, z, t) := (h(x), h(y), h(z), h(t)).$$

The operator $\mathfrak{G}_s^{[m]}$ involves all the inverse branches h_a and is defined by

$$\mathfrak{G}_s^{[m]}[L] := \sum_{a\in\Sigma} \mathfrak{H}_s^{[m]}[h_a]L \circ V[h_a].$$

Since conditional probabilities (canonical or general) $p^\star_{[i|w]}$, $p_{[i|w]}$ are expressible with various secants,

$$p^\star_{[i|w]} = p_i^\star \frac{H[h_w](h_i(0), h_i(1))}{H[h_w](0, 1)}, \qquad \frac{p_{[i|w]}}{p^\star_{[i|w]}} = \frac{H[F \circ h_w](h_i(0), h_i(1))}{H[F \circ h_w](0, 1)},$$

the series $\Lambda^{[m]}(s)$ can be expressed by means of the quasi-inverse $(I - \mathfrak{G}_s^{[m]})^{-1}$ of $\mathfrak{G}_s^{[m]}$ applied to the multisecant $\mathfrak{H}_s^{[m]}[F]$ of the initial distribution F, as follows:

$$\Lambda^{[m]}(s) = \sum_{i\in\Sigma} p_i^{\star m}\left(I - \mathfrak{G}_s^{[m]}\right)^{-1}\left[\mathfrak{H}_s^{[m]}[F]\right](0, 1, h_i(0), h_i(1)). \tag{8}$$

Remark that the multisecant operators $\mathfrak{G}_s^{[m]}$ constitute extensions of both the classical Ruelle operators and the generalized operators in two senses: first, if ℓ is the diagonal of L, i.e., $\ell(u) := L(u, u, u, u)$, then the operators $\mathfrak{G}_s^{[m]}$ and $\mathcal{G}_s$ satisfy

$$\mathfrak{G}_s^{[m]}[L](u, u, u, u) := \mathcal{G}_s[\ell](u);$$

second, the operators $\mathfrak{G}_s^{[m]}$ are generalizations of $\mathbf{G}_s$ in the sense that for all m, the multisecant $\mathfrak{H}_s^{[m]}$ generalizes the secant H

$$\mathfrak{H}_s^{[m]}[h](x, y, x, y) = H[h]^s(x, y).$$

4.3. Analytic Properties of Operators.

The Mellin transforms of the mean values of the parameters obtained in Section 3.3 involve Dirichlet series $\Lambda(s)$ and $\Lambda^{[m]}(s)$ defined in (3) and (4). These series are related by (7) and (8) to the quasi-inverses of the operators $\mathbf{G}_s$ and $\mathfrak{G}_s^{[m]}$. First, we recall the main properties of the classical Ruelle operator, and then we extend these properties to the generalized operators. We finally deduce the main analytical properties of their quasi-inverses.

If condition (d) of the definition of Section 4.1 holds, then we can prove the following: for $\Re(s) > \mu$, the Ruelle operator $\mathcal{G}_s$ acts on the Banach space $A_\infty(\mathcal{V})$ formed with all functions f that are holomorphic in the domain $\mathcal{V}$ and are continuous on the closure $\overline{\mathcal{V}}$, endowed with the sup-norm. It is compact and even more nuclear in the sense of Grothendieck [11, 12]. Furthermore, for real values of parameter s, it has positive properties that entail (via theorems of Perron–Frobenius style due to Krasnosel'skij [20]) the existence of dominant spectral objects: there exists a unique dominant eigenvalue $\lambda(s)$ positive, analytic for $s > \mu$, a dominant eigenfunction denoted by ψ_s, and a dominant projector e_s. Under the normalization condition $e_s[\psi_s] = 1$, these last two objects are also unique. Then, compacity entails the existence of a spectral gap between the dominant eigenvalue and the remainder of the spectrum, that separates the operator $\mathcal{G}_s$ in two parts $\mathcal{G}_s = \lambda(s)\mathcal{P}_s + \mathcal{N}_s$, where $\mathcal{P}_s$ is the projection of $\mathcal{G}_s$ onto the dominant eigenspace and involves the dominant spectral objects $\lambda(s)$, ψ_s, and e_s under the form $\mathcal{P}_s[h](x) := \psi_s(x)e_s[h]$; the operator $\mathcal{N}_s$ is relative to the remainder of the spectrum, so that its spectral radius is strictly smaller than the dominant eigenvalue.

For $s = 1$, the classical Ruelle operator is a density transformer, and this property entails explicit values of some spectral objects. In particular, $\lambda(1) = 1$ and $e_1[f] = \int_0^1 f(x)\,dx$. The operator $I - \mathcal{G}_s$ is invertible in the plane $\Re(s) > 1$ and near $s = 1$, the operator $(I - \mathcal{G}_s)^{-1}$ decomposes as

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

so that it has a simple pole at $s = 1$. More precisely, its residue at $s = 1$ satisfies, for a function f positive on $[0, 1]$ and $x \in [0, 1]$,

$$\lim_{s\to 1}(s - 1)(I - \mathcal{G}_s)^{-1}[f](x) = \frac{-1}{\lambda'(1)}\psi_1(x)\int_0^1 f(t)\,dt.$$

Two kinds of situations on the line $\Re(s) = 1$ need to be distinguished depending on the periodicity of the source. A source is said to be periodic if the dominant eigenvalue function $s \to \lambda(s)$ is periodic (that is $\lambda(s + u) = \lambda(s)$ for some u).

- **(i)** In the aperiodic case, the operator $(I - \mathcal{G}_s)^{-1}$ has no other poles on the line $\Re(s) = 1$.
- **(ii)** In the periodic case, the operator $(I - \mathcal{G}_s)^{-1}$ has simple poles on the line $\Re(s) = 1$ that are regularly distributed, and there is a strip on the left of the line $\Re(s) = 1$ that is free of poles.

We now describe the properties of the generalized operators $\mathbf{G}_s$ and $\mathfrak{G}_s^{[m]}$, and we denote by $\mathfrak{G}_s$ one of these possible extensions of $\mathscr{G}_s$. Then the order d of the extension $\mathfrak{G}_s$ is 2 for G_s and 4 for the operators $\mathfrak{G}_s^{[m]}$.

The operator $\mathfrak{G}_s$ acts on the Banach space $\mathfrak{B}_\infty(\mathscr{V})$ formed with all functions L that are holomorphic in the domain $\mathscr{V}^d$ and are continuous in the closure $\overline{\mathscr{V}}^d$, endowed with the sup-norm. The operator is compact and its spectrum is discrete. All the operators $\mathfrak{G}_s$ relative to the same value of parameter s have the same spectrum, denoted by $\mathscr{S}\mathscr{P}(s)$ and the multiplicity of a given eigenvalue in $\mathscr{S}\mathscr{P}(s)$ only depends on the order d of the extension. The dominant eigenvalue $\lambda(s)$ is the same for all the extensions, and positive properties entail the existence of a dominant eigenfunction denoted by Ψ_s, and a dominant projector E_s that are easily related to the spectral objects of $\mathscr{G}_s$, namely, the dominant eigenfunction ψ_s and the dominant projector e_s, via the generalization properties,

$$\Psi_s(u, \dots, u) = \psi_s(u), \quad E_s[L] = e_s[\ell], \quad \text{if } \ell \text{ is the diagonal of } L.$$

The operator $I - \mathfrak{G}_s$ is invertible in the plane $\Re(s) > 1$ and near $s = 1$, the operator $(I - \mathfrak{G}_s)^{-1}$ has a simple pole at $s = 1$. More precisely, its residue at $s = 1$ satisfies, for a function L positive on $[0, 1]^d$ and $x \in [0, 1]^d$

$$\lim_{s \to 1}(s-1)(I - \mathfrak{G}_s)^{-1}[L](x) = \frac{-1}{\lambda'(1)} \Psi_1(x) \int_0^1 \ell(t)\, dt,$$

where ℓ is the diagonal mapping of L.

As previously mentioned, two different situations may happen for the quasi-inverse $(I - \mathfrak{G}_s)^{-1}$ on the line $\Re(s) = 1$, depending on the periodicity of the source.

4.4. Analytic Properties of Dirichlet Series

We now transfer the properties in the previous paragraph to properties of the Dirichlet series $\Lambda^{[m]}(s)$. We then consider analytic properties of Dirichlet series relative to size and path length of Patricia tries.

Each function $\Lambda^{[m]}(s)$ is analytic on the plane $\Re(s) > 1$. At $s = 1$, $\Lambda^{[m]}(s)$ has a pole of order 1, with a residue

$$r_m = \frac{-1}{\lambda'(1)} K^{[m]}(\mathscr{S}) \quad \text{with} \quad K^{[m]}(\mathscr{S}) = \sum_{i \in \Sigma} p_i^{\star m}\ \Psi_1^{[m]}(0, 1, h_i(0), h_i(1)).$$

Here, the derivative $-\lambda'(1)$ coincides with the entropy $h(\mathscr{S})$ of the source. The constant $K^{[m]}(\mathscr{S})$ is related to the dominant eigenfunction $\Psi_1^{[m]}$ of the operator $\mathfrak{G}_1^{[m]}$. The equality (valid for $a, b, c, d \in [0, 1]$)

$$\Psi_1^{[m]}(a, b, c, d) = \lim_{k \to \infty} \left(\mathfrak{G}_1^{[m]}\right)^k [1](a, b, c, d),$$

provides another expression for $K^{[m]}(\mathscr{S})$, that involves the canonical fundamental probabilities $p_w^\star$,

$$K^{[m]}(\mathscr{S}) = \lim_{k \to \infty} \sum_{w \in \Sigma^k} p_w^\star \sum_{i \in \Sigma} \left(p_{[i|w]}^\star\right)^m.$$

Remark that $K^{[m]}(\mathscr{S})$ satisfies the inequality $K^{[m]}(\mathscr{S}) \le 1$. Furthermore, it follows from the equality $K^{[1]}(\mathscr{S}) = 1$ that the singular expansion of $\Lambda(s) = \Lambda^{[1]}(s)$ is of the form

$$\Lambda(s) \asymp \frac{-1}{\lambda'(1)(s-1)} + C(\mathscr{S}, f), \tag{9}$$

where $C(\mathscr{S}, f)$ is a constant depending on the source $\mathscr{S}$ and the initial density f.

At $s = 1$, the Dirichlet series relative to size and path length of Patricia tries satisfy, via Eqs. (5) and (6)

$$\Lambda_S(1) = r_1 - \sum_{m\ge 2} \frac{r_m}{m(m-1)},$$
$$\Lambda_L(1) = -\sum_{m\ge 2} \frac{r_m}{m-1},$$

provided that the series defined in the previous two equations are convergent. Since the inequality $K^{[m]}(\mathscr{S}) \le 1$ holds, the first series is always convergent. However, it is not a priori true for the second series. Here, Rényi's condition (d_4) of definition of dynamical sources provides a general framework where such a result is valid.

Proposition 4. *Dynamical sources satisfy the uniformity condition* (U): *there exists a constant* $\rho < 1$ *such that for all* $w \in \Sigma^\star$, *all* $i \in \Sigma$, *one has* $p_{[i|w]} \le \rho$.

Proof. We essentially use the condition (d_4) together with several applications of the mean value theorem. First,

$$p_{[j|w]} = \frac{p_{w\cdot j}}{p_w} = \frac{|F(h_w(h_j(0))) - F(h_w(h_j(1)))|}{|F(h_w(0)) - F(h_w(1))|}$$
$$= \frac{|F'(c)|}{|F'(d)|}\frac{|h'_w(e)|}{|h'_w(f)|} p_j^\star$$

for some c, d, e, f in $]0, 1[$. Since $F' = f$ is strictly positive and analytic, there exists a constant L that bounds the ratio $f(c)/f(d)$. For any word $w = a_1 \cdots a_n$, the derivative $h'(w)$ of $h_w = h_{a_1} \circ \cdots \circ h_{a_n}$ satisfies

$$h'_w(x) = h'_{a_1}(s_1(x)) \times \cdots \times h'_{a_n}(s_n(x)),$$

with $s_k(x) := h_{a_{k+1}} \circ \cdots \circ h_{a_n}(x)$ for $1 \le i \le n-1$ and $s_n(x) = x$, so that

$$\log \frac{|h'_w(e)|}{|h'_w(f)|} = \sum_{k=1}^{n} (\log h'_{a_k}(s_k(e)) - \log h'_{a_k}(s_k(f)))$$
$$= \sum_{k=1}^{n} \frac{|h''_{a_k}(c_k)|}{|h'_{a_k}(c_k)|} |s_k(e) - s_k(f)| \le K\delta^{n-k}.$$

Here, c_k is a point in $]s_k(e), s_k(f)[$, and the last bound is provided by the conditions (d_2) and (d_4). Finally, there exists a constant $K' := L\exp(K1/(1-\delta)) > 1$ such that for all $w \in \Sigma^*$ and $j \in \Sigma$,

$$\frac{1}{K'} p_j^* \le p_{[j|w]} \le K' p_j^*,$$

$$\text{so that} \quad 1 - p^*_{[i|w]} = \sum_{j\neq i} p^*_{[j|w]} \ge \frac{1}{K'}(1 - p_i^*) \ge \frac{1}{K'}(1 - p^*), \quad \text{with } p^* = \max p_i^*,$$

and the result is thus obtained with $\rho = 1 - (1/K')(1 - p^*)$. ∎

Then, the uniformity condition provides the bound $K^{[m]}(\mathcal{S}) \leq \rho^{m-1}$, so that we can prove the following.

Proposition 5. *For a dynamical source that satisfies Rényi's condition, the two limits*

$$\lim_{k\to\infty} \sum_{w\in\Sigma^k} p^\star_w \sum_{i\in\sigma} (1-p^\star_{[i|w]})|\log(1-p^\star_{[i|w]})|, \quad \lim_{k\to\infty} \sum_{w\in\Sigma^k} p^\star_w \sum_{i\in\Sigma} p^\star_{[i|w]}|\log(1-p^\star_{[i|w]})|$$

exist and define two constants $1 - C_1(\mathcal{S})$ *and* $C_2(\mathcal{S})$ *that can be also expressed with dominant spectral objects of generalized Ruelle operators. Moreover, the two Dirichlet series relative to size and path length of Patricia tries satisfy at* $s = 1$

$$\Lambda_S(1) = \frac{-1}{\lambda'(1)}(1 - C_1(\mathcal{S})), \quad \Lambda_L(1) = \frac{1}{\lambda'(1)} C_2(\mathcal{S}).$$

5. ASYMPTOTIC ANALYSIS OF SIZE AND PATH LENGTH

We can now come back to the analysis of additive parameters of tries. First, we give the main result in the case when the source is a general dynamical source. Then, we sharpen the result in the case of three specific sources: memoryless sources, Markovian sources, and the continued fraction source.

5.1. The Main Result

The singular expansions (9) of the Dirichlet series $\Lambda(s)$ and the expression of $\Lambda_S(1)$, $\Lambda_L(1)$ of Proposition 5 together with the singular expansion of the function $\Gamma(s)$ at $s = 0$ or $s = -1$ provide the singular expansion of the Mellin transforms near $s = -1$. Moreover, under the uniformity condition (U), the Eqs. (5) and (6) define two analytic functions at $s = 1$. In fact, since the spectrum $\mathcal{SP}(s)$ of the operator $\mathfrak{G}_s^{[m]}$ does not depend on m (see [3]), there exists a disk $\mathcal{D}$ where all the functions $(s-1)\Lambda^{[m]}(s)$ are analytic and form a normal family of analytic functions.

Due to the fast decrease of the function $\Gamma(s)$ toward $\pm i\infty$, Mellin analysis applies on the strip $\langle -2, -1\rangle$ and this entails the following expressions for the average values of additive parameters of tries. Finally, basic depoissonization techniques enable us to obtain the asymptotic expressions of the mean values in the Bernoulli model. These formulas involve the entropy $h(\mathcal{S})$ of the source and three constants $C_1(\mathcal{S})$, $C_2(\mathcal{S})$, and $C(\mathcal{S})$. The last constant $C(\mathcal{S})$ depends both on the mechanism of the source and the initial density f. The first three constants $h(\mathcal{S})$, $C_1(\mathcal{S})$, and $C_2(\mathcal{S})$ only depend on the mechanism of the source $\mathcal{S}$ and are expressible by means of dominant spectral objects of the Ruelle operators, or alternatively, as limits that involve canonical fundamental probabilities.

Theorem 1. *Let* $(\mathcal{B}_n, \mathcal{S})$ *be the Bernoulli model of size* n *relative to a dynamical source* $\mathcal{S}$ *with an initial density* f. *The average values of size and path length of tries and Patricia tries involve the entropy* $h(\mathcal{S})$ *of the source and the three constants* $C_1(\mathcal{S})$, $C_2(\mathcal{S})$, *and* $C(\mathcal{S})$

$$h(\mathcal{S}) = -\lambda'(1) = \lim_{k\to\infty} \sum_{w\in\Sigma^k} p^\star_w |\log p^\star_w|,$$

$$C_1(\mathcal{S}) = 1 - \lim_{k\to\infty} \sum_{w\in\Sigma^k} p^\star_w \sum_{i\in\Sigma} (1 - p^\star_{[i|w]})|\log(1 - p^\star_{[i|w]})|,$$

$$C_2(\mathcal{S}) = \lim_{k\to\infty} \sum_{w\in\Sigma^k} p^\star_w \sum_{i\in\Sigma} p^\star_{[i|w]}|\log(1 - p^\star_{[i|w]})|.$$

Two situations arise depending on the periodicity of the source.

(i) *When the source is aperiodic, the expectations of size and path length of tries and Patricia tries are*

$$\widehat{S}(n) = \frac{1}{h(\mathcal{S})}n + o(n), \widehat{S_P}(n) = \frac{1 - C_1(\mathcal{S})}{h(\mathcal{S})}n + o(n),$$

$$\widehat{L}(n) - \frac{1}{h(\mathcal{S})}n \log n = n\left[\frac{\gamma}{h(\mathcal{S})} + C(\mathcal{S}, f)\right] + o(n),$$

$$\widehat{L_P}(n) - \frac{1}{h(\mathcal{S})}n \log n = n\left[\frac{\gamma - C_2(\mathcal{S})}{h(\mathcal{S})} + C(\mathcal{S}, f)\right] + o(n).$$

(ii) *When the source is periodic, the expectations of size and path length of tries and Patricia tries are*

$$\widehat{S}(n) = \frac{1}{h(\mathcal{S})}n[1 + Q(\log(n))] + o(n^{1-\alpha}),$$

$$\widehat{S_P}(n) = \frac{1 - C_1(\mathcal{S})}{h(\mathcal{S})}n + nQ_S(\log(n)) + o(n^{1-\alpha}),$$

$$\widehat{L}(z) - \frac{1}{h(\mathcal{S})}n \log n = n\left[\frac{\gamma}{h(\mathcal{S})} + C(\mathcal{S}, f) + Q(\log(n))\right] + o(n^{1-\alpha}),$$

$$\widehat{L_P}(z) - \frac{1}{h(\mathcal{S})}n \log n = n\left[\frac{\gamma - C_2(\mathcal{S})}{h(\mathcal{S})} + C(\mathcal{S}, f) + Q_P(\log(n))\right] + o(n^{1-\alpha}).$$

The functions $Q(u)$, $Q_S(u)$, and $Q_P(u)$ depend on the source $\mathcal{S}$ and are of very small amplitude; α *is a positive constant, satisfying* $0 < \alpha < 1$, *that is determined by the width of the region of s such that the spectrum* $\mathcal{SP}(s) \cap \{1\}$ *is empty.*

5.2. Memoryless Sources

Memoryless sources are defined in Section 4.2. These are sources built on a finite or infinite alphabet Σ, where symbol m always occurs with probability p_m. The standard Ruelle operator associated to the system is

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

The initial probabilities p_i^* equal p_i and the dominant eigenfunction is the constant function 1.

All the constants that intervene in the expectations of size and path length of standard and Patricia tries are expressible in terms of the probabilities p_i.

Proposition 6. *Consider a memoryless source S with finite or denumerable alphabet* Σ *and probabilities* $\{p_i\}_{i\in\Sigma}$. *The average values of the parameters involve the four* *constants* $h(\mathscr{S})$, $C(\mathscr{S})$, $C_1(\mathscr{S})$, *and* $C_2(\mathscr{S})$

$$h(\mathscr{S}) = \sum_{i\in\Sigma} p_i\,|\log p_i|, \quad C(\mathscr{S}) = \frac{1}{2}\frac{\sum_{i\in\Sigma} p_i \log^2 p_i}{(\sum_{i\in\Sigma} p_i \log p_i)^2},$$

$$C_1(\mathscr{S}) = 1 - \sum_{i\in\Sigma}(1-p_i)\,|\log(1-p_i)|, \quad C_2(\mathscr{S}) = \sum_{i\in\Sigma} p_i\,|\log(1-p_i)|$$

under the form

$$S(n) \approx \frac{1}{h(\mathscr{S})}n, \quad L(n) - \frac{n\log n}{h(\mathscr{S})} \approx \left[C(\mathscr{S}) + \frac{\gamma}{h(\mathscr{S})}\right]n.$$

$$S_P(n) \approx \frac{1-C_1(\mathscr{S})}{h(\mathscr{S})}n, \qquad L_P(n) - \frac{n\log n}{h(\mathscr{S})} \approx \left[C(\mathscr{S}) + \frac{\gamma - C_2(\mathscr{S})}{h(\mathscr{S})}\right]n.$$

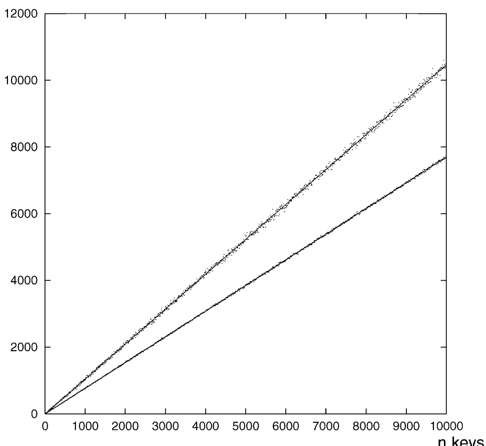
Some fluctuation terms appear whenever the source is periodic.

These expressions for the constants confirm those obtained for unbiased memoryless sources by Rais, Jacquet, and Szpankowski [25].

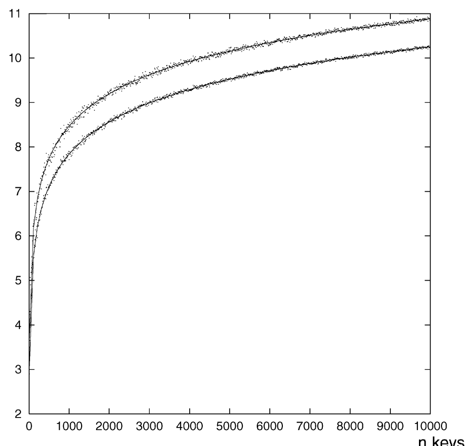
Figure 3 shows a plot of the formula (black curves) compared to simulation results (one dot per trie) for tries (highest curves) and Patricia tries (lowest curves). The experimentation is performed for size and depth (i.e., $L(n)/n$) for tries and Patricia tries built on sets of n words (up to 10000 words) emitted by a memoryless source of probabilities $(\frac{1}{7}, \frac{2}{7}, \frac{4}{7})$.

Remark. Quite often, fluctuations that occur in the case of periodic sources are not taken into account by the authors. Fayolle et al. [6], Pollicott [23], and Vallée [28] are the first authors who consider these fluctuations. We recall that the class of periodic sources includes all unbiased memoryless sources; we give here some other examples of periodic memoryless sources:

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



Size of tries and Patricia tries



Depth of tries and Patricia tries

Fig. 3.

5.3. Markov Chains

Here, the alphabet Σ is finite, of cardinality r, and the words are produced by a Markov chain of order 1, with transition probabilities $p_{[i|j]}$. The matrix Π_s whose general term is defined by $p^s_{[i|j]}$ plays a central role. For $s = 1$, it equals the transition matrix of the Markov chain. The operator $\mathcal{G}_s$ is then a matrix operator $r \times r$ that acts on vectors of functions. The dominant eigenvalue of the operator $\mathcal{G}_s$ is exactly the dominant eigenvalue of the matrix Π_s, and components of the associated eigenfunction are constant. For $s = 1$, the dominant eigenvalue $\lambda(s)$ equals 1 and the components of the normalized dominant eigenfunction correspond to the stationary probabilities.

Proposition 7. *Consider a Markov chain S with a finite alphabet Σ, transition probabilities $\{p_{[i|j]}\}_{i,\, j\in\Sigma}$, and stationary probabilities $\{\pi_k\}_{k\in\Sigma}$. The average values of the parameters involve the three main constants $h(\mathcal{S})$, $C_1(\mathcal{S})$, and $C_2(\mathcal{S})$*

$$h(\mathcal{S}) = -\sum_{k\in\Sigma} \pi_k \sum_{i\in\Sigma} p_{[i|k]}\,|\log p_{[i|k]}|,$$

$$C_1(\mathcal{S}) = 1 - \sum_{k\in\Sigma} \pi_k \sum_{i\in\Sigma} (1 - p_{[i|k]})\,|\log(1 - p_{[i|k]})|,$$

$$C_2(\mathcal{S}) = \sum_{k\in\Sigma} \pi_k \sum_{i\in\Sigma} p_{[i|k]}\,|\log(1 - p_{[i|k]})|$$

under the form

$$S(n) \approx \frac{1}{h(\mathcal{S})} n, \quad L(n) - \frac{n\log n}{h(\mathcal{S})} \approx \left[C(\mathcal{S}) + \frac{\gamma}{h(\mathcal{S})}\right] n,$$

$$S_P(n) \approx \frac{1 - C_1(\mathcal{S})}{h(\mathcal{S})} n, \quad L_P(n) - \frac{n\log n}{h(\mathcal{S})} + \left[C(\mathcal{S}) \approx \frac{\gamma - C_2(\mathcal{S})}{h(\mathcal{S})}\right] n.$$

5.4. The Continued Fraction Source

The continued fraction source is a dynamical source on the infinite alphabet $\mathbf{N}\backslash\{0\}$, where the dependency of past history is unbounded. This source was extensively studied by Flajolet and Vallée in [9] and [10]. In this case, the Ruelle operator is called the Ruelle–Mayer operator,

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

and convergence is granted for any complex s satisfying $\Re(s) > 1/2$. The entropy of the source is related to Lévy's constant that intervenes in the metric theory of continued fractions and the analysis of the Euclidean algorithm [29]. The dominant eigenfunction of $\mathcal{G}_1$, known as Gauss' density, is $1/(\log 2(1 + x))$.

Proposition 8. *Consider the continued fraction source $\mathcal{S}$ with uniform initial density. The asymptotic behavior of parameters for tries and Patricia tries involves the four main constants $h(\mathcal{S})$, $C(\mathcal{S})$, $C_1(\mathcal{S})$, and $C_2(\mathcal{S})$.*

The first two constants admit a closed form: they are Levy's constant and a variation of Porter's constant,

$$h(\mathcal{S}) = \frac{\pi^2}{6\log 2}, \quad C(\mathcal{S}) = 12\frac{\gamma \log 2}{\pi^2} + 9\frac{(\log 2)^2}{\pi^2} - 72\frac{\zeta'(2)\log 2}{\pi^4} - \frac{1}{2}.$$

The other constants involve the function $p_i(x) := (1+x)/((i+x)(i+1+x))$ *under the form*

$$C_1(\mathcal{S}) = 1 - \sum_{i\geq 1}\int_0^1 \frac{1}{\log 2(1+x)}(1-p_i(x))\,|\log(1-p_i(x))|\,dx,$$

$$C_2(\mathcal{S}) = \sum_{i\geq 1}\int_0^1 \frac{1}{\log 2(1+x)} p_i(x)\,|\log(1-p_i(x))|\,dx.$$

Proof. The inverse branch relative to symbol m is a linear fractional transformation (LFT) of the form $h_m(z) := 1/(m+z)$, and it is clear that Rényi's condition holds. For a prefix $w := a_1, \ldots, a_k$ of length k, the inverse branch $h_w := h_{a_1} \circ \cdots \circ h_{a_k}$ is a LFT that can be expressed by means of continuants P_k and Q_k (see [9])

$$h_w(z) = \frac{P_k + P_{k-1}z}{Q_k + Q_{k-1}z}, \qquad \text{with } \det(h_w) := P_k Q_{k-1} - P_{k-1}Q_k = (-1)^k.$$

This entails a nice expression for the fundamental probabilities $p^\star_w = |h_w(0) - h_w(1)|$

$$p^\star_w = \left[Q_k^2\left(1 + \frac{Q_{k-1}}{Q_k}\right)\right]^{-1}, \quad p^\star_{w\cdot i} = \left[Q_k^2\left(i + \frac{Q_{k-1}}{Q_k}\right)\left(1 + i + \frac{Q_{k-1}}{Q_k}\right)\right]^{-1}.$$

Then, the conditional probability $p^\star_{[i|w]}$ only depends on symbol i and rational Q_{k-1}/Q_k whose continued fraction expansion is relative to the mirror $\hat{w} = a_k, \ldots, a_1$ of word w,

$$p^\star_{[i|w]} = \frac{1 + (Q_{k-1}/Q_k)}{(i + (Q_{k-1}/Q_k))(1 + i + (Q_{k-1}/Q_k))}.$$

The classical relation between continuants, i.e., the equality $P_k(w) = Q_{k-1}(\hat{w})$ entails that

$$(I - \mathcal{G}_s)^{-1}[f](0) = \sum_{w\in\Sigma^*} \frac{1}{Q_k^{2s}} f\left(\frac{P_k}{Q_k}\right) = \sum_{w\in\Sigma^*} \frac{1}{Q_k^{2s}} f\left(\frac{Q_{k-1}}{Q_k}\right).$$

Thus, the Dirichlet series $\Lambda(s)$ and $\Lambda^{[m]}(s)$ defined in (7) and (4) are expressible in terms of the Ruelle–Mayer operator $\mathcal{G}_s$

$$\Lambda(s) = (I - \mathcal{G}_s)^{-1}\left[\frac{1}{(1+x)^s}\right](0), \quad \Lambda^{[m]}(s) = (I - \mathcal{G}_s)^{-1}[f_s^{[m]}(x)](0),$$

where the functions $f_s^{[m]}(x)$ are defined by

$$f_s^{[m]}(x) := \frac{1}{(1+x)^s} \sum_{i\geq 1} p_i(x)^m, \qquad \text{with } p_i(x) := \frac{1+x}{(i+x)(i+1+x)}.$$

Note that $p_i(x)$ can be viewed as the probability of emitting symbol i once the infinite word w corresponding to the mirror of $M(x)$ has been emitted. Finally, the constants $1 - C_1(\mathcal{S})$ and $C_2(\mathcal{S})$ are of the form

$$C_1(\mathcal{S}) = 1 - \frac{1}{\log 2}\int_0^1 F_S(t)\,dt, \quad C_2(\mathcal{S}) = \frac{1}{\log 2}\int_0^1 F_L(t)\,dt$$

with

$$F_S(x) = \frac{1}{(1+x)} \sum_{i\geq 1}(1-p_i(x))|\log(1-p_i(x))|,$$

$$F_L(x) = \frac{1}{(1+x)} \sum_{i\geq 1} p_i(x)|\log(1-p_i(x))|.$$

The constant $C(\mathcal{S})$ has been determined by Flajolet and Vallée in [10] in their study of standard tries. They prove that $C(\mathcal{S})$ is a variant of the Porter's constant. ∎

Remark that the expressions of $C_1(\mathcal{S})$ and $C_2(\mathcal{S})$ confirm the general form of Theorem 1. We can get approximation of the constants

$$1 - C_1(\mathcal{S}) \approx 0.87, \quad C_2(\mathcal{S}) \approx 0.276.$$

The first approximation proves that a Patricia trie built on the continued fraction source contains about 13% less nodes than its associated trie.

5.5. Some Open Questions

Our analysis of the path length requires Rényi's condition while the corresponding study of the size does not need this condition.

We ask the following question: *Is it possible that the correcting term appears in the main term of the asymptotic expansion of the path length?* This situation may only occur when the uniformity condition (U) is not fulfilled. We are not aware of any natural sources for which the uniformity condition does not hold.

ACKNOWLEDGMENTS

I thank Brigitte Vallée and Wojciech Szpankowski for their valuable advice which improve the quality of this article and Julien Clément for his useful discussions. I also thank the anonymous referees for their precious remarks.

REFERENCES

[1] T. Bell, I. Witten, and J. Cleary, Modelling for text compression, ACM Computing Surveys 21(4) (1989), 557–591.

[2] J. Clément, Arbres digitaux et sources dynamiques, Thèse de doctorat de l'Université de Caen, 2000.

[3] J. Clément, P. Flajolet, and B. Vallée, Dynamical sources in information theory: A general analysis of trie structures, Algorithmica 29(1) (2001), 307–369.

[4] L. Devroye, A study of trie-like structures under the density model, The Annals of Applied Probability 2(2) (1992), 402–434.

[5] S. Edelkamp, Dictionary automaton in optimal space, Technical Report of the Computer Science Department, University of Freiburg, 29, 1999.

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

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

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

[9] P. Flajolet and B. Vallée, Continued fraction algorithms, functional operators, and structure constants, Theoretical Computer Science 194(1/2) (1998), 1–34.

[10] P. Flajolet and B. Vallée, Continued fraction, comparison algorithms and fine structure constants, Conference Proceedings, Canadian Mathematical Society, 2000, pp. 1–30.

[11] A. Grothendieck, Produit tensoriels topologiques et espaces nucléaires, Memories of the American Mathematical Society, 16 (1955), 1–140.

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

[13] D. Gusfield, Algorithms on strings, trees, and sequences. Computer science and computational biology, vol xviii, Cambridge University Press, Cambridge, (1997), pp. 1–534.

[14] G. Held and T. Marshall, Data and image compression, 4th ed., Wiley, New York, 1996.

[15] P. Jacquet and M. Régnier, Trie partitioning process: limiting distributions, Lecture notes in computer science, vol 214, Springer Verlag, New York, 1986, pp. 196–210.

[16] P. Jacquet and W. Szpankowski, Analytical depoissonization and its applications, Theoretical Computer Science 201(1/2) (1998), 1–62.

[17] P. Jacquet and W. Szpankowski, Asymptotic behavior of the lempel-ziv parsing scheme and digital search trees, Theoretical Computer Science 144(1/2) (1995), 161–197.

[18] P. Kirschenhofer, H. Prodinger, W. Szpankowski, On the balance property of patricia tries: external path length viewpoint, Theoretical Computer Science 68(1) (1989), 1–17.

[19] D.E. Knuth, The art of computer programming volume 3: sorting and searching, 3rd ed., Addison-Wesley, Reading, MA, 1998, pp. 1–736.

[20] M.A. Krasnosel'skij, Positive Solutions of operator equations, P. Noordhoff, Groningen, The Netherlands, 1964, pp. 1–381. [Translated from the Russian by Richard E. Flaherty. Edited by Leo F. Boron].

[21] H.M. Mahmoud, Evolution of random search trees, New York, Series X Wiley-Interscience, 1992, 1–324.

[22] D.R. Morrison, PATRICIA–Practical algorithm to retrieve information coded in alphanumeric, Journal of the ACM 15(4) (1968), 514–534.

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

[24] R. Sedgewick, Algorithms in *C* 3rd ed., Parts 1–4: Fundamentals, Sorting, Searching, and Strings, Addison-Wesley, Reading, MA, 1998, 1–720.

[25] B. Rais, P. Jacquet, and W. Szpankowski, Limiting distribution for the depth in PATRICIA tries, SIAM Journal of Discrete Mathematics 6(2) (1993), 197–213.

[26] W. Szpankowski, Average case analysis of algorithms on sequences, Wiley, New York (2001) 1–576.

[27] W. Szpankowski, Patricia tries again revisited, Journal of the ACM, 37(4) (1990), 691–711.

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

[29] B. Vallée, Digits and continuants in euclidean algorithms. Ergodic versus tauberian theorems, Journal de Théorie des Nombres de Bordeaux 12 (2000), 531–570.

[30] J. Ziv and A. Lempel, A universal algorithm for sequential data compression, IEEE Transactions on Information Theory 23(3) (1977), 337–343.

[31] J. Ziv and A. Lempel, Compression of individual sequences via variable-rate coding, IEEE Transactions on Information Theory 24(5) (1978), 530–536.