



HAL
open science

Statistics for dynamical sources

Brigitte Vallée

► **To cite this version:**

| Brigitte Vallée. Statistics for dynamical sources. 2008. hal-00210594

HAL Id: hal-00210594

<https://hal.science/hal-00210594>

Preprint submitted on 21 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistics for dynamical sources

Brigitte VALLÉE

GREYC, CNRS and University of Caen

14032 Caen Cedex, France

Email: brigitte.vallee@info.unicaen.fr

Abstract—A quite general model of source that comes from dynamical systems theory is introduced. Within this model, basic problems of algorithmic information theory contexts are analysed. The main tool is a new object, a (generalized) transfer operator, which can be viewed as a “generating” operator for “fundamental” probabilities. Its dominant spectral objects are related to important parameters of the source, such as the entropy, and play a central rôle in all the results.

I. DYNAMICAL SOURCES.

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

A. Dynamical sources

A probabilistic dynamical source is defined by two objects: a symbolic mechanism and a density. The mechanism is described by a dynamical system and associates an infinite word $M(x)$ to a real number x of the $[0, 1]$ interval. It can be viewed as a generalization of numeration systems, the binary expansion of a real x , or the continued fraction expansion of the real x being well-known instances. [see Fig 1]

Definition. [Dynamical system]. A dynamical system (\mathcal{I}, T) is defined by four elements:

- an alphabet Σ , finite or denumerable
- a topological partition of $\mathcal{I} :=]0, 1[$ with disjoint open intervals $\mathcal{I}_m, m \in \Sigma$,
- an encoding mapping σ which is constant and equal to m on each \mathcal{I}_m ,
- a shift mapping T whose restriction to \mathcal{I}_m is a bijection of class C^2 from \mathcal{I}_m to $\mathcal{J}_m := T(\mathcal{I}_m)$. The local inverse of $T|_{\mathcal{J}_m}$ is denoted by h_m .

Such a dynamical system can be viewed as a “dynamical source”, since, on an input x of \mathcal{I} , it outputs the word $M(x)$ formed with the sequence of symbols $\sigma T^j(x)$, i.e., $M(x) := (\sigma x, \sigma T x, \sigma T^2 x, \dots)$. [see Fig. 1] The k -th prefix of the word is just $P_k(x) := (\sigma x, \sigma T x, \sigma T^2 x, \dots, \sigma T^{k-1} x)$.

Once the mechanism has been fixed, the density f on the interval \mathcal{I} can vary. This induces then different probabilistic behaviours for source words.

Definition. [Probabilistic dynamical sources.] Let $\mathcal{S} = (\mathcal{I}, T)$ be a dynamical system and let f be a real density on interval \mathcal{I} . Let F be the associated distribution function. The pair (\mathcal{S}, F) defines a source [denoted by \mathcal{S}_F], which emits words from $\Sigma^\infty = M(\mathcal{I})$ with the probability induced from F by M .

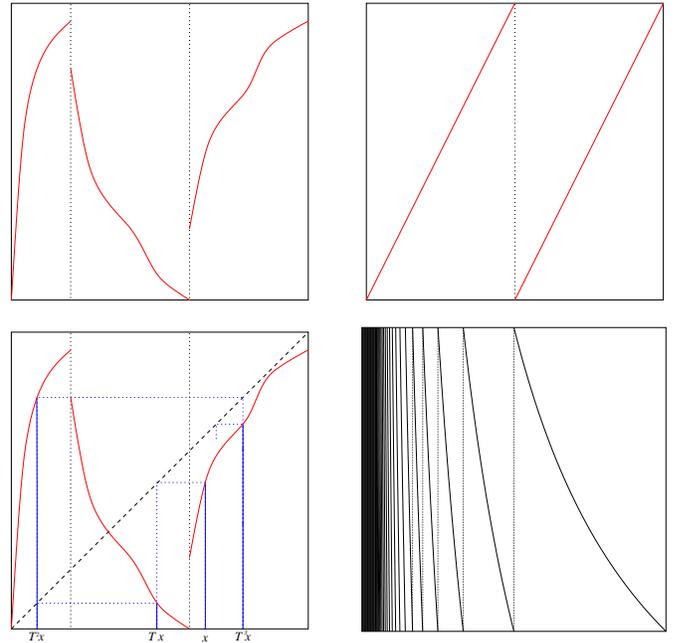


Fig. 1. On the left: A dynamical system, with a partition coded with $\Sigma = \{a, b, c\}$ and a word produced by a dynamical system $M(x) = (c, b, a, c, \dots)$. On the right: two classical sources resp. related to the binary expansion, or the continued fraction expansion.

Such sources may possess a high degree of correlations, due to the geometry of the branches and also to the shape of branches. The geometry of the branches is defined by the respective positions of “horizontal” intervals \mathcal{I}_m with respect to “vertical” intervals $\mathcal{J}_\ell := T(\mathcal{I}_\ell)$ and allows to describe the set \mathcal{S}_m formed with symbols which can be possibly emitted after symbol m . The geometry of the system then provides a first access to the correlation between successive symbols.

The system is *complete* if, for any $m \in \Sigma$, the interval \mathcal{J}_m is the whole interval I . Then all the symbols of Σ can be emitted after any symbol m . The system is *markovian*, if, for any m , the interval \mathcal{J}_m is an union of intervals \mathcal{I}_ℓ for $\ell \in \mathcal{K}_m$. In this case, the set \mathcal{S}_m equals \mathcal{K}_m . In the case when the system is not Markovian, it is sometimes possible to obtain a refinement of the partition, for which the new system becomes Markovian.

But, this is not always possible, and, in the case when it is not possible, the set S_m cannot be characterized when considering only bounded parts of the previous history. In all the cases, the following property [always true for a complete system] is essential:

(P1) [Topologically mixing] *For any pair of symbols (b, e) , there exists $n_0 \geq 1$ such that, for any $n \geq n_0$, there is a word of length n which begins with symbol b and finishes with symbol e [i.e., $\mathcal{I}_b \cap T^{-n}(\mathcal{I}_e) \neq \emptyset$].*

The shape of the branches, and more precisely, the behavior of derivatives h'_m has also a great influence on correlations between symbols. The two following properties play an important rôle:

(P2) [Expansiveness] *There exist constant $\delta_m, \delta, \gamma_m, s_0 < 1$ such that, for any $m \in \Sigma$, for any $x \in \mathcal{I}$, for any $s > s_0$, one has : $0 < \gamma_m < |h'_m(x)| \leq \delta_m \leq \delta < 1$, with $\sum_{m \in \Sigma} \delta_m^s < +\infty$.*

(P3) [Bounded distorsion]

$$\exists K, \forall m \in \Sigma, \forall x \in \mathcal{I}, |h''_m(x)| \leq K |h'_m(x)|.$$

[Note that (P3) and the second part of (P2) are always true for a finite alphabet].

Classical sources. For instance, the classical sources (memoryless sources, or Markov chains) are related to particular dynamical sources; more precisely, a memoryless source corresponds to a complete system, with affine branches, and a uniform initial density. A Markov chain corresponds to a markovian system, with affine branches and an initial density which is piecewise constant on the partition \mathcal{I}_m .

B. Dirichlet series of fundamental measures.

The branches of T^k , and also its inverse branches, are then indexed by Σ^k , and, for any $w = m_1 \dots m_k \in \Sigma^k$, the mapping $h_w := h_{m_1} \circ h_{m_2} \circ \dots \circ h_{m_k}$ is a C^2 bijection from \mathcal{J}_w onto \mathcal{I}_w . It is possible that the word w cannot be produced by the source: this means that \mathcal{J}_w is empty, and the inverse branch h_w does not exist. All the words that begin with the same prefix w correspond to real numbers x that belong to the same interval \mathcal{I}_w . The probability p_w that a word begins with prefix w is called the *fundamental probability* of w .

The Dirichlet series of fundamental probabilities, denoted by $\Lambda(s), \Lambda(k, s)$, are defined (for complex s) as

$$\Lambda_k(\mathcal{S}_F, s) := \sum_{w \in \Sigma^k} p_w^s, \quad \Lambda(\mathcal{S}_F, s) := \sum_{w \in \Sigma^*} p_w^s \quad (1)$$

and will play a fundamental rôle in the analysis of all the probabilistic features of the source \mathcal{S}_F [see Section II].

We now define the generalized transfer operator relative to a dynamical system \mathcal{S} , and show how it generates the Dirichlet series of fundamental measures.

The density transformer \mathcal{G} of a dynamical system describes the evolution of densities on \mathcal{I} under iterations of T : if X is a random variable with density f , then TX has density $\mathcal{G}[f]$. The density transformer \mathcal{G} can be expressed as

$$\mathcal{G}[f](x) = \sum_{m \in \Sigma} |h'_m(x)| \cdot f \circ h_m(x) \cdot \mathbb{1}_{\mathcal{J}_m}(x), \quad (2)$$

and acts on functions of one variable $x \in \mathcal{I}$. Here, we use extensions of the density transformer and we give them the rôle of generating operators.

C. Generation of p_w^s .

We wish to generate the quantities p_w^s , for some (complex s). The new tool is the generalized transfer operator that involves secants of inverse branches

$$H(x, y) := \left| \frac{h(x) - h(y)}{x - y} \right| \quad (3)$$

instead of tangents $|h'(x)|$ of inverse branches as in (2). Each component operator, denoted by $\mathbf{G}_{s,[w]}$, now defined with the secant H_w relative to branch h_w , acts on functions L of two variables $(x, y) \in \mathcal{I} \times \mathcal{I}$ in the following way

$$\mathbf{G}_{s,[w]}[L](x, y) := H_w^s(x, y) L(h_w(x), h_w(y)). \quad (4)$$

If now the function L is the secant of the distribution F ,

$$L(x, y) = \left| \frac{F(x) - F(y)}{x - y} \right|, \quad (5)$$

the relation

$$p_w^s = |F(h_w(0)) - F(h_w(1))|^s = \mathbf{G}_{s,[w]}[L^s](0, 1)$$

proves that $\mathbf{G}_{s,[w]}$ generates the quantity p_w^s , so that $\mathbf{G}_{s,[w]}$ is called the generating operator of w .

D. Generating operators

The generating operator relative to a collection \mathcal{L} of words is defined as the sum of all the generating operators relative to the words of \mathcal{L} . For two prefixes w, w' , the relation $p_{w.w'} = p_w p_{w'}$ is no longer true when the source has some memory, and is replaced by the following composition property

$$\mathbf{G}_{s,[w.w']} = \mathbf{G}_{s,[w']} \circ \mathbf{G}_{s,[w]}, \quad (6)$$

so that unions and Cartesian products of collections of words translate into sums and compositions of the associated generating operators. The generating operator \mathbf{G}_s of alphabet Σ

$$\mathbf{G}_s := \sum_{m \in \Sigma} \mathbf{G}_{s,[m]}, \quad (7)$$

plays a fundamental rôle here. Equation (6) entails that the k -th \mathbf{G}_s^k generates all the words of length k , and the quasi-inverse $(I - \mathbf{G}_s)^{-1}$ generates all the finite words. Then, alternative expressions for the Dirichlet series of fundamental probabilities (1) hold:

$$\Lambda_k(\mathcal{S}_F, s) = \mathbf{G}_s^k[L^s](0, 1), \quad (8)$$

$$\Lambda(\mathcal{S}_F, s) = (I - \mathbf{G}_s)^{-1}[L^s](0, 1).$$

Moreover, the operator \mathbf{G}_s extends the density transformer as follows: If ℓ denotes the diagonal mapping of L , defined by $\ell(x) := L(x, x)$, one obtains

$$H_w(x, x) = |h'_w(x)|, \quad \mathbf{G}_1[L](x, x) = \mathcal{G}[\ell](x). \quad (9)$$

E. Spectral properties of the transfer operators.

Asymptotic properties of powers \mathbf{G}_s^k of the transfer operator are thus needed and they are closely related to *dominant spectral properties* of operator \mathbf{G}_s . In the same vein, asymptotic analysis of coefficients of the quasi-inverse $(I - \mathbf{G}_s)^{-1}$ is dependent on the location of poles of \mathbf{G}_s . Such poles arise from values of s where $(I - \mathbf{G}_s)^{-1}$ is singular, that is, values s for which 1 is an eigenvalue of \mathbf{G}_s . In this way, the poles also relate to the spectral properties of the transfer operator. Under quite general hypotheses, and on a convenient functional space, the density transformer \mathcal{G} [and then \mathbf{G}_1 via (9)] admits $\lambda = 1$ as an eigenvalue of largest modulus. But, generally speaking, this is not a unique dominant eigenvalue isolated from the remainder of the spectrum.

Definition. [Decomposable source] *A dynamical system \mathcal{S} is said to be decomposable with respect to a Banach space \mathcal{L} iff the transfer operator \mathbf{G}_s of system \mathcal{S} , when acting on \mathcal{L} , possesses for real $s > s_0$ (with $s_0 < 1$) a unique dominant eigenvalue $\lambda(s)$ (real positive) separated from the remainder of the spectrum by a spectral gap, i.e., $\rho(s) := \sup\{|\lambda| ; \lambda \in \text{Sp } \mathbf{G}_s, \lambda \neq \lambda(s)\} < \lambda(s)$. A source \mathcal{S}_F formed with a decomposable system \mathcal{S} [with respect to \mathcal{L}] and a function F whose secant belongs to \mathcal{L} is said to be decomposable.*

Proposition 1. [Main properties for decomposable sources]. *Suppose that \mathcal{S}_F is decomposable. Then, the following is true:*

(a) *When s is near the real axis, $\Lambda_k(\mathcal{S}_F, s)$ asymptotically behaves as the k -th power of $\lambda(s)$. There exists $\mu < 1$, such that, when s is near a real s_0*

$$\Lambda_k(\mathcal{S}_F, s) = a(s) \cdot \lambda(s)^k \cdot [1 + O(\mu^k)], \quad \text{with a uniform } O.$$

(b) *The function $s \mapsto \Lambda(\mathcal{S}_F, s)$ is analytic in the plane $\Re(s) > 1$, and it has a simple pole at $s = 1$, with a residue equal to $-\lambda'(1)$.*

Proof. Due to the existence of the spectral gap, the operator \mathbf{G}_s decomposes into two parts, namely $\mathbf{G}_s = \lambda(s)\mathbf{P}_s + \mathbf{N}_s$, where \mathbf{P}_s is the projection of \mathbf{G}_s onto the dominant eigenspace generated by the dominant eigenfunction φ_s , and \mathbf{N}_s , relative to the remainder of the spectrum, has a spectral radius equal to $\rho(s)$, which is strictly less than $\lambda(s)$. This decomposition extends to the iterates \mathbf{G}_s^n and the quasi-inverse $(I - \mathbf{G}_s)^{-1}$, and proves the result.

The operator \mathbf{N}_1 describes the correlations of the source. A decomposable dynamical source is ergodic and mixing with an exponential rate equal to $\rho(1)$. Most of the classical sources, namely all the memoryless sources, and all the primitive Markov chains, are easily proven to be decomposable. We now present sufficient conditions (which involve properties described in I.A) under which a general dynamical source will be proven to be decomposable.

Proposition 2. [Sufficient conditions for decomposability] *A markovian dynamical system on a finite alphabet which satisfies (P1) and (P2) is decomposable, with respect to the space $\mathcal{C}_p^1(\mathcal{I})$ of piecewise functions of class \mathcal{C}^1 . A complete*

dynamical system on a (possible infinite) alphabet which satisfies (P2), (P3) is decomposable, with respect to the space $\mathcal{C}^1(\mathcal{I})$ of functions of class \mathcal{C}^1 .

In our results of Section II, some sources will be “exceptional”. There are two kinds of exceptional sources, both related to exceptional properties of the pressure $\Lambda(s) := \log \lambda(s)$. The first one arises when the pressure $\Lambda(s)$ is periodic and intervenes as an exception for Results C, E. The second one arises when the pressure $\Lambda(s)$ is affine and intervenes as an exception for Results B, E. We prove that exceptional sources are conjugated to the two simpler sources, memoryless sources or Markov chains, where all branches are affine and we show that exceptional sources can only be conjugated to one of these simpler models: a “complex” source cannot be exceptional.

Proposition 3. [Exceptional sources.] *The two following assertions are equivalent:*

- $\Lambda(s)$ is affine,
- $\exists s$ for which $\Lambda''(s) = 0$.

In this case, the source is said to be affine. The two following assertions are equivalent :

- $\exists t \neq 0$ for which $\Lambda(s)$ is periodic with period t ,
- $\exists t \neq 0$ for which $\Lambda(s + it) = 1$.

In this case, the source is said to be periodic. Any source affine or periodic is called exceptional. An exceptional source is conjugated to a source with affine branches.

II. MAIN STATISTICS FOR DECOMPOSABLE SOURCES.

A. Entropy

The entropy $h(\mathcal{S}_F)$ of the source \mathcal{S}_F is defined as the limit, if it exists, of a quantity that involves probabilities p_w ,

$$\begin{aligned} h(\mathcal{S}_F) &:= \lim_{k \rightarrow \infty} \frac{-1}{k} \sum_{w \in \Sigma^k} p_w \log p_w \\ &= \lim_{k \rightarrow \infty} \frac{-1}{k} \frac{d}{ds} \Lambda_k(\mathcal{S}_F, s) \Big|_{s=1}. \end{aligned} \quad (10)$$

Result A. *The entropy $h(\mathcal{S}_F)$ of a decomposable source \mathcal{S}_F equals $-\lambda'(1)$. It only depends on the mechanism \mathcal{S} .*

B. Distribution of prefixes with fixed length k .

We describe the distribution of the fundamental probabilities of the words of Σ^k . More precisely, we study the random variable ℓ_k defined as a step function, constant on any fundamental interval \mathcal{I}_w of depth k and equal to p_w . Since the fundamental intervals of depth k form a quasi-partition of \mathcal{I} , the random variable ℓ_k is almost everywhere defined on \mathcal{I} . We examine characteristics of its distribution when x is distributed over the interval \mathcal{I} according to density f , and use its moment generating function $\mathbb{E}[\exp(s \log \ell_k)] = \mathbb{E}[\ell_k^s]$ which is closely related to $\Lambda_k(\mathcal{S}_F, s)$:

$$\mathbb{E}[\ell_k^s] = \sum_{w \in \Sigma^k} p_w^s p_w = \sum_{w \in \Sigma^k} p_w^{1+s} = \Lambda_k(\mathcal{S}_F, 1 + s). \quad (11)$$

Then, when \mathcal{S}_F is decomposable, the moment generating function $\mathbb{E}[\exp(s \log \ell_k)]$ behaves nearly like the k -th power of function $\lambda(1 + s)$. The central limit theorem of probability

theory asserts that exact large powers induce Gaussian laws in the asymptotic limit. Here, we shall use an extension of the central limit theorem to “quasi-powers” which has been developed in a general setting by Hwang [13]. This extension is valid provided that $\Lambda''(1)$ is not zero. We know that this is the case for most of the decomposable sources, but not for those that behave like an unbiased Bernoulli source.

Result B. *The distribution of probabilities of words of Σ^k follows asymptotically a log-normal law, for all not affine sources. This proves a strong “equipartition property”, in the flavour of the Shannon–MacMillan–Breimann Theorem. The dominant behaviours of the mean and the variance involve the first two derivatives of $\log \lambda(s)$ at $s = 1$: the mean is asymptotic to $k\lambda'(1)$ and the variance is asymptotic to $k[\lambda''(1) - \lambda'(1)^2]$.*

C. Asymptotic behaviour of the number of the most probable prefixes.

Are then many words whose probability is at least x ? We study here the asymptotic behaviour of functions B, A defined as

$$B(x) = \sum_{w: p_w \geq x} 1, \quad A(y) := B(e^{-y}),$$

when x tends to 0 [or y tends to ∞]. The central rôle is now played by the function $\Lambda(\mathcal{S}_F, s)$ which is both the Mellin transform [see paragraph E] of function B , and the Laplace transform of function A . Thus, the location of poles of $\Lambda(\mathcal{S}_F, s)$ provides a precise knowledge on the asymptotics of B near 0.

Result C. *The number $B(x)$ of finite prefixes satisfies, for “most” of the sources,*

$$B(x) \simeq \frac{1}{h(\mathcal{S})} \frac{D(x)}{x} \quad \text{for } x \rightarrow 0.$$

If \mathcal{S} is not periodic, one has $D(x) = 1$; if \mathcal{S} is periodic, $D(x)$ is a function which satisfies $0 < D_1 \leq D(x) \leq D_2$ for some constants D_1 and D_2 .

D. Longest common prefix of two words

For two words $M(x)$ and $M(y)$, one defines $C(x, y)$ to be the length of their longest common prefix,

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

Then C is a random variable almost everywhere defined on the square $\mathcal{I} \times \mathcal{I}$. The event $[C(x, y) \geq k]$ equals $\{(x, y) \in \mathcal{I} \times \mathcal{I}; P_k(x) = P_k(y)\}$. Then x and y belong to the same fundamental interval of depth k , and

$$[C \geq k] = \bigcup_{w \in \Sigma^k} \mathcal{I}_w \times \mathcal{I}_w.$$

In the case when the two words are independently drawn from the same probabilistic dynamical source \mathcal{S}_F , the density $g(x, y)$ on the square equals the product of densities $f(x) \cdot f(y)$ so that

$$\mathbb{P}[C \geq k] = \sum_{w \in \Sigma^k} p_w^2 = \Lambda_k(\mathcal{S}_F, 2).$$

Again, the Dirichlet series of fundamental measures appear, now at $s = 2$; More general drawings, which involve a density function on the square of the form $g(x, y) = h(x, y) \cdot |x - y|^r$ with $h(x, x) \neq 0$ and $r > -1$ are called to be of valuation r . In this case, the distribution of coincidence C deals with the Dirichlet series $\Lambda_k(\mathcal{S}_F, 2 + r)$.

Result D. *The coincidence between two source words, i.e., the length of their longest common prefix, follows a geometric law. The ratio of the geometric law depends on the drawing of the words. If the two words are independently drawn, then the ratio equals $\lambda(2)$. The quantity $c(\mathcal{S})$ is called the coincidence probability of the source. It only depends on the mechanism of the source. More generally, the ratio equals $\lambda(2 + r)$ where $r > -1$ is a parameter which is related to the valuation of drawing of the words.*

E. Average shape of a trie.

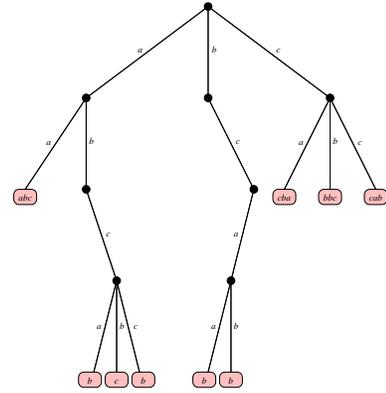


Fig. 2. An example of a trie on alphabet $\Sigma = \{a, b, c\}$ built on the set of words $\{aaabc, abca, abcb, abcc, bcaa, bcab, ca, cb, cc\}$.

In pattern matching problems, an essential data structure is a digital tree, called trie, which plays the rôle of a dictionary. The internal nodes are used for directing the search, and the leaves contain the words of the dictionary. There are as many leaves as words in the dictionary. A node can be labelled by the path which relates it to the root [see Fig 2]. Trie analysis aims at describing the average shape of a trie [size S , external path length P , height H]; it is extensively done in the case of classical sources [see [16] for instance]. In our context, a trie built on a set of n words is characterized by the set X of reals $\{x_1, x_2, \dots, x_n\}$ which have created the words [via the function M of Section I.A. It is denoted by $T(X)$. Such a trie is completely defined by the internal nodes which *actually* exist, and these nodes are labelled by prefixes w for which the interval \mathcal{I}_w contains at least two elements of X . We denote by N_w the cardinality of $\mathcal{I}_w \cap X$. The probability that node n_w exists is just $\mathbb{P}[N_w \geq 2]$ while the contribution of node n_w to the external path length is $\mathbb{E}[N_w \mid N_w \geq 2]$. If we wish the contributions of set X to two disjoint intervals \mathcal{I}_w and $\mathcal{I}_{w'}$ be independent, we are led to first work in a Poisson model: first, one draws the cardinality N of the set X with respect to a Poisson law, with parameter z ,

$$\mathbb{P}[N = n] = e^{-z} \frac{z^n}{n!},$$

then one independently draws the n reals of set X with respect to density f . In this case, the random variable N_w itself follows a Poisson law of parameter $p_w z$, and the expectation of the size S and the external path length P admit the following expressions [z refers to the parameter of the Poisson model]:

$$\begin{aligned} \mathbb{E}[P_{(z)}] &= \sum_{w \in \Sigma^*} p_w z (1 - e^{-p_w z}), \\ \mathbb{E}[S_{(z)}] &= \sum_{w \in \Sigma^*} 1 - e^{-p_w z} (1 + p_w z). \end{aligned}$$

The two previous expressions are harmonic sums, and the convenient tool for studying such sums is the Mellin transform (see [10]) because the Mellin transform of an harmonic sum is a product of two factors:

$$\text{if } A(z) = \sum_w g(p_w z), \quad \text{then } \hat{A}(s) = \hat{g}(s) \cdot \left(\sum_w p_w^{-s} \right).$$

Then, the Mellin transforms of $\mathbb{E}[P_{(z)}], \mathbb{E}[S_{(z)}]$ involve the Dirichlet series $\Lambda(\mathcal{S}_F, s)$, and particularly its singular behaviour near $\Re s = 1$, notably its residue $h(\mathcal{S})$.

A third interesting parameter is the height, which is at most equal to k provided that there do not exist internal nodes n_w relative to $w \in \Sigma^k$. Due to the independence in the Poisson model, one has

$$\mathbb{P}[H_{(z)} \leq k] = \prod_{w \in \Sigma^k} \mathbb{P}[N_w \leq 1] = \prod_{w \in \Sigma^k} e^{-p_w z} (1 + p_w z)$$

$$\text{so that } \log \mathbb{P}[H_{(z)} \leq k] = \sum_{w \in \Sigma^k} (-p_w z + \log(1 + p_w z)).$$

Suppose that the two following approximations can be used

$$-p_w z + \log(1 + p_w z) \sim \frac{1}{2} p_w^2 z^2, \quad \sum_{w \in \Sigma^k} p_w^2 \sim a \lambda(2)^k.$$

Then, the mean value of the height is

$$\mathbb{E}[H_{(z)}] \sim \sum_{k \geq 0} \left(1 - \exp\left(-\frac{az^2}{2} \lambda^k(2)\right) \right)$$

which is an harmonic sum. The relative Dirichlet series

$$\sum_{k \geq 0} \lambda(2)^{-ks} = \frac{1}{1 - \lambda(2)^{-s}}$$

has a simple pôle at $s = 0$, with a residue which involves $|\log \lambda(2)|$.

All these facts can be actually proven (in the Poisson model). Then, it is possible to return to the Bernoulli model (where the cardinality n of the number of words is fixed). One obtains:

Result E. *The three shape parameters of a trie [size, external length path, height] built on n words [independently drawn from the source] admit as asymptotic mean (for $n \rightarrow \infty$) the following quantities which involve the entropy or the coincidence probability*

$$\mathbb{E}[S_n] \sim \frac{n}{h(\mathcal{S})}, \quad \mathbb{E}[P_n] \sim \frac{n \log n}{h(\mathcal{S})} \quad \mathbb{E}[H_n] \sim \frac{\log n}{2|\log c(\mathcal{S})|}.$$

For a periodic source, the main asymptotic terms of $\mathbb{E}[S_n]$ and $\mathbb{E}[H_n]$ involve a periodic function of $\log n$ of small amplitude.

Bibliographic notes. A nice book about recent research in dynamical systems is [2]. Dynamical sources have been introduced in [17], where all the first four results of Section II are proven. Some hypotheses for proving decomposability are made precise in [5], and a general decomposability criterion is proven in [6]. Tries analysis is performed in [8], [7]. Devroye [9] had already observed the dependence of the shape of tries under the distribution of reals. Coincidence of words is motivated by an algorithm which compares rationals via their continued fraction expansions [1], [3], [14] and is first analysed in [18]. Initial statement of the Shannon-MacMillan-Breimann Theorem can be found in [15]. The continued fraction expansion motivates the introduction of dynamical sources. For a specific treatment of this case, see [11], [12], where a relation between the shape of tries and the Riemann hypothesis is exhibited...

REFERENCES

- [1] AVNAIM, F., BOISSONNAT, J.-D., DEVILLERS, O., PREPARATA, F., AND YVINEC, M. Evaluation of a new method to compute signs of determinants. In *Eleventh Annual ACM Symposium on Computational Geometry* (1995), pp. C16–C17. Full paper in *Algorithmica*, 1997.
- [2] BEDFORD, T., KEANE, M., AND SERIES, C., Eds. *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*. Oxford University Press, 1991.
- [3] BEELER, M., GOSPER, R. W., AND SCHROEPPLE, R. HAKMEM. Memorandum 239, M.I.T., Artificial Intelligence Laboratory, Feb. 1972.
- [4] BILLINGSLEY, P. *Ergodic Theory and Information* John Wiley & Sons, 1965.
- [5] CHAZAL, F., MAUME-DESCHAMPS, V., VALLÉE, B. Erratum to “Dynamical sources in information theory: fundamental intervals and word prefixes”, *Algorithmica* 38 pp 591–596 (2004).
- [6] CHAZAL, F., MAUME-DESCHAMPS, V. Statistical of Markov dynamical sources: applications to information theory *Discrete Mathematics and Computer Science*, 6, 2004, 283–314.
- [7] CLÉMENT, J., FLAJOLET P., VALLÉE, B. Dynamical sources in Information Theory: Analysis of general tries, *Algorithmica* (2001), vol 29 (1/2) pp 307–369
- [8] CLÉMENT, J. Arbres digitaux et sources dynamiques, PhD, University of Caen (2000)
- [9] DEVROYE, L. A probabilistic analysis of the height of tries and of the complexity of triesort. *Acta Informatica*, vol 21, pp 229–237
- [10] FLAJOLET, P., GOURDON, X., AND DUMAS, P. Mellin transforms and asymptotics : Harmonic sums. *Theoretical Computer Science* 144, 1–2 (June 1995), 3–58.
- [11] FLAJOLET, P., VALLÉE, B. Continued fraction algorithms, functional operators, and structure constants, *Theoretical Computer Science* (1998), vol 194, 1–2, pp 1–34.
- [12] FLAJOLET, P., VALLÉE, B. Continued Fractions, Comparison Algorithms, and Fine Structure Constants, *Constructive, Experimental et Non-Linear Analysis*, Michel Thera, Editor, Proceedings of Canadian Mathematical Society, Vol 27 (2000), pages 53–82
- [13] HWANG, H.-K. On convergence rates in the central limit theorems for combinatorial structures, *European Journal of Combinatorics* 19 (1998) 329–343.
- [14] KNUTH, D.E. Volume D of *Computers and Typesetting*, MF: the program Addison Wesley, Reading, Massachusetts, 1986.
- [15] SHANNON, C. A mathematical theory of communication. *Bell System Tech. J.* (27) 379–423, 623–656 (1948)
- [16] SZPANKOWSKI, W. *Average-case analysis on sequences* John Wiley and sons, New-York, 2001.
- [17] VALLÉE, B. Dynamical sources in Information Theory: Fundamental Intervals and Word prefixes, *Algorithmica* (2001), vol 29 (1/2) pp 262–306
- [18] VALLÉE, B. Algorithms for computing signs of 2×2 determinants: dynamics and average-case algorithms, Proceedings of ESA’97, pp 486–499, LNCS 1284, Springer Verlag.