Typical Depth of a Digital Search Tree built on a general source

Kanal Hun, Brigitte Vallée

Typical Depth of a Digital Search Tree built on a general source*

Kanal Hun† Brigitte Vallée‡

Abstract

The digital search tree (dst) plays a central role in compression algorithms, of Lempel-Ziv type. This important structure can be viewed as a mixing of a digital structure (the trie) with a binary search tree. Its probabilistic analysis is thus involved, even in the case when the text is produced by a simple source (a memoryless source, or a Markov chain). After the seminal paper of Flajolet and Sedgewick (1986) [11] which deals with the memoryless unbiased case, many papers, due to Drmota, Jacquet, Louchard, Prodinger, Szpankowski, Tang, published between 1990 and 2005, dealt with general memoryless sources or Markov chains, and perform the analysis of the main parameters of dst's–namely, internal path length, profile, typical depth– (see for instance [7, 15, 14]). Here, we are interested in a more realistic analysis, when the words are emitted by a general source, where the emission of symbols may depend on the whole previous history. There exist previous analyses of text algorithms or digital structures that have been performed for general sources, for instance for tries ([3, 2]), or for basic sorting and searching algorithms ([22, 4]). However, the case of digital search trees has not yet been considered, and this is the main subject of the paper. The idea of this study is due to Philippe Flajolet and the first steps of the work were performed with him, during the end of 2010.

1 Introduction

The trie and the digital search tree are two tree structures which contain words (namely, strings or keys) and are used as dictionaries, in compression algorithms for instance. They are important data structures in Computer Science. They are built in a recursive way, and the words are directed towards the various subtrees according to their first symbol. However, in a trie, the words are only placed in the external nodes, and the trie does not depend on the arrival time of the words, whereas the words are placed in the internal nodes of the dst, in a way which depends on their arrival time (See Figure 1). Then, the dst is a more compact structure than the trie; as it can be viewed as an hybrid between the trie and the binary search tree, its analysis is more involved than the trie analysis. The complexity of many algorithms that use these trees as the main underlying data structures can be expressed with various tree parameters, namely, the profile, or the typical depth... A main question is: Is the dst actually more efficient than the trie? One assumes that there is a processus, called a source, which emits infinite words built on a (finite) alphabet. The sequence of words contained in the tree (trie or dst) is formed with n independently chosen words produced by this source, and the analysis aims describing the asymptotic probabilistic behavior of the main tree parameters when the number n of words becomes large. The probabilistic behaviour of these parameters strongly depends on the probabilistic properties of the source which emits the words.

For simple sources, namely memoryless sources (where the symbols are independently drawn) or Markov chains (where the dependency between symbols is bounded), the probabilistic behaviour of the main tree parameters (for tries and dst's) has already been deeply analyzed, even if the analyses are more difficult for the dst. The book of Szpankowski [20] provides a complete review of these results, which are due to a large number of people (already cited in the abstract). They involve various types of simple sources –periodic, aperiodic, diophantine sources– which are defined in Section 3.4 and are summarized as follows:

THEOREM 1.1. [Classical results] *Consider a simple source. The following holds for the depth of the trie or the dst built on a random sequence of n words independently drawn from the source:*

(*a*) *The mean and the variance satisfy*

$$\begin{aligned} \mathbb{E}[D_n] &= \mu \log n + \mu_1 + R_1(n) \\ \mathrm{Var}[D_n] &= \nu \log n + \nu_1 + R_2(n) \end{aligned}$$

The constants μ, ν depend on the source, but not on the type of tree. The only case where $\nu = 0$ arises for an unbiased memoryless source. The constants μ_1, ν_1 depend both on the source and on the type of tree, and the inequality $\mu_1^{(D)} < \mu_1^{(T)}$ holds.

(*b*) *The functions $R_i(n)$ are of the same type for both tries and dst's, and this type depends on the source:*

(*b*1) *If the source is periodic, then*

*Thanks to the Agence Universitaire de la Francophonie (AUF) for the scholarship of K.H, and also to the Agence Nationale de la Recherche for the two projects: ANR Magnum (ANR 2010 BLAN 0204) and ANR Boole (ANR 2009 BLAN 0011)

†GREYC, Université de Caen, 14032 Caen, France.

‡GREYC, CNRS and Université de Caen, 14032 Caen, France.

$R_i(n) = \delta_i(n) + O(n^{-\alpha})$, *for some* $\alpha > 0$*;*
Here, $\delta_i(n)$ *are periodic functions of* $\log n$*.*

($b2$) *If the source is aperiodic diophantine, then* $R_i(n) = O\left(\exp[-(\log n)^{\beta}]\right)$ *for some* $\beta > 0$.

(c) *If the source is not an unbiased memoryless source, the depth* D_n *asymptotically follows a Gaussian law.*

Of course, such simple sources are not realistic, and it is interesting to extend results of this type to more general sources, where the emission of symbols may depend on the whole previous history. Our main result provides an extension of Theorem 1.1 (which only holds for simple sources) to two large classes of sources, the `UNI` Class and the `DIOP` Class, defined in Sections 4.3, and 4.4. We then study two types of sources, and two types of trees (tries and dst's), then we obtain four types of results. The result for tries built on `UNI` sources is already known [2], but the other three types of results are new.

THEOREM 1.2. *Consider a stationary*[1] *source* `UNI` *or* `DIOP`*. The following holds for the depth of the trie or the dst built on a random sequence of* n *words independently drawn from the source:*

(a) *The mean and the variance satisfy*

$$\begin{array}{rcccccc} \mathbb{E}[D_n] & = & \mu \log n & + & \mu_1 & + & R_1(n) \\ \mathrm{Var}[D_n] & = & \nu \log n & + & \nu_1 & + & R_2(n) \end{array}$$

The constants μ, ν *does not depend on the type of tree. They only depend on the source, and can be expressed with the dominant eigenvalue* $\lambda(s)$ *of the source (see Section 4.2 and (8.32)). Here, the inequality* $\nu > 0$ *holds. The constants* μ_1, ν_1 *depend both on the source and on the type of tree, and the inequality* $\mu_1^{(D)} < \mu_1^{(T)}$ *holds.*

(b) *The functions* $R_i(n)$ *are of the same type for both tries and dst's, and this type depends on the source*

($b1$) *If the source is* `UNI`*, then* $R_i(n) = O(n^{-\alpha})$*, for some* $\alpha > 0$

($b2$) *if the source is* `DIOP`*, then* $R_i(n) = O\left(\exp[-(\log n)^{\beta}]\right)$ *for some* $\beta > 0$

(c) *The depth* D_n *asymptotically follows a Gaussian law. The speed of convergence is* $O((\log n)^{-1/2})$ *in the case of a* `UNI` *source.*

Our methods may be also of independent interest, as we provide in Section 2 a new point of view for a general source, where it is possible to study *both* tries and dst's. Even if Section 3 focusses here on the dst analyses, the methods can be applied for trie analyses, and we also obtain new results for tries, when they are built on a `DIOP` source. We explain the similarities of the behaviors of the two structures –tries and dst's–, by the similarities of their Dirichlet series, related to a plain quasi-inverse for the trie, and to an infinite product of quasi-inverses for the dst (see Remark in Section 4.1). We also mention an explicit formula for the mean `dst` depth in Proposition 3.2 which seems to be new, and only known in the memoryless unbiased case, where it is obtained via q-calculus.

Plan of the paper. Section 2 presents the trees and the source, and Section 3 describes the main steps of our method. Then Section 4 focuses on particular sources, where the analytic part of our method can be performed, and leads in Section 5 to the asymptotic gaussian law. Sections 6, 7, 8 are devoted to proofs.

2 Digital Search Trees and Sources

Here we introduce the main actors of the study: first the tree structures (digital search tree and tries), their parameters of interest, together with the main generating functions; second, the mechanism which produces the words, called the source.

2.1 Tree structures. Consider an alphabet Σ, here assumed to be finite and of the form $\Sigma := \{a_1, a_2, \dots a_r\}$. Let $\mathcal{Y}$ be a sequence of infinite words of $\Sigma^{\mathbb{N}}$. Denote by $\mathcal{Y}_{(a_j)}$ the subsequence of $\mathcal{Y}$ formed with the words of $\mathcal{Y}$ which begin with a_j, from which the symbol a_j is removed.

The tree $\mathtt{dst}(\mathcal{Y})$ is defined as follows: If $\mathcal{Y}$ is empty, then $\mathtt{dst}(\mathcal{Y})$ is empty. Otherwise:
– The root of $\mathtt{dst}(\mathcal{Y})$ contains the first word First $(\mathcal{Y})$.
– There are r subtrees built with the sequence $\underline{\mathcal{Y}} := \mathcal{Y} \setminus \{\mathrm{First}(\mathcal{Y})\}$, and the j–th subtree is $\mathtt{dst}\ (\underline{\mathcal{Y}}_{(a_j)})$

The tree $\mathtt{trie}(\mathcal{Y})$ is defined as follows: If $\mathcal{Y}$ is empty, then $\mathtt{trie}(\mathcal{Y})$ is empty; if $\mathcal{Y}$ contains only one word, the tree $\mathtt{trie}(\mathcal{Y})$ is an external node which contains this word. Otherwise:
– The root of $\mathtt{trie}(\mathcal{Y})$ is an internal node.
– There are r subtrees and the j-th subtree is $\mathtt{trie}(\mathcal{Y}_{(a_j)})$.

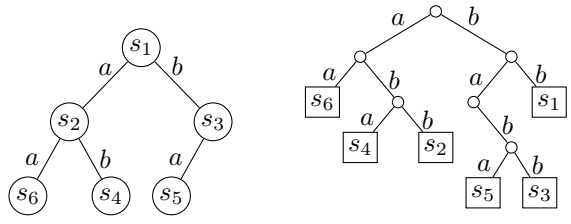

Figure 1: A dst (left) and a trie (right) built from the sequence $s_1 = bbabb\dots; s_2 = abbaa\dots; s_3 = babba\dots, s_4 = ababb\dots; s_5 = babab\dots; s_6 = aaaab\dots;$

[1] The results for the trie hold even if the source is non stationary.

A node which contains a word is said to be full. In a dst, all the nodes are full, whereas, in a trie, only the external nodes are full. When the sequence $\mathcal{Y}$ is composed with n words, the $\mathtt{dst}(\mathcal{Y})$ and the $\mathtt{trie}(\mathcal{Y})$ have exactly n full nodes, and the size equals n. The level of any node is the number of nodes from the root to this node (the level of the root equals 0) and $d_{n,i}$ is the level of the node which contains the i-th word.

The sequence $b_{n,k}$, defined as the number of full nodes at level k in a tree of size n, satisfies

$$b_{n,k} = \sum_{i=1}^{n} [\![d_{n,i} = k]\!],$$

where $[\![.]\!]$ denotes the Iverson bracket. This sequence is called the profile and $B_{n,k} := \mathbb{E}[b_{n,k}]$ is the average profile. The depth, denoted by D_n, is defined as the level of a random full node, via the equalities

$$\Pr[D_n = k] := \frac{1}{n} \sum_{i=1}^{n} \Pr[d_{n,i} = k] = \frac{1}{n} B_{n,k}.$$

This is the main object of the present study.

We mainly use two generating functions of the profile, first the probability generating function $B_n(u)$,

$$B_n(u) := \sum_{k \geq 0} B_{n,k}\, u^k, \tag{2.1}$$

second, the Poisson bivariate generating function $B(z,u)$, together with its normalized version

$$B(z,u) = e^{-z} \sum_{n \geq 0} B_n(u) \frac{z^n}{n!}, \quad \underline{B}(z,u) := \frac{B(z,u) - z}{u - 1}. \tag{2.2}$$

These generating functions are closely related to the probability generating function G_n of the depth D_n,

$$G_n(u) := \mathbb{E}[u^{D_n}] = \sum_{k \geq 0} \Pr[D_n = k]\, u^k = \frac{1}{n} B_n(u). \tag{2.3}$$

2.2 General sources. The probabilistic properties of a digital search tree depend on the probabilistic features of the mechanism which produces the words it contains.

A general source $\mathcal{S}$ built on the alphabet Σ produces at each discrete time $t = 0, t = 1, \ldots$ a symbol from Σ. If X_n is the symbol emitted at time $t = n$, a source produces the infinite word $(X_0, X_1, \ldots X_n, \ldots)$. For any finite prefix $w \in \Sigma^\star$, the probability p_w that a word produced by the source $\mathcal{S}$ begins with the finite prefix w is called the fundamental probability of prefix w. The set $\{p_w, \ \ w \in \Sigma^\star\}$ completely defines the source $\mathcal{S}$.
Such a source $\mathcal{S}$ defines a sequence of "shifted" sources $\mathcal{S}_{(u)}$ (for $u \in \Sigma^\star$), as it is now described: the source $\mathcal{S}_{(u)}$ gathers all the words of $\mathcal{S}$ which begin with $u \in \Sigma^\star$, from which the prefix u is removed or "hidden". The source $\mathcal{S}_{(u)}$ exists as soon as the probability p_u is non zero and is completely defined by all the fundamental (conditional) probabilities p_w/p_u, when w is any finite prefix which begins with u (we denote this situation by the inequality $u \leq w$). In this case, w can be written as $w = u \cdot v$ and the conditional probability $p_w/p_u = p_{u \cdot v}/p_u$ is just the fundamental probability relative to prefix v in the source $S_{(u)}$. It is also denoted as $q_{v|u}$, and we prefer this notation since it shows the dependence with respect to the "visible" prefixes v emitted by the source $\mathcal{S}_{(u)}$.

We associate to the source $\mathcal{S}$ an infinite matrix $\mathbf{P}$, whose rows and columns are indexed by $\Sigma^\star$: The coefficients at the row w which are possibly non-zero are located at the columns $w.i$ (for $i \in \Sigma$) and equal $p_{w.i}/p_w = q_{i|w}$. The related graph admits, as vertices, all the sources $\mathcal{S}_{(u)}$ associated to prefixes u for which $p_u \neq 0$, and there is an edge from $\mathcal{S}_{(u)}$ to $\mathcal{S}_{(v)}$ if and only if ($v = u \cdot i$ for $i \in \Sigma$) and ($q_{i|u} = p_{u \cdot i}/p_u$ is non-zero). An example of the graph associated to the source $\mathcal{S}$ is shown in Figure 2.

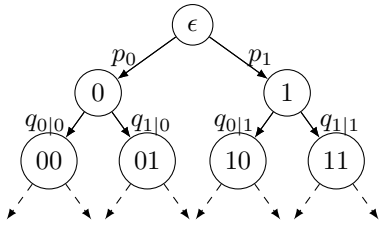

Figure 2: The graph of matrix $\mathbf{P}$ for $\Sigma = \{0, 1\}$.

For any $s \in \mathbb{C}$, the matrix $\mathbf{P}_s$ is obtained from the matrix $\mathbf{P}$ by raising each coefficient to the power s. We denote by $\mathcal{B}(\Sigma^\star)$ the Banach space of the bounded functions $X : \Sigma^\star \to \mathbb{C}$ endowed with the sup-norm. The operator $\mathbf{P}_s$ operates on $\mathcal{B}(\Sigma^\star)$ in a natural way, and transforms a function $X \in \mathcal{B}(\Sigma^\star)$ into a function $Y \in \mathcal{B}(\Sigma^\star)$ as follows:

$$Y(w) := \mathbf{P}_s[X](w) := \sum_{i \in \Sigma} q_{i|w}^s\, X(w \cdot i). \tag{2.4}$$

2.3 Pruning the transition matrix. This representation is quite redundant for simple sources, where the correlations between emitted symbols are "weak". All the sources $\mathcal{S}_{(u)}$ are not needed for the describing the source $\mathcal{S}$, and we define an equivalence relation on sources as

$$\big[\mathcal{S}_{(u)} \equiv \mathcal{S}_{(v)} \quad \Longleftrightarrow \quad \forall w \in \Sigma^\star, \quad q_{w|u} = q_{w|v}\big].$$

The pruned graph is obtained by keeping only one representant in each equivalence class. We describe two instances (See Figure 3.)

For a memoryless source, there is only one equivalence class, and $\mathbf{P}_s$ is a matrix of order 1, with a coefficient $p_1^s + \ldots + p_r^s$, where p_i is the probability of emitting the symbol a_i.

In a Markov chain of order k, there are two types of sources:

– first, the "initial" sources $S_{(u)}$ related to a prefix u of length strictly less than k;

– then, all the sources $S_{(u)}$ related to prefixes u with the same suffix of length k are all equivalent.

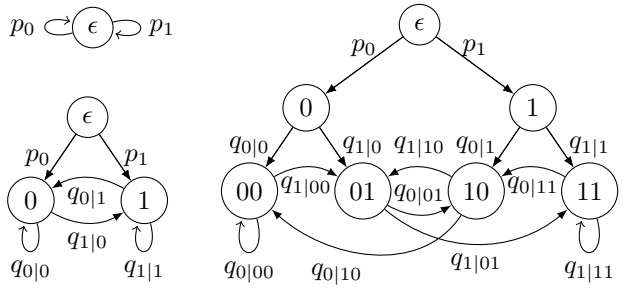

Figure 3: The graph for a memoryless source (left top) – the graph of a Markov chain of order 1, 2.

3 Main principles of our method.

We wish to obtain an alternative expression of the generating function $\underline{B}(z,u)$ of the profile. We first derive in Proposition 3.1 a system of equations. Next, we perform an algebraic study, which first provides in Proposition 3.2 an exact expression for the probability generating function of the depth and introduces a central object in our study, the `dst` Dirichlet series $\Delta(s,u)$, for which we obtain an alternative expression in Proposition 3.3. Finally we explain the main principles for the analytic study, centered on the Rice Formula, which will lead to an asymptotic estimate of the probability generating function of the depth, and finally to the asymptotic gaussian law.

3.1 The basic recurrence and the system of functional equations.

We deal with the sequence of all the sources $\mathcal{S}_{(w)}$ associated to the initial source $\mathcal{S}$. We focus here on the case of $\Sigma = \{0,1\}$.

The variable $b_{n,k}^{(w)}$ is the `dst` profile for the source $\mathcal{S}_{(w)}$, defined as in Section 2.1. It satisfies the basic recurrence,

$$b_{n,k}^{(w)} = b_{K_n,k-1}^{(w.0)} + b_{n-1-K_n,k-1}^{(w.1)}, \qquad (\text{for } n,k \geq 1)$$

where $K_n = K_n^{(w)}$ is the number of nodes in the left subtree. As the variable $K_n^{(w)}$ follows a binomial law of parameters $n-1$ and $q_{0|w}$, the expectations $B_{n,k}^{(w)} := \mathbb{E}[b_{n,k}^{(w)}]$ satisfy the recurrence, for $n,k \geq 1$,

$$B_{n+1,k}^{(w)} = \sum_{j=0}^{n} \binom{n}{j} q_{0|w}^j \, q_{1|w}^{n-j} \left(B_{j,k-1}^{(w\cdot 0)} + B_{n-j,k-1}^{(w\cdot 1)} \right),$$

with $B_{n,0}^{(w)} = 1, B_{0,k}^{(w)} = 0 \quad$ for any $w \in \Sigma^\star, n \geq 1, k \geq 0$.

Dealing with the associated generating functions, we obtain the following result.

Proposition 3.1. *Consider a source $\mathcal{S}$ and its shifted sources $\mathcal{S}_{(w)}$. Then, the modified Poisson generating functions $\underline{B}^{(w)}(z,u)$ of the* `dst` *profiles relative to the sources $S_{(w)}$ and defined as in (2.2) are solutions of the system of functional equations*

$$\frac{d}{dz}\underline{B}^{(w)}(z,u) + \underline{B}^{(w)}(z,u) = z + u \sum_{i \in \Sigma} \underline{B}^{(w\cdot i)}(q_{i|w} z, u).$$

3.2 General strategy for the algebraic study.

This system of functional equations involves three operations: (i) the differentiation d/dz with respect to z; (ii) the change of variables $z \mapsto qz$; (iii) the shift on words $w \mapsto w.i$.

In comparison, the derivation does not occur in the case of tries, (see Section 6.4) and it creates one of the main difficulties in the `dst` analysis. There are two main transforms with which the derivation "disappears": the Laplace transform and the Mellin transforms. This is why we mainly deal with these transforms, together with a third main tool, the Rice formula.

Our method is composed with three main steps, each dedicated to the use of one of the three main tools. We limit ourselves to a smooth source.

Definition 3.1. *A source is smooth if all the fundamental probabilities p_w are strictly positive and if there exists $p < 1$ for which $q_{i|w} \leq p$ for any $(i,w) \in \Sigma \times \Sigma^\star$.*

(a) We first use the Laplace transform, as in [9], which provides, in the case when the source is smooth, an exact expression, first for all the bivariate series $\underline{B}^{(w)}(z,u)$, then for the series $B_n^{(w)}(u)$ (See Sections 6.1 and 6.2 for the proofs).

Proposition 3.2. *Let $\mathcal{S}$ be a p–smooth source with fundamental probabilities p_w. The* `dst` *bivariate Dirichlet series, defined as*

$$\Delta(s,u) := \sum_{v \in \Sigma^\star} \delta(v,u) p_v^s, \tag{3.5}$$

$$\text{with} \quad \delta(v,u) := \frac{1}{p_v} \sum_{w \geq v} u^{|w|} p_w \prod_{\substack{\alpha \in [\epsilon, w] \\ \alpha \neq v}} \frac{1}{1 - p_v p_\alpha^{-1}},$$

exists for $\Re s > 1$ and $|u| \le 1$. Moreover, the probability generating functions of the profile and typical depth admit the following expression

$$\frac{B_n(u)-n}{u-1} = \sum_{\ell=2}^{n} (-1)^\ell \binom{n}{\ell} \Delta(\ell, u), \tag{3.6}$$

$$\frac{\mathbb{E}[u^{D_n}]-1}{u-1} = \frac{1}{n} \sum_{\ell=2}^{n} (-1)^\ell \binom{n}{\ell} \Delta(\ell, u),$$

(*b*) As these exact expressions are binomial sums, we use the Rice formula, described in Section 3.3, which transforms a binomial expression into an integral on a vertical line of the complex plane. If we know the "tameness" of the series $s \mapsto \Delta(s,u)$, namely, its behaviour when $(\Re s, u)$ is close to (1, 1), it is possible to shift the contour of the integral to the left, and obtain asymptotic estimates for $B_n(u)$ and $\mathbb{E}[u^{D_n}]$.

(*c*) Via the Mellin transform of the series $z \mapsto B(z,u)$, we obtain, as in [15], an alternative expression for $\Delta(s,u)$, that will be central in the study of its tameness (See Section 6.3 for a proof).

PROPOSITION 3.3. *Consider a p–smooth source $\mathcal{S}$ and its generalized matrix $\mathbf{P}_s$. Then, the infinite product*

$$\mathbf{Q}(s,u) := (I - u\mathbf{P}_s)^{-1} \cdot \ldots \cdot \ldots (I - u\mathbf{P}_{s+k})^{-1} \ldots,$$

exists for $\Re s > 1$ and $|u| \le 1$. Moreover, the `dst` *Dirichlet series admits the following alternative expression which involves these infinite products,*

$$\Delta(s,u) = {}^t\mathbf{E}\, \mathbf{Q}(s,u) \cdot \mathbf{Q}(2,u)^{-1}\, \mathbf{1}, \tag{3.7}$$

where $\mathbf{1}$ is the vector whose all the components equal 1 and ${}^t\mathbf{E}$ equals $(1, 0, \ldots)$.

3.3 General strategy for the analytic study. The Rice Formula [16, 17] transforms a binomial sum into an integral in the complex plane. One has

$$n \cdot \frac{\mathbb{E}[u^{D_n}]-1}{u-1} = \sum_{\ell=2}^{n} (-1)^\ell \binom{n}{\ell} \Delta(\ell, u) = \frac{(-1)^{n+1}}{2i\pi} \int_{\Re s = \sigma_1} L_n(s,u)\, ds, \tag{3.8}$$

for any real $\sigma_1 \in]1, 2[$,

$$\text{with} \quad L_n(s,u) = \frac{n!\, \Delta(s,u)}{s(s-1)\ldots(s-n)}.$$

Then, along general principles in analytic combinatorics described in[12, 13], the integration line can be pushed to the left, as soon as $L_n(s,u)$ –closely related to $\Delta(s,u)$– has good analytic properties: we need a region $\mathcal{R}$ on the left of $\Re s = 1$, where $\Delta(s,u)$ is meromorphic, with polynomial growth (for $|\Im s| \to \infty$). We finally obtain a residue formula

$$(-1)^{n+1} n \cdot \frac{\mathbb{E}[u^{d_n}]-1}{u-1} = \sum_{s} \operatorname{Res}\left[L_n(s,u)\right] + \frac{1}{2i\pi} \int_{\mathcal{C}_2} L_n(s,u)\, ds, \tag{3.9}$$

where $\mathcal{C}_2$ is a curve of class $\mathcal{C}^1$ enclosed in $\mathcal{R}$ and the sum is extended to all poles s of $L_n(s,u)$ inside the domain delimited by the vertical line $\Re s = \sigma_1$ and the curve $\mathcal{C}_2$.

The first term in (3.9) provides the asymptotic behaviour, and the remainder integral is estimated using the polynomial growth of $s \mapsto \Delta(s,u)$ for $|\Im(s)| \to \infty$. As Proposition 3.3 shows that $\Delta(s,u)$ involves quasi-inverses $(I - u\mathbf{P}_s)^{-1}$, its behaviour is dictated by the spectral properties of the operator $\mathbf{P}_s$.

3.4 Case of simple sources. Here, the operator $\mathbf{P}_s$ is a matrix of finite order. When s is real, and the source p-smooth, the matrix $\mathbf{P}_s$ has all its coefficients strictly positive, and it has a unique dominant eigenvalue, denoted by $\lambda(s)$. (In the memoryless case, the equality $\lambda(s) = p_1^s + \ldots + p_r^s$ holds and involves the probability p_i of emitting the symbol a_i). We consider the set $\mathcal{Z}$ of complex numbers s for which the spectrum $\operatorname{Sp}\mathbf{P}_s$ contains 1. This set always contains $s = 1$ and there are two main situations for the intersection $\mathcal{Z} \cap \{\Re s = 1\}$:

(*P*) The intersection $\mathcal{Z} \cap \{\Re s = 1\}$ contains another point, distinct from $s = 1$. This happens iff, for any pair of cycles $\mathcal{C}, \mathcal{K}$, of respective probabilities $p(\mathcal{C}), p(\mathcal{K})$, all the ratios $\log p(\mathcal{C})/\log p(\mathcal{K})$ are rational numbers. Then, the intersection $\mathcal{Z} \cap \{\Re s = 1\}$ is a set of the form $\{s_k = 1 + kit_0, \quad k \in \mathbb{Z}\}$, for some $t_0 > 0$, and the mapping $s \mapsto \operatorname{Sp}\mathbf{P}_s$ is periodic of period it_0. In this case, the source itself is said to be periodic.

(*A*) The intersection $\mathcal{Z} \cap \{\Re s = 1\}$ is reduced to $s = 1$, and the source is said to be aperiodic. The position of $\mathcal{Z}$ with respect to the line $\Re s = 1$ depends on the arithmetic properties of the ratios $\log p(\mathcal{C})/\log p(\mathcal{K})$, as we now explain.

An irrational number x is diophantine if its irrationality exponent[2] is finite. A simple aperiodic source is diophantine if there exists a ratio $\log p(\mathcal{C})/\log p(\mathcal{K})$ which

[2] We recall that the irrationality exponent of a irrational number x is defined by $\omega(x) :=$
$\sup\left\{\alpha, \left|x - \frac{p}{q}\right| \le \frac{1}{q^{2+\alpha}} \quad \text{for an infinite number of pairs } (p,q)\right\}$.

is diophantine. In this case, the distance between $\mathcal{Z}_t := \mathcal{Z} \cap \{0 < t_0 \leq |\Im s| \leq t\}$ and the vertical line $\Re s = 1$ is $\Omega(|t|^{-\beta})$ (where β is related with the irrationality exponent, see [10]) and there exists an hyperbolic region, of the form $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$, with

$$\mathcal{R}_1 := \left\{ s = \sigma + it;\ |t| \geq B,\ \sigma > 1 - \frac{A}{|t|^\beta} \right\} \tag{3.10}$$

$$\mathcal{R}_2 := \left\{ s = \sigma + it;\ |t| \leq B, \sigma > 1 - \frac{A}{B^\beta} \right\},$$

where $(I - \mathbf{P}_s)^{-1}$ is meromorphic, admits $s = 1$ as the only pole, and $||(I - \mathbf{P}_s)^{-1}||$ is $O(|t|^\beta)$ when $|\Im s| \to \infty$.

3.5 Tameness. For general sources, the notion of tameness is introduced to "copy" the behaviour of simple aperiodic sources. However, the functional space on which $\mathbf{P}_s$ acts has to be made precise. In an informal way, a source is said to be tame if there exist

(a) a space $\mathcal{F}$ on which the operator $\mathbf{P}_s$ acts,

(b) a region $\mathcal{R}$ located on the left of the line $\Re s = 1$,

such that the quasi-inverse $(I - \mathbf{P}_s)^{-1}$ fulfills two main properties :

(i) it is meromorphic on $\mathcal{R}$ with a unique pole at $s = 1$,

(ii) it is of polynomial growth on $\mathcal{R}$ (when $|\Im s| \to \infty$).

For distributional studies, we need a reinforcement of this notion, in fact a "uniform perturbation" of it, as we deal with quasi-inverses $(I - u\mathbf{P}_s)^{-1}$, with u close to 1. This will lead to the notion of uniform tameness.

We will see that, for general sources, the region $\mathcal{R}$ is not always hyperbolic, but may be a vertical strip for sources that are "quite different" from simple sources.

4 Extension of regular sources, operators, and tameness

We introduce a general framework where sources can be proven tame. We first explain how a smooth stationary source can be extended into a dynamical system. Then, the secant transfer operator $\mathbb{H}_s$ provides a convenient extension of the transition operator $\mathbf{P}_s$, and Proposition 4.1 provides an extension of Proposition 3.3, where the `dst` Dirichlet series is now expressed in terms of the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$. Next, we consider two classes of dynamical sources, the `UNI` Class, and the `DIOP` Class for which the quasi-inverses are tame. It will be then possible to apply the Rice formula to obtain the asymptotic gaussian law in Section 5.

4.1 Extension of sources and operators. There are three main steps:

Step 1. First, we consider the mirror operation which reverses the finite prefixes and then the (finite) past. This transforms the operator $\mathbf{P}_s$ into the operator $\widehat{\mathbf{P}}_s$

Step 2. Second, if the source is regular enough, it can be extended into a source which possesses an infinite past. Moreover, there exists a unique invariant distribution under the shift "towards the past".

Step 3. With a convenient parameterization, the reverse past of a stationary source leads to a dynamical system, with surjective branches [see Section 7.3]. When it is of class $\mathcal{C}^2$, its secant transfer operator $\mathbb{H}_s$ provides a good extension of $\widehat{\mathbf{P}}_s$.

Finally, we obtain the following result, proven in Section 7.

PROPOSITION 4.1. *Consider a smooth stationary source, whose reverse past leads to a complete dynamical system $(\mathcal{I}, T)$ of the unit interval, of class $\mathcal{C}^2$. Denote by $\mathcal{H}$ the set of the inverse branches of T. Consider the infinite product*

$$\mathbb{K}(s,u) := (I - u\mathbb{H}_s)^{-1} \circ \cdots \circ (I - u\mathbb{H}_{s+2})^{-1} \circ \ldots \tag{4.11}$$

defined with the secant operator $\mathbb{H}_s$ of the dynamical system,

$$\mathbb{H}_s[F](x,y) := \sum_{h \in \mathcal{H}} S^s[h](x,y)\, F(h(x), h(y)), \tag{4.12}$$

which involves the secants

$$S[h](x,y) := \left| \frac{h(x) - h(y)}{x - y} \right| \tag{4.13}$$

of the inverse branches $h \in \mathcal{H}$. Then, the `dst` *Dirichlet series $\Delta(s,u)$ admits an alternative expression, as an infinite product,*

$$\Delta(s,u) = (I - u\mathbb{H}_s)^{-1} \circ \mathbb{R}(s,u)[1](0,1) \tag{4.14}$$

$$\text{with} \quad \mathbb{R}(s,u) = \mathbb{K}(s+1,u) \circ \mathbb{K}(2,u)^{-1}.$$

Remark. For $s = 1$, the factor $\mathbb{R}(1,u)$ "disappears". As the `trie` Dirichlet series $\Lambda(s,u)$ is $s(I - u\mathbb{H}_s)^{-1}[1](0,1)$ (see Section 6.4), this explains the similarity between the betaviour of the two trees.

4.2 The `Good` Class. Since the source is both of class $\mathcal{C}^2$ and smooth, the shift T is expansive, and the source belongs to the so–called `Good` Class, introduced in [21]. Then, the secant operator $\mathbb{H}_s$ acts on the functional space $\mathcal{C}^1([0,1]^2))$, with dominant spectral properties when s is close to the real axis, namely a dominant eigenvalue $\lambda(s)$, together with a spectral gap. Moreover, as Dolgoyat explains it in [5, 6], it is convenient to deal with the operator $\mathbb{H}_s$ via a norm $||\cdot||_{(1,t)}$ which depends on $t := \Im s$, defined as $||F||_{(1,t)} = ||F||_0 + (1/|t|)||DF||_0$, where $||F||_0$ is the sup-norm.

Then, for $(\Re s, u)$ close to (1, 1), the following holds for the operators of (4.14):

(a) For (s, u) close to (1, 1) the operator $s \mapsto (I - u\mathbb{H}_s)^{-1}$ is meromorphic and admits a simple pôle at $s = 1 + \sigma(u)$, defined by the relation $\lambda(1 + \sigma(u)) = 1/u$.

(b) For $(\Re s, u)$ close to (1, 1), the norm $(1, t)$ of the infinite product $\mathbb{R}(s, u)$ is uniformly bounded.

(c) It remains to deal with the norm of the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$ when $(\Re s, u)$ is close to (1, 1), and when $|\Im s| \to \infty$. We wish to obtain a polynomial growth there.

We define two (large) subclasses of the `Good` Class –the `UNI` Class, the `DIOP` Class– for which such a (uniform) polynomial growth of the norm $||(I - u\mathbb{H}_s)^{-1}||_{(1,t)}$ can be proven. This will entail (uniform) tameness properties for the function $\Delta(s, u)$.

4.3 The `UNI` Class. The `UNI` Condition is a geometric condition, studied by Dolgopyat [5] which expresses that the dynamical system is very different from a system with affine branches. First, one defines $\rho(h, k)$ as a measure of the difference between the "shape" of the two branches h, k of $\mathcal{H}^n$,

$$\rho(h,k) = \inf_{x \in \mathcal{I}} |\Psi'_{h,k}(x)| \quad \text{with} \quad \Psi_{h,k}(x) = \log \left| \frac{h'(x)}{k'(x)} \right| .$$

Then, one considers the "natural" probability $\Pr_n$ defined on each set $\mathcal{H}^n \times \mathcal{H}^n$, by $\Pr_n\{(h, k)\} := |h(\mathcal{I})| \cdot |k(\mathcal{I})|$, where $|\mathcal{J}|$ denotes the length of the interval $\mathcal{J}$. The condition `UNI` expresses that the distance ρ is "not too often too small":

Definition 4.1. [Condition `UNI`]. *A p-smooth dynamical system $(\mathcal{I}, T)$ of class $\mathcal{C}^2$ is of* `UNI` *type if there exists $K > 0$ such that, for any q with $p < q < 1$, for any integer n, one has $\Pr_n[\ \rho \leq q^n] \leq K q^n$.*

The "distance" ρ is always zero for a simple source, and such a source never satisfies the Condition `UNI`. The Condition `UNI` is sufficient to imply *uniform* tameness *in a vertical strip*. This was proven by Dolgopyat [5], rewritten by Baladi-Vallée[1], and extended to the secant operator by Cesaratto-Vallée [2].

Theorem 4.1. [Dolgopyat, Baladi-Vallée, Cesaratto-Vallée] *For a source of* `UNI` *type, there exists a complex neighborhood $\mathcal{U}$ of $u = 1$ and a vertical strip $\mathcal{R}$ on the left of the vertical line $\Re s = 1$, such that, for any $u \in \mathcal{U}$,*

(a) The operator $s \mapsto (I - u\mathbb{H}_s)^{-1}$ is meromorphic in $\mathcal{R}$, with a unique pole at $s = 1 + \sigma(u)$

(b) The norm $||(I - u\mathbb{H}_s)^{-1}||_{(1,t)}$ is of polynomial growth when $|\Im s| \to \infty$, uniform with respect to u.

4.4 The `DIOP` classes. The `DIOP` Conditions are arithmetic conditions which "copy" the definition of simple diophantine sources, and extends it to a general source. For an inverse branch h, one denotes by $h^\star$ its unique fixed point (such a point exists and is unique for a system of the Good Class), by $p(h)$ its depth, and one lets, for h, k, ℓ in $\mathcal{H}^\star$,

$$c(h) = \frac{\log |h'(h^\star)|}{p(h)}, \quad c(h,k) = \frac{c(h)}{c(k)}, \quad c(h,k,\ell) = \frac{c(h) - c(k)}{c(h) - c(\ell)}.$$

The definition of diophantine dynamical sources deals with these ratios:

Definition 4.2. [`DIOP2` and `DIOP3`]. *A p-smooth dynamical system $(\mathcal{I}, T)$ of class $\mathcal{C}^2$ is* `DIOP2` *if there exist two branches h et k of $\mathcal{H}^\star$ for which the ratio $c(h, k)$ is diophantine. It is* `DIOP3` *if there exist three branches h, k and ℓ of $\mathcal{H}^\star$ for which the ratio $c(h, k, \ell)$ is diophantine.*

These conditions are sufficient to entail hyperbolic tameness of sources. This was proven by Dolgopyat [6], and extended to the secant operator by Roux and Vallée [18, 19].

Theorem 4.2. [Dolgopyat, Roux-Vallée] For a `DIOP` source, there exists an hyperbolic region $\mathcal{R}$ on the left of $\Re s = 1$ and a real neighborhood $\mathcal{T}$ of 0, such that, for any $u = e^{i\theta}$ with $\theta \in \mathcal{T}$,

(a) the operator $s \mapsto (I - u\mathbb{H}_s)^{-1}$ is meromorphic in $\mathcal{R}$, with a unique pole at $s = 1 + \sigma(u)$,

($b3$) For a `DIOP3` source, the norm $||(I - u\mathbb{H}_s)^{-1}||_{(1,t)}$ is of polynomial growth for $s \in \mathcal{R}, |\Im s| \to \infty$, uniformly with respect to θ.

($b2$) For a `DIOP2` source, and $\theta \in \mathcal{T} \cap \mathbb{Q}$, the norm $||(I - u\mathbb{H}_s)^{-1}||_{(1,t)}$ is of polynomial growth for $s \in \mathcal{R}, |\Im s| \to \infty$. This polynomial growth depends on the denominator of the rational θ, with an upper bound of the type $O(\mathrm{den}(\theta))\, |t|^\alpha$ for some $\alpha > 0$.

5 Gaussian laws for the depth of a digital search tree.

It is now possible to obtain gaussian laws for the typical depth D_n, with (if possible) a speed of convergence. The proof can be found in Section 8. There will be two cases:

(a) The first case occurs for `UNI` sources when G_n is well-behaved in a complex neighborhood of $u = 1$, and we will use the moment generating function $M_n(w) := G_n(e^w)$ and the Quasi-Powers theorem, which also provides a speed of convergence.

(b) The second case occurs, for `DIOP` sources, when G_n is only well-behaved on (a part of) the circle $|u| = 1$. We then use the characteristic function $M_n(i\theta) := G_n(e^{i\theta})$ and the Goncharov theorem. However, the study is more involved in the `DIOP2` case.

6 Proofs for Section 3.

We provide proofs for the main results of Section 3.

6.1 Proof of Proposition 3.2. The use of Laplace transfer is not very usual in the digital trees analyses. There are some instances, in particular in [9]. We consider the Laplace transform $\mathcal{L}$ which transforms $\underline{B}^{(w)}(z,u)$ into $C^{(w)}(t,u)$, defined as

$$C^{(w)}(t,u) := \int_0^\infty e^{-tx}\underline{B}^{(w)}(x,u)dx.$$

With Proposition 3.1, the generating functions $\widehat{C}^{(w)}(t,u) := t^2 C^{(w)}(t,u)$ satisfy the system of functional equations

(6.15)
$$(t+1)\widehat{C}^{(w)}(t,u) = 1 + u\sum_{i\in\Sigma}\frac{1}{q_{i|w}}\widehat{C}^{(w\cdot i)}\left(\frac{t}{q_{i|w}},u\right).$$

We first focus on $\widehat{C}(t,u)$ relative to the source $\mathcal{S} := \mathcal{S}_{(\epsilon)}$. Iterating Relation (6.15), and using the formula of conditional probabilities, one obtains

$$\widehat{C}(t,u) = \sum_{w\in\Sigma^\star} u^{|w|}p_w \prod_{v\le w}\frac{1}{1+tp_v^{-1}} \tag{6.16}$$

Using decomposition into partial fractions

$$\prod_{v\le w}\frac{1}{1+tp_v^{-1}} = \sum_{v\le w}\frac{r(v,w)}{1+tp_v^{-1}},$$

$$\text{with}\quad r(v,w) := \prod_{\alpha\in[\epsilon,w]\setminus\{v\}}\frac{1}{1-p_vp_\alpha^{-1}},$$

we let

$$\delta(v,u) := \frac{1}{p_v}\sum_{w\ge v} r(v,w)u^{|w|}p_w$$

$$= \frac{1}{p_v}\sum_{w\ge v} p_w \prod_{\alpha\in[\epsilon,w]\setminus\{v\}}\frac{1}{1-p_vp_\alpha^{-1}}. \tag{6.17}$$

We will see later in Lemma 6.1 that the series which defines $\delta(v,u)$ is absolutely convergent. Then, it is possible to change the order of summations, and this leads to an alternative expression for

$$C(t,u) = \frac{1}{t^2}\widehat{C}(t,u) = \frac{1}{t^2}\sum_{v\in\Sigma^\star}\delta(v,u)\frac{p_v}{1+tp_v^{-1}}. \tag{6.18}$$

We now apply the inverse Laplace transform on both sides of (6.18) to recover first the Poisson generating function $\underline{B}(z,u)$ relative to the source $\mathcal{S} = \mathcal{S}_\epsilon$,

$$\underline{B}(z,u) = \sum_{v\in\Sigma^\star}\delta(v,u)[e^{-zp_v} - 1 + zp_v], \tag{6.19}$$

then the expression of $B(z,u)$ itself, via Relation (2.2),

$$e^zB(z,u) = ze^z + (u-1)\sum_{v\in\Sigma^\star}\delta(v,u)[e^{z(1-p_v)} - e^z + e^zzp_v].$$

Extracting coefficients in (2.2) leads to

$$B_n(u) := n![z^n]e^zB(z,u)$$
$$= n + (u-1)\sum_{v\in\Sigma^\star}\delta(v,u)[(1-p_v)^n - 1 + np_v],$$

and finally, with binomial expansion,

$$B_n(u) = n + (u-1)\sum_{\ell=2}^{n}(-1)^\ell\binom{n}{\ell}\sum_{v\in\Sigma^\star}\delta(v,u)p_v^\ell.$$

6.2 Study of the Dirichlet vector $\boldsymbol{\Delta}(s,u)$. The bivariate vectorial `dst` Dirichlet series $\boldsymbol{\Delta}(s,u)$, whose the component of index w is the bivariate Dirichlet series $\Delta_{(w)}(s,u)$ relative to the source $\mathcal{S}_{(w)}$ plays an important role in the sequel. We then prove the following:

LEMMA 6.1. *If the source $\mathcal{S}$ is p-smooth, the bivariate vectorial* `dst` *Dirichlet series $\boldsymbol{\Delta}(s,u)$, whose component of index w is the bivariate Dirichlet series of the source $\mathcal{S}_{(w)}$ belongs to the space $\mathcal{B}(\Sigma^\star)$ of the bounded functions $X:\Sigma^\star\to\mathbb{C}$ endowed with the sup-norm, and is analytic on the domain $\mathcal{B} :=< 1,+\infty > \times\{u;\ |u|\le 1\}$.*

Proof. First, the norm $||\mathbf{P}_s||$ satisfies, for $\sigma := \Re s$,

$$||\mathbf{P}_s|| \le \mu(\sigma)\quad\text{with}\quad \mu(\sigma) = \sup\left\{\sum_{i\in\Sigma}q_{i|w}^\sigma;\ w\in\Sigma^\star\right\}.$$

The term $\delta(v,u)$ decomposes into two factors,

$$\delta(v,u) = \beta(v)\cdot\gamma(v,u),\ \text{with}$$

$$\beta(v) = \prod_{\alpha\in[\epsilon,v[}\frac{1}{1-p_{v|\alpha}},\qquad\text{and}\qquad \gamma(v,u) = \gamma[\mathcal{S}_{(v)},u],$$

$$\text{where}\quad \gamma[\mathcal{S},u] = 1 + \sum_{w\in\Sigma^+}u^{|w|}p_w\prod_{\alpha\in]\epsilon,w]}\frac{1}{1-p_\alpha^{-1}}$$
$$= 1 + \sum_{w\in\Sigma^+}(-u)^{|w|}p_w\left[\prod_{\alpha\in]\epsilon,w]}\frac{p_\alpha}{1-p_\alpha}\right].$$

Using the two inequalities

$$\beta(v)\le\prod_{i=1}^{k}\frac{1}{1-p^i}\quad\text{for } v\in\Sigma^k,\qquad\left|\sum_{v\in\Sigma^k}p_v^s\right|\le\mu(\sigma)^k,$$

the series of general term $\beta(v)p_v^s$ satisfies

$$\left|\sum_{v\in\Sigma^\star}\beta(v)p_v^s\right| \le \sum_{v\in\Sigma^\star}\beta(v)p_v^\sigma \le \sum_{k\ge0}\mu(\sigma)^k\prod_{i=1}^{k}\frac{1}{1-p^i}$$

and, with a classical equality (due to partitions),

$$\sum_{k\geq 0}\mu^k\prod_{\ell=1}^{k}\frac{1}{1-p^\ell}=\prod_{\ell\geq 0}\frac{1}{1-\mu p^\ell},$$

which holds for $\mu, p<1$, we obtain, for $\Re s>1$

$$\left|\sum_{v\in\Sigma^\star}\beta(v)p_v^s\right|\leq\prod_{\ell\geq 0}\frac{1}{1-\mu(\sigma)p^\ell}.$$

We now study $\gamma(\mathcal{S},u)$, which is expressed as a series,

$$\gamma(\mathcal{S},u)=\sum_{k\geq 0}\gamma_k(\mathcal{S},u),$$

$$\text{with}\quad \gamma_k(\mathcal{S},u)=\sum_{w\in\Sigma^k}(-u)^{|w|}p_w\prod_{v\in\mathcal{P}_w}\frac{p_v}{1-p_v}.$$

We compare $|\gamma_k(\mathcal{S},u)|$ and $|\gamma_{k+1}(\mathcal{S},u)|$,

$$|\gamma_{k+1}(\mathcal{S},u)|\leq\sum_{i\in\Sigma}\sum_{w\in\Sigma^k}|u|^{|w\cdot i|}p_{w\cdot i}\prod_{v\in\mathcal{P}_{w.i}}\frac{p_v}{1-p_v}$$

$$=\sum_{w\in\Sigma^k}|u|^{|w|}p_w\prod_{v\in\mathcal{P}_w}\frac{p_v}{1-p_v}\sum_{i\in\Sigma}|u|q_{i|w}\frac{p_{w.i}}{1-p_{w.i}}.$$

Consider $\theta<1$. As soon as p satisfies

$$p^{k+1}\leq\theta/(1+\theta),$$

the quotient $p_{w.i}/(1-p_{w.i})$ is less than θ, and $|\gamma_{k+1}(\mathcal{S},u)|\leq\theta|u|\gamma_k(\mathcal{S},u)$.

This ends the proof of Lemma 6.1 and Proposition 3.2.

6.3 Proof of Proposition 3.3. We extend here the approach of [14] and use some well-known properties of the Mellin transform (see [8]).
With (6.19), the function $z\mapsto\underline{B}(z,u)$ is expressed with an harmonic sum which involves the basis function $f(z):=e^{-z}-1+z$. The function f satisfies

$$f(z)=O(z^2),\quad (z\to 0^+),\quad f(z)=O(z)\quad (z\to\infty)\ ,$$

and its Mellin transform $f^*(s)$ exists in the fundamental strip $<-2,-1>$ and coincides there with the function $\Gamma(s)$. Furthermore, the function $\Delta(s,u)$ defined in (3.5) exists in the domain $\mathcal{B}$ defined in Lemma 6.1, and the following factorization holds for

$$Z(s,u):=\mathcal{M}[z\mapsto\underline{B}(z,u);\,s]=\tag{6.20}$$

$$\Gamma(s)\left(\sum_{v\in\Sigma^\star}\delta(v,u)\,p_v^{-s}\right)=\Gamma(s)\,\Delta(-s,u),$$

when s belongs the fundamental strip $<-2,-1>$ and $|u|\leq 1$.

We consider more generally the sequence of the Mellin transforms

$$Z^{(w)}(s,u)=\mathcal{M}[z\mapsto\underline{B}^{(w)}(z,u);\,s]$$

that exists for s in the fundamental strip $<-2,-1>$ and $|u|\leq 1$ and satisfies the system of equations, deduced from the initial system of Proposition 3.1,

$$-(s-1)Z^{(w)}(s-1,u)+Z^{(w)}(s,u)=u\sum_{i\in\Sigma}q_{i|w}^{-s}\,Z^{(w\cdot i)}(s,u).\tag{6.21}$$

With a factorization analog to (6.20), the functions $\Delta_{(w)}(s,u)$ satisfy the system

$$\Delta_{(w)}(s+1,u)=\Delta_{(w)}(s,u)-u\sum_{i\in\Sigma}q_{i|w}^{s}\Delta_{(w\cdot i)}(s,u).$$

Introducing the matrix $\mathbf{P}_s$ and using the vectorial Dirichlet series $\mathbf{\Delta}(s,u)$ we finally obtain the matrix equation

$$\mathbf{\Delta}(s+1,u)=(I-u\mathbf{P}_s)\mathbf{\Delta}(s,u)$$

$$i.e.\quad \mathbf{\Delta}(s,u)=(I-u\mathbf{P}_s)^{-1}\mathbf{\Delta}(s+1,u).$$

We now prove that the infinite product is convergent.

LEMMA 6.2. *Denote by $\mathcal{B}(\Sigma^\star)$ the Banach space formed with the bounded complex functions $\Sigma^\star\to\mathbb{C}$, endowed with the norm $||X||:=\sup|X_{(w)}|$. For a smooth source, the following holds:*

(*i*) *The infinite product*

$$\mathbf{Q}(s,u):=(I-u\mathbf{P}_s)^{-1}\cdot\ldots\cdot(I-u\mathbf{P}_{s+k})^{-1}\ldots$$

is convergent on the subset $\{(s,u);\quad |u|\mu(\sigma)<1\}$, and defines an operator which acts on $\mathcal{B}(\Sigma^\star)$ and is invertible.

(*ii*) *Consider the vector $\mathbf{1}$ whose all components equal 1. Then, the equality $\mathbf{\Delta}(2,u)=\mathbf{1}$ holds.*

Proof. **Proof of** (*i*)**.** One has: $\|u\mathbf{P}_s\|\leq|u|\mu(\sigma)<1$, and, for $k\geq 0$, $\|u\mathbf{P}_{s+k}\|\leq|u|\mu(\sigma)p^k<1$. Thus, the quasi-inverses $(I-u\mathbf{P}_{s+k})^{-1}$ are well-defined for $\sigma>1$ and any $k\geq 0$, and their norms satisfy

$$\|(I-u\mathbf{P}_{s+k})^{-1}\|\leq\frac{1}{1-|u|\mu(\sigma)p^k}=1+\frac{|u|p^k\mu(\sigma)}{1-|u|\mu(\sigma)p^k},$$

We remark the inequality $\mu(\sigma)p^k<1/2$ for $k\geq k_0$, so that for $k\geq k_0$, one has

$$\|(I-u\mathbf{P}_{s+k})^{-1}\|\leq 1+\frac{2}{2-|u|}p^k.$$

Since the series of general term p^k is convergent, the infinite product $\mathbf{Q}(s,u)$ is normally convergent and defines an analytic function on the domain $\{(s,u); |u|\mu(\sigma) < 1\}$.

Proof of (ii). The assertion $\mathbf{\Delta}(2,u) = \mathbf{1}$ is proven by using the following result shown in [15] :

Proposition A. *Consider a sequence* $\{f_n\}_{n=0}^{\infty}$ *whose Poisson generating function* F *is an entire function, whose Mellin transform* $F^*(s)$ *exists in the strip* $< -2, -1 >$ *and is factored as* $\Gamma(s)\cdot\gamma(-s)$, *where* $\gamma(s)$ *is analytical for* $s \in < 1, \infty >$. *Then*

$$\gamma(n) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k f_n, \quad \textit{for } n \geq 2$$

We apply the result with $n = 2$ to each function $z \mapsto \underline{B}^{(w)}(z,u)$, for $w \in \Sigma^\star$. With the equalities

$$B_0^{(w)}(u) = 0, \quad B_1^{(w)}(u) = 1, \quad B_2^{(w)}(u) = 1 + u,$$

$$\text{one then obtains:} \quad \sum_{k=0}^{2} \binom{2}{k} (-1)^k B_k^{(w)}(u) = u - 1,$$

$$\text{and:} \qquad \Delta_{(w)}(2,u) = \sum_{k=0}^{2} \binom{2}{k} (-1)^k \underline{B}_k^{(w)}(u) = 1.$$

Now, the equality $\mathbf{\Delta}(2,u) = \mathbf{1}$ holds, and this ends the proof of Lemma 6.2.

We now end the proof of Proposition 3.3. The equality $\mathbf{\Delta}(2,u) = \mathbf{1} = \mathbf{Q}(2,u)\,\mathbf{\Delta}(\infty,u)$ implies the equality $\mathbf{\Delta}(\infty,u) = \mathbf{Q}(2,u)^{-1}\mathbf{1}$, and finally

$$\mathbf{\Delta}(s,u) = \mathbf{Q}(s,u)\cdot\mathbf{Q}(2,u)^{-1}\mathbf{1}. \tag{6.22}$$

We now focus on the first component $\Delta(s,u) := \mathbf{\Delta}_{(\epsilon)}(s,u)$ of the vector $\mathbf{\Delta}(s,u)$, and we derive an exact expression for the Mellin transform of the Poisson generating function $\underline{B}(z,u)$.

6.4 Comparison with the analysis for Tries. It is interesting to compare the Poisson generating functions relative to a digital search tree to their analogs for a trie. The recurrences for tries are easier since there is no word in internal nodes. One has for $n \geq 2$ and $k \geq 1$,

$$B_{n,k}^{(w)} = \sum_{j=0}^{n} \binom{n}{j} q_{0|w}^{j}\, q_{1|w}^{n-j} \left(B_{j,k-1}^{(w\cdot 0)} + B_{n-j,k-1}^{(w\cdot 1)} \right),$$

$$\text{with, for any } w \in \Sigma^\star,\ B_{n,0}^{(w)} = 0 \ \text{ for } n \neq 1; \quad B_{1,0}^{(w)} = 1,$$

$$B_{0,k}^{(w)} = 0 \quad \text{for } k \geq 0; \qquad B_{1,k}^{(w)} = 0 \quad \text{for } k \geq 1.$$

Then, the modified Poisson generating function of the profile,

$$\underline{B}_T(z,u) := \frac{1}{u-1}[B_T(z,u) - z]$$

satisfies $\underline{B}_T^{(w)}(z,u) = z(1-e^{-z}) + u\sum_{i\in\Sigma} \underline{B}_T^{(w\cdot i)}(q_{i|w}z, u)$.

Then, a simple iteration provides the explicit expression

$$\underline{B}_T(z,u) = \sum_{v\in\Sigma^\star} u^{|v|} z p_v (1 - e^{-zp_v}),$$

and the Mellin transform of $z \mapsto \underline{B}_T(z,u)$ involves

$$\Lambda(s,u) := \sum_{v\in\Sigma^\star} u^{|v|} p_v^s = {}^t\mathbf{E}\,(I - u\mathbf{P}_s)^{-1}\mathbf{1}$$

under the form $-\Gamma(s+1)\Lambda(-s,u) = -s\Gamma(s)\,\Lambda(-s,u)$. The analog $\mathbf{\Delta}_T$ of $\mathbf{\Delta}$ satisfies $\mathbf{\Delta}_T(s,u) = s(I - u\mathbf{P}_s)^{-1}\mathbf{1}$, and finally $\Delta_T(s,u) = s\Lambda(s,u)$.

7 Proof for Proposition 4.

7.1 Step 1. The mirror operation. The mirror operation ϕ will play an important role in the sequel. It is defined on the set $\Sigma^\star$ and transforms the finite word $w = w_1 w_2 \dots w_{k-1} w_k$ into its mirror $\phi(w) = w_k w_{k-1} \dots w_2 w_1$. We also denote by ϕ the mirror operation induced on $\mathcal{B}(\Sigma^\star)$ and defined by the equality $\phi(X)(w) := X(\phi(w))$, and by $\widehat{p}_w$ the mirror probability $\widehat{p}_w := p_{\phi(w)}$. We furthermore define the g-function on $\Sigma \times \Sigma^\star$ by the equalities

$$g(i\cdot w) := q_{i|\phi(w)} = \frac{\widehat{p}_{i\cdot w}}{\widehat{p}_w}. \tag{7.23}$$

The mirror operation appears in a natural way, as we now explain: when the symbol X_n has to be emitted, it "looks at" (from its relative point of view), its immediate neighbors, which form the word $X_{n-1}, X_{n-2}, \dots, X_1, X_0$ (in this order), namely the mirror of the prefix $X_0, X_1, \dots, X_{n-1}$. The prefix $\phi(w)$ defines the reverse past history.

When the operator $\mathbf{P}_s$ transforms X into Y, the conjugate $\widehat{\mathbf{P}}_s$ of the operator $\mathbf{P}_s$ via the mirror operation ϕ, transforms, (by definition) $\phi(X)$ into $\phi(Y)$. Since the two vectors $\mathbf{1}$ and $\mathbf{E}$ are invariant under the mirror ϕ, and the two words w and $\phi(w)$ have the same length, the relation (3.7) can be re-written as

$$\Delta(s,u) = {}^t\mathbf{E}\,\widehat{\mathbf{Q}}(s,u)\cdot\widehat{\mathbf{Q}}(2,u)^{-1}\,\mathbf{1}, \tag{7.24}$$

where the operator $\widehat{\mathbf{Q}}(s,u)$ is defined as

$$\widehat{\mathbf{Q}}(s,u) := (I - u\widehat{\mathbf{P}}_s)^{-1}\cdot(I - u\widehat{\mathbf{P}}_{s+1})^{-1}\cdot\dots. \tag{7.25}$$

When $Y := \mathbf{P}_s[X]$ is defined as in Eq. (2.4), its transform $\phi(Y)$ satisfies

$$\phi(Y)(w) = Y(\phi(w)) = \sum_{i\in\Sigma} q_{i|\phi(w)}^{s}\, X(\phi(w)\cdot i)$$

$$= \sum_{i\in\Sigma} g(i\cdot w)^s\, \phi(X)(i\cdot w).$$

Then, if T denotes the shift on $\Sigma\times\Sigma^\star$ which associates to the finite word $w = w_1w_2\dots w_k$, the shifted word $T(w) = w_2\dots w_k$, the mapping $\widehat{\mathbf{P}}_s$ which associates $\phi(Y)$ to $\phi(X)$ is defined as the mirror of (2.4),

$$(7.26)\quad Y = \widehat{\mathbf{P}}_s[X] \Longleftrightarrow Y(w) = \sum_{i\in\Sigma} g(i\cdot w)X(i\cdot w)$$

$$\Longleftrightarrow Y(w) = \sum_{\substack{v\\ T(v)=w}} g(v)^s\, X(v).$$

Under this form, the mapping $\widehat{\mathbf{P}}_s$ resembles the transfer operator of the dynamical system $(\Sigma^\star, T)$ relative to the function g^s. This system describes the past of the source when reversing the time, which will be called in the sequel the "reverse past" of the source. However, the shift T is only defined on $\Sigma\times\Sigma^\star$, (not on the whole $\Sigma^\star$) and $\Sigma^\star$ is not compact.

In the following, we will extend the mapping $\widehat{\mathbf{P}}_s$ into a mapping which acts on functions defined on the compact space $\Sigma^{\mathbb{N}}$. This space is a metric space, whose definition is now recalled. First, the coincidence $\gamma(u,v)$ between two words u and v of $\Sigma^{\mathbb{N}}$, is defined as the length of their longest common prefix,

$$\gamma(u,v) = \max\{k;\quad u_i = v_i,\quad \forall i\le k\}.$$

With a real $\theta\in]0,1[$, the coincidence defines a distance $d_\theta(u,v) = \theta^{\gamma(u,v)}$ and the set $\Sigma^{\mathbb{N}}$ endowed with this distance is denoted by $\Sigma^{\mathbb{N}}_\theta$.

7.2 Step 2. Extension towards the past. The coincidence may also be defined between two words u and v of $\Sigma^\star$, via the addition of a ending symbol which does not belong to the initial alphabet. The g-functions of Markov chains admit the following characterization:

$$\mathcal{S} \text{ is a Markov chain of order } k$$
$$\Longleftrightarrow \qquad (\quad \gamma(u,v)\ge k+1 \quad\Longrightarrow\quad g(u) = g(v)\quad).$$

Then, it is natural to consider "good" sources, where the g-functions are continuous or even Hölder with exponent α. Namely, assume, that for some $\alpha>0$, one has

$$[\text{Hölder}]\quad \forall u,v\in\Sigma^\star,\quad |g(u)-g(v)|\le d_\theta(u,v)^\alpha.$$

Since the space $\Sigma^\star$ is a dense subset of $\Sigma^{\mathbb{N}}_\theta$, the function g can be extended to $\Sigma^{\mathbb{N}}_\theta$ "by continuity" and its extension $\underline{g}$ is also a Hölder function on $\Sigma^{\mathbb{N}}_\theta$ with the same exponent as g. Via the extension $\underline{g}$, the source $\mathcal{S}$ is extended into a source $\underline{\mathcal{S}}$, and the extended source $\underline{\mathcal{S}}$ has given an "infinite" past, described by the family $\underline{g}(i\cdot v)$ for $v\in\Sigma^{\mathbb{N}}$. More precisely, $\underline{g}(i\cdot v)$ is the probability of emitting i when the reverse past history has just emitted the infinite word v.

Consider the space $\mathcal{H}_\alpha(\Sigma^{\mathbb{N}}_\theta)$ formed with the Hölder functions $X:\Sigma^{\mathbb{N}}_\theta\to\mathbb{C}$ with exponent α. Then, via the extension $\underline{g}$ of g, the operator $\widehat{\mathbf{P}}_s$ defined in (7.26) is extended on $\mathcal{H}_\alpha(\Sigma^{\mathbb{N}}_\theta)$ into an operator $\underline{\widehat{\mathbf{P}}}_s$ defined as

$$\underline{\widehat{\mathbf{P}}}_s[X] = Y$$
$$\Longleftrightarrow \quad Y(v) = \sum_{i\in\Sigma} \underline{g}(i\cdot v)^s\, X(i\cdot v) = \sum_{\substack{u\\ T(u)=v}} \underline{g}(u)^s\, X(u)$$

Here, T is the shift towards the past, defined on the reverse infinite past $\Sigma^{\mathbb{N}}$ by

$$T(i\cdot w) = w,\quad \text{for}\quad i\in\Sigma,\ w\in\Sigma^{\mathbb{N}}.$$

Now, the operator $\underline{\widehat{\mathbf{P}}}_s$ is the (true) transfer operator of the system $(\Sigma^{\mathbb{N}}, T)$ relative to $\underline{g}^s$. Via classical results, this operator $\underline{\widehat{\mathbf{P}}}_s$ admits on $\mathcal{H}_\alpha(\Sigma^{\mathbb{N}}_\theta)$ a unique invariant measure, denoted by ν, which satisfies

$$\underline{g}(v)\, d\nu(Tv) = d\nu(v),\quad \text{or}\quad \underline{g}(i\cdot v)d\nu(v) = d\nu(i\cdot v).$$

We extend this expression to any pair (u,v) of infinite words: we define the interval $[u,v]$ as the set of all the infinite words t which satisfy $u\le t\le v$ for the lexicographic order on $\Sigma^{\mathbb{N}}$, and we consider the probability $\underline{g}(i\cdot[u,v])$ of emitting i knowing that the infinite reverse past belongs to the interval $[u,v]$, namely

$$\underline{g}(i\cdot[u,v]) := \frac{\int_u^v \underline{g}(i\cdot t)d\nu(t)}{\int_u^v d\nu(t)},$$

so that $\quad \underline{g}(i\cdot[w^-,w^+]) = \underline{g}(i\cdot w)$.

Remark that it is not completely clear if $\underline{g}(i\cdot[u,v])$ is always well-defined: we need the source, and its g-functions, to be more regular. In order to define in an easier way these extra regularity assumptions, we change our point of view and consider the problem on the unit interval of the real line.

We first insist on a particularity of a stationary source, related to the stationary measure ν.
For $w = w_0w_1\dots w_k$, denote by

$$\widehat{\pi}_w = \Pr_\nu[X_0 = w_k,\dots,X_k = w_0],$$

so that, for any $i\in\Sigma$,

$$\widehat{\pi}_{w\cdot i} = \Pr_\nu[X_0 = i, X_1 = w_k,\dots,X_{k+1} = w_0].$$

This implies, for a stationary source, the equality

$$(7.27)\qquad \sum_{i\in\Sigma}\widehat{\pi}_{w\cdot i} = \Pr_\nu[X_1 = w_k,\dots,X_{k+1} = w_0]$$
$$= \Pr_\nu[X_0 = w_k, X_2 = w_{k_1}\dots X_k = w_0] = \widehat{\pi}_w.$$

This will be central to define a parameterization of the reverse past, in the case of a stationary source.

7.3 Parameterization of the source – Dynamical system $\mathcal{D}$ of the unit interval. Here, the finite alphabet $\Sigma := \{a_1, a_2, \dots, a_r\}$ is listed in the increasing order, and the reverse past history $\Sigma^\star$ is ordered with the lexicographic order induced from the order on Σ. As the source is p–smooth:

(i) For any $w \in \Sigma^\star$, the probability $\widehat{\pi}_w$ is strictly positive. This means that all the words of $\Sigma^{\mathbb{N}}$ are emitted by the source.

(ii) The supremum $\sup\{\widehat{\pi}_w : w \in \Sigma^k\} \leq p^k$ tends to 0, as $k \to \infty$.

For any prefix $w \in \Sigma^\star$ of the reverse past, we denote by $|w|$ the length of w (i.e., the number of the symbols that it contains) and b_w, c_w, $\widehat{\pi}_w$ the probabilities that a word of the reverse past begins with a prefix α of the same length as w, which satisfies $\alpha < w$, $\alpha \leq w$, or $\alpha = w$, meaning

(7.28)
$$b_w := \sum_{\substack{\alpha, |\alpha|=|w|, \\ \alpha<w}} \widehat{\pi}_\alpha, \quad c_w := \sum_{\substack{\alpha, |\alpha|=|w|, \\ \alpha\leq w}} \widehat{\pi}_\alpha, \quad \widehat{\pi}_w = c_w - b_w.$$

Then, the equality (7.27) entails the inclusions $[b_{w\cdot i}, c_{w\cdot i}] \subset [b_w, c_w]$ for any $i \in \Sigma$.

Given an infinite word of the reverse past $v \in \Sigma^{\mathbb{N}}$, denote by v_k its prefix of length k. The sequence $(b_{v_k})_{k\geq 0}$ is increasing, the sequence $(c_{v_k})_{k\geq 0}$ is decreasing, and $c_{v_k} - b_{v_k} = \widehat{\pi}_{v_k}$ tends to 0 for $k \to \infty$. Thus a unique real $N(v) \in [0,1]$ is defined as the common limit of (b_{v_k}) and (c_{v_k}), and $N(v)$ can be viewed as the probability that an infinite word u of the reverse past be smaller than v. The mapping $N : \Sigma^{\mathbb{N}} \to [0,1]$ is strictly increasing outside the exceptional set formed with words of $\Sigma^{\mathbb{N}}$ which end with an infinite sequence of the smallest symbol a_1 or with an infinite sequence of the largest symbol a_r. More precisely, one has $N(u) = N(v)$ with $u > v$ if and only if there exists $w \in \Sigma^\star$ and $i \in [2..r]$ for which $u = w \cdot a_i \cdot a_1^\infty$ with and $v = w \cdot a_{i-1} \cdot a_r^\infty$.

Conversely, almost everywhere, except on the set $\{b_w, w \in \Sigma^\star\}$, there is a mapping M which associates, to a number x of the interval $\mathcal{I} := [0,1]$, a word $M(x) \in \Sigma^{\mathbb{N}}$ of the reverse past, for which $N(M(x)) = x$. Hence, the probability that an infinite word u of the reverse past be smaller than $M(x)$ equals x. The lexicographic order on the reverse past is then compatible with the natural order on the interval $\mathcal{I}$. The interval $\mathcal{I}_w := [b_w, c_w]$, of length $\widehat{\pi}_w$, gathers (up to a denumerable set) all the reals x for which the word $M(x)$ of the reverse past begins with the finite prefix w. This is the fundamental interval of the prefix w.

Denote by T the shift on $\Sigma^{\mathbb{N}}$ (here the shift towards the past of the reverse past) and by $\check{T}$ the shift induced by T on $[0,1]$ via conjugation of mappings N, M, namely

$$\check{T}(x) := N[T(M(x))],$$

As, by definition, one has $T(i \cdot v) = v$, the equality $N(T(i \cdot v)) = N(v) = \check{T}(N(i \cdot v))$ holds; Then, if we let $x := N(v)$, the real $y = N(i \cdot v)$ satisfies $\check{T}(y) = x$; this is an antecedent of x by $\check{T}$, completely defined by the pair (i, x) and is denoted by $h_i(x)$; For each map h_i, the image $h_i(]0,1[)$ coincides with the set $i \cdot \Sigma^{\mathbb{N}} \setminus \{i^-, i^+\}$. More generally, if we let $x := N(v)$, the real $y = N(w\cdot v)$, for $w = w_1 w_2 \dots w_k \in \Sigma^k$ satisfies $\check{T}^k(y) = x$; this is an antecedent of x by $\check{T}^k$, equal to $h_w(x)$, where $h_w = h_{w_1} \circ h_{w_2} \circ \dots \circ h_{w_k}$. Finally, the pair $([0,1], \check{T})$ gives rise to a complete dynamical system $\mathcal{D}$ of the interval $[0,1]$ on the alphabet Σ, as we now define it:

A complete dynamical system of interval $\mathcal{I} := [0,1]$ relative to an alphabet Σ is defined by a mapping $\check{T} : \mathcal{I} \to \mathcal{I}$ (called the shift) for which

a) there exists a topological partition of $\mathcal{I}$ with disjoint open intervals $\mathcal{I}_J$ for $j \in \Sigma$, i.e. $\mathcal{I} = \cup_{j\in\Sigma} \bar{\mathcal{I}}_j$.

b) the restriction $\check{T}|_{\mathcal{I}_j}$ is a continuous bijection from $\mathcal{I}_j$ to $]0,1[$, whose inverse is denoted by h_j.

As $N(u)$ coincides with the probability that an infinite word be smaller than u, the following equalities relate the g–functions relative to the stationary distribution, denoted by $\underline{g}$, and the secants of inverse branches, defined in (4.13),

(7.29)
$$\underline{g}(i\cdot w) = \left| \frac{N(i\cdot w^+) - N(i\cdot w^-)}{N(w^+) - N(w^-)} \right| = S[h_i](h_w(0), h_w(1)).$$

The second equality comes from the definition of M, and the equalities $(w^- = M(h_w(0)), w^+ = M(h_w(1)))$.

7.4 More regular sources. An important subclass of sources is formed by regular sources for which all the branches h_i are of class $\mathcal{C}^2$. In this case, the (secant) transfer operator, introduced in [21] and defined as

$$\mathbb{H}_s[F](x,y) := \sum_{i\in\Sigma} \left| \frac{h_i(x) - h_i(y)}{x - y} \right|^s F(h_i(x), h_i(y))$$

play an important role, as in many studies in dynamical sources. It will provide the good extension for the operator $\widehat{\mathbf{P}}_s$ in the stationary case, as we now explain. As the branches h_i are of class $\mathcal{C}^2$, the secant $(x,y) \mapsto S[h](x,y)$ is of class $\mathcal{C}^1([0,1]^2)$, and the secant operator acts on $\mathcal{C}^1([0,1]^2)$.

We consider the subset $\mathcal{F}$ of functions $X \in \mathcal{B}(\Sigma^\star)$ which are associated to a function F of $\mathcal{C}^1([0,1]^2)$, by the relation $X(w) = F(h_w(0), h_w(1))$. Then, the subset

$\mathcal{F}$ is invariant under the action of $\widehat{\mathbf{P}}_s$. Indeed, if $X \in \mathcal{F}$ is associated to F, then the function $Y := \widehat{\mathbf{P}}_s[X]$ is associated to the function $G = \mathbb{H}_s[F]$ which also belongs to $\mathcal{C}^1([0,1]^2)$. This is due to the relation $Y(w) =$

$$\sum_{i\in\Sigma} \underline{g}(i\cdot w)^s\, X(i\cdot w) = \sum_{i\in\Sigma} \underline{g}(i\cdot w)^s F(h_{i\cdot w}(0), h_{i\cdot w}(1)).$$

The equality $h_{i\cdot w} = h_i \circ h_w$, together with Relation (7.29) entails the equality

$$\begin{aligned} Y(w) &= \sum_{i\in\Sigma} |S[h_i\circ h_w](0,1)|^s \cdot F(h_i\circ h_w(0), h_i\circ h_w(1)) \\ &= \mathbb{H}_s[F](h_w(0), h_w(1). \end{aligned}$$

Then, the two operators, the operator $\mathbb{H}_s$, when acting on $\mathcal{C}^1([0,1]^2)$, and the operator $\widehat{\mathbf{P}}_s$, when acting on $\mathcal{F}$ are conjugate, and the Dirichlet series satisfies

$$\begin{aligned} \Delta(s,u) &= \mathbb{K}(s,u)\circ\mathbb{K}(2,u)^{-1}[1](0,1) \\ &= (I - u\mathbb{H}_s)^{-1} \circ \mathbb{K}(s+1,u) \circ \mathbb{K}(2,u)^{-1}[1](0,1), \end{aligned}$$

where $\mathbb{K}(s,u)$ denotes the infinite product

$$\mathbb{K}(s,u) := (I - u\mathbb{H}_s)^{-1}\circ(I - u\mathbb{H}_{s+1})^{-1}\circ(I - u\mathbb{H}_{s+2})^{-1}\circ\dots$$

The operator $\mathbb{H}_s$, when acting on $\mathcal{C}^1([0,1]^2)$ has nice properties, and its quasi-inverse is deeply studied. It will make possible to derive sufficient conditions on the system $\mathcal{D}$ under which the `dst` Dirichlet series $\Delta(s,u)$ will be tame.

8 Sketch of the proof for the main Theorem.

The Rice Formula provides an asymptotic estimate for the probability generating function $G_n(u) := \mathbb{E}[u^{D_n}]$: The pole $s = 1+\sigma(u)$ of $\Delta(s,u)$ provides the main term and the Rice integral provides the remainder term.

8.1 Main term. For $(\Re s, u)$ close to (1, 1), there are two poles for the function $s \mapsto L_n(s,u)$, with

$$L_n(s,u) := \Delta(s,u)\frac{n!}{s(s-1)(s-2)\dots(s-n)}, \tag{8.30}$$

a pole at $s = 1$ and a pole at $s = 1 + \sigma(u)$, where $\sigma(u)$ is defined for u close to 1, by the equations $\sigma(1) = 0$ and $\lambda(1 + \sigma(u)) = 1/u$, which involve the dominant eigenvalue $\lambda(s)$ of the source. The contribution of these two poles provides the main term of the asymptotic estimate for the ratio $n(G_n(u) - 1)/(u - 1)$. One has

$$\mathrm{Res}\,(L_n(s,u); s = 1) = (-1)^{n-1}\frac{n}{1-u}$$

for $|u| < 1$ and by analytic continuation, the relation holds for any $u \neq 1$, so that the main term of the asymptotic estimate of $G_n(u)$ for u close to 1 is due to the pole $s = 1 + \sigma(u)$, and this leading term is

$$(u-1)r(u)\Gamma(-1-\sigma(u))n^{\sigma(u)}, \tag{8.31}$$

where $r(u)$ is the residue of the function $s \mapsto \Delta(s,u)$ at $s = 1 + \sigma(u)$. In the following, we let $U(w) = \sigma(e^w)$, so that the function U is analytic in a neighborhood $\mathcal{W}$ of 0, where it admits the following Taylor expansion,

$$U(w) = \mu w + \nu\frac{w^2}{2} + O(w^3), \qquad \text{with} \tag{8.32}$$

$$\mu = U'(0) = -\frac{1}{\lambda'(1)}, \quad \nu = U''(0) = \frac{\lambda'(1)^2 - \lambda''(1)}{\lambda'(1)^3}.$$

We will see the constants μ, ν respectively appear in the leading terms of the estimates of $\mathbb{E}[D_n], \mathrm{Var}[D_n]$.

8.2 Estimates for the Rice integral. For the remainder terms, the needed (and somewhat classical) estimates on the Rice integral are summarized in the following proposition which is proven for instance in [2].

Proposition R. *For a function $\Delta(s,u)$ defined when $(\Re s, u)$ is close to (1, 1), the following estimates hold for integrals which involve the function $L_n(s,u)$ defined in (8.30):*

(i) *Consider a vertical line $\Re(s) = \alpha$ with $\alpha \notin \mathbb{N}$ and assume that there is a domain $\mathcal{U}$ such that, for any $u \in \mathcal{U}$, the function $s \mapsto \Delta(s,u)(s)$ be continuous on $\Re(s) = \alpha$ and of uniform polynomial growth there: $\Delta(s,u) = O(s^r)$ as $|s| \to \infty$ on $\Re(s) = \alpha$. Then, the integral of $L_n(s,u)$ on the vertical line $\Re s = \alpha$ admits the uniform estimate, as $n \to \infty$,*

$$\int_{\Re s=\alpha} L_n(s,u)ds = O\left(n^\alpha\right).$$

(ii) *Consider a curve ρ of hyperbolic type, namely of the form $\rho := \rho_0 \cup \rho_1$, with*

$$\rho_0 := \{s = \sigma + it; \quad |t| \geq B, \quad \sigma = 1 - \frac{A}{|t|^{\beta_0}}\}$$

$$\rho_1 = \{s = \sigma + it; \quad \sigma = 1 - \frac{A}{B^{\beta_0}}, |t| \leq B\},$$

for some strictly positive constants (A, B, β_0), and assume that there is a domain $\mathcal{U}$ such that, for any $u \in \mathcal{U}$, the function $s \mapsto \Delta(s,u)$ be continuous on ρ and of uniform polynomial growth there.
Then, with $\beta < (1+\beta_0)^{-1}$, the integral of $L_n(s,u)$ on the curve ρ admits the uniform estimate

$$\int_\rho L_n(s,u)ds = n\cdot O\left(\exp[-(\log n)^\beta]\right), \quad (n\to\infty)$$

8.3 Case of the UNI class. In this case, the probability generating function G_n is well-behaved in a complex neighborhood of $u = 1$, and we will use the moment generating function $M_n(w) := G_n(e^w)$ and the Quasi-Powers theorem, which also provides a speed of convergence.

Theorem D. [Quasi-Powers Theorem (Hwang)] *Consider a sequence of variables D_n, defined on probability space $(\Omega_n, \mathbb{P})$ and their moment generating functions $M_n(w) := G_n(e^w) = \mathbb{E}[e^{wD_n}]$. Suppose that the functions $M_n(w)$ are analytic in a complex neighborhood $\mathcal{W}$ of zero, and satisfy*

$$M_n(w) = \exp[\beta_n U(w) + V(w)]\left(1 + O(\kappa_n^{-1})\right), \tag{8.33}$$

where the O-term is uniform on $\mathcal{W}$. Moreover, $U(w)$ and $V(w)$ are analytic on $\mathcal{W}$ and the sequences β_n, κ_n tend to ∞ (for $n \to \infty$).
Then, the mean and the variance satisfy

$$\begin{array}{rcccccc} \mathbb{E}_n[D_n] & = & U'(0)\,\beta_n & + & V'(0) & + & O(\kappa_n^{-1}), \\ \mathrm{Var}_n[D_n] & = & U''(0)\,\beta_n & + & V''(0) & + & O(\kappa_n^{-1}) \end{array}$$

Furthermore, if $U''(0) \neq 0$, the distribution of D_n on Ω_n is asymptotically Gaussian, with speed of convergence $O(\kappa_n^{-1} + \beta_n^{-1/2})$.

We let $\beta_n = \log n$, $\kappa_n := n^\alpha$, $U(w) := \sigma(e^w)$, $V(w) = \log\left[(e^w - 1)r(e^w)\Gamma(-1 - \sigma(e^w))\right]$.

The functions U and V are analytic in a neighborhood of $w = 0$ and the first two derivatives of U at $w = 0$ satisfy (8.32). Therefore, the mean and variance of the depth D_n satisfy

$$\begin{array}{rcccccc} \mathbb{E}_n[D_n] & = & U'(0)\,\log n & + & V'(0) & + & O(n^{-\alpha}), \\ \mathrm{Var}_n[D_n] & = & U''(0)\,\log n & + & V''(0) & + & O(n^{-\alpha}), \end{array}$$

for some $\alpha > 0$. The constants c_1 and c_2 are expressed with derivatives of functions σ and $w \mapsto r(e^w)$ at $w = 0$. If, moreover, the function $s \mapsto \log \lambda(s)$ is strictly convex, the depth D_n asymptotically follows a Gaussian law with speed of convergence $O((\log n)^{-1/2})$.

8.4 Case of the DIOP Class. Here we deal with the characteristic function and use the Goncharov theorem.

Theorem E. [Goncharov] *Consider a sequence of random variables D_n, and denote by $G_n(u) := \mathbb{E}[u^{D_n}]$ the probability generating function of D_n. Furthermore, let $\mu_n := \mathbb{E}[D_n]$ and $\nu_n := (\mathrm{Var}[D_n])^{1/2}$. If the characteristic function $\widetilde{M}_n(i\theta)$ of the variable $(D_n - \mu_n)/\nu_n$,*

$$\widetilde{M}_n(i\theta) := \exp\left[-i\theta\frac{\mu_n}{\nu_n}\right] \cdot G_n(e^{i\theta/\nu_n}), \tag{8.34}$$

tends to $e^{-\theta^2/2}$ for any real θ, then, the variables D_n asymptotically follow a Gaussian law.

For applying this Theorem, we first need to estimate the expectation and the variance of D_n. Previously, in the UNI case, this was directly given by Theorem D, but this is no longer the case for DIOP sources.

Direct study for the mean and the variance. With the same methods as these described in the paper, we obtain the estimates provided in Assertion (a) of our main Theorem. First, one has

$$\mathbb{E}[D_n] = B_n'(1), \qquad \mathbb{E}[D_n^2] = B_n''(1) + B_n'(1),$$

and, second, there exist analogs of (3.6), namely

$$\begin{aligned} B_n'(1) &= -\frac{1}{n}\sum_{\ell=2}^{n}(-1)^\ell \binom{n}{\ell}\Delta(\ell), \\ B_n''(1) &= -\frac{2}{n}\sum_{\ell=2}^{n}(-1)^\ell \binom{n}{\ell}\widetilde{\Delta}(\ell), \end{aligned}$$

where the functions

$$\Delta(s) := \Delta(s, 1), \quad \widetilde{\Delta}(s) := \partial/\partial u \Delta(s, u)|_{u=1}$$

involve also infinite products, namely the product

$$(I - \mathbb{H}_s)^{-1} \circ \mathbb{R}(s, 1) \qquad \text{for } \Delta(s),$$

and a sum of infinite products for $\widetilde{\Delta}(s)$, of the form

$$(I - \mathbb{H}_s)^{-2} \circ \mathbb{H}_s \circ \mathbb{R}(s, 1) + (I - \mathbb{H}_s)^{-1} \circ \frac{\partial}{\partial u}\mathbb{R}(s, u)|_{u=1},$$

where the last derivative is itself a sum of infinite products. Then, there is a double pole at $s = 1$ for $\Delta(s)/(s-1)$ and thus a leading term for $\mathbb{E}[D_n]$ of order $\log n$. In the same vein, there is a triple pole at $s = 1$ for $\widetilde{\Delta}(s)/(s-1)$, and thus a leading term for $\mathbb{E}[D_n^2]$ of order $\log^2 n$, equal to the square of the leading term for $\mathbb{E}[D_n]$. Then, there is a telescoping, and $\mathrm{Var}[D_n]$ is of order $\log n$. With more precise computations, we obtain the estimates of Assertion (a) of the main Theorem,

$$\begin{array}{rcl} \mu_n & = & \mu \log n \quad (1 + O(1/\log n)) \\ \nu_n & = & \nu \log n \quad (1 + O(1/\log n)) \end{array} \tag{8.35}$$

End of the study for the DIOP3 case. The expression of the leading term of $G_n(u)$ given in (8.31), together with the estimate, for complex w close to 0,

$$(1 - e^w)\, r(e^w)\, \Gamma(-1 - \ U(w)) = 1 + O(w),$$

prove that the leading term of $G_n(e^w)$ is $n^{U(w)}\left[1 + O(|w|)\right] + O(|w|)$.

Now, the Taylor expansion of U given in (8.32) taken at $i\theta/\nu_n$, together with the expression of μ_n and ν_n given in (8.35) prove the final expression for the leading term of $\widetilde{M}_n(i\theta)$, for $\theta^3/\nu_n \to 0$, namely

$$e^{-\theta^2/2}\left[1 + O\left(\frac{\theta}{\nu_n}\right) + O\left(\frac{\theta^2}{\nu_n^2}\right) + O\left(\frac{\theta^3}{\nu_n}\right)\right] + O\left(\frac{\theta}{\nu_n}\right). \tag{8.36}$$

In the DIOP3 case, we choose as the curve $\mathcal{C}$ the hyperbolic curve which "borders" the asymptotic region, and we use Assertion (ii) of Proposition R. This ends the proof of our main theorem in the DIOP3 case.

Particularities of the DIOP2 case. The proof is more involved because the norm of the quasi-inverse $||(I - e^{i\theta/\nu_n}\mathbb{H}_s)^{-1}||_{(1,t)}$ is proven to be of polynomial growth in the hyperbolic region $\mathcal{R}$ when $|\Im s| \to \infty$, only when θ/ν_n is a rational number of the form p_n/q_n, and, in this case, the upper bound on the norm is $O(q_n)|t|^\alpha$, for some $\alpha > 0$.

The idea is to replace θ/ν_n by a rational of the form $p_n/q_n = \theta'/\nu_n$ for which the following holds:

$$\left|\frac{\theta}{\nu_n} - \frac{\theta'}{\nu_n}\right| \leq \frac{1}{q_n}; \quad |G_n(i\theta/\nu_n) - G_n(i\theta'/\nu_n)| \leq \frac{\mu_n}{q_n},$$

$$\left|e^{-\theta^2/2} - e^{-\theta'^2/2}\right| \leq K\frac{\nu_n}{q_n}.$$

We then choose $q_n = (\log n)^\gamma$ with $\gamma > 1$ such that the last two differences tend to zero.

We apply the Rice Formula and obtain for $\widetilde{M}_n(i\theta')$ a good estimate since θ'/ν_n is a rational number. The leading term is of the same form as in (8.36) (with θ' instead of θ) and the remainder term is, with $\beta' < \beta$.

$$(\log n)^\gamma O\left(\exp\left[-(\log n)^\beta\right]\right) = O\left(\exp\left[-(\log n)^{\beta'}\right]\right).$$

This ends the sketch of the proof of our main Theorem.

Conclusion. We have provided a new point of view for a general source, where it is possible to study *both* tries and dst's. We have explained the similarities of the behaviors of the two structures, tries and dst's, by the similarities of their Dirichlet series. We have exhibited two particular classes of sources, the class UNI and the class DIOP where the typical depth of tries and dst's asymptotically follows a gaussian law.

References

[1] Baladi, V., and Vallée, B. Euclidean algorithms are Gaussian. *Journal of Number Theory 110* (2005), 331–386.

[2] Cesaratto, E. and Vallée, B. Gaussian distribution of trie depth for dynamical sources. *Submitted*

[3] Clément, J., Flajolet, P., and Vallée, B. Dynamical sources in information theory: A general analysis of trie structures. *Algorithmica 29*, 1/2 (2001), 307–369.

[4] Clément, J., Nguyen-Thi, T-H., and Vallée, B. A general framework for the realistic analysis of sorting and searching algorithms. Application to some popular algorithms. *To appear in Combinatorics, Probability and Computing*

[5] Dolgopyat, D. On decay of correlations in Anosov flows, *Ann. of Math.* 147 (1998) 357-390.

[6] Dolgopyat, D. Prevalence of rapid mixing (I) *Ergodic Theory and Dynamical Systems* 18 (1998) 1097-1114.

[7] Drmota, M. and Szpankowski, W. The Expected Profile of Digital Search Trees, *J. Combinatorial Theory*, Ser. A, 118, 1939-1965, 2011.

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

[9] Flajolet, P. and Richmond, B. Generalized Digital Trees and their difference differential equations *Random Structures and Algorithms* vol 3 (3), (1992)

[10] Flajolet, P., Roux, M. and Vallée, B. Digital trees and memoryless sources: from arithmetics to analysis Proceedings of AofA'10, *DMTCS*, proc AM, pp 231–258 (2010)

[11] Flajolet, P., and Sedgewick, R. Digital Search Trees revisited *Siam J. Comput.* (1986)

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

[13] Flajolet, P., and Sedgewick, R. *Analytic Combinatorics*. Cambridge University Press, 2008.

[14] Jacquet, P., Szpankowski, W., and Tang, J. Average Profile of the Lempel-Ziv Parsing Scheme for a Markovian Source, *Algorithmica*, 31, 318-360, 2001.

[15] Louchard, G., Szpankowski, W., Tang, J. Average profile of the generalized digital search tree and the generalized Lempel-Ziv algorithm, *SIAM J. Computing*, 28, 935-954, 1999.

[16] Nörlund, N. E. Leçons sur les équations linéaires aux différences finies. In *Collection de monographies sur la théorie des fonctions*. Gauthier-Villars, Paris, 1929.

[17] Nörlund, N. E. *Vorlesungen über Differenzenrechnung*. Chelsea Publishing Company, New York, 1954.

[18] Roux, M. Séries de Dirichlet, Théorie de l'information, et Analyse d'algorithmes, *PhD thesis*, University of Caen, 2011.

[19] Roux, M. and Vallée, B. Information theory : Sources, Dirichlet series, and realistic analysis of data structures, Proceedings of Words, 11, *Electronic Proceedings of Theoretical Computer Science*, Volume 63, pp 199-214 (2011)

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

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

[22] Vallée, B., Clément, J., Fill, J. A., and Flajolet, P. The number of symbol comparisons in QuickSort and QuickSelect. In *ICALP 2009*, vol. 5555 of *Lecture Notes in Computer Science*, pp. 750–763.