



Analysis of depth of digital trees built on general sources

Kanal Hun

► To cite this version:

Kanal Hun. Analysis of depth of digital trees built on general sources. Computer Science [cs]. Université de Caen Basse-Normandie, 2014. English. ⟨NNT : ⟩. ⟨tel-01136777⟩

HAL Id: tel-01136777

<https://hal.science/tel-01136777v1>

Submitted on 28 Mar 2015

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.



HAL Authorization



Université de Caen Basse-Normandie

Ecole doctorale SIMEM

Thèse de doctorat

présentée et soutenue le 15 décembre 2014

par

Kanal HUN

en vue de l'obtention du

Doctorat de l'Université de Caen Basse-Normandie

Spécialité : Informatique

Analysis of depth of digital trees built on general sources

Jury:

Hsien-Kuei HWANG	Directeur de recherche, Academia Sinica, Taiwan	<i>Rapporteur</i>
Ralph NEININGER	Professeur, Université de Frankfort	<i>Rapporteur</i>
Pierre ARNOUX	Professeur, Institut de mathématique de Luminy, Marseille	<i>Examineur</i>
Julien CLÉMENT	Chargé de recherche CNRS, Université de Caen	<i>Examineur</i>
Philippe JACQUET	Directeur de recherche, Bell Labs, Alcatel-Lucent, Paris	<i>Examineur</i>
Brigitte VALLÉE	Directrice de recherche CNRS, Université de Caen	<i>Directrice de thèse</i>

Acknowledgments

First of all, I would like to express my deep thanks to my PhD supervisor, Brigitte Vallée, for having proposed me this interesting research topic and provided me a good supervision. This thesis could not have been realized without her patience, guidance, assistance, support and encouragement. She has helped me through extremely difficult times over the course of the analysis and the writing of the dissertation and for that I sincerely thank her for her confidence in me. Deep thanks also go to Philippe Flajolet (1948-2011) who proposed the project with Brigitte Vallée. I still remembered his several phone calls discussing about that project, especially his wishes for my wedding's day. I wish I could have been under his co-supervision.

My thanks next go to the main supporter, the Agence universitaire de la Francophonie (AUF) to Cambodia for providing me the scholarship. For this, I shall thank Dr. Seam Ngonn, the correspondent at the Royal University of Phnom Penh, who helped with some administrative documents. I wish to thank also the Doctoral School SIMEM for having admitted my PhD candidature and their help.

I shall express my warm thanks to Hsien-Kuei Hwang and Ralph Neininger who give me the great honor to accept to be my reviewers. Many thanks go to Pierre Arnoux, Philippe Jacquet, and Julien Clément for being the member of the jury. Additional thanks go to Julien Clément for helping editing my first chapter and for useful discussions at the beginning of my research.

I shall thank all the professors at the Department of Mathematics, Royal University of Phnom Penh, who had provided me some valuable mathematics backgrounds for my advance studies. I would additionally like to thank the International Center for Pure and Applied Mathematics (ICPAM) for promoting the program of Master of Science in Mathematics in Cambodia and for providing (or seeking) some financial supports for Cambodian students to study abroad. I shall thank to all the (french) professors involved with this Master program for their great lectures, good advice and recommendation, especially encouraging me to pursue advanced studies.

I would also like to extend my appreciation to all the members of the GREYC laboratory and the Computer Science department for their kind hospitality and helps during my stay. I can not forget to thank GREYC (especially for complementary grant) and CNRS for their financial support of all my missions, .

Finally, I would like to extend my deepest gratitude to my parents, Sim Hun and Lak Hong, for giving birth to me, their hard working, love, and (financial and spiritual) support , and also to my wife, without whose love, patience, support and understanding I could never have completed this doctoral degree.

Dedicated to my parents, wife and my son.
Dedicated also to Philippe Flajolet's memory.

Contents

Introduction	10
1 Tries and digital search trees.	11
1.1 Binary Search Trees	13
1.1.1 Definition	14
1.1.2 Main operations	14
1.2 Tries	15
1.2.1 Definition	16
1.2.2 Dynamic Structure	16
1.2.3 Suffix trees	18
1.3 Digital Search Trees	18
1.3.1 Definition	19
1.3.2 Dynamic Structure	19
1.3.3 Relation between tries and DSTs	20
1.3.4 A unified view for parameters for tries and dst's.	20
1.4 Symbol-Comparisons BST	21
1.5 Data Compression	22
1.5.1 Coding	22
1.5.2 Huffman coding	22
1.5.3 Arithmetic Coding	24
1.5.4 Dictionary Compression Algorithms	26
1.5.5 LZ 77	26
1.5.6 LZ78	27
2 Sources	31
2.1 Generalities on sources.	32
2.1.1 $\mathbb{N}$ -history and $\mathbb{Z}$ -history	32
2.1.2 Fundamental probabilities and generating function of the source.	32
2.1.3 Entropy of the source.	33
2.2 Simple sources	33
2.2.1 Memoryless sources.	33
2.2.2 Markov chains.	34
2.3 General sources: three models.	37
2.3.1 Sequence of sources and generalized transition matrix.	37
2.3.2 A dynamical point of view.	40
2.3.3 Parameterization of a source.	44
2.4 Relations between the three points of view.	46
2.4.1 The mirror operation and the operator $\hat{\mathbf{P}}_s$	47

2.4.2	Extension to “infinite past”. The source $\underline{S}$.	49
2.4.3	The dynamical system $\mathcal{D}$ and the transfer operator $\mathbb{H}_s$	50
3	Tameness of Sources	55
3.1	Generalities.	56
3.1.1	Role of the quasi-inverse $\mathbb{T}(s, u)$	56
3.1.2	The trie operator and the dst operator.	57
3.1.3	Our needs for using the Rice methodology.	57
3.1.4	Definitions of tameness.	58
3.1.5	Plan of the chapter.	58
3.2	Periodicity and Quasi-periodicity.	59
3.2.1	Case of a memoryless source.	59
3.2.2	Case of a Markov chain.	61
3.2.3	Case of a dynamical source.	63
3.2.4	Quasi periodicity is exceptional.	64
3.2.5	An intermediary notion : p -periodicity	64
3.2.6	Contrapositive versions of the statements about periodicity.	65
3.3	Main principles for tameness.	65
3.3.1	Geometric conditions on branches. The UNI Class.	65
3.3.2	Arithmetical conditions on branches. The DIOP Classes.	66
3.3.3	Shape of tameness regions.	67
3.4	Instances of tameness.	68
3.4.1	The periodic case. Tameness in vertical strips with an infinite number of poles.	68
3.4.2	The UNI Class. Tameness in vertical strips.	68
3.4.3	The DIOP Classes. Tameness in hyperbolic regions.	69
3.4.4	Conclusion of tameness study for classical sources.	70
3.4.5	A small piece of history.	70
4	Main tools from analysis of algorithms	73
4.1	Bernoulli and Poisson models.	74
4.1.1	Bernoulli model.	74
4.1.2	Poisson model.	75
4.1.3	Relation between the two models. The Poisson Generating Function	75
4.1.4	Interest of the Poisson model.	75
4.1.5	Binomial recurrence.	76
4.1.6	Return from the Poisson model to the Bernoulli model	77
4.2	Mellin Transform	77
4.2.1	Basic Properties	77
4.2.2	Functional properties.	79
4.3	The Poisson-Mellin-Newton-Rice cycle	80
4.3.1	The Valuation-Degree Condition	80
4.3.2	Existence and expression of the analytic lifting $\varpi(s)$	80
4.4	Rice’s methodology	81
4.4.1	First step: An integral form.	81
4.4.2	Second step: shifting to the left.	83
4.4.3	Tameness of $\varpi(s)$.	83
4.4.4	Asymptotic estimates for B_n .	84
4.5	Proof of the remainder estimates in the Rice method.	86

4.5.1	The statement	86
4.5.2	Estimates near the real axis.	86
4.5.3	Far from the real axis.	87
4.5.4	Proof of Proposition 4.16.	88
4.6	Depoissonization	89
4.6.1	A general depoissonization result.	89
4.6.2	The $\mathcal{JS}$ conditions.	90
4.6.3	Analytic extension of the sequence B_n	90
4.7	Laplace Transform	91
5	Profile and depth for simple sources – Algebraic analyses	93
5.1	Main objects of the algebraic study	94
5.1.1	Parameters of a digital tree.	94
5.1.2	Generating functions.	94
5.1.3	Outline of the method.	95
5.2	Functional equations for Poisson generating functions	96
5.2.1	Case of memoryless sources	96
5.2.2	Case of Markov chains	98
5.3	Analytic liftings via Mellin transforms of Poisson generating functions.	99
5.3.1	Memoryless case.	99
5.3.2	Markov Chains case.	100
5.4	Exact expressions of the expectations in the Bernoulli model	103
5.4.1	Expressions as binomial sums	103
5.4.2	Integral expressions for the probability generating functions.	103
5.4.3	An additive expression for $\Delta_D(s, u)$ in the unbiased memoryless case.	104
5.4.4	An exact expression for the profile in the unbiased memoryless case.	105
6	Profile and depth for a general source – Algebraic Analysis	107
6.1	General strategy	108
6.1.1	A sequence of shifted sources.	108
6.1.2	A sequence of generating functions.	108
6.1.3	General strategy for the algebraic study.	109
6.2	The basic recurrence and the system of functional equations.	110
6.3	The expression of the analytic lifting as a product.	112
6.4	Expression of the analytic lifting as a series.	114
6.5	Final expressions in the Bernoulli model.	117
6.5.1	Expression as binomial sums.	117
6.5.2	Expression as Rice's integrals.	118
6.5.3	Expression of the sequence $B_{n,k}$	118
6.5.4	Towards to the next chapter	119
7	Distribution of the depth	121
7.1	Statements of our main results.	122
7.1.1	Asymptotic estimates for the mean and the variance.	122
7.1.2	Asymptotic Gaussian law.	123
7.2	Probabilistic theorems for the study of asymptotic distribution.	123
7.2.1	Goncharov Theorem	123
7.2.2	Speed of convergence towards the normal law	124
7.2.3	Quasi-Power Theorem	124

7.2.4	Our study	124
7.3	Beginning the study	125
7.3.1	The mixed Dirichlet series.	125
7.3.2	Tameness of the mixed Dirichlet series on the right of $\Re s = 1$	125
7.3.3	An integral expression for the generating function $G_n(u)$	126
7.4	Applying the Rice method.	127
7.4.1	Properties of the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$ for (s, u) close to $(1, 1)$	127
7.4.2	Main functions of the analysis.	128
7.4.3	Various types of tameness.	128
7.5	Three estimates.	129
7.5.1	An expression for the moment generating function of the depth D_n	129
7.5.2	An expression for the cumulant generating function	131
7.5.3	An expression for the normalized characteristic function.	133
7.6	Proofs of the main results	134
7.6.1	Asymptotic estimates for the mean value $\mathbb{E}[D_n]$ and the variance $\mathbb{V}[D_n]$	134
7.6.2	Asymptotic Gaussian law for the depth	134
7.6.3	Speed of convergence.	135
7.6.4	Expression of the subdominant constants.	137
7.6.5	Conclusion of the Chapter.	138
Conclusion		1
A Analytic and meromorphic functions		5
A.1	Analytic functions	5
A.2	Meromorphic Functions and Residues.	6
A.3	Gamma function.	7
A.4	Dirichlet series	8
A.4.1	Classical Dirichlet series.	8
A.4.2	Our Dirichlet series	9
Bibliography		15

Introduction

This thesis is devoted to the analysis of digital trees (tries and digital search trees) when they are built on words that are independently emitted by a general source. Digital trees constitute a fundamental structure in algorithmics and theoretical computer science and they also play a central role in compression algorithms of Lempel-Ziv type. We here perform a probabilistic analysis of these two structures, and we are mainly interested in two specific parameters: the (typical) depth that is the length of a random branch of the tree, and the profile, which counts the number of nodes at each level. Such parameters, and notably the depth, well describe the shape of these digital trees, and thus play an important role in the analysis of the performance of the algorithms which use these trees as underlying structures.

A (very) short summary of these results has been presented in the ANALCO14 conference and appeared in the proceedings of this conference [29].

Context of the study.

Digital trees. Tries and digital search trees are two different data structures which share some important principles and implement a digital tree. The structure of digital tree is central in theoretical computer science. It contains words and plays the same role as a dictionary, as it explicitly uses the representation of a word as a sequence of symbols. Such a structure underlies the main algorithms which deal with words for searching and sorting, or performing data compression. It also appears in many other applications. It is thus important to better understand its “shape” which has a great influence on the algorithms which use it as an underlying structure. This could lead to important algorithmic improvements.

The two structures share many important features: They both contain words, they are defined in a recursive way, and the words are directed towards the various subtrees according to their first symbol. Both structures can be viewed as a tool that “distinguishes” a set of words, and they are “dynamic” as they evolve in an efficient way, when the set of words itself evolves. Their efficiency is closely related to the “compactness” of their shape (see figure 1). However, the two digital trees are not built on the same principles: in a digital search tree, the words are placed at internal nodes, whereas, in a trie, only the external nodes contain words (the internal nodes are just used to direct the words). The dst can be viewed as an hybrid between the trie and the binary search tree, and it is a more compact structure than the trie and it seems more efficient than the trie, at least from a theoretical point of view.

The efficiency of a digital tree also depends on the process that creates the words it contains. As mentioned above, a digital tree aims at “distinguishing” a set of words and if the process creates words that are very “similar”, the task of building the digital tree for efficient searching and maintenance becomes harder. This is why a digital tree cannot be studied independently of the process which creates the words. Such a process is called a source.

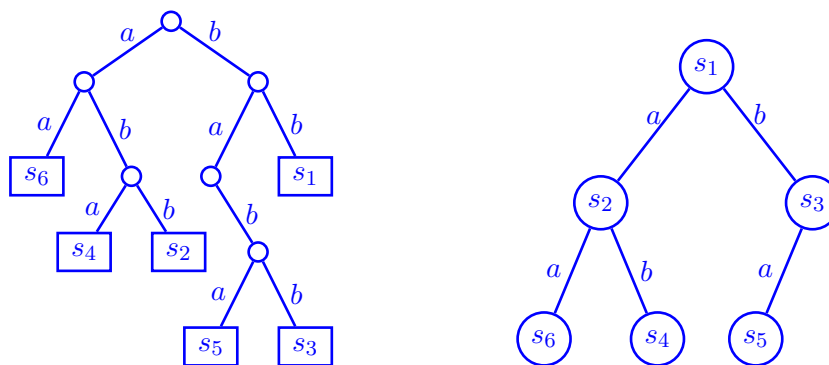


Figure 1: A trie on the left and a dst on the right

Sources. In information theory contexts, a source is a process which emits symbols, one symbol at each discrete time unit. One often deals with simple sources (memoryless sources or Markov chains) where the correlation between successive symbols is the weakest possible. Memoryless sources (where the symbols are independently drawn) or Markov chains (where the dependency between symbols is bounded). We have already mentioned that the efficiency of a digital tree depends on the possible correlations between symbols. This is why it is natural to perform the study of a digital tree on more correlated sources, where the emission of a given symbol may depend on the whole previous history.

Probabilistic analysis. Probabilistic analysis of data structures aims to study the probabilistic behaviour of such structures, in relation with the algorithms that use them as underlying structures. One often isolates a parameter of the data structure, which plays a central role in the related algorithms. For trees, this is often the length of a (random) branch of the tree, often called the (typical) depth. Then, for digital trees, a main parameter is the depth, together with a closely related parameter, called the profile, which counts the number of nodes at each level. As this parameter (depth, profile) becomes a random variable, one conducts a probabilistic study when the size n of the problem (here the number of words contained in the digital structure) becomes large. The average-case analysis focuses on the mean (and often the variance) and aims at obtaining asymptotic estimates. Distributional analysis describes the asymptotic law followed by the parameter, and estimates the speed of convergence towards this law, when n becomes large.

Analytic combinatorics. We work in the analytic combinatorics domain which is based on the use of generating functions. There are two main steps in analytic combinatorics, as it is described in the book of Flajolet and Sedgewick [24]: an *algebraic* or *combinatorial* part and an *asymptotic* and *analytic* step. The first part computes a mathematical formal object (most of the time a generating function) with algebraic and exact tools, and the second part deals with the generating function viewed as a function of the complex variable, studies its singularities, and transfers this knowledge about its singularities into an asymptotic expansion.

State of the art and motivations

We now describe the main results which exist in the probabilistic analysis of digital trees and use methods from Analytic Combinatorics. Even with this restriction, there exist many, many

papers, which form a kind of jungle. We use the term “jungle” in its literal sense: a very dense landscape, where there are so many paths that it is not easy to choose a path.

An attempt for a possible classification. It is thus very difficult to describe the existing results in a synthetic way, (without omitting not too many works), because the performed analysis may depend on (at least...) six main types of choice:

- the type of the *digital tree* : a trie? or a digital search tree?
- the type of *parameter* : the depth? the path length? the profile? the size (for a trie)?
- the type of *source* : a memoryless source ? biased or not biased ? a Markov chain? a more general source?
- the type of the *probabilistic study* which is conducted : a study on the mean? on the variance? on the asymptotic law? does it include the speed of convergence?
- the *methodology* which is used : Rice’s method? Depoissonnization techniques? Laplace transform?
- and finally the possible *application* to a specific algorithmic problem: for instance, the compression algorithms of Lempel-Ziv type?

A very incomplete review of the existing results. The story started with the work of De La Briandais [10] and Coffman and Eve [38]. Then Knuth performed an important study which is included in his *Art of Computer Programming, Volume III* [43]. Then, there was the seminal paper due to Flajolet and Sedgewick (1986) [23] who conducted the analysis of the two structures (together with the Patricia trie) with modern tools. Even if their analysis was restricted to the binary source and focusses on the study of the analysis of the mean depth, it provided all the main ideas that were used later on. In particular, they introduce the Rice method into the domain of analytic combinatorics. After that, many average-case analysis studies and then distributional studies had been extensively carried out. The book of Szpankowski [71] provides a complete review of these results, which are due to a large number of people. For simple sources (memoryless sources or Markov chains) the probabilistic behaviour of depth or other main tree parameters (for tries and dst’s) is now well analyzed.

Tries. Tries are easier to analyze than digital search trees, and this explains why the first analyses deal with tries. The distributional analysis of the depth for the unbiased memoryless source was first performed by Jacquet et Régnier [31, 32, 59]. They performed the limiting distributions for the size, the depth and the height. Then Szpankowski [69] analyzed the depth and path length for general memoryless sources. He obtained all the factorial moments of the depth. After that, in 1991, the paper of Jacquet and Szpankowski [33] performed the distributional analysis of the depth in the Markov chain model. It exhibited an asymptotic gaussian law and obtained asymptotic estimates for the mean and variance. These results are mainly based on the *inclusion-exclusion rule*. There also exist analyses of other trie parameters such as size, path length, profile, height . . . , that are studied in [40, 46, 55, 65]. The paper [40] provided a further analysis of trie for an unbiased memoryless source which revisited the papers of Knuth [43] and Flajolet and Sedgewick [23]. The paper [65] studied limiting distributions for the costs of partial match retrievals in multidimensional tries. The paper [55] performed a distributional analysis of the profile in the memoryless case, according to different possible ranges for the level. Finally, the paper [46] studied the path length under the Markov model.

Dst’s. Even if the dst is more difficult to deal with, there are also many papers which analyze the dst, due to Drmota, Jacquet, Louchard, Prodinger, Szpankowski, Tang, published between

1990 and 2005. They dealt with general memoryless sources or Markov chains, and perform the analysis of the main parameters of dst's—namely, path length, profile, typical depth—(see for instance [21, 26, 37, 41, 48, 49, 57, 70]). For simple sources, they provided asymptotic estimates for the mean and variance of the depth and they also exhibited an asymptotic Gaussian law for the depth. The papers [37, 49] related the analysis of the Lempel-Ziv algorithms to the analysis of several parameters of digital search trees.

The *depth* seemed to be first analyzed in 1986 by Kirschenhofer and Prodinger [40]. They provided the asymptotics of the mean and variance for an unbiased memoryless source. After that, in 1994, again for an unbiased memoryless source, Kirschenhofer and al. in [41] revisited the paper [23], and provided the asymptotics of the variance of the path length and thus the depth. Then, Louchard [47] began the distributional study. He proved the limiting distribution in the unbiased memoryless case. Next, the paper of Louchard and Szpankowski ([48], 1995) extended the distributional analysis to a general memoryless source. They exhibited the Gaussian limiting distribution with an explicit computation of the asymptotic estimates of the mean and the variance. In fact, this result gave rise to another result in the Lempel-Ziv parsing algorithm, where they showed that the typical phrase length is asymptotically normally distributed. An extension to the Markov chain case has been done by Jacquet, Szpankowski and Tang [37]. They obtained asymptotic estimates for the mean and the variance of the phrase length. The results also led to bounds on the average redundancy of the Lempel-Ziv code.

It is worthwhile to mention a few other early works for the depth in the 1970's and 1980's, even if some of them deal with different contexts – extendible hashing or bucket sorting–: Konheim and Newman [44]), Yao [75], Fagin, Nivergelt, and Pippenger [16], Devroye [11], Mendelson [51], and Pittel [56].

The *profile* of dst seemed to be first analyzed in 2009 by Knessel and Szpankowski [42]. They study the average profile in the case of an unbiased memoryless source. They provided the asymptotic expansion of the average profile for various ranges of levels. Then, in 2011, Drmota and Szpankowski [14] extended the analysis to a general memoryless source. They obtained an exact expression of the mean profile in the unbiased case, then, in the biased case, they presented a precise analysis of the average profile by analytic methods, and provided an asymptotic estimate of the average profile. Their results also led to an “unusual” Local Limit Theorem for the depth of a dst in the memoryless case.

It is also worth to mention the analysis of other dst parameters as the path length, height, size For example, Fuchs, Lee and Prodinger [26] performed the analysis of the variance of path length for the unbiased memoryless source. Then, Hubalek [28] studied the variance of the dst path length in the general memoryless case. He gave the asymptotics of the mean and the variance and the method was based on the Mellin convolution approach. In relation to the Lempel-Ziv model, Jacquet, Szpankowski and Louchard [36] proved an asymptotic normal law for the path length in the memoryless case. The result was obtained by ‘renewal equation’ which related the number of phrases in the Lempel-Ziv78 through the path length of the associated digital search tree built over a fixed number of independent strings.

General sources. There exist also analyses which deal with general sources. The first work is due to Clément, Flajolet, Vallée [5, 7] who conducted an average-case analysis of tries, followed by the paper [2] which focuses on Patricia tries. Finally, Cesaratto and Vallée in [4] performed a distributional analysis of the trie depth for particular dynamical sources which satisfy a strong tameness property¹.

¹We will return to this notion which will be fundamental in the thesis.

All these previous results paved the way for a more complete and unified study

- The existing analyses deal with two similar structures, but they are not performed in a unified way: we here introduce a unified point of view for the two structures.
- The existing analyses deal with various types of sources (memoryless sources, Markov chains), but they are not performed in a unified way: we need a unified point of view for these two models of sources, and we wish to extend to a more general model of sources.
- The existing results are not precise enough for the remainder terms in the asymptotics of the mean and the variance. These remainder terms do not always seem completely proven, and perhaps sometimes inexact.
- The asymptotic gaussian law is always expected but is not always obtained in an explicit way. The speed of convergence is not always exhibited.
- The methods used are not completely transparent. There are two schools in Analytic Combinatorics: The Rice School and the DePoisson School which each strongly defends the methodology to be dealt with. But it is not clear to evaluate the advantages and the drawbacks of each method. Are they easy to compare?

Our framework.

Our main choices are as follows: We wish to analyze

- two types of *digital tree* (trie and dust) in a unified way.
- with a special focus on the typical depth and the profile.
- when the words are emitted by a *general* source.
- with a *complete probabilistic* study: mean, variance, asymptotic law, speed of convergence.
- with *the Rice methodology*.
- *We do not succeed* to extend our analysis to the compression algorithms.

Our analyses are based on five specificities.

Two steps in the analysis. The first step of the analysis (the combinatorial step) builds a central object of the analysis, that is called the *mixed Dirichlet series*², as it characterizes the behavior of the digital tree with respect to the source. Here, this series is a bivariate generating function, as It is usual in distributional analyses, denoted by $\Delta(s, u)$ and is of Dirichlet type “in s ”. Then, in its second step (the analytic step), our analysis uses the Rice method on this mixed Dirichlet series. However, the Rice method needs $\Delta(s, u)$ to be *tame*, and in particular to be of polynomial growth for $|s| \rightarrow \infty$. And, as $\Delta(s, u)$ depends on the source, its tameness will strongly depend on the tameness of the source.

Modelling a general source. We propose three points of view for modelling a source. In particular, we model a general source with an (infinite) matrix $\mathbf{P}$ which can be viewed as the transition matrix of the source. Then, the algebraic studies are performed with the infinite matrix $\mathbf{P}_s$ which extends the matrix $\mathbf{P}$, and we obtain exact formulae for our mixed Dirichlet series

²There are other important occurrences of such a mixed Dirichlet series in the recent works about the “realistic” analyses of sorting and searching algorithms, where the mixed Dirichlet series characterizes the behavior of the algorithm with respect to the source [6, 9, 74] .

which involve quasi-inverses of the form $(I - u\mathbf{P}_s)^{-1}$ and are exact generalisations of formulae which exist for a source with bounded memory, where the matrix $\mathbf{P}$ is of finite dimension. However, when we wish to perform analytic studies, we need “good” spectral properties of the operator $\mathbf{P}_s$, and we restrict ourselves to more specific sources, namely “good” dynamical sources. In this case, we obtain exact formulae for our mixed Dirichlet series which now involve quasi-inverses of the form $(I - u\mathbb{H}_s)^{-1}$ where $\mathbb{H}_s$ is now the secant transfer operator of the source. For good dynamical sources, such quasi-inverses can be precisely studied.

Two similar structures. Of course, there are two mixed Dirichlet series $\Delta(s, u)$, one for each digital tree –for trie, this is $\Delta_T(s, u)$, and for the dst, this is $\Delta_D(s, u)$ –. As the two structures are different, their mixed Dirichlet are not equal, but they both involve, as we already said, quasi-inverses. There is an infinite product of quasi-inverses for the dst and only one quasi-inverse for the trie: the difference between these two forms well reflect the difference of the two data structures. However, the two series are of the same type, as they only involve quasi-inverse, and they can be analyzed in parallel. And, at $s = 1$ which is the “key-point” for the analysis, the two series coincide. This explains the similarity of their probabilistic behaviours.

Central role played by the source tameness. Since scientists of analytic combinatorics study digital trees, they know that (a priori unexpected) arithmetical phenomena due to the source may play an important role in the analysis. A “periodicity” phenomenon occurs for a binary source with probabilities (p, q) when the ratio $\log p / \log q$ is rational, and creates a periodic term in the asymptotics. This phenomenon is well-known as it occurs for the simplest source, namely the unbiased binary source. The situation is more complex, and much less studied when the source is *not* periodic. In this case, the position of complex numbers s which are solutions of the equation $p^s + q^s = 1$ with respect to the vertical line $\Re s = 1$ is central in the analysis, and has a direct influence on the nature of the remainder terms of the asymptotics. This is why *non periodic* sources are much *more difficult* to deal with than periodic sources, and yet a “random” source is *almost always* non periodic.

The *tameness* of a source is closely related to these phenomena. This is a notion which is not yet popular in Analytic Combinatorics, even in average-case analyses, and this is yet more true for distributional analysis. Here, for instance, the position of complex numbers s solutions of the equation $u(p^s + q^s) = 1$ (with u close to 1) has to be studied. We thus need to *perturb* the previous notions of tameness in order to obtain a notion of uniform tameness. This is needed to conduct distributional studies when the source is not periodic. Finally, we exhibit three main types of tameness which arise in a natural way, already for simple sources, and also for dynamical sources, as it was proven in relation with the works of Dolgopyat [12, 13]. These types are characterized by the shape of the tameness region, which may be a vertical strip (*S*-shape), a vertical strip with “holes” (which arises in the periodic case and called *P*-shape) or a hyperbolic region (*H*-shape).

Rice’s methodology revisited. Many analyses are easier to perform in the Poisson model (where the cardinality of the inputs follows a Poisson law) than in the Bernoulli model, where this cardinality is fixed. This is due to nice independence properties of the Poisson model, and this is the case here, for the combinatorial step of our analysis. However, it is needed to return to the Bernoulli model that is much more natural for algorithmic issues, and this return leads to the Rice method. This method, described in [53, 54] is a classical tool in mathematics which transforms a discrete object (a binomial sum) into a continuous object (an integral). It was introduced by Flajolet and Sedgewick into the analytic combinatorics domain [25], and

it is now widely used in many analyses of the domain, in particular in the return Poisson $\rightarrow$ Bernoulli that has been just mentioned. There is another method which may perform this return, called the DePoissonization method. Amongst the previous works that we have described, the DePoissonization method is perhaps more frequently used, even though the Rice method has also its supporters. And this thesis chooses the Rice method because it seems to be “cleaner” to use: the Rice method is central in our analyses.

Our version of the Rice method relies *explicitly* on two tools, the *Poisson-Mellin-Newton-Rice Cycle* [20], and the *Valuation-Degree Condition* which are very often used in an *implicit* way. We think our version of the Rice method is easy to understand and especially easy to use.

A unified framework for the probabilistic analysis. We then deal with two types of digital trees and three types of tameness. The generating functions used in probabilistic contexts (the moment generating function, the cumulant generating function, or the characteristic function) are then expressed in terms of the mixed Dirichlet series, and the probabilistic study will depend on the tameness of $\Delta(s, u)$, itself closely related to the tameness of the source. For two types of tameness (*S*-shape, or *P*-shape), the probabilistic study can be done with a plain application of the Quasi-Powers Theorem of Hwang [30]. However, the case when the source is tame with a hyperbolic shape (*H*-shape) is the most difficult to deal with³, and we need an extension of the Quasi-Powers framework. We define such an extension which makes possible a unified analysis in the six cases (two types of digital trees, three types of sources).

Our results.

An exact expression for the mean profile. We have said that the mixed Dirichlet series $\Delta(s, u)$ of the dst is expressed as the product of two infinite products of quasi inverses of the form $(I - u\mathbb{H}_s)^{-1}$. And this is a convenient form for the analysis, as both the singularities and the tameness become apparent. However, it is not easy to extract the coefficients of u^k in such an expression. When using the Laplace transform, we get an alternative expression of the Dirichlet series, which is expressed as a power series (with respect to u). This gives rise to an exact (but involved) formula for the mean profile $B_{n,k}$ (the mean number of nodes at level k in a dst of size n .)

A formula which provides an extension for a q -analog formula. Then, we have two different expressions for the same Dirichlet series (one under an additive form, and the other one under multiplicative form). They are thus equal, and this gives rise to an identity which does not seem trivial (at all!). In the very particular case of the unbiased memoryless source, this identity exactly coincides with an identity in the domain of q -analogs,

$$\frac{1}{(1-v)(1-vq)(1-vq^2)\dots(1-vq^j)\dots} = \sum_{k \geq 0} \frac{v^k}{(1-q)(1-q^2)\dots(1-q^k)}.$$
⁴

We have thus discovered (?) an extension of such a formula, which appears to be new.

Asymptotic estimates for the mean and the variance.

Consider a stationary source $\mathcal{S}$, whose reverse past is a dynamical source which is assumed to be tame with a type $Y \in \{S, P, H\}$. Consider any of the two types of digital trees –trie (type T)

³In an informal way, the tameness region is smaller, and thus the remainder term is larger.

⁴A simple proof of this identity can be founded in Knuth [43]

or dst (type D)– built on a random sequence of n words independently drawn from the source. Then, the mean and the variance of their depth D_n admit the following asymptotic expansions, for any $X \in \{T, D\}$ and $Y \in \{S, P, H\}$,

$$\begin{aligned}\mathbb{E}[D_n] &= \mu \log n + \mu_X + R_Y(n) \\ \text{Var}[D_n] &= \nu \log n + \nu_X + R_Y(n)\end{aligned}\tag{1}$$

[Dominant constants]. The constants μ, ν are expressed with the dominant eigenvalue $\lambda(s)$ of the source, as

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

The only case where $\nu = 0$ arises for an unbiased memoryless source.

[Sub-dominant constants]. The constants μ_X, ν_X depend both on the source and on the type $X \in \{T, D\}$ of digital tree and the inequality $\mu_T > \mu_D$ holds. There exist explicit (but involved) expressions for these constants.

[Remainder terms]. The type of functions $R_Y(n)$ only depends on the type Y of source tameness. The functions $R_Y(n)$ admit the general form :

$$R_Y(n) = \begin{cases} O(n^{-\delta}) & \text{if the source is } S\text{-tame} \\ O(\exp[-(\log n)^\rho]) & \text{if the source is } H\text{-tame} \\ \Pi(n) + O(n^{-\delta}) & \text{if the source is } P\text{-tame} \end{cases}\tag{2}$$

Here, δ is the width, ρ is related to hyperbolic exponent ρ_0 via the relation $\rho < 1/(1 + \rho_0)$ and $\Pi(n)$ is a periodic function of $\log n$, whose period η is the period of the source.

Asymptotic Gaussian law.

Consider a stationary source S , whose reverse past is a dynamical source which is assumed to be tame of any type. Assume moreover S not to be conjugated to an unbiased memoryless source. Consider a digital tree (trie or dst) built on n words independently drawn from the source. Then,

(a) the depth D_n of the digital tree asymptotically follows a Gaussian law

$$\frac{D_n - \mathbb{E}[D_n]}{\sqrt{\text{Var}[D_n]}} \xrightarrow{d} \mathcal{N}(0, 1).\tag{3}$$

(b) Moreover, the speed of convergence towards the Gaussian law is of order $(\log n)^{-1/2}$.

Plan of the thesis

The thesis is divided into two main parts. The first part contains the first four chapters (Chapters 1, 2, 3, 4) and describes the general framework of the study, and the second part that contains the next three chapters (5, 6, 7) is devoted to the proofs of the main results. Chapter 5 can be viewed as a transition Chapter, as we will explain. (See the graph of the chapters in Figure 2 page 10).

Part I.- Description of our framework.

Chapter 1 is devoted to the description of the two digital data structures under study : the trie and the dst. We explain why they are “dynamic” structures which perform in an efficient way all the usual operations that are expected from a dictionary : search, insertion, deletion. The

depth measures the complexity of these operations. We then review the main algorithms for data compression and mention that the trie, and especially, the dst are underlying structures for the compression algorithms of Lempel-Ziv type.

Chapter 2 describes possible models for sources. We first recall the two models of simple sources (memoryless sources and Markov chains) and introduce three possible models for a general source: via a (generalized) transition matrix, a dynamical system or a parameterization of the unit interval. Next, we describe relations between these three points of view. The modelisation by a transition matrix will be well adapted to the algorithmic (and combinatorial) study, but we restrict ourselves to more specific sources in the analytic step. Given a regular source, we build its reverse past and we ask it to give rise to a “good” dynamical source.

Chapter 3 introduces the *tameness* notion. We first give a general definition for tameness. Then, we first recall some classical facts on periodicity and quasi-periodicity, and explain why tameness can be viewed as a re-enforcement of non-periodicity or non-quasi-periodicity. This gives some intuition on this notion, and explains how to obtain sufficient conditions which may entail such a tameness. Then, we present three types of tameness which arise in a natural way for simple sources, but also for dynamical sources, in relation with deep results due to Dolgopyat [12, 13] and here adapted to our context.

Chapter 4 describes the main tools we will use in our analysis. We first define two probabilistic models for the inputs (the words to be put in the digital trees), namely the Bernoulli and the Poisson models. Next, we explain the Rice methodology, and insist on two main aspects of the method, the Poisson-Mellin-Newton-Rice Cycle, and the Valuation Degree Condition. We say a few words on the Mellin transform and Depoissonization method, and conclude with a short description of the Laplace method that will be used just once (but leads to a nice result).

Part II: Proofs of our results.

Chapter 5 revisits combinatorial steps which are usually performed in the case of simple sources. This chapter is not actually original, but we prepare ourselves (and the reader) to the following chapter. We set up the stage, with the main steps of the combinatorial study. We begin to deal with the Poisson model, then we return to the usual Bernoulli model, and we exhibit the two mixed Dirichlet series, that arise in a natural way under a multiplicative form. We mention that, in the case of an unbiased memoryless source, there exists an additive form for the Δ series of dst which is not known in other cases of simple sources.

Chapter 6 is a clean extension of Chapter 5, where the (finite) transition matrix is replaced by its generalized counterpart, which extends it in the case of a source with unbounded memory. The mixed Dirichlet $\Delta(s, u)$ arises, with its multiplicative expression in terms of quasi-inverses which involves the generalized transition matrix. Then, with Laplace transform, we obtain an additive form for the Δ Dirichlet series which appears to be an exact generalization of the previous form, only obtained for unbiased memoryless sources. This leads to an exact expression for the average profile.

Chapter 7 performs the last step of the analysis. We now restrict to our “specific sources”, and using a good knowledge of their “generating” operator (singularities, tameness), we apply the second step of the Rice method, and obtain precise asymptotic estimates for the three types of generating functions that are used in the probabilistic study (moment generating function, cumulant generating functions, and characteristic functions). We then deduce asymptotic estimates for the mean and variance, we exhibit the asymptotic Gaussian law with the (optimal) speed of convergence.

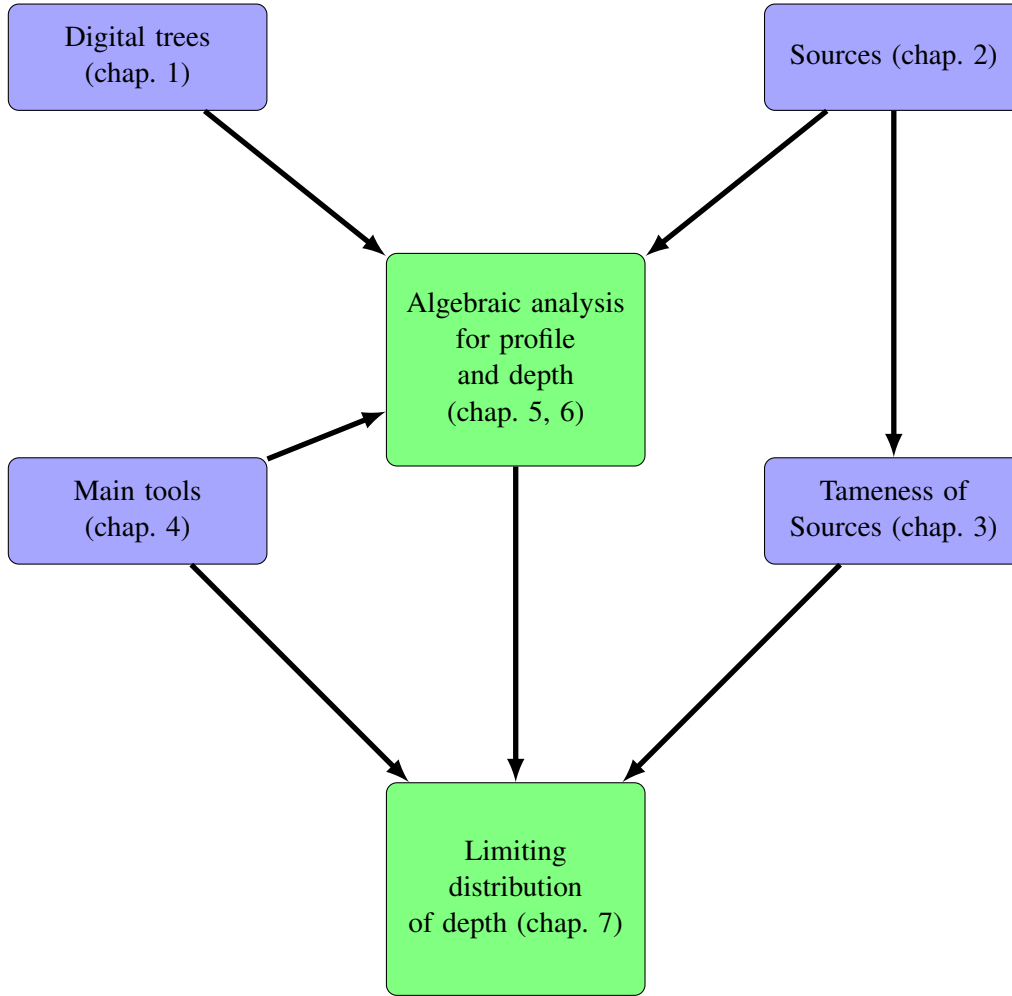


Figure 2: The graph of dependence for the Chapters of the Thesis.

Relations with other works. This thesis was carried out in the GREYC Laboratory. During the same time, another PhD student in the GREYC, Thu Hien Nguyen Thi, prepared her thesis [52] under the joint supervision of Julien Clément and Brigitte Vallée, on the subject “Towards a realistic analysis of sorting and searching algorithms”. In fact, even if the subjects of the two theses are different, they share two main important objects: general sources and Poisson-Rice methodology. This is why there exist Chapters in our theses which are very closely related: The present Chapter 2 shares many sections with Chapter 2 of Thu Hien’s thesis (but Chapter 3 is different, as Thu Hien does not deal with distributional studies) – And our Chapter 4 is almost the same as Chapter 4 of Thu Hien’s thesis.

During the preparation of the thesis, Eda Cesaratto and Brigitte Vallée wrote the last version of their paper [4] and we worked together for the precise proof of Proposition 4.16 which finally appeared in the paper [4], in the present thesis and also in Thu Hien’s thesis [52].

Chapter 1

Tries and digital search trees.

Contents

1.1	Binary Search Trees	13
1.1.1	Definition	14
1.1.2	Main operations	14
1.2	Tries	15
1.2.1	Definition	16
1.2.2	Dynamic Structure	16
1.2.3	Suffix trees	18
1.3	Digital Search Trees	18
1.3.1	Definition	19
1.3.2	Dynamic Structure	19
1.3.3	Relation between tries and DSTs	20
1.3.4	A unified view for parameters for tries and dst's.	20
1.4	Symbol-Comparisons BST	21
1.5	Data Compression	22
1.5.1	Coding	22
1.5.2	Huffman coding	22
1.5.3	Arithmetic Coding	24
1.5.4	Dictionary Compression Algorithms	26
1.5.5	LZ 77	26
1.5.6	LZ78	27

This first chapter is devoted to describing the two main digital trees which are the main actors of the thesis. We wish to study these two different structures – the trie and the digital search tree (see figure 1.1)– and compare them. We first give a formal definition, and explain how they support the important operations that are needed for a dynamic structure: Search, Insert, and Delete. In fact, the digital search tree can be viewed as a mixing of two other trees–the trie and the binary search tree–. Even though the binary search tree is not a digital structure, the main ideas that are used for implementing the dynamic operations have the same flavour and it is thus nice to first define and study the binary search tree. We also explain how it can be viewed as a digital structure In Section 1.4.

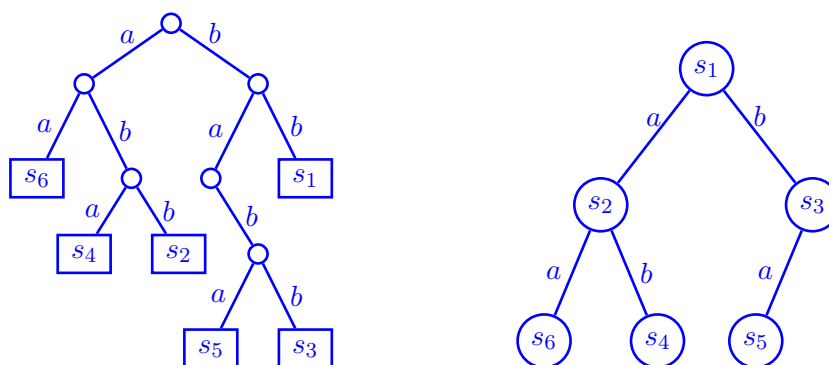


Figure 1.1: A trie on the left and a DST on the right

Digital trees like tries or digital search trees are important in theoretical computer science, either as data structures or as models for data splitting. Flajolet describes them as ubiquitous structures (See for instance the paper [19]). But they are also central in various applications like data compression, pattern matching or hashing. For example, the popular Lempel-Ziv compression schemes are strongly related to tries or digital search trees, according to the version used (see [71] for example). This is why we review, in the second part of the chapter, the main algorithms which deal with data compression, as they use in a deep way the studied digital structure, as least from a conceptual point of view.

Computer science is mainly concerned with the fundamental question of managing information and this information is usually represented under a digital form. There is no question that there are more and more data available everyday, a trend which seems not likely to falter. This calls for a need not only for efficient data structures tailored for these huge amounts of data, but also for an urge to analyze data structures in order to predict their behaviour, possibly detect flaws and get better designs.

There is a whole hierarchy for data structures. At the top level, we have abstract data structures defined by their functionalities. For instance, the abstract data structure dictionary aims at storing dynamically a set of words, and provides the basic operations searching for a word, inserting a word and deleting a word. Then the algorithmic approach provides implementations for this data structure specifying more precisely how operations are to be done. This specific aspect can be described with the help of a formal language (going from pseudo-code to actual code in some programming language). Depending on the choice of a data structure, operations may be more or less efficient with respect to space or speed. Thus the choice of a particular data structure is usually tailored to the application which will use them.

In this thesis, we are mainly interested in data structures that are used to store and retrieve efficiently data when they are viewed as words. Tries and digital search trees (DST for short) are both *digital tree structures* based upon the same basic principle: the data is viewed as a word, that means a sequence of symbols from a certain alphabet, and all the operations on these data structures will strongly rely on this representation.

These data structures are often presented as an efficient solution to the dictionary problem mentioned before. In this chapter, we introduce these two digital data structures, *tries* and *digital search trees*. We also introduce another data structure, that is not a digital structure, namely the *binary search tree*, as it may help to a better understanding of tries and dst, which are our primary subject of interest. However binary search trees do not belong to the family of digital trees and

are based on comparisons of key values (say for instance integers or floating numbers), while tries and DST use the digital representation of these keys. However, we will describe in Section 1.4 a version of the BST structure which can be also viewed as a digital structure.

In computer science, data has to be stored in a clever way. To do this, the *dictionary* is defined as an abstract data type for storing words (sometimes with associated values). This abstract data type has to be made precise (and then implemented) by specifying a data structure together with procedures allowing operations such as: Search, Insert, Delete, etc. To *search* is to find a previously stored key, to *insert* a key is to store it in the data structure for possible later retrieval; and to *delete* a key is to remove it from the data structure. Throughout this thesis, we consider (infinite) words built from an alphabet Σ . This does not seem to be a natural definition, but it leads to clearer definitions and appears more convenient for the analysis.

A commonly used structure for implementation of dictionaries is a *tree*. We first give a definition of a r -ary tree.

Definition 1.1. Consider an integer $r \geq 2$. A r -ary tree is a rooted plane tree, where each node is either an external node (i.e., it has no successor) or it has r successors and is called an internal node. External nodes are also called leaves. Each internal node has exactly r children. In the case when $r = 2$, this is called binary tree.

Remark. We can give a recursive definition of r -ary trees as: A r -ary tree $\mathcal{B}$ is either just an external node or an internal node (the root) with exactly r subtrees that are again r -ary trees. This can be represented via the following combinatorial equation

$$\mathcal{B} = \square + (\bullet \times \mathcal{B}^r), \quad (1.1)$$

where $\square$ denotes an external node (or leaf) and $\bullet$ an internal node (see Fig. 1.2 for an illustration). Since the children are ordered and form a sequence, the tree is said to be plane since it can be uniquely drawn on the plane.

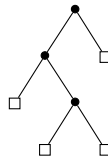


Figure 1.2: A binary tree with 3 internal nodes and 4 leaves (or external nodes).

1.1 Binary Search Trees

A binary search tree (BST) [43] is a data structure used for storing keys whose values lie in an ordered domain. It is a binary tree and also a search tree. It is defined in a recursive way by the fact that the key contained in the root of a non empty BST is greater than any values in the left subtree and less (or equal if equal values are permitted in the input) than any values in the right subtree. Moreover left and right subtrees are also binary search trees. We note that keys are stored only in internal nodes. The binary search tree is a very well-known data structure which admits very efficient implementations.

1.1.1 Definition

A binary search tree is defined as follows.

Definition 1.2. Consider a sequence $\mathcal{Y}$ of different keys. The BST tree structure is recursively constructed as follows:

- If $|\mathcal{Y}| = 0$, the tree $\text{BST}(\mathcal{Y})$ is an (empty) external node $\square$.
- Otherwise, the tree $\text{BST}(\mathcal{Y})$ is

$$\text{BST}(\mathcal{Y}) = \langle \text{First}(\mathcal{Y}), \text{BST}(\mathcal{Y}_{<\text{First}(\mathcal{Y})}), \text{BST}(\mathcal{Y}_{\geq \text{First}(\mathcal{Y})}) \rangle,$$

where $\text{First}(\mathcal{Y})$ is the first element of $\mathcal{Y}$ and $\mathcal{Y}_{<\alpha}$ (respectively $\mathcal{Y}_{\geq\alpha}$) is the subsequence of $\mathcal{Y}$ formed with keys which are less than α (respectively greater or equal to α).

Remark. In figures (like Fig. 1.3), external (void) nodes are not drawn to get a lighter representation (but are implicitly present since it is a binary tree).

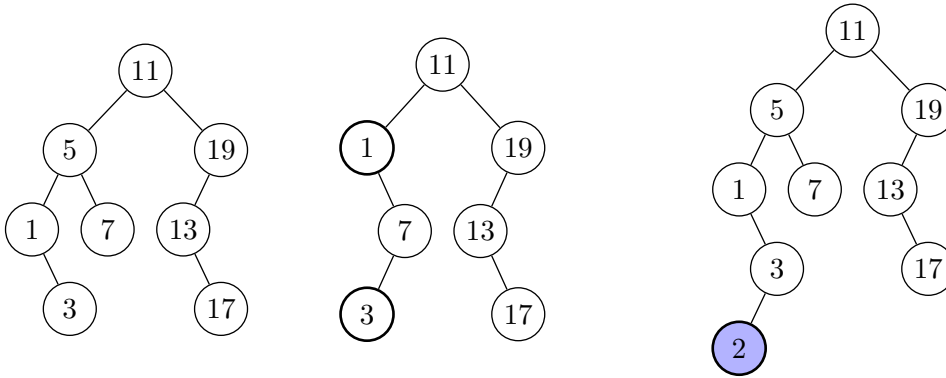


Figure 1.3: A BST (left) built from a list of $\{11, 5, 19, 13, 1, 7, 3, 17\}$; (middle) delete 5 from the list; and (right) insert 2 into the tree.

1.1.2 Main operations

The BST is a dynamic structure because it is easy to maintain the structure, even though the content of the dictionary is subject to frequent changes (insertions or deletions).

Searching. The search algorithm tells if a record containing a given key is stored in the tree. The search is *successful* when the key sought is found otherwise the search is *unsuccessful*.

Function $\text{BSTSearch}(\text{Element } y, \text{ Bst } \mathcal{Y})$

/ $\mathcal{Y}$ is a BST, y is the value of the key we are looking for. */*
/ Returns nil, if $y \notin \mathcal{Y}$, and a subtree of $\mathcal{Y}$ whose root contains y otherwise. */*

if $\mathcal{Y}$ is void **then return** nil

if $\text{Key}(\text{Root}(\mathcal{Y})) = y$ **then return** $\mathcal{Y}$

if $\text{Key}(\text{Root}(\mathcal{Y})) < y$ **then**

 | **return** $\text{BSTSearch}(y, \text{left}(\mathcal{Y}))$

else

 | **return** $\text{BSTSearch}(y, \text{right}(\mathcal{Y}))$

The procedure begins its search at the root and traces a path downward the tree as shown in Figure 1.3. For each node of the tree $\mathcal{Y}$ it encounters, it compares y with the $\text{Key}(\text{Root}(\mathcal{Y}))$ (the key at the root of the tree). If the two keys are equal, the search terminates. If y is smaller than $\text{Key}(\text{Root}(\mathcal{Y}))$, the search continues in the left subtree; if y is larger or equal to $\text{Key}(\text{Root}(\mathcal{Y}))$, the search continues in the right subtree. We follow this recursive process until termination. For example, suppose we are to search for a key 7 in Figure 1.3. First, we compare it to the key 11 at the root and since $7 < 11$, we move to the left. Again, we compare with 5 and since $7 > 5$ we move to the right. Finally, we reach a key 7 to be sought for.

Inserting a key. There are several ways to insert a key so that the structure of the binary search tree is preserved. We present here the simplest technique where insertion is made at leaves.

```

Function Insert (Element  $x$ , Bst  $\mathcal{Y}$ )
/*  $\mathcal{Y}$  is a bst,  $x$  is the value of the key we want to insert. */
if  $\mathcal{Y}$  is void then
    | return CreateNode( $x$ )
if  $\text{Key}(\text{Root}(\mathcal{Y})) < x$  then
    | left( $\mathcal{Y}$ ) = Insert ( $x$ , left( $\mathcal{Y}$ ))
else
    | right( $\mathcal{Y}$ ) = Insert ( $x$ , right( $\mathcal{Y}$ ))
return  $\mathcal{Y}$ 

```

To *insert* a new key with value x , we search, then replace the left or right available external node that caused termination with the new key x . Suppose we are to insert 2 into the tree (left of Figure 1.3). We search for the key 2 and we reach the node containing the key 3. This node has two children (external nodes which are not represented) and we place the key 2 in the left node subtree since $2 < 3$. See Fig. 1.3 for an illustration.

Cost of the main operations in a BST. Basic operations on a binary search tree of size n take, in the worst-case, a time that is proportional to the height of the tree. On average, this takes a time proportional to the length of a random branch. Under the uniform permutation model, it is of order $O(\lg n)$. The path length of a random binary search tree is exactly the number of comparisons needed to sort a random permutation of $\{1, 2, \dots, n\}$ with QuickSort, which is of order $O(n \log n)$ (see for example [74]).

1.2 Tries

Now, the keys are (infinite) words built on a finite alphabet Σ , and we can make use of their representation as a sequence of symbols. That is the purpose of digital trees.

A trie [5, 27, 43, 50, 66] (pronounced as “try”, derived from the word *retrieval*) was introduced in 1959 by Fredkin and by de la Briandais[10]. It is a data structure that uses the decomposition of words as sequences of symbols to organize and search the dictionary. The trie is itself built on symbols and compares words using the lexicographic order on their prefixes to discriminate them. This structure is quite different from the one we have already encountered. Firstly, the trie is now an r -ary tree (if r is the cardinality of the alphabet). Secondly, links between a node and its children are labelled by symbols of the alphabet. Finally, data are stored in the leaves of the tree.

1.2.1 Definition

Now, when the alphabet Σ is of cardinality r , the trie will be a r -ary tree :

Definition 1.3. Let us consider a finite alphabet $\Sigma = \{a_1, \dots, a_r\}$ (with $r \geq 2$) and a sequence $\mathcal{Y}$ of distinct words over Σ . The trie associated with $\mathcal{Y}$ is a r -ary tree defined recursively as:

- If $|\mathcal{Y}| = 0$, the trie is empty: $\text{trie}(\mathcal{Y}) = \emptyset$.
- If $|\mathcal{Y}| = 1$, the trie is a leaf (external node) containing the unique word in $\mathcal{Y}$
- If $|\mathcal{Y}| > 1$,

$$\text{trie}(\mathcal{Y}) = \langle \bullet, \text{trie}(\mathcal{Y}_{(a_1)}), \dots, \text{trie}(\mathcal{Y}_{(a_r)}) \rangle,$$

where $\bullet$ denotes an internal node and $\mathcal{Y}_{(\alpha)}$ is the subsequence of $\mathcal{Y}$ formed with the words of $\mathcal{Y}$ which begin with α , from which the first symbol α is removed. The internal node is connected to a non-empty subtrie $\text{trie}(\mathcal{Y}_{(\alpha)})$ by a link labelled α .

Note that a trie structure does not depend on the order of insertion of words.

In a trie, all words (or more accurately suffixes of words) are stored in external nodes while internal nodes are branching nodes used to direct keys to their destinations. Internal nodes do not contain any part of the data. The trie structure compares words via prefixes: it separates the words according to symbols encountered. We usually identify in a trie (and later on in a digital search tree) nodes of the tree with words labelling branches from the root to these nodes. Each external node contains a suffix of a word, while the prefix is found with the labels which are on the path from the root to this external node. More precisely, the total word u is written as $w \cdot v$, where w is the prefix w labelling the corresponding branch, and the suffix v of the word u is stored in the external node labelled w .

For instance, in Figure 1.4, we see the decomposition of the six words of the trie. Each word decomposes as a prefix which labels the branch and a suffix which is stored in the external nodes.

$$\begin{array}{llll} S_1 = ab \cdot s_1, & S_2 & = abb \cdot s_2, & S_3 = babb \cdot s_3, \\ S_4 = aba \cdot s_4, & S_5 & = baba \cdot s_5, & S_6 = aa \cdot s_6. \end{array}$$

Each prefix is the longest prefix to be shared with another word of the trie.

1.2.2 Dynamic Structure

A trie allows the same basic operations as binary search trees: Search, Insert and Delete.

Searching. To search for a key with value y in a binary trie $\mathcal{Y}$, we perform the following operations $\text{Search}(y, \mathcal{Y})$ until termination:

Function <code>trieSearch(Word y, Trie $\mathcal{Y}$)</code>	
<code>/* $\mathcal{Y}$ is a Trie, y is a word.</code>	<code>*/</code>
<code>/* Returns False, if $y \notin \mathcal{Y}$, and True otherwise.</code>	<code>*/</code>
if $\mathcal{Y}$ is void then return False	
if $\mathcal{Y}$ is a full node containing word v then return $y = v$	
if $y = \varepsilon$ then return False	
Decompose y as $y = \alpha \cdot y'$ (with $\alpha \in \Sigma, y' \in \Sigma^*$)	
return <code>trieSearch(y', $\mathcal{Y}_{(\alpha)}$)</code>	

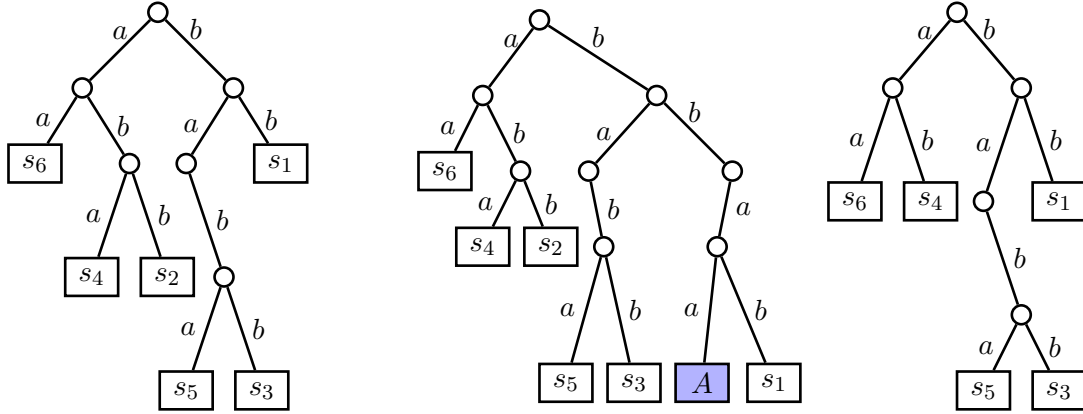


Figure 1.4: A TRIE (left) built from the sequence $S_1 = bbabb\dots$; $S_2 = abbaa\dots$; $S_3 = babba\dots$; $S_4 = ababb\dots$; $S_5 = babab\dots$; $S_6 = aaaab\dots$; (middle) insert a word $A = bbaab\dots$, and (right) delete the word S_2 .

Searching for a key is very much like looking up a word in a dictionary. To search for a key in the tree, we start at the root and follow a path down the trie until we either fall off the trie (i.e., we follow a *null* pointer in a branch node) or we reach a full node (node containing a key). The path we follow is determined by the digits from left to right of the key sought for. Suppose, for example, we are to search for a key $s_2 = abbaa\dots$ in Figure 1.4. We use the first symbol, a , in the key s_2 to move from the root node to the left (internal) node, named B . Since B is a branch node, we use the next symbol, b , of the key to move further down the trie to the right and we reach another branch node, say C . Then we use the next symbol b to move to the next level and we reach a full node. We compare the key sought for with the key in the reached node. We perform the comparison and we get the match. See Figure 1.4 on the left. Observe that only one key-comparison is made to complete the search. The access path from the root to an external node (a leaf of a trie) is the minimal prefix of the information contained in this external node; it is minimal in the sense that this prefix is not a prefix of any other strings.

After an unsuccessful search it is sometimes desirable to enter a new key into the tree.

Inserting a key. To insert a key x , we search for it and if found, we replace the found key by x . Otherwise, we have two cases. Case 1: If the search for the key x ends by falling off the trie from the branch node (an internal node) q , then we simply add a child to the node q . Case 2: If the search for x ends at a full node p , then one key together with the one we want to insert share a common prefix p (if we associate nodes with finite words as mentioned before). We use these two keys to construct a subtrie which replaces the node p . For example, we want to insert the word A into the trie. The search for a key A terminates at the full node d containing a suffix of s_1 but this suffix is not equal to the corresponding suffix of A to be inserted. We then build a subtrie by adding branching nodes until we reach the first digit at which the two words differ. See trie in the middle on Figure 1.4.

Function `trieInsert(Word  $x$ , Trie  $\mathcal{Y}$ )`

```

/*  $\mathcal{Y}$  is a Trie,  $x$  is a word which is not a prefix of any word in
    $\mathcal{Y}$ . */
if  $\mathcal{Y}$  is void then
    return CreateNode( $x$ )
Decompose  $x$  as  $x = \alpha \cdot x'$  (with  $\alpha \in \Sigma$ ,  $x' \in \Sigma^*$ )
 $\mathcal{Y}_{(\alpha)} = \text{trieInsert}(x', \mathcal{Y}_{(\alpha)})$ 
return  $\mathcal{Y}$ 

```

We see that large amount of unnecessary internal nodes (which don't help the navigation in the tree) may be added to the trie when inserting a new key. Thus, a trie has usually more internal nodes than external nodes, differing in that aspect from almost all other search trees. It is possible to eliminate a number internal nodes in the trie constructed with n keys so that it has just $n - 1$ internal nodes by collapsing one-way branches, and hence guarantee that every internal node will have non-null descendants. This resulting data-structure is called the *Patricia trie* [43].

Deleting a key. To delete a key x from the tree, we retrace the path from the root to x (similarly to a search) and discard the full node corresponding to x and all the branching nodes that are roots of subtrees that have only one key in them. For example, when s_2 is removed from the tree, we discard also another branching node in this case, see Figure 1.4 on the right. When deleting a key in a trie, one removes not only the key but also possibly some other internal nodes.

Observation. A trie takes memory space because of its unnecessary internal nodes. For example, a trie requires 12 iterations to distinguish between the words *computation* and *computations*. In this case, it would be better to make use of the mirrors of words (here *noitatupmoc* and *snoitatupmoc* which are distinguished with just one comparison).

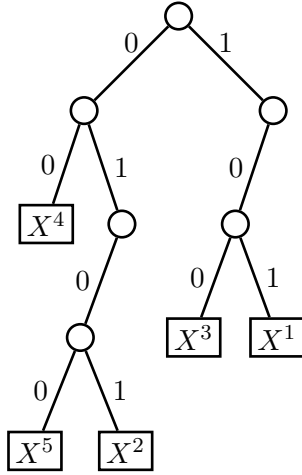
1.2.3 Suffix trees

A suffix tree is a trie constructed from the suffixes of a *given* string. Given a string $X = x_1x_2 \dots$ built from an alphabet, denote by $X^j = x_jx_{j+1}x_{j+2} \dots$ the j th suffix of X starting at the position j th. In this case the strings X^j for $j = 1, \dots, n$ strongly depend on each other while in a trie the strings of $\mathcal{Y}$ might be completely independent.

Suffix trees have numerous applications. When collapsing unary branches (resulting in compact suffix trees), it has a linear number of nodes (in the length of the input words) and thus is quite efficient. It is one of the most important conceptual data structure in text algorithmics although in practice, other data structures (e.g. suffix arrays) are usually used (because they are easier to program and more efficient).

1.3 Digital Search Trees

The “digital search tree” data structure, originated from the paper of Coffman and Eve in 1970 [38], is used to perform the search in a similar way as tries, scanning one symbol at a time from left to right in the key. This DST [23, 43, 66] structure can be seen as a mixing between BST and tries in the sense that every node stores data and it uses symbols of the key to decide whether to navigate in the left or right tree structure.



In the binary case, for $X = 1010010001 \dots$,
We can build the suffixes of X as follow.

$$\begin{aligned} X^1 &= X = 1010010001 \dots, \\ X^2 &= 010010001 \dots, \\ X^3 &= 10010001 \dots, \\ X^4 &= 0010001 \dots, \\ X^5 &= 010001 \dots \end{aligned}$$

$\vdots$

Figure 1.5: A suffix tree built from first five suffixes of $X = 1010010001 \dots$.

1.3.1 Definition

Definition 1.4. Let us consider a finite alphabet $\Sigma = \{a_1, \dots, a_r\}$ (with $r \geq 1$) and a sequence of distinct infinite words $\mathcal{Y}$ over Σ . The tree $\text{DST}(\mathcal{Y})$ is defined as follows:

- If $|\mathcal{Y}| = 0$, then $\text{DST}(\mathcal{Y})$ is empty : $\text{DST}(\mathcal{Y}) = \emptyset$
- If $|\mathcal{Y}| \geq 1$, the root of $\text{DST}(\mathcal{Y})$ contains the first word $\text{First}(\mathcal{Y})$, and there are r subtrees built with the sequence $\underline{\mathcal{Y}} := \mathcal{Y} \setminus \{\text{First}(\mathcal{Y})\}$, and the j th subtree is $\text{DST}(\underline{\mathcal{Y}}_{(a_j)})$.

1.3.2 Dynamic Structure

In a DST, all the nodes are full. In contrast to tries, the insertion depends on the order of the keys (as in a BST). In a DST, strings are directly stored in nodes. The search for an available node follows the prefix structure of a string as in a trie. An example of a digital search tree is shown in Figure 1.6.

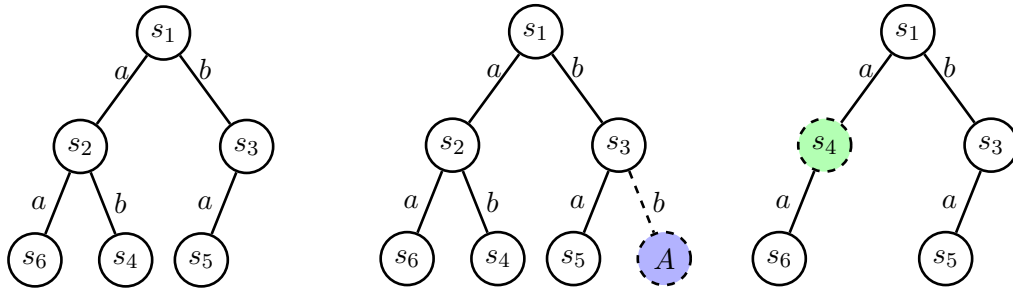


Figure 1.6: A DST (left) built from the sequence $S_1 = bbabb \dots$; $S_2 = abbaa \dots$; $S_3 = babba \dots$, $S_4 = ababb \dots$; $S_5 = babab \dots$; $S_6 = aaaab \dots$; (middle) insert a word $A = bbaab \dots$ and (right) delete a word s_2

In Figure 1.6, we see the decomposition of each of the six words as a prefix which labels the branch and the suffix which is stored in the node.

$$S_1 = s_1, \quad S_2 = a \cdot s_2, \quad S_3 = b \cdot s_3, \quad S_4 = ab \cdot s_4, \quad S_5 = ba \cdot s_5, \quad S_6 = aa \cdot s_6$$

Searching for keys in a DST is faster since one has to search for the longest prefix of the key. To search for a key s_4 in Fig 1.6, first we compare it with the key s_1 at the root of the tree. Since there is no match and since the first symbol of s_4 is a , we move to the left and compare with s_2 . Since there is no match and since the second symbol of s_4 is b , we move to the right and compare with s_4 . The key is found!

To insert a new word to the tree, we follow the same method as in a BST (i.e., the null pointer which caused termination is replaced by a pointer to the new word).

1.3.3 Relation between tries and DSTs

There is a simple way to relate a trie and a DST. Suppose that links in the tree are elastic links¹ and that when you release an external node of the trie it will go up on its path to the root and occupy the available internal node of minimal level of the tree on this path. For instance the first word of the sequence will end up at the root, etc. If we do this process for each key of the sequence in the order of the input sequence, we obtain the corresponding DST.

Because the keys have been lifted to the levels of the internal nodes, the level of nodes should be at least as small as in tries. The extra space occupied by the internal nodes in tries is eliminated in digital search trees – each key occupies one node, and the number of nodes allocated is precisely the number of keys (as in binary search trees).

1.3.4 A unified view for parameters for tries and dst's.

The cost of the main operations on a digital structures can be read on the shape parameters of the tree. When dealing with an alphabet Σ of cardinality r , both trees are r -ary trees. But, if we wish to give an unified point of view on both structures, we have to distinguish two types of nodes : nodes that contain (some part of) data or nodes which do not contain any part of the data. The first ones are said to be *full*. This is why the following definition will be important:

Definition 1.5. *The following notations and definitions will be used throughout this thesis for both types of digital tree (trie or dst)*

- A full node is a node containing data (either a word or some part of it).
- The level of a node is the length of the path from the root to it. The level of a node is the length of the prefix which labels the node.
- The size of a tree is equal to the number of full nodes it contains. This is also the number of words it contains.
- The profile $b_{n,k}$ is the number of full nodes at level k in a tree of size n .
- The height of a tree is the maximum level among all the full nodes in the tree.
- The typical depth is the level of a randomly selected full node.

Relations between these parameters. This sequence $b_{n,k}$ is called the profile. For a tree of size n (i.e., with n full nodes), denote $d_{n,i}$ as the level of the full node containing (a part of) the i -th key. We observe the relation

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

¹An image borrowed to Nicolas Broutin.

where $\llbracket \cdot \rrbracket$ denotes the Iverson bracket². Consider any probabilistic model and denote by $B_{n,k}$ the average profile, then the following equality holds

$$B_{n,k} := \mathbb{E}[b_{n,k}] = \sum_{i=1}^n \Pr[d_{n,i} = k].$$

The typical depth (called in the thesis the depth), denoted by D_n , is defined as the level of a random full node. Its distribution is defined via the equalities

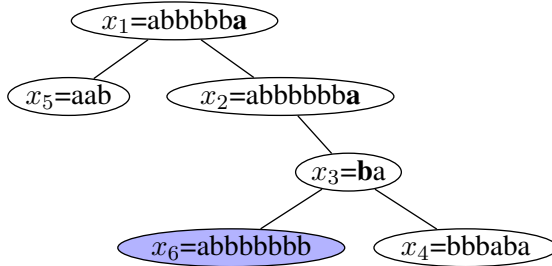
$$\Pr[D_n = k] := \frac{1}{n} \sum_{i=1}^n \Pr[d_{n,i} = k] = \frac{1}{n} B_{n,k}.$$

This is the main object of the present study and it is closely related to the average profile.

Algorithmic meaning of the parameters. The depth provides a good indication of the typical cost needed when performing a positive search. The height is a measure of the worst-case performance of a positive search. The profile describes the general shape of the tree. In the two cases, the height is the maximum number of symbol-comparisons needed to separate any two words. The (typical) depth is the mean number of symbol-comparisons needed to separate any two words.

1.4 Symbol-Comparisons BST

We now present another point of view on BST which transforms it into a digital structure. Now the keys are viewed as words, and the cost of a comparison between two keys (now two words) is the number of symbol-comparisons which are needed to distinguish the two words. This BST is constructed in the same way as before except that one makes comparisons with symbols instead of values. This is illustrated in Figure 1.7 .



To insert x_6 in the tree, instead of 3 key-comparisons, one needs 16 symbol-comparisons:

7 comparisons with x_1
 8 comparisons with x_2
 1 comparison with x_3

Figure 1.7: A BST built from a sequence of alphabet $\{a, b\}$, $x_1 = abbbbba$, $x_2 = abbbbbbba$, $x_3 = ba$, $x_4 = bbbaba$, $x_5 = aab$, $x_6 = abbbbbbb$; considering symbol comparisons.

Note that in this different cost model where the unitary cost is the cost of comparing two symbols, we define the search cost of a BST as the symbol path length which measures the total number of symbol comparisons to construct the BST. The complexity of searching is measured by the number of comparisons performed during the searching process. It is proven in [74] that the mean number of symbol-comparisons needed to perform a successful search in a BST becomes now proportional to $O(\lg^2 n)$.

²Defined as

$$\llbracket P \rrbracket = \begin{cases} 1, & \text{if } P \text{ is true;} \\ 0, & \text{otherwise.} \end{cases}$$

1.5 Data Compression

A common compression problem involves finding an efficient algorithm to reduce the size of a certain type of data. Examples of data are text, audio, image and video etc. Given data, we wish to compress it into smaller representation from which the original data can be exactly recovered at a later time. This process is called *data compression*. Data compression [58] plays a crucial role in computer science, telecommunication and computer networks of electronic engineering. For example, people use data compression software such as zip, gzip and WinZip to reduce the file size before storing or transferring it in media.

We explain in this section why digital trees are particularly useful in the context of data compression.

There are two major families of compression techniques when considering the possibility of reconstructing exactly the original source. They are called *lossless* and *lossy* compression. In a *lossless* data compression, the original data can be reconstructed exactly from the compressed version, i.e., there is no loss of any information during the compression process. A practical example of lossless data is *textual data* (printed books, database information, numerical data, electronic mail, etc.), where it is important to preserve the original data. By contrast, data such as: audio, image, video data etc, where some information is allowed to be lost during the compression process is called *lossy*. In this chapter, we focus only on lossless data compression schemes and we consider three main coding (compression) techniques which are described below. Each one is related to some aspects of this thesis.

Digital trees have played a crucial role in data compression not only in the way of data structures (usually to find regularity or repetitions in data in order to allow for compression) but also to model some aspects of compression schemes [63, 64, 68].

1.5.1 Coding

The fundamental representation of data is ASCII code which consists of a set of *fixed length* (8 bit) codewords. It is also possible to represent an alphabet (used for representing original data) by a set of *variable length* codewords over another alphabet. The code is then called a variable length code since different symbols will be coded by codewords of different length. In usual situations the alphabet for the code is binary.

Suppose that the alphabet of a source data is $\mathcal{S} = (s_1, s_2, \dots, s_n)$. The (binary) digital representation of the symbol set is called the *code* $\mathcal{C} = (c_1, c_2, \dots, c_n)$ and the representation c_i of each symbol is called the *codeword* for symbol s_i , where $i = 1, 2, \dots, n$. The process of assigning codewords to each symbol in a source is called *encoding*. The reverse process to reconstruct the sequence of symbols in the source is called *decoding*. Since the aim of compression is to obtain the shortest code, to compress (decompress) a message means to encode data (decode). Usually the encoding process produce the code thanks to a model for the data it wish to compress.

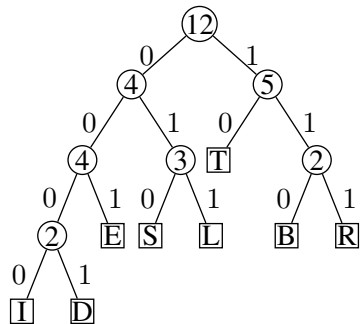
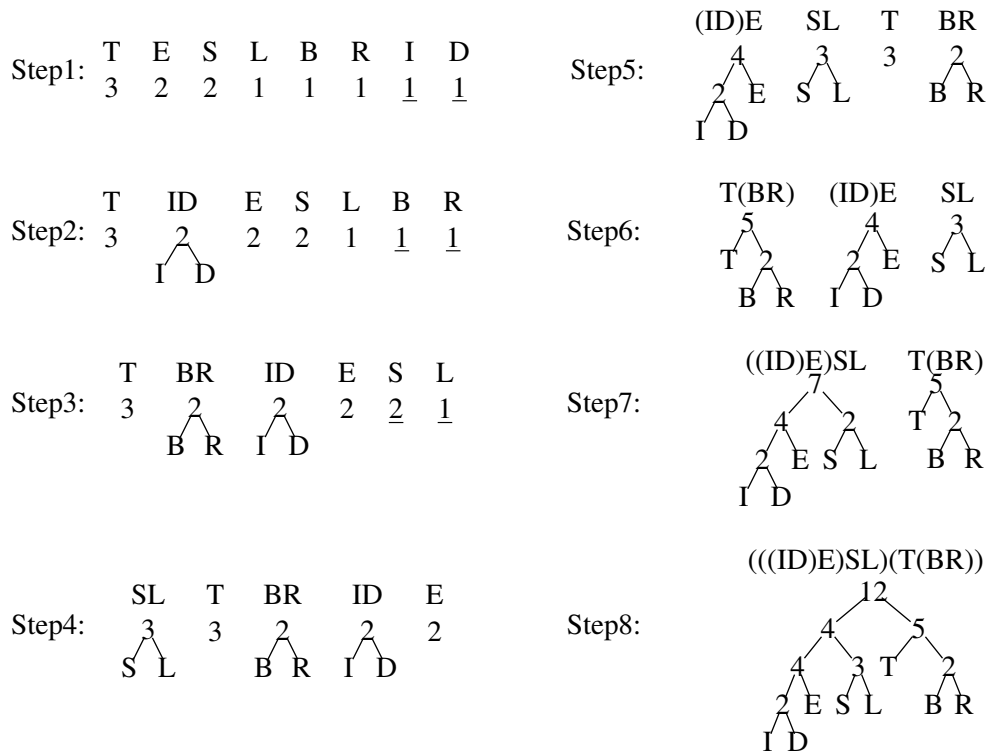
A code is *uniquely decodable* if there is only one possible way to decode encoded messages. A family of codes allowing this is called *prefix codes*, i.e., none of the codewords of this code is the prefix of another codeword. Prefix codes are a subset of the uniquely decodable codes.

We wish to stress out that these prefix codes are very important in source coding.

1.5.2 Huffman coding

The Huffman coding, introduced in 1952 by D. Huffman, is a method for compressing data with variable-length codes. In English text, some letters are used much more frequently than

others. Huffman's idea is, instead of using a fixed-length code, to present shorter codes for more frequently occurring symbols (with a higher probability) and the longer codes to the symbols that occur less often. Huffman code can be obtained by constructing a binary tree from the bottom-up (and the bits of each codeword are constructed from right to left).



Symbols can be encoded from the Huffman tree on the left. Actually, this is a binary trie where each symbol represent an external node. Decoding can be also obtained from the Huffman tree. For instance, the sequence 0100011010 can be decoded as SETT. The decoder scans bit by bit of the sequence from left to right to decode 010 as 'S', 001 as 'E' etc.

Figure 1.8: Steps of building codes of Huffman for the string LEBSTTRIEDST. From the Huffman tree, we can encode the symbols as T=10, E=001, S=010, L=011, B=110, R=111, I=0000, D=0001.

The Huffman encoding and decoding algorithms can be easily implemented in practice since they use prefix codes.

Example 1.6. Suppose we want to compress the string LEBSTTRIEDST.

We construct a table.

Symbol:	T	E	S	L	B	R	I	D
Sorted list frequency:	3	2	2	1	1	1	1	1

This source consists of symbols from an alphabet (T, E, S, L, B, R, I, D) with the recurrent statistics (3, 2, 2, 1, 1, 1, 1, 1). We want to assign a variable length prefix code to the alphabet, i.e., one codeword of some length for each symbol.

Encoding. The Huffman encoding algorithm begins by constructing a list of all the alphabet symbols in descending order of their probabilities. It then constructs from the bottom up, a binary tree with a symbol at every leaf. Once we have the binary tree, it is easy to assign a 0 to the left branch and a 1 to the right branch for each internal node of the tree. Each symbol corresponds to a leaf in the tree. The ‘bottom-up’ binary tree is built from the leaves to the root as the following steps in each iteration:

1. Combine the last two items which have the minimum frequencies.
2. The combined item, which represents a subtree, is placed according to its combined frequency on the sorted list.

The prefix codes can be written from the Huffman tree as shown in Figure 1.8.

The Huffman code produced by procedure is proved to provide an optimal code in the following sense. Given a text t of length n over a alphabet of size r and with frequencies of symbols $(f_1, \dots, f_r)$ ($\sum_i f_i = n$), the Huffman code gives codewords $(c_1, \dots, c_r)$ over $\{0, 1\}$ which minimizes the length of the compressed text $\sum_i f_i |c_i|$.

Decoding. Before starting the compression of a data file, the compressor (encoder) has to determine the codes. It does that based on the probabilities (or frequencies of occurrence) of the symbols. In a static context, the probabilities or frequencies have to be written, as information side, on the output, so that the Huffman tree can be reconstructed on the decoder side and then allows to decompress the data.

Decoding is done using the Huffman tree where the symbol is obtained by “walking down” from the root to a leaf for each symbol using the bit-string for encoded data.

Shannon-Fano coding. Another well-known coding approach that is similar to the Huffman coding is Shannon-Fano coding which was discovered in the late 1940’s by C. E. Shannon and R. M. Fano. They differ in that the Shannon-Fano coding is generated by constructing a ‘top-down’ binary tree instead of ‘bottom-up’ tree as in Huffman tree. We refer more details to [58] for example.

Observe that Shannon-Fano codes are *prefix codes* and are thus uniquely decodable [58], just as Huffman codes. Fano-Shannon codes are not optimal in the sense which was given for Huffman codes.

1.5.3 Arithmetic Coding

We have just mentioned two famous static models of Huffman and Shannon-Fano coding techniques. The word ‘static’ means the probability distribution remains unchanged during the process of encoding and decoding. These codes are amenable to a dynamic context where the statistics are gathered along the encoding process and the coding trees are adopted to this changing statistics. In this section, we introduce another famous coding method called the *arithmetic coding*. In contrast to previous coding techniques, instead of replacing every single input symbol with a codeword, the arithmetic approach encodes a stream of input symbols with a single fraction as the compressed output. The arithmetic method is based on the cumulative probability

of a symbol sequence corresponds to a unique subinterval of the initial $[0, 1]$. This process is no longer related to digital trees and more to the process of modelling a source of Chapter 2 and transmits the parameter corresponding to the input word (the stream of symbols).

We consider here an arithmetic coding for binary alphabet (A, B) . The input symbols A, B are independently generated with the probability p_A, p_B . Suppose the length of the string of symbols is given and known to encoder and decoder.

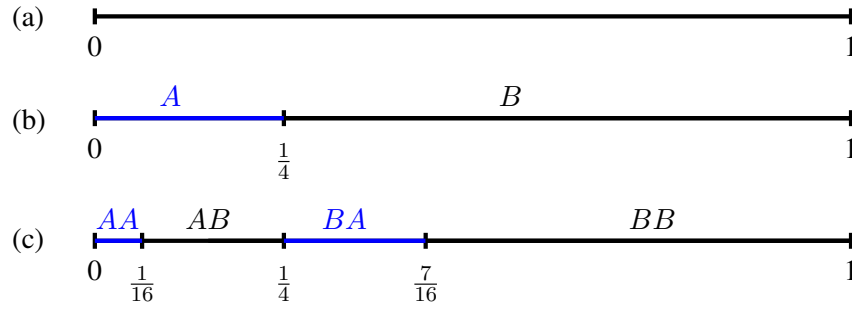


Figure 1.9: Encoding for one and two-symbol string.

Encoding. The goal is to assign a unique interval to each symbol sequence of a known length. The arithmetic encoder reads a sequence of source symbols one symbol at a time. Each time a new subinterval is derived according to the probability of the input symbol. This process, is iterated until the end of the input symbol sequence. Then, the arithmetic coder outputs a chosen *real number* within the final subinterval for the entire input symbol sequence. The chosen real decimal must assure that each codeword is short.

Example 1.7. Consider an alphabet $\Sigma = \{A, B\}$ with the probability of independent event $p_A = 1/4, p_B = 3/4$.

An input sequence of a single symbol A, B is assigned to real number in the subinterval $[0, 1/4), [1/4, 1)$ respectively. This is done by dividing the interval $[0, 1)$ into two subintervals of size which is proportional to the probability of each symbol. The subinterval $[0, 1/4)$ is divided according to the probability of p_{AA}, p_{AB} into two new subintervals $[0, 1/16)$ and $[1/16, 1/4)$ respectively. Similarly, the subinterval $[1/4, 1)$ is divided according to the probability of p_{BA}, p_{BB} into two new subintervals $[1/4, 7/16)$ and $[7/16, 1)$ respectively. Thus, the string of two symbols AA, AB, BB, BA can be encoded as a real number in the intervals $[0, 1/16), [1/16, 1/4), [1/4, 7/16)$ and $[7/16, 1)$ respectively. Note that in this setting we have to transmit the length of the data input.

Decoding. From example 1.7, a string of two symbols can be decoded from a codeword $x = 0.1$ as the following. First we must know the length of the original text (here equal to two). First, read 0.1 in the current interval according to the probability p_A, p_B . Since $0 < 0.1 < 1/4$, thus the corresponding output symbol is A . Next, read 0.1 for the second symbol in the new subinterval according to the probability p_{AA} or p_{AB} . Now, $1/16 < 0.1 < 1/4$ and the corresponding symbol is B . Therefore the decoded string of length two is AB .

To be effectively implemented, the coding/decoding process must be carefully done to “emulate” the infinite precision floating number supposed by this technique.

1.5.4 Dictionary Compression Algorithms

In typical texts, the same substrings of the text (called phrases) may appear several times. *Dictionary-based methods* maintain a dictionary containing bits and pieces of the data that have been already seen (The “Déjà Vu” Principle) and can be therefore referred later with a short reference inside the dictionary. As a string of data symbols is read from the input, the algorithm searches the dictionary for the longest match to the string. Once a match is found, the string is compressed by replacing it with a pointer to the dictionary. Dictionary algorithms are widely used in applications (in commercial software programs and other disciplines). For instance, in UNIX or Linux, commands *compress*, *uncompress*, *gzip* and *gunzip* have all used the dictionary compression methods.

Dictionary compression algorithms are not based on a statistical model like those of Huffman coding and Arithmetic coding. The dictionary is used to store the string patterns seen before and the indexes are used to encode the repeated patterns. Dictionary-based methods are adaptive (dynamic) in nature because the dictionary is updated during the compression and decompression process. The content of the dictionary changes depending on the input sequence of the text to be compressed. In theory, the efficient implementation of these algorithms involves digital trees See [71]. In practice, hash tables (for tries), suffix arrays (for suffix trees) and other techniques are used. Of course a static dictionary is also possible if we know in advance a well-adapted dictionary for a given class of texts.

We introduce two most fundamental popular dictionary algorithms: LZ77 and LZ78. General principle of LZ algorithms is to partition a string into substrings called *phrases*. Each substring portion can then be referred via a pointer in the dictionary.

The LZ algorithm uses the dictionary method to perform compression. The principle of *dictionary-based compression* is based on the fact that parts of data tend to appear several times in a given data file. For example, some words or phrases may occur more frequently than the others in a text file. The dictionary method works as follows. First, it maintains a dictionary that contains bits and pieces of the data. Next, when a string of data symbols is read from the input, the algorithm searches the dictionary for the longest match to the string. Once a match is found, the string is compressed by replacing it with a pointer to the dictionary. The field of dictionary-based compression is based on the work of Jacob Ziv and Abraham Lempel in the LZ77 and LZ78 papers [76, 77].

The two algorithms use digital structures, as underlying data structures. The LZ77 is closely related to the suffix tree (trie) and the LZ78 to the digital search tree. An important theoretical property of these algorithms is their optimality. When the input text is generated by a “nice” source, compression is asymptotically optimal as the size of the input increases. That is, these two algorithms will code an indefinitely long string in the minimum size dictated by the entropy of the source. Then, it would appear that any one of the two algorithms LZ77 or LZ78 is the (optimal) solution to the text compression problem. The optimality, however, occurs as the size of the input tends to infinity.

1.5.5 LZ 77

The main idea is to use part of the previously-processed input as the dictionary.

At a given point of the encoding process, the basic idea is to find in the input text the longest prefix already seen. More precisely, let us assume that the first n symbols X_1^n of the text are given to the encoder and decoder. This initial string is sometimes called the “database string”. Then we search for the longest prefix X_{n+1}^{n+l} of X_{n+1}^∞ that is repeated in X_1^n , that is,

Let I_n be the largest ℓ such that $X_{n+1}^{n+\ell} = X_m^{m+\ell-1}$ for some prescribed range of m and ℓ .

In general, a code built for LZ77 consists of the triple (m, ℓ, char) , where char is the symbol $X_{m+\ell}$.

Usually, the LZ77 technique uses a sliding window, that we use only a *portion* of the previously seen input file. A set of suffixes of new phrases are inserted into the dictionary so that the corresponding structure is a trie. The encoder examines the input sequence through a sliding window, where the window consists of two parts: a *search buffer* that contains a portion of the recently encoded sequence, and a *look-ahead buffer* that contains the next portion of the sequence to be encoded.

The heart of all versions of the Lempel-Ziv schemes is the algorithm that finds the longest prefix of length I_n that occurs in the database string of length n . It turns out that the suffix tree discussed in Section 1.2.3 can be used to efficiently find such a factor. For example, consider a sequence $X = 1010010001\dots$ and assume X_1^4 is the database string. The suffix tree built over X_1^4 is shown in Figure 1.10.

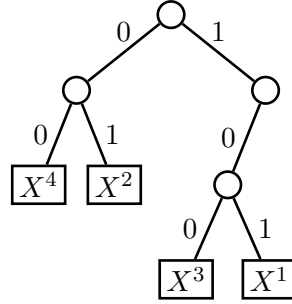


Figure 1.10: A suffix tree built from first four suffixes of $X = 1010010001\dots$.

Let us now look for I_4 , that is, the longest prefix of X_5^∞ that occurs in the database X_1^4 . In the “growing database implementation” it is X_5^8 since it is equal to X_2^5 . This can be seen by inserting the fifth suffix of X into the suffix tree from Figure 1.10– which actually leads to the suffix tree shown in Figure 1.5.

1.5.6 LZ78

The process is different in the algorithm called Lempel-Ziv78. The idea is again to cut the text into phrases. Each phrase is encoded with the index of a phrase that has occurred previously, followed by an explicit character (symbol) extended it. The notation (i, α) means copying the phrase of index i followed by character α . The text will thus be encoded by a sequence of pairs (**pointer, symbol**). The pointer indicates a previously parsed substring which is stored in a dictionary. (A phrase containing only one symbol is coded with index equal to zero.)

Example 1.8. Consider a string $X = abbaaabaabaab$ over the alphabet $\Sigma = \{a, b\}$.

Index:	1	2	3	4	5	6	7
Parsed phrase:	(a)	(b)	(ba)	(aa)	(baa)	(ab)	(aab)

The string X is parsed into 7 phrases shown in the above table and its code is $0a0b2a1a3a1b4b$ (no separator between phrases) corresponding the decomposition $(a)(b)(ba)(aa)(baa)(ab)(aab)$ of the input text. The pair $3a$ indicates that this phrase consists of the third phrase in the dictionary (ba) followed by the letter a . It is worth to observe that we need 3 ($=\lceil \log_2 7 \rceil$) bits to code

a phrase and then 7 phrases need 21 bits. Every terminal symbol requires one ($=\lceil \log_2 2 \rceil$) bit and there are 7 symbols in the code, hence the total length of the code is $21 + 7 = 28$.

The most time consuming part of the algorithm is finding the next phrase, that is searching the dictionary. A way to do this is to use a digital search tree to build the dictionary. For example, the string 11000101011011101 is partitioned into (1), (10), (0), (01), (010), (11), (011), (101). This parsing into phrases can be effectively done by inserting these phrases into a DST. Assume that the first phrase of the Lempel-Ziv scheme is the empty phrase that is stored in the root (and with index 0). To create a new phrase, the search starts at the root of the DST and proceeds down the tree as directed by the input symbols (exactly in the same manner as in the digital search tree construction). The search continues until we have to create a new node which will be the next phrase. Indexes of phrases are stored directly in the nodes of the tree. Each index is thus associated with its phrase seen as the word labelling the path from the root to the node.

We leave the root empty (or put the empty phrase into it) in this case. All the other phrases of the Lempel-Ziv algorithm are stored in internal nodes. When a new phrase is created, the search starts at the root and proceeds down the tree as directed by the input symbols exactly in the same manner as in the DST construction. The phrase is just a concatenation of symbols leading from the root to this node, which also stores the phrase.

An example showing how the phrases are inserted in a digital search tree structure is shown in Figure 1.11.

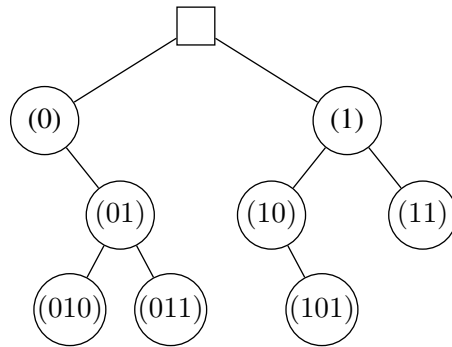


Figure 1.11: A digital tree representation of the Lempel-Ziv parsing for the string 11000101011011101.

The LZ78 method does not use any search buffer, look-ahead buffer, or sliding window. It is based on dictionary of previously-encountered strings. This dictionary starts empty, and its size is limited only by the amount of available memory. The encoder has two-field forms of outputs. The first field is a pointer to the dictionary; the second is the code of a symbol. Nothing is ever deleted from the dictionary, which is both an advantage over LZ77 (since future strings can be compressed even by strings seen in the distant past) and a liability (because the dictionary tends to grow rapidly and to fill up the entire available memory).

An attractive practical feature of LZ78 is that searching can be implemented efficiently by inserting each phrase in a DST structure. Each node in the DST contains the parsed phrase of the string. The process of inserting a new phrase will yield the longest phrase previously seen, so for each input character the encoder must traverse just one arc down the DST.

We observe the differences between digital search trees and the LZ parsing scheme. In the LZ scheme, we consider a word of *fixed length*, say n , while for DST we deal with a fixed number of strings, say m , resulting in a DST consisting of exactly m nodes. The number of nodes in the associated DST is equal to the number of phrases generated by the LZ algorithm. Then the

probabilistic models are different and the DST is just a first step to understand the complexity of the LZ process.

Conclusion of the Chapter.

We have defined the two main data structures which will be analyzed in the thesis. Of course, the shape of such structures heavily depends on the process which emits the words that will be further inserted in these digital trees. For instance, if the words share long common prefixes, it would be more difficult to distinguish them: the digital trees will have long branches, and their shape will be not compact. This is why the process that emits the words, that is called a source, will be the second main actor of the thesis. The next Chapter is devoted to present the process.

Chapter 2

Sources

Contents

2.1	Generalities on sources.	32
2.1.1	$\mathbb{N}$ -history and $\mathbb{Z}$ -history	32
2.1.2	Fundamental probabilities and generating function of the source.	32
2.1.3	Entropy of the source.	33
2.2	Simple sources	33
2.2.1	Memoryless sources.	33
2.2.2	Markov chains.	34
2.3	General sources: three models.	37
2.3.1	Sequence of sources and generalized transition matrix.	37
2.3.2	A dynamical point of view.	40
2.3.3	Parameterization of a source.	44
2.4	Relations between the three points of view.	46
2.4.1	The mirror operation and the operator $\hat{\mathbf{P}}_s$.	47
2.4.2	Extension to “infinite past”. The source $\underline{\mathcal{S}}$.	49
2.4.3	The dynamical system $\mathcal{D}$ and the transfer operator $\mathbb{H}_s$	50

In information theory contexts, data items are (infinite) words that are produced by a common mechanism, called a *source*. In this chapter, we first introduce in Section 2.1 main definitions, together with a main tool, the generating function of the source, the Λ series. Then, Section 2.2 describes the *simple* sources, where the correlations between successive symbols are weak, namely the memoryless sources and the Markov chains. In each case, the expression of the Λ series is provided. In Section 2.3, we introduce three different models for describing more general source. The first model, which is based on an extension of the transition matrix, appears to be new and will be very useful in the algebraic analysis of Chapters 5 and 6. However, in the case when the source is correlated, the transition matrix is infinite, and we have to find sufficient conditions to perform the analytic analysis (see Chapter 7). This is why we relate this model to two other models of sources that are well-adapted to analysis of algorithms, namely the model of dynamical source, introduced in 2001 by Vallée in [72] and now well-known, and the notion of parameterized source, introduced for the first time in 2009 in [74], and largely used in some other recent papers (see for instance [8]). We explain in Section 2.4 how to relate these three points of view, and we finally define the model of source we deal with, when we wish to analyze our two digital trees built on words emitted by a source.

2.1 Generalities on sources.

Consider a mechanism, called a *probabilistic source*, which produces symbols one at each time unit from a finite alphabet Σ , of cardinality r , denoted in the following as $\Sigma := \{0, 1, \dots, r-1\}$. When discrete time evolves, a source produces infinite words. More precisely, if X_n is the symbol emitted at time $t = n$, the source is defined by the sequence (X_n) of random variables with values in Σ . The sequence (X_n) may be *bi-infinite* and produces words in $\Sigma^{\mathbb{Z}}$; it may be only *right-infinite* and produces words in $\Sigma^{\mathbb{N}}$.

2.1.1 $\mathbb{N}$ -history and $\mathbb{Z}$ -history

There are thus two kinds of histories, which are made precise in the following definition:

Definition 2.1. *There are two kinds of sources:*

- (i) [$\mathbb{N}$ -history sources]. *A source is said to have a $\mathbb{N}$ -history if it is a right-infinite sequence of random variables of the form $(X_0, X_1, \dots, X_n, X_{n+1}, \dots) \in \Sigma^{\mathbb{N}}$.*
- (ii) [$\mathbb{Z}$ -history sources]. *A source is said to have a $\mathbb{Z}$ -history if it is a bi-infinite sequence of random variables of the form $(\dots, X_{-n}, \dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots, X_n, \dots) \in \Sigma^{\mathbb{Z}}$.*

A bi-infinite sequence defines two sequences, its positive history $(X_0, X_1, X_2, \dots, X_n, \dots)$ and its negative history, which we call its “past” history, namely $(\dots, X_{-n}, \dots, X_{-2}, X_{-1}, X_0)$. We will see that we are led to consider the “reverse past”, namely the sequence

$$(Y_0, Y_1, Y_2, \dots, Y_n, \dots) = (X_0, X_{-1}, X_{-2}, \dots, X_{-n}, \dots)$$

defined with $Y_i := X_{-i}$ for $i \geq 0$.

The bi-infinite framework (with indices $n \in \mathbb{Z}$ and words in $\Sigma^{\mathbb{Z}}$) is easier to deal with for *probabilistic* studies. The right-infinite framework (with indices $n \in \mathbb{N}$ and words in $\Sigma^{\mathbb{N}}$) is more natural to deal with for *algorithmic* studies.

There is a compromise which may be found: there exists a $\mathbb{Z}$ -history but only the positive part of the history is “shown”, whereas the negative part of the history is produced but remains “hidden”. It may have an influence on the positive part of the history. We adopt the point of view in the sequel. We always begin dealing with $\mathbb{N}$ -history sources, but, looking at “the past”, we build in natural way (see Section 2.4) a natural $\mathbb{Z}$ history for our sources,

In Sections 2.2 and 2.3, the sources have a positive history, and we consider source on the alphabet Σ , with an origin for time $t = 0$. Such a source is described by a sequence of random variables $(X_0, X_1, \dots, X_n, X_{n+1}, \dots)$, where each variable X_i takes its values in the alphabet Σ .

2.1.2 Fundamental probabilities and generating function of the source.

We now define the main objects of the source: the cylinders, the fundamental probabilities and the generating functions.

Definition 2.2. [Cylinders and fundamental probabilities]. *For $w \in \Sigma^*$, the subset $\mathcal{C}_w := w \cdot \Sigma^{\mathbb{N}}$ of $\Sigma^{\mathbb{N}}$ which gathers all the words which begin with the prefix w is called the cylinder of w . A source $\mathcal{S}$ is completely defined by the probabilities of its cylinders, which are called the fundamental probabilities. The probability of the cylinder $\mathcal{C}_w$ is denoted by p_w and is called the fundamental probability of the prefix w . This is the probability that an infinite word $X \in \Sigma^{\mathbb{N}}$ begins with a prefix w .*

Definition 2.3. The generating functions of the source $\mathcal{S}$ are series of Dirichlet type. The univariate generating series $\Lambda_k(s), \Lambda(s)$ are defined as

$$\Lambda_k(s) = \sum_{w \in \Sigma^k} p_w^s, \quad \Lambda(s) = \sum_{w \in \Sigma^*} p_w^s = \sum_{k \geq 0} \Lambda_k(s). \quad (2.1)$$

The bivariate generating series $\Lambda(s, u)$ are defined as

$$\Lambda_k(s, u) = \sum_{w \in \Sigma^k} u^{|w|} p_w^s, \quad \Lambda(s, u) = \sum_{w \in \Sigma^*} u^{|w|} p_w^s = \sum_{k \geq 0} \Lambda_k(s, u). \quad (2.2)$$

As the equality $\Lambda_k(1) = 1$ holds for any $k \geq 0$, the series $\Lambda(1, u)$ equals $1/(1 - u)$, and many interesting probabilistic properties of the source can be “seen” on the Dirichlet series $\Lambda(s, u)$ near the point $(1, 1)$.

2.1.3 Entropy of the source.

In information theory, the entropy is a central object which measures the quantity of information of the source. It plays a central role in the thesis.

Definition 2.4. The entropy $h(\mathcal{S})$ relative to a probabilistic source $\mathcal{S}$ is defined as the limit (if it exists) of a sequence that involves the fundamental probabilities,

$$h(\mathcal{S}) := \lim_{k \rightarrow \infty} \frac{-1}{k} \sum_{w \in \Sigma^k} p_w \log p_w. \quad (2.3)$$

Many probabilistic properties of the source are related to the behavior of the series $\Lambda(s), \Lambda_k(s)$ when s is close to 1. For example, the entropy admits the alternative expression

$$h(\mathcal{S}) = \lim_{k \rightarrow \infty} \frac{-1}{k} \frac{d}{ds} \Lambda_k(s)_{|s=1}. \quad (2.4)$$

2.2 Simple sources

The simple sources are those for which there are only *weak* correlations between successive symbols. For us, a simple source is a memoryless or a good¹ Markov chain. We review the properties of these sources, and describe in each case, their Λ series and their entropy.

2.2.1 Memoryless sources.

This is the simplest model, where the random variables are independent with the same distribution.

Definition 2.5. A source $\mathcal{S}$ is said to be memoryless if the variables X_k are independent with the same distribution. It is defined by the set p_i of probabilities, where p_i is the probability of emitting the symbol $i \in \Sigma$ at any time k , namely $p_i := \Pr[X_k = i]$ for any k . In the case when all the probabilities p_i are equal, the source is called unbiased

For a prefix $w = w_1 w_2 \cdots w_k \in \Sigma^k$, the fundamental probability p_w in the memoryless model satisfies the multiplicativity property $p_w = p_{w_1} p_{w_2} \cdots p_{w_k}$.

This leads to an exact expression of the Dirichlet series:

¹For us, and throughout the thesis, “good” means “irreducible and aperiodic”, as we explain in Section 2.2.2.

Lemma 2.6. *In the memoryless case, the Λ Dirichlet series defined in (2.1, 2.2), are expressed as a function of*

$$\lambda(s) = \sum_{i \in \Sigma} p_i^s$$

as

$$\Lambda_k(s) = \lambda(s)^k, \quad \Lambda(s) = \frac{1}{1 - \lambda(s)}, \quad \Lambda(s, u) = \frac{1}{1 - u\lambda(s)}$$

and the entropy equals

$$h(\mathcal{S}) = - \sum_{i \in \Sigma} p_i \log p_i = -\lambda'(1).$$

Proof. We begin with the multiplicative property of p_w which extends to p_w^s . Then

$$\sum_{w \in \Sigma^k} p_w^s = \sum_{w \in \Sigma^k} \prod_{i=1}^k p_{w_i}^s = \prod_{i=1}^k \sum_{w_i \in \Sigma} p_{w_i}^s = \lambda(s)^k.$$

This proves the first relation. When summing over k ; we get the other two relations. The expression of the entropy is easily deduced from (2.4), as one has

$$\frac{1}{k} \frac{d}{ds} \Lambda_k(s) = \frac{1}{k} \left(k \lambda'(s) \lambda^{k-1}(s) \right).$$

At $s = 1$, the equality $\lambda(1) = 1$ leads to the result. $\square$

In the case of the alphabet $\Sigma = \{0, 1, \dots, r-1\}$, the unbiased memoryless source is associated with probabilities defined as $p_i = 1/r$. For a finite word w , the probability p_w equals $1/r^{|w|}$. This unbiased source is related to the expansion in base r of a real that is uniformly drawn from the unit interval $[0, 1]$.

2.2.2 Markov chains.

A Markov chain (of order 1) is a source where the correlations between symbols may exist but are in a sense the “weakest” possible, as the emitted symbol can only be correlated with the previous symbol.

Definition 2.7. *A source on the alphabet Σ is a Markov chain of order 1 if and only if it satisfies the following : at each time k , and for each pair (i, j) of symbols of Σ^2 , the conditional probability of emitting j knowing that the previously emitted symbol is i does not depend on the time k , and is denoted by $p_{j|i}$,*

$$\forall k \in \mathbb{N}, \quad \Pr[X_{k+1} = j | X_k = i] = p_{j|i}.$$

A Markov source is then completely defined by the vector V of initial probabilities $(v_i)_{i \in \Sigma}$ together with the transition matrix $\mathbf{R} := (p_{j|i})_{(i,j) \in \Sigma^2}$.

The matrix $\mathbf{R}$ is *stochastic*: the sum of each row is equal to 1. This entails that the matrix $\mathbf{R}$ has an eigenvalue equal to 1 with the eigenvector $\mathbf{1}$ whose all the components are equal to 1.

The following result provides an expression of the Λ -Dirichlet series as a function of the basic objects of the Markov chain:

Lemma 2.8. For a Markov chain, the Λ series defined in (2.1, 2.2) are expressed with the matrix $\mathbf{R}_s$ of general coefficient $p_{j|i}^s$, together with the vector V_s of components v_i^s ,

$$\Lambda_k(s) = {}^tV_s \cdot \mathbf{R}_s^{k-1} \cdot \mathbf{1}, \quad \Lambda(s) = 1 + {}^tV_s \cdot (I - \mathbf{R}_s)^{-1} \cdot \mathbf{1}, \quad \Lambda(s, u) = 1 + u {}^tV_s \cdot (I - u\mathbf{R}_s)^{-1} \cdot \mathbf{1},$$

where $\mathbf{1}$ is the column vector whose all components are equal to 1.

Proof. For a finite prefix w of the form $w = w_1 w_2 \dots w_k \in \Sigma^k$, the fundamental probability p_w in this case is

$$p_w = v_{w_1} p_{w_2|w_1} p_{w_3|w_2} \dots p_{w_k|w_{k-1}}.$$

For $k \geq 1$, the coefficient of the matrix $\mathbf{R}_s^{k-1}$ at the index (i, j) is the sum of all the terms of the form

$$p_{i_2|i_1}^s \cdot p_{i_3|i_2}^s \dots p_{i_k|i_{k-1}}^s, \quad \text{with } i_1 = i, i_k = j, \text{ and } (i_2, \dots, i_{k-1}) \text{ any vector of } \Sigma^{k-2}.$$

Then, the matrix ${}^tV_s \mathbf{R}_s^{k-1}$ (that has one row and r columns) has in its j -th column the sum of all the terms of the form

$$v_i^s \cdot p_{i_2|i}^s \cdot p_{i_3|i_2}^s \dots p_{j|i_{k-1}}^s, \quad \text{with } (i, i_2, \dots, i_{k-1}) \text{ any vector of } \Sigma^{k-1}.$$

This is the sum of all the terms of the form p_w^s where the prefix w of length k ends with the symbol j . To obtain the sum over all the prefixes of length k , we apply the matrix with the column vector, whose all its coefficients equal to 1, namely the vector $\mathbf{1}$. We have then proven the first equality of the Lemma,

$$\Lambda_k(s) = {}^tV_s \cdot \mathbf{R}_s^{k-1} \cdot \mathbf{1}.$$

We sum over k to get the other relations. □

The coefficient (i, j) of the matrix $\mathbf{R}^k$, denoted by $\mathbf{R}_{j|i}^k$, represents the probability of going from the state i to the state j in k time units. A Markov chain (X_n) is *irreducible* if and only if every state can be reached from every other state, that is for all $i, j \in \Sigma$, there exists $n > 0$ such that $\mathbf{R}_{j|i}^n > 0$. A Markov chain (X_n) is *aperiodic* if for any state $i \in \Sigma$, the possible times to go from i to i [namely the possible lengths of cycles which contain i] have a greatest common divisor 1, that is $\gcd\{n > 0 : \mathbf{R}_{i|i}^n > 0\} = 1$.

In the thesis, the term “aperiodic” is used with another meaning. This is why we never use the term *aperiodic* with the previous meaning. A Markov chain whose transition matrix is irreducible and aperiodic will be said to be a good Markov chain. This will be said for the transition matrix too.

Definition 2.9. A Markov chain which is irreducible and aperiodic is said to be good.

The Perron-Frobenius theorem states the following: Consider a matrix $\mathbf{T}$ with positive coefficients which is moreover good. Then, the matrix $\mathbf{T}$ has a unique dominant eigenvalue λ , and a unique dominant eigenvector Π with positive components π_i whose sum equals 1.

We apply this theorem to the matrices ${}^t\mathbf{R}_s$ and $\mathbf{R}_s$ for any real s . Then, the two matrices ${}^t\mathbf{R}_s$ and $\mathbf{R}_s$ have the same unique dominant eigenvalue $\lambda(s)$. The matrix ${}^t\mathbf{R}_s$ has a unique dominant eigenvector Π_s with positive components $\pi_s^{(j)}$. The matrix $\mathbf{R}_s$ has a unique dominant eigenvector Ψ_s with positive components $\psi_s^{(j)}$. Since the matrix $\mathbf{R}$ is stochastic, the dominant eigenvalue $\lambda(s)$ satisfies $\lambda(1) = 1$, and the matrix ${}^t\mathbf{R} = {}^t\mathbf{R}_1$ has a unique (normalized) fixed vector $\Pi := \Pi_1$ with positive components $\pi^{(j)}$, whose sum equals 1.

With these properties, we prove the following:

Lemma 2.10. *In the case of a good Markov chain, with a transition matrix $\mathbf{R} = (p_{j|i})$, a dominant eigenvalue $\lambda(s)$, and a fixed (normalized) fixed vector $\Pi := (\pi^{(j)})$ of ${}^t\mathbf{R}$, the following holds*

(a) *The Λ -series satisfies for (s, u) near $(1, 1)$,*

$$\Lambda(s, u) \asymp \frac{1}{1 - u\lambda(s)}$$

(b) *The entropy satisfies*

$$h(\mathcal{S}) = -\lambda'(1) = - \sum_{(i,j) \in \Sigma^2} \pi^{(i)} p_{j|i} \log p_{j|i}. \quad (2.5)$$

Proof.

(a) With the Perron-Frobenius theorem, and for real values of s , the matrix $\mathbf{R}_s$ decomposes as

$$\mathbf{R}_s = \lambda(s)\mathbf{G}_s + \mathbf{N}_s$$

where $\mathbf{G}_s$ is the projection on the dominant eigenspace, and $\mathbf{N}_s$ is the remainder matrix, whose spectral radius $\rho(s)$ satisfies $\rho(s) := \max\{|\lambda|; \lambda \in \text{Sp}\mathbf{R}_s\} < |\lambda(s)|$. In this case, the dominant projector is defined by the relation $\mathbf{G}_s[\mathbf{x}] = \langle \Pi_s, \mathbf{x} \rangle \cdot \Psi_s$ which involves the scalar product $\langle \cdot, \cdot \rangle$ defined on $\mathbb{R}^r$. These matrices satisfy $\mathbf{G}_s \cdot \mathbf{N}_s = \mathbf{N}_s \cdot \mathbf{G}_s = 0$, so that the previous decomposition extends to any $k \geq 1$, namely

$$\mathbf{R}_s^k = \lambda^k(s)\mathbf{G}_s + \mathbf{N}_s^k, \quad \text{and thus} \quad (I - \mathbf{R}_s)^{-1} = \frac{\lambda(s)}{1 - \lambda(s)}\mathbf{G}_s + (I - \mathbf{N}_s)^{-1}. \quad (2.6)$$

This entails the following asymptotic behavior for the Λ Dirichlet series

$$\Lambda_k(s) = \lambda^{k-1}(s) [{}^tV_s \cdot \mathbf{G}_s \cdot \mathbf{1}] + {}^tV_s \cdot \mathbf{N}_s^k \cdot \mathbf{1} = \lambda^k(s)w_s [1 + O(\rho^k)]$$

for some nonzero constant w_s and some $\rho < 1$. This also proves the estimate

$$\Lambda(s, u) = 1 + \frac{u\lambda(s)}{1 - u\lambda(s)} [{}^tV_s \cdot \mathbf{G}_s \cdot \mathbf{1}] + {}^tV_s \cdot (I - u\mathbf{N}_s)^{-1} \cdot \mathbf{1}$$

which leads to Assertion (a)

(b) We first prove the equality $h(\mathcal{S}) = -\lambda'(1)$. This is obtained by taking the derivative of the estimate $\Lambda_k(s) = \lambda^k(s)w_s [1 + o(\rho^k)]$ with respect to k , namely

$$\frac{1}{k} \frac{d}{ds} \Lambda_k(s) \sim_{k \rightarrow \infty} \lambda'(s) \lambda^{k-1}(s) w_s \quad \text{and then} \quad \frac{1}{k} \frac{d}{ds} \Lambda_k(s)|_{s=1} \sim_{k \rightarrow \infty} \lambda'(1),$$

since the equality $w_1 = 1$ holds.

We then obtain an alternative expression for the derivative $\lambda'(1)$. Taking the derivative (with respect to s) of the equality ${}^t\Pi_s \cdot \mathbf{R}_s = \lambda(s) {}^t\Pi_s$ leads at $s = 1$ to the equality

$${}^t\Pi_1 \cdot \mathbf{R}'_1 \cdot \mathbf{1} + {}^t\Pi'_1 \cdot \mathbf{R}_1 \cdot \mathbf{1} = \lambda'(1) {}^t\Pi_1 \cdot \mathbf{1} + \lambda(1) {}^t\Pi'_1 \cdot \mathbf{1}.$$

Moreover, since the matrix $\mathbf{R}$ is stochastic, the equality $\mathbf{R}_1 \mathbf{1} = \mathbf{1}$ holds. This entails the expression for the entropy of the source given in (2.5). $\square$

The two expressions given in Lemmas 2.6 and 2.8 are the first steps to study the analytical properties of the Dirichlet series, which will be central in the analysis of the probabilistic properties of the source. We wish to describe natural extensions of these simple sources, which have possibly stronger correlations between their symbols, but where the generating function $\Lambda(s, u)$ admits an alternative expression which makes possible its analytic study. More precisely, we are interested in *general* sources for which the $\Lambda(s, u)$ series admits an alternative expression of the same type as (2.6, 2.8), from which the position and the nature of its singularities become apparent.

2.3 General sources: three models.

We introduce in this Section three possible models for a general source $\mathcal{S}$. First, we consider a source $\mathcal{S}$ as a sequence of *shifted* sources and extend the notion of transition matrix of a Markov chain. Second, we adopt a dynamical point of view, and introduce *dynamical sources* which extend simple sources in a natural way. Finally, we consider a natural *parameterization* of the source, which “replaces” cylinders by fundamental intervals and associates a real of the unit interval with an infinite word emitted by the source. The second model (dynamical sources) was proposed by Vallée in [72] and the third point of view (the notion of parameterization) was introduced in [73]. We begin with the first model, which has not yet been used in Analysis of Algorithms.

2.3.1 Sequence of sources and generalized transition matrix.

A source defines a sequence of sources, which are related via a directed graph whose matrix provides a generalization of the transition matrix of a Markov chain.

Sequence of sources. We now explain how the source $\mathcal{S}$ defines a sequence of (conditional) sources $\mathcal{S}_{(u)}$ (for $u \in \Sigma^*$), as it is now described:

Definition 2.11. *For a prefix u whose fundamental probability p_u is non zero, the source $\mathcal{S}_{(u)}$ gathers all the words of $\mathcal{S}$ which begin with $u \in \Sigma^*$, from which the prefix u is removed. The source $\mathcal{S}_{(u)}$ is completely defined by the fundamental (conditional) probabilities $p_{w|u} := p_w/p_u$, when w is any finite prefix which begins with u .*

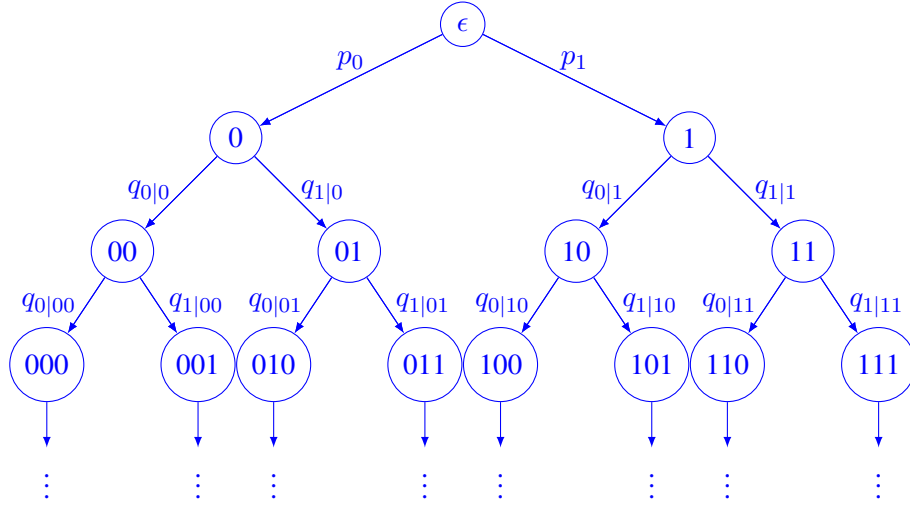
In the case when w is a prefix which begins with u (we denote this situation by the inequality $w \geq u$), the prefix w is written as $w = u \cdot v$; since the prefix u is removed for building the source $\mathcal{S}_{(u)}$, the prefix v is the prefix which remains “visible” in the source $\mathcal{S}_{(u)}$. Then, 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 $\mathcal{S}_{(u)}$. 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 $\mathcal{S}_{(u)}$.

Graph and generalized transition matrix. The relations between these different sources are described by a (directed) graph.

Definition 2.12. *The graph of a source $\mathcal{S}$ is the weighted directed graph which admits as vertices the sources $\mathcal{S}_{(u)}$ associated with prefixes u for which $p_u \neq 0$. Furthermore, there is an edge from $\mathcal{S}_{(u)}$ to $\mathcal{S}_{(v)}$ if and only the three conditions hold:*

$$\text{there exists a symbol } i \in \Sigma \text{ for which } v = u \cdot i, \quad p_u \neq 0, \text{ and } \quad p_v \neq 0.$$

The weight of this edge is the probability $p_v/p_u = q_{i|u}$.

Figure 2.1: The graph of the transition matrix $\mathbf{P}$ of the source $\mathcal{S}$.

The matrix of this graph will play a central role in this thesis. It extends the transition matrix of a Markov chain and this is why it is called the transition matrix of the source.

Definition 2.13. The matrix $\mathbf{P}$ of the source is an infinite matrix, whose rows and columns are indexed by Σ^* . The coefficients at the row u which are possibly non-zero are located at the columns $u \cdot i$ (for $i \in \Sigma$) and equal $p_{u,i}/p_u = q_{i|u}$.

The matrix $\mathbf{P}_s$ is obtained from the matrix $\mathbf{P}$ by raising all the non-zero coefficients of the matrix $\mathbf{P}$ to the power s

Full expression of the Λ Dirichlet series. We now describe an alternative expression for the Λ -series of the source, of the same flavor as expressions provided in Lemma 2.6 or in Lemma 2.8:

Lemma 2.14. The Λ series of the source $\mathcal{S}$ are expressed with the transition matrix $\mathbf{P}_s$ of the source, under the form

$$\Lambda_k(s) = {}^t\mathbf{E} \cdot \mathbf{P}_s^k \cdot \mathbf{1}, \quad \Lambda(s) = {}^t\mathbf{E} \cdot (I - \mathbf{P}_s)^{-1} \cdot \mathbf{1}, \quad \Lambda(s, u) = {}^t\mathbf{E} \cdot (I - u\mathbf{P}_s)^{-1} \cdot \mathbf{1},$$

where $\mathbf{1}$ is the column vector indexed by Σ^* whose all components are equal to 1, and ${}^t\mathbf{E}$ is the row vector indexed by Σ^* whose all components are equal to 0 except the component relative to the empty prefix ϵ , that is equal to 1.

Proof. For any $k \in \mathbb{N}$, the only coefficients of matrix $\mathbf{P}_s^k$ at the row u that are possibly non-zero are located at the columns $u \cdot w$ (for $w \in \Sigma^k$) and equal to $(p_{u,w}/p_u)^s$. Then, we find all the probabilities p_w relative to a prefix $w \in \Sigma^k$ in the row of index ϵ of the matrix $\mathbf{P}_s^k$. Then, the series $\Lambda_k(s)$ is expressed under the form

$$\Lambda_k(s) := \sum_{w \in \Sigma^k} p_w^s = {}^t\mathbf{E} \cdot \mathbf{P}_s^k \cdot \mathbf{1},$$

where the two vectors $\mathbf{1}$ and $\mathbf{E}$ are defined in Lemma 2.14. □

Pruning the graph and the matrix. This graph takes into account the correlations between symbols and may be redundant for simple sources, where the correlations between emitted symbols are “weak”. In this case, all the sources $\mathcal{S}_{(u)}$ are not needed for the description of the source $\mathcal{S}$. Generally speaking, we say that two sources are equivalent if and only if they have the same distribution (i.e., the same fundamental probabilities). Then, the equivalence relation on “shifted” sources is defined by

$$\mathcal{S}_{(u)} \equiv \mathcal{S}_{(v)} \iff \forall w \in \Sigma^*, \quad q_{w|u} = q_{w|v}.$$

Definition 2.15. The pruned graph is obtained by keeping only one representative in each equivalence class. The set of its vertices defines a subset $\Sigma_0 \subset \Sigma^*$ which is called the pruned language. The pruned matrix of the matrix $\mathbf{P}_s$ is called the pruned matrix of the source and it is denoted by $\mathbf{A}_s$. Its rows and its columns are indexed by Σ_0 .

We now give three examples of the resulting pruned graph, represented in Figure 2.2.

Memoryless source. All the sources $\mathcal{S}_{(u)}$ are equivalent, for any $u \in \Sigma^*$, and there is only one equivalence class. The pruned graph has only one vertex, The pruned language is $\{\epsilon\}$ and $\mathbf{A}_s$ is a matrix of order 1, with a (unique) coefficient equal to $\lambda(s) = p_0^s + p_1^s + \dots + p_{r-1}^s$, where p_i is the probability of emitting the symbol i .

Markov chain of order k . There are two types of sources:

- first, the “initial” sources $\mathcal{S}_{(u)}$ related to a prefix u of length strictly less than k ;
- then, all the sources $\mathcal{S}_{(u)}$ related to a prefix u whose suffix of length k equals v are all equivalent to source $\mathcal{S}_{(v)}$.

There is a close connection between the transition matrix $\mathbf{A}_s$ and the usual data of a Markov chain, described by the matrix $\mathbf{R}_s$ and the initial vector V_s whose components are v_i^s . In the case of a Markov chain of order 1, the language Σ_0 is $\Sigma \cup \{\epsilon\}$ of cardinality $r + 1$ the transition matrix $\mathbf{P}_s$ (of order $r + 1$) is written as

$$\mathbf{A}_s := \begin{pmatrix} 0 & {}^tV_s \\ \mathbf{0} & \mathbf{R}_s \end{pmatrix}$$

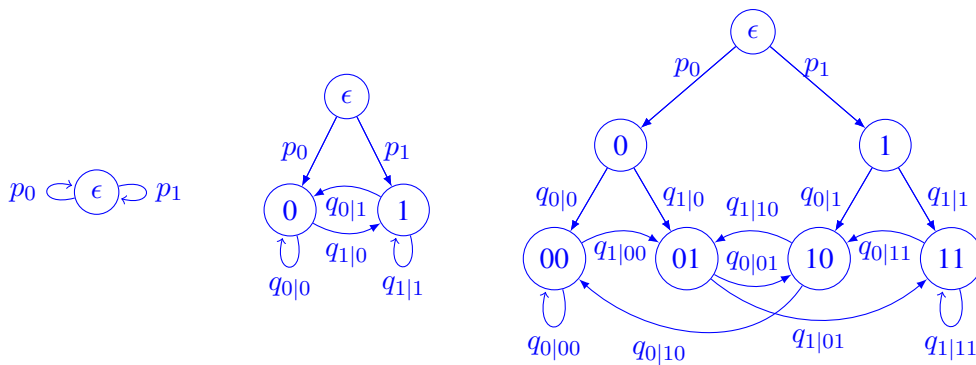


Figure 2.2: The graph for a memoryless source (on the left) has only one equivalence class, the graph (middle) for a Markov chain of order 1 with 2 equivalence classes, and the graph (right) for a Markov chain of order 2 with 4 equivalence classes.

An instance of an infinite pruned graph: the comb. This is an instance of a VLMC (Variable Length Markov Chain). We consider an *intermittent* source on the binary alphabet $\{0, 1\}$ where the dependency from the past is unbounded. Such a source has two regimes, depending whether

it emits 0 or 1. Two sources $S_{(u)}$ and $S_{(v)}$ are equivalent if and only there exists $k \geq 0$ for which the two prefixes u and v finish with a sequence of exactly k occurrences of 0. Each equivalence class is then labelled by the integer k , which equals the maximal length of the last sequence of 0. And, for $k \geq 1$, the minimal source of the equivalence class is the source $S_{(0^k)}$; for $k = 0$, there is another equivalence class formed with all the sources $S_{(u)}$ for which u finishes with 1, and the minimal source is the source $S_{(1)}$.

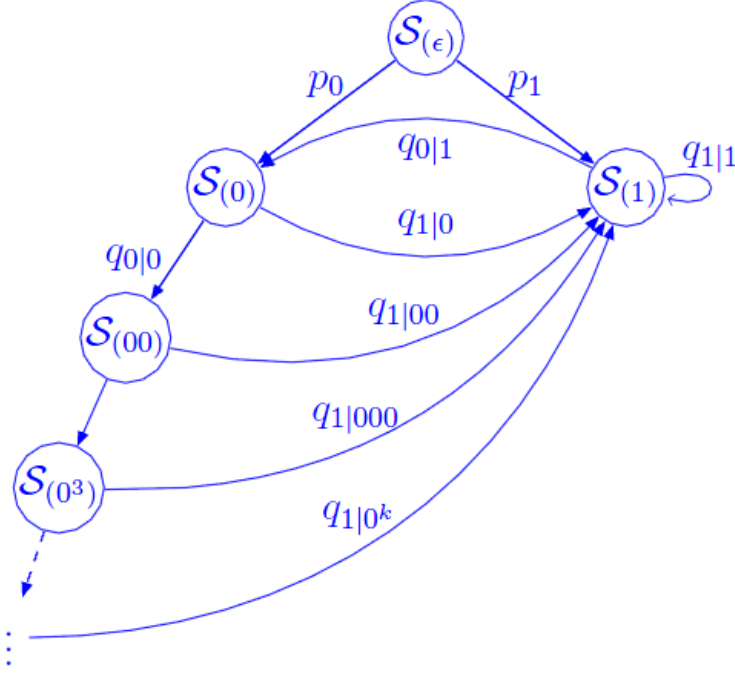


Figure 2.3: The pruned graph of the intermittent source called “comb”.

This gives rise to a pruned graph which remains *infinite*, with vertices $S_{(\epsilon)}$, $S_{(1)}$ and $S_{(0^k)}$ for $k > 0$, and the pruned language is $\Sigma_0 = \{\epsilon, 1\} \cup \{0^k; k \geq 1\}$. All the edges labeled with 1 return to the source $S_{(1)}$. See Figure 2.3.

Pruned expression of the Λ Dirichlet series.

Lemma 2.16. *The Λ series of the source S are expressed with the pruned matrix $\mathbf{A}_s$ of the source, under the form*

$$\Lambda_k(s) = {}^t\mathbf{E} \cdot \mathbf{A}_s^k \cdot \mathbf{1} \quad \Lambda(s) = {}^t\mathbf{E} \cdot (I - \mathbf{A}_s)^{-1} \cdot \mathbf{1}, \quad \Lambda(s, u) = {}^t\mathbf{E} \cdot (I - u\mathbf{A}_s)^{-1} \cdot \mathbf{1},$$

where $\mathbf{1}$ is now the column vector indexed by Σ_0 whose all components are equal to 1, and $\mathbf{E}$ is the row vector indexed by Σ_0 whose all components are equal to 0 except the component relative to the empty prefix ϵ , that is equal to 1.

Proof. Clear from the definition. □

2.3.2 A dynamical point of view.

There is another point of view which extends the notion of simple sources. This is a *dynamical point of view*, and we consider *dynamical* sources which are built with dynamical systems of

the interval. Informally speaking, simple sources are recovered as dynamical sources with *affine* branches. However, the model of dynamical sources provides an actual extension of these simple sources: when the derivatives of the branches of the dynamical system are no longer constant, this creates correlations between successive symbols and the sources are no longer simple.

Dynamical system of the unit interval. We first recall the definition of a dynamical system, as it is used in the present context (see [72]).

Definition 2.17 (Dynamical system of the interval). *A dynamical system of interval $\mathcal{I} := [0, 1]$ is defined by a mapping $T : \mathcal{I} \rightarrow \mathcal{I}$ (called the shift) for which*

(a) *there exists a finite alphabet Σ , and a topological partition of $\mathcal{I}$ with disjoint open intervals $\mathcal{I}_m$, $m \in \Sigma$, i.e. $\bar{\mathcal{I}} = \bigcup_{m \in \Sigma} \bar{\mathcal{I}}_m$.*

(b) *The restriction of T to each $\mathcal{I}_m$ is a $\mathcal{C}^2$ bijection from $\mathcal{I}_m$ to $T(\mathcal{I}_m)$.*

The system is complete when each restriction is surjective, i.e., $T(\mathcal{I}_m) = \mathcal{I}$.

The system is Markovian when each interval $T(\mathcal{I}_m)$ is a union of intervals $\mathcal{I}_j$.

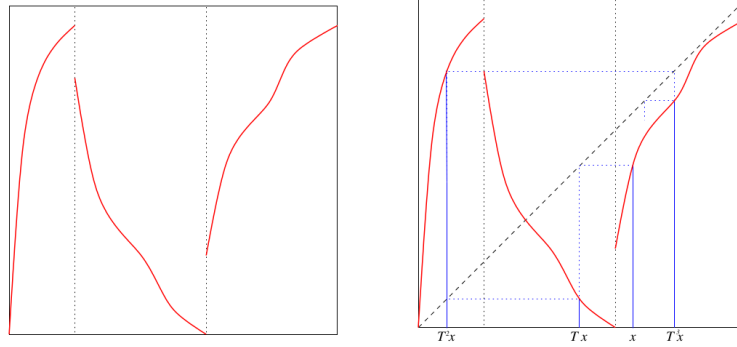


Figure 2.4: A dynamical system with $\Sigma = \{a, b, c\}$ and a word $M(x) = (c, b, a, c, \dots)$

A dynamical system, together with a distribution G on the interval $\mathcal{I}$, defines a probabilistic source, which is called a *dynamical source* and is now described. The map T is used as a shift mapping, and the mapping τ whose restriction to each $\mathcal{I}_m$ is equal to m , is used for coding. The words are emitted as follows. To each real x , (except for a denumerable set), one associate the trajectory

$$\mathcal{T}(x) = (x, T(x), T^2(x), \dots, T^j(x), \dots);$$

which gives rise, via the mapping τ , to the word $M(x) \in \Sigma^{\mathbb{N}}$ defined as

$$M(x) = (m_1(x), m_2(x), m_3(x), \dots, m_j(x), \dots), \quad \text{with} \quad m_j(x) = \tau(T^{j-1}(x)). \quad (2.7)$$

Inverse branches. In the case of a complete system, one denotes by h_j the local inverse of T restricted to $\mathcal{I}_j$ and by $\mathcal{H} := \{h_j, j \in \Sigma\}$ the set of all local inverses. Each local inverse of the k -th iterate T^k is then associated to a word $w = m_1 m_2 \dots m_k \in \Sigma^k$; it is of the form $h_w := h_{m_1} \circ h_{m_2} \dots \circ h_{m_k}$, and

$$\mathcal{I}_w = h_w(\mathcal{I}), \quad p_w = |G(h_w(1)) - G(h_w(0))|. \quad (2.8)$$

The set of all the inverse branches of T^k is $\mathcal{H}^k = \{h_w; w \in \Sigma^k\}$. For $h \in \mathcal{H}^k$, the number k is called the *depth* of h and it is denoted by $p(h)$. The set $\mathcal{H}^* := \bigcup_{k \geq 0} \mathcal{H}^k$ is thus the set of all inverse branches (of any depth).

Simple sources. Simple sources can be recovered as particular cases of dynamical sources.

A memoryless source is a dynamical source which satisfies the two properties:

- (a) It is associated to a complete dynamical system where each restriction $T_{[m]}$ of T to $\mathcal{I}_m$ is affine (increasing or decreasing)
- (b) Its initial density is uniform.

A Markov chain is a dynamical source which satisfies the two properties:

- (a) It is associated to a Markovian dynamical system whose each restriction $T_{[m]}$ of T to $\mathcal{I}_m$ is affine (increasing or decreasing)
- (b) Its initial density is constant on each $\mathcal{I}_m$.

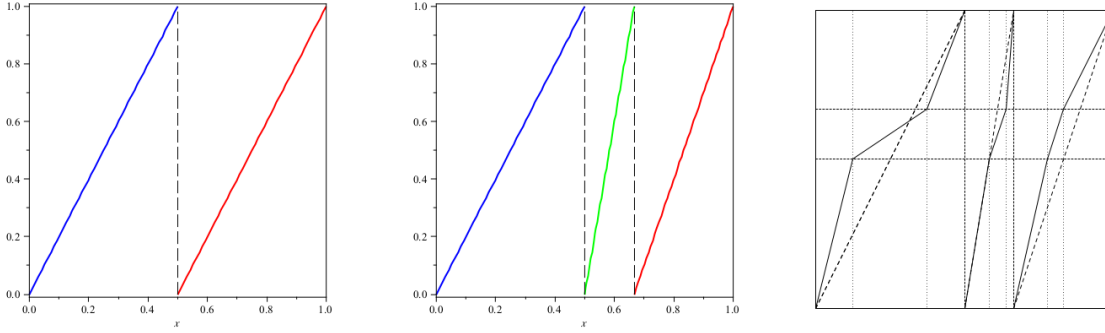


Figure 2.5: Memoryless or Markov chain

Plain transfer operator. One of the main tools in dynamical system theory is the transfer operator introduced by Ruelle [62], denoted by $\mathbf{H}_s$. It generalizes the density transformer $\mathbf{H}$ that describes the evolution of the density.

We here consider the case of a complete dynamical system: if $f = f_0$ denotes the initial density on $\mathcal{I}$, and f_1 the density on $\mathcal{I}$ after one iteration of T , then f_1 can be written as $f_1 = \mathbf{H}[f_0]$, where $\mathbf{H}$ is defined by

$$\mathbf{H}[f](x) := \sum_{h \in \mathcal{H}} |h'(x)| f \circ h(x) = \sum_{i \in \Sigma} |h'_i(x)| f \circ h_i(x).$$

The transfer operator extends the density transformer; it depends on a complex parameter s , acts on the functions f defined on the unit interval $\mathcal{I}$, and of class $\mathcal{C}^1$; it is defined as

$$\mathbf{H}_s[f](x) := \sum_{h \in \mathcal{H}} |h'(x)|^s f \circ h(x) = \sum_{i \in \Sigma} |h'_i(x)|^s f \circ h_i(x), \quad (2.9)$$

and coincides with $\mathbf{H}$ when $s = 1$.

Secant operator. The secant transfer operator $\mathbb{H}_s$ introduced by Vallée in [72], is a generalized version of the (classical) transfer operator. This operator involves the secant of inverse branches (instead of their derivatives), it acts on functions F of two variables defined on $\mathcal{I}^2$ and of class $\mathcal{C}^1$; it is defined as

$$\mathbb{H}_s[F](x, y) := \sum_{h \in \mathcal{H}} \left| \frac{h(x) - h(y)}{x - y} \right|^s F(h(x), h(y)) = \sum_{i \in \Sigma} \left| \frac{h_i(x) - h_i(y)}{x - y} \right|^s F(h_i(x), h_i(y)). \quad (2.10)$$

Denote by $\text{diag } F$ the function defined by $\text{diag } F(x) := F(x, x)$. The equality $\mathbb{H}_s[F](x, x) = \mathbf{H}_s[\text{diag } F](x)$ holds on the diagonal $x = y$ and shows that the secant operator is an extension of the plain transfer operator. Moreover, multiplicative properties of secants entail the relation

$$\mathbb{H}_s^k[F](x, y) = \sum_{h \in \mathcal{H}^k} \left| \frac{h(x) - h(y)}{x - y} \right|^s F(h(x), h(y)). \quad (2.11)$$

Expression of the Λ -series. In our regular dynamical systems, the branches h are of class $\mathcal{C}^2$ and their secant (as their tangents) are of class $\mathcal{C}^1$. This secant operator is very useful for generating fundamental probabilities. This is why the Λ -series can be expressed with this secant operator.

Lemma 2.18. *Consider a complete dynamical source, defined by the pair $(\mathcal{I}, T)$, a distribution G , and the secant L of the distribution G , defined by*

$$L(x, y) := \frac{G(x) - G(y)}{x - y}. \quad (2.12)$$

Then, the Λ -series of the source defined in (2.1, 2.2) admit alternative expressions which involve the quasi-inverse of the secant operator defined in (2.10), applied to the function L^s , where L is the secant of the distribution G ,

$$\Lambda_k(s) = \mathbb{H}_s^k[L^s](1, 0), \quad \Lambda(s) = (1 - \mathbb{H}_s)^{-1}[L^s](1, 0), \quad \Lambda(s, u) = (I - u\mathbb{H}_s)^{-1}[\mathbf{1}](1, 0).$$

Proof. With Equation (2.8), we write, for $w \in \Sigma^k$,

$$p_w^s = |G(h_w(1)) - G(h_w(0))|^s = \left| \frac{h_w(1) - h_w(0)}{1 - 0} \right|^s \times \left| \frac{G(h_w(1)) - G(h_w(0))}{h_w(1) - h_w(0)} \right|^s$$

Now, if L is the secant of the distribution G , defined in (2.12), then the Λ -series defined in (2.1, 2.2) can be expressed as

$$\Lambda_k(s) := \sum_{w \in \Sigma^k} p_w^s = \mathbb{H}_s^k[L^s](1, 0), \quad \Lambda(s) := \sum_{w \in \Sigma^*} p_w^s = (1 - \mathbb{H}_s)^{-1}[L^s](1, 0). \quad (2.13)$$

□

Good Class. Here we consider particular complete dynamical systems, for which it is possible to prove that the quasi-inverse has nice spectral properties on a convenient functional space. This will entail nice properties for the $\Lambda(s)$ series, which are reviewed in Appendix A.

Definition 2.19. *A dynamical system of the interval $(\mathcal{I}, T)$ belongs to the Good Class if it is complete, with a set $\mathcal{H}$ of inverse branches which is uniformly contracting, i.e, there exists a constant $\rho < 1$ (called the contraction ratio) for which*

$$\forall h \in \mathcal{H}, \quad \forall x \in \mathcal{I}, \quad |h'(x)| \leq \rho.$$

When the dynamical system belongs to the Good Class, the transfer operator (acting on a convenient functional space) behaves as a finite matrix, and all what we have said for the transition matrix $\mathbf{P}_s$ of a good Markov chain can be extended to the transfer operator $\mathbb{H}_s$: It admits dominant spectral properties for s near the real axis, together with a spectral gap. This implies the following :

Lemma 2.20. *For a dynamical source of the Good Class, the function $\Lambda(s)$ is analytic on the half plane $\{\Re s > 1\}$ and admits a simple pole at $s = 1$ with a residue equals to $-1/\lambda'(1)$. Moreover, the entropy of the source is well defined and expressed with the dominant eigenvalue $\lambda(s)$ of the transfer operator $\mathbb{H}_s$ as $h(\mathcal{S}) = -\lambda'(1)$.*

2.3.3 Parameterization of a source.

The parameterization of a source is based on the same principle as those used for arithmetic coding (See Chapter 1) in compression frameworks. It also aims at extending what we have done in the case of a dynamical system: we have associated to a real $x \in \mathcal{I}$ a word $M(x)$ that is the encoding of the trajectory $\mathcal{T}(x)$ [See (2.7)]. We have also built in (2.8) a family of fundamental intervals $\mathcal{I}_w$, whose measures are the fundamental probabilities of the source. Then, for each depth k , the fundamental intervals $\mathcal{I}_w$ form a topological partition of the interval $\mathcal{I}$.

A general source is completely defined by the family (p_w) . We will now adopt the inverse point of view and wish to associate to this source a family of fundamental intervals $\mathcal{I}_w$ whose probability is equal to p_w and a parameterization $M : \mathcal{I} \rightarrow \Sigma^{\mathbb{N}}$, which will extend the point of view of dynamical sources.

Non ambiguous sources. Here, and in all the sequel, we restrict ourselves to a class of particular sources, that gathers *non ambiguous* sources.

Definition 2.21. Let Σ be alphabet of cardinality r which will be ordered. A source over the alphabet Σ produces infinite words of $\Sigma^{\mathbb{N}}$, and is specified by the fundamental probabilities p_w , $w \in \Sigma^*$, where p_w is the probability that an infinite word begins with the finite prefix w . When the two following properties hold,

(i) $p_w > 0$ for any $w \in \Sigma^*$, (ii) $\pi_k := \max\{p_w : w \in \Sigma^k\}$ tends to 0, as $k \rightarrow \infty$, the source is said to be non-ambiguous.

In the sequel, all the sources are assumed to be non-ambiguous.

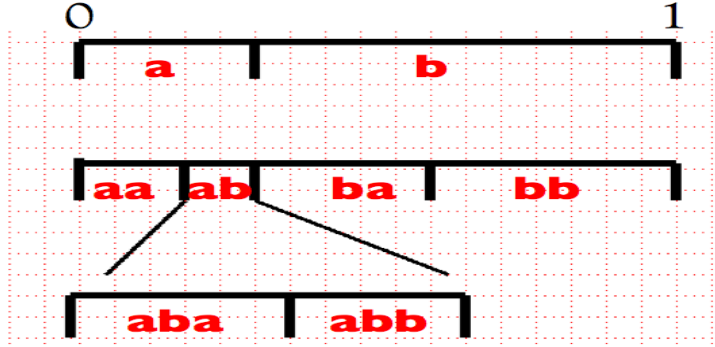


Figure 2.6: The parameterization of a source. One associates with a real $x \in [0, 1]$ an infinite word denoted by $M(x)$. The first symbol is the index of the fundamental interval of depth 1 that contains x . the prefix of length 2 is the index of the fundamental interval of depth 2 that contains x . The process is repeated for the prefixes of increasing length by considering the successive depths. When x is uniformly chosen in $[0, 1]$, the probability that $M(x)$ begins with w is equal to p_w .

The mapping $N : \Sigma^{\mathbb{N}} \rightarrow [0, 1]$. For any prefix $w \in \Sigma^*$, we denote by $|w|$ the length of w (i.e., the number of the symbols that it contains) and b_w, c_w, p_w the probabilities that a word begins

with a prefix α of the same length as w , which satisfies $\alpha < w$, $\alpha \leq w$, or $\alpha = w$, meaning

$$b_w := \sum_{\substack{\alpha, |\alpha|=|w|, \\ \alpha < w}} p_\alpha, \quad c_w := \sum_{\substack{\alpha, |\alpha|=|w|, \\ \alpha \leq w}} p_\alpha, \quad p_w = c_w - b_w. \quad (2.14)$$

Then, the equality (2.26) entails

$$[b_w, c_w] = \bigcup_{i \in \Sigma} [b_{w \cdot i}, c_{w \cdot i}], \quad [b_{w \cdot i}, c_{w \cdot i}] \subset [b_w, c_w] \quad \text{for any } i \in \Sigma.$$

Given an infinite word $v \in \Sigma^\mathbb{N}$, 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 $b_{v_k} - c_{v_k} = p_{v_k}$ tends to 0 when k tends to infinity. 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 be smaller than v . The mapping $N : \Sigma^\mathbb{N} \rightarrow [0, 1]$ is strictly increasing outside the exceptional set formed with words of $\Sigma^\mathbb{N}$ which end with an infinite sequence of the smallest symbol 0 or with an infinite sequence of the largest symbol $(r-1)$. More precisely, one has $N(u) = N(v)$ with $u > v$ if and only if there exists $w \in \Sigma^*$ and $i \in [1..(r-1)]$ for which $u = w \cdot i \cdot 0^\infty$ with $v = w \cdot (i-1) \cdot (r-1)^\infty$.

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

Dynamical system built by the parameterization. Consider the shift on $\Sigma^\mathbb{N}$ (denoted by $\tilde{T}$), By definition of $\tilde{T}$, for any $i \in \Sigma$ and $u \in \Sigma^\mathbb{N}$ one has $\tilde{T}(i \cdot u) = u$. Consider the map $\tilde{T} : [0, 1] \rightarrow [0, 1]$ which is almost everywhere defined by the relation

$$\tilde{T}(x) := N[\tilde{T}(M(x))].$$

Consider, for $v \in \Sigma^\mathbb{N}$ and $i \in \Sigma$, the two infinite words v and $i \cdot v$. The two real numbers defined by $x := N(v)$, and $y = N(i \cdot v)$ satisfy

$$x = N(v) = N[\tilde{T}(i \cdot v)] = N[\tilde{T}(M(y))] = \tilde{T}(y).$$

Then y is an antecedent of x by $\tilde{T}$, completely defined by the pair (i, x) and is denoted by $h_i(x)$; For each symbol i , the image $h_i([0, 1])$ coincides with the image by N of the cylinder $i \cdot \Sigma^\mathbb{N}$ (up to a denumerable set). 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 $\tilde{T}^k(y) = x$; this is an antecedent of x by $\tilde{T}^k$, equal to $h_w(x)$, where $h_w = h_{w_1} \circ h_{w_2} \circ \dots \circ h_{w_k}$.

Parameterisation conditions. Consider a sequence (p_w) defined on Σ^* . The whole process is well defined as soon as the following three conditions hold :

$$(II) \quad \left\{ \begin{array}{ll} (i) & p_w > 0; \\ (ii) & \pi_k := \max\{p_w : w \in \Sigma^k\} \text{ tends to } 0, \text{ as } k \rightarrow \infty; \\ (iii) & \sum_{i \in \Sigma} p_{w \cdot i} = p_w. \end{array} \right. \quad (2.15)$$

Relation between parameterized sources and dynamical sources. Finally, the pair $([0, 1], \tilde{T})$ gives rise to a complete dynamical system $\mathcal{D}$ of the interval $[0, 1]$ on the alphabet Σ . Of course, the branches h_i are not generally speaking of class $\mathcal{C}^2$ and the dynamical system $\mathcal{D}$ is not generally a dynamical system of class $\mathcal{C}^2$.

For a complete dynamical source of class $\mathcal{C}^2$, with increasing branches, there are two shifts: the shift on the unit interval (denoted by T) and the shift on $\Sigma^{\mathbb{N}}$ (denoted by $\tilde{T}$). By definition of the word $M(x)$ in dynamical sources, one has $\tilde{T}(M(x)) = M(T(x))$. This means that in the case of a dynamical source (with increasing branches) of class $\mathcal{C}^2$, the shift T of the dynamical system coincides with the shift $\tilde{T}$ defined by the parameterization.

2.4 Relations between the three points of view.

We have just described a relation between dynamical sources and parameterized sources. We are led to mix these three points of view, as we now explain.

The two Dirichlet series. We will see later in Chapters 5 and 6 that two Dirichlet series play a central rôle in our analyses of digital trees. They are denoted in a generic way by $\Delta(s, u)$ but there is a different expression for each type of digital tree (trie = T , and dst = D). Both involve the transition matrix $\mathbf{P}_s$ of the source via its quasi-inverses

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) &= s\Lambda(s, u) = s^t \mathbf{E}(I - u\mathbf{P}_s)^{-1} \mathbf{1} \\ \Delta_D(s, u) &= {}^t \mathbf{E} \mathbf{Q}(s, u) \cdot \mathbf{Q}(2, u)^{-1} \mathbf{1}. \end{cases} \quad (2.16)$$

Both series involve two vectors indexed by Σ^* : the vector $\mathbf{1}$ has all its components equal to 1 and the vector $\mathbf{E}$ has all its components equal to 1 except the first one relative to the empty prefix $\epsilon \in \Sigma^*$, that equals 1. Moreover, the dst series $\Delta_D(s, u)$ involves the infinite product

$$\mathbf{Q}(s, u) := (I - u\mathbf{P}_s)^{-1} \cdot (I - u\mathbf{P}_{s+1})^{-1} \cdot \dots \cdot (I - u\mathbf{P}_{s+k})^{-1} \cdot \dots \quad (2.17)$$

There are similar expressions that involve the pruned transition matrix $\mathbf{A}_s$ of the source, where the two vectors $\mathbf{1}$ and $\mathbf{E}$ are now indexed by the finite language Σ_0 .

Two main cases for a source. There are two cases for the source. The pruned transition matrix becomes a finite matrix (and the source is a memoryless source or a Markov chain), or it remains infinite (and the source has an unbounded memory).

(a) *Bounded memory.* The pruned matrix $\mathbf{A}_s$ can be directly studied as a function of the matrix $\mathbf{R}_s$ of the Markov chain together with the initial conditions. In particular, if we assume the Markov chain to be good, the matrix $\mathbf{R}_s$ satisfies the Perron Frobenius Theorem. We directly study the quasi-inverse $(I - u\mathbf{R}_s)^{-1}$ of the transition matrix.

(b) *Unbounded memory.* Here, the functional analysis study is more difficult since the operator $\mathbf{P}_s$ acts on functional spaces of infinite dimension, and we have to choose a convenient functional space, where $\mathbf{P}_s$ fulfills properties which will be similar to those of a finite good matrix. We already know that, in the case of a good dynamical source, the secant operator $\mathbb{H}_s$ has a similar behaviour as a finite matrix. It seems clear that, in the case of a dynamical source, the operator $\mathbf{P}_s$ and $\mathbb{H}_s$ play the same role, as the two expressions in Lemmas 2.14 and 2.18 seem to show it. And, in tries study, we can use these two expressions. However, for dst studies, we only have the expression of the Dirichlet series as a function of the infinite product, and we may expect that it can be “transformed” into an expression of the same type, which would involve

the quasi-inverse of the secant operator $(I - u\mathbb{H}_s)^{-1}$. In a more formal setting, this would be possible if the two operators $\mathbf{P}_s$ and $\mathbb{H}_s$ were conjugate. However, the previous statement is not true.... and the following of the Chapter provides a proof of another conjugation....

The main steps for comparing $\mathbf{P}_s$ and $\mathbb{H}_s$. We first remark that in the case when the source has an unbounded memory, this memory is defined by the “reverse past”. We first introduce in Section 2.4.1 the mirror operation, closely related to the notion of g -functions and explain why it is natural. The mirror matrix $\hat{\mathbf{P}}_s$ resembles to a transfer operator which operates on Σ^* (viewed as the “finite past” of the source). Then, Section 2.4.2 extends the operator $\hat{\mathbf{P}}_s$ into an operator $\hat{\mathbf{P}}_s$ which now operates on the set $\Sigma^{\mathbb{N}}$ of infinite words, viewed as the infinite past of the source. The source itself, (provided that it be regular enough and *stationary*) may be extended in a source $\underline{\mathcal{S}}$ which has now an “infinite past”. Finally, in Section 2.4.3, using extra regularity assumptions, we can view the source $\underline{\mathcal{S}}$ as a dynamical source and then it is now possible to relate the secant transfer operator $\mathbb{H}_s$ and the operator $\hat{\mathbf{P}}_s$.

2.4.1 The mirror operation and the operator $\hat{\mathbf{P}}_s$.

Mirror operation on prefixes and g -functions. The mirror operation is used to reverse the finite prefixes. It will play an important role in the sequel. 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 $X_0X_1 \dots X_{n-1}$. The prefix $\phi(w)$ defines the reverse past history.

Definition 2.22. [Mirror operation on prefixes and g -functions].

(i) The mirror operation $\phi : \Sigma^* \rightarrow \Sigma^*$ associates to a finite prefix w its mirror $\phi(w)$: if the finite prefix w is of the form $w = w_1w_2 \dots w_{k-1}w_k$, then its mirror $\phi(w)$ is defined as $\phi(w) = w_kw_{k-1} \dots w_2w_1$.

(ii) The probability $p_{\phi(w)}$ is denoted by $\hat{p}_w$.

(iii) The g -function on $\Sigma \times \Sigma^*$ is defined by the equalities

$$g(i \cdot w) := q_{i|\phi(w)} = \frac{\hat{p}_{i \cdot w}}{\hat{p}_w}. \quad (2.18)$$

Operator $\hat{\mathbf{P}}_s$. We denote by $\mathcal{B}(\Sigma^*)$ the Banach space of the bounded functions $X : \Sigma^* \rightarrow \mathbb{C}$ endowed with the sup-norm. The matrix $\mathbf{P}_s$ is viewed as an operator which acts on $\mathcal{B}(\Sigma^*)$ in a natural way: it transforms a function $X \in \mathcal{B}(\Sigma^*)$ into a function $Y \in \mathcal{B}(\Sigma^*)$ as follows:

$$Y(w) := \mathbf{P}_s[X](w) := \sum_{i \in \Sigma} q_{i|w}^s X(w \cdot i). \quad (2.19)$$

Definition 2.23. The mirror operation induced by ϕ on $\mathcal{B}(\Sigma^*)$ is defined by the equality $\phi(X)(w) := X(\phi(w))$. We denote by $\hat{\mathbf{P}}_s$ the conjugate of the operator $\mathbf{P}_s$ via the mirror operation ϕ .

Remark. The matrix $\hat{\mathbf{P}}_s$ is the matrix transition of the mirror graph of the source, that is obtained from the initial graph with reversing labels of the vertices.

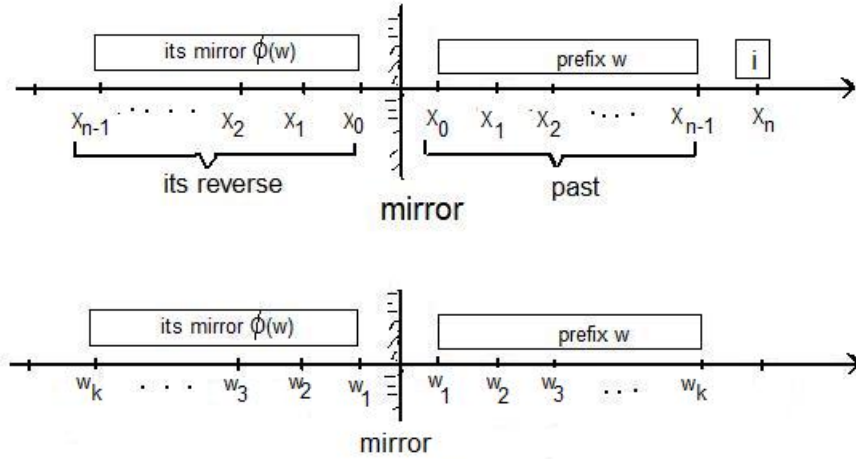


Figure 2.7: Visualization of the mirror operation

Expression of the Dirichlet series Λ and Δ with $\hat{\mathbf{P}}_s$. 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, we obtain the following relation

Proposition 2.24. *The Dirichlet series are expressed in terms of the quasi-inverse of the operator $\hat{\mathbf{P}}_s$. Consider the infinite product $\hat{\mathbf{Q}}(s, u)$*

$$\hat{\mathbf{Q}}(s, u) := (I - u\hat{\mathbf{P}}_s)^{-1} \cdot (I - u\hat{\mathbf{P}}_{s+1})^{-1} \cdot (I - u\hat{\mathbf{P}}_{s+2})^{-1} \cdot \dots \quad (2.20)$$

Then the series $\lambda(s, u)$, $\Delta(s, u)$ admit the following expressions

$$\Lambda(s, u) = {}^t\mathbf{E} \cdot (I - u\hat{\mathbf{P}}_s)^{-1} \cdot \mathbf{1}, \quad \Delta(s, u) = {}^t\mathbf{E} \hat{\mathbf{Q}}(s, u) \cdot \hat{\mathbf{Q}}(2, u)^{-1} \mathbf{1}. \quad (2.21)$$

The operator $\hat{\mathbf{P}}_s$ viewed as a transfer operator. When the operator $\mathbf{P}_s$ transforms X into Y , the conjugate $\hat{\mathbf{P}}_s$ of the operator $\mathbf{P}_s$ via the mirror operation ϕ , transforms, (by definition) $\phi(X)$ into $\phi(Y)$. In fact, when $Y := \mathbf{P}_s[X]$ is defined as in Eq. (2.19), its transform $\phi(Y)$ satisfies

$$\begin{aligned} \phi(Y)(w) &= Y(\phi(w)) = \sum_{i \in \Sigma} q_{i|\phi(w)}^s X(\phi(w) \cdot i) \\ &= \sum_{i \in \Sigma} g(i \cdot w)^s \phi(X)(i \cdot w). \end{aligned} \quad (2.22)$$

Then, if $\check{T}$ denotes the shift on $\Sigma \times \Sigma^*$ defined by the equality $\check{T}(i \cdot w) = w$, the mapping $\hat{\mathbf{P}}_s$ which associates $\phi(Y)$ to $\phi(X)$ is defined as the mirror of (2.19),

$$\begin{aligned} \phi(Y) = \hat{\mathbf{P}}_s[\phi(X)] &\iff \phi(Y)(w) = \sum_{i \in \Sigma} g^s(i \cdot w) \phi(X)(i \cdot w) \\ &\iff \phi(Y)(w) = \sum_{\substack{v \\ \check{T}(v)=w}} g(v)^s \phi(X)(v), \end{aligned} \quad (2.23)$$

Under this form, the mapping $\hat{\mathbf{P}}_s$ resembles the transfer operator of the dynamical system $(\Sigma^*, \check{T})$ 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 $\check{T}$ is only defined on $\Sigma \times \Sigma^*$, (not on the whole Σ^*) and Σ^* is not compact.

2.4.2 Extension to “infinite past”. The source $\underline{\mathcal{S}}$.

In the following, we will extend the mapping $\widehat{\mathbf{P}}_s$ into a mapping which acts on functions defined on the compact space $\Sigma^{\mathbb{N}}$.

The metric space $\Sigma^{\mathbb{N}}$. This space is a metric space, whose definition is now recalled.

Definition 2.25. 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,

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

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 defines a metric space that is denoted by $\Sigma_\theta^{\mathbb{N}}$.

The coincidence may also be defined between two words u and v of Σ^* , via the addition of an ending symbol which does not belong to the initial alphabet.

Extension of the g -functions in the Hölder case. We remark that in a Markov chain of order k , the sources $\mathcal{S}_{(u)}$ are $\mathcal{S}_{(v)}$ are equivalent as soon as the two *prefixes* u and v finish with the *same suffix* of order k . In the mirror graph, the sources $\widehat{\mathcal{S}}_{(u)}$ are $\widehat{\mathcal{S}}_{(v)}$ are equivalent as soon as the two prefixes u and v satisfy the inequality $\gamma(u, v) \geq k + 1$. This means that, in a Markov chain of order k , the g -functions are constant on the cylinders of depth k . It is then natural to consider “good” sources, where the g -functions are continuous or even Hölder with exponent α . Namely, assume, that for some $\alpha > 0$, the following regularity assumption holds

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

Since the space Σ^* is a dense subset of $\Sigma^{\mathbb{N}}$, the function g can be extended to a function $\underline{g}$ defined on $\Sigma^{\mathbb{N}}$ which is also Hölder.

Extension of the operator $\widehat{\mathbf{P}}_s$ into $\widehat{\mathbf{P}}_s$. We are now ready to define the operator $\widehat{\mathbf{P}}_s$. Consider the space $\mathcal{H}_\alpha(\Sigma_\theta^{\mathbb{N}})$ formed with the Hölder functions $X : \Sigma_\theta^{\mathbb{N}} \rightarrow \mathbb{C}$ with exponent α . Then, via the extension $\underline{g}$ of g , the operator $\widehat{\mathbf{P}}_s$ defined in (2.23) is extended on $\mathcal{H}_\alpha(\Sigma_\theta^{\mathbb{N}})$ into an operator $\widehat{\mathbf{P}}_s$ defined as

$$\widehat{\mathbf{P}}_s[X] = Y \quad \Longleftrightarrow \quad Y(v) = \sum_{i \in \Sigma} \underline{g}(i \cdot v)^s X(i \cdot v) = \sum_{\substack{u \\ T(u)=v}} \underline{g}(u)^s X(u). \quad (2.24)$$

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

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

and the operator $\widehat{\mathbf{P}}_s$ is the (true) transfer operator of the system $(\Sigma^{\mathbb{N}}, \tilde{T})$ relative to $\underline{g}^s$.

Extension of the source $\mathcal{S}$ into the source $\underline{\mathcal{S}}$ with an infinite past. Moreover, the operator $\widehat{\mathbf{P}}_1$ is *quasi-compact* [the definition of this notion can be found in [1]] and admits on $\mathcal{H}_\alpha(\Sigma_\theta^{\mathbb{N}})$ a unique invariant measure, denoted by ν , which satisfies

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

The measure ν is thus invariant by the shift $\tilde{T}$. This means that the source is stationary with respect to this measure.

Assume now that our initial source $\mathcal{S}$ is itself *stationary*. 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 $\underline{g}(v)$ for $v \in \Sigma^{\mathbb{N}}$, where $\underline{g}(i \cdot v)$ is the probability of emitting i when the reverse past history has just emitted the infinite word v , namely

$$\underline{g}(i \cdot v) = \Pr_{\nu}[X_0 = i | X_{-1} = v_1, X_{-2} = v_2, \dots, X_{-\ell} = v_{\ell} \dots]. \quad (2.25)$$

As the source is stationary, one also has, for any $k \in \mathbb{Z}$,

$$\underline{g}(i \cdot v) = \Pr_{\nu}[X_k = i | X_{k-1} = v_1, X_{k-2} = v_2, \dots, X_{k-\ell} = v_{\ell} \dots].$$

We will insist on a particularity of this stationary source related to the invariant measure ν . For a finite word $w = w_0 w_1 \dots w_k$, denote by $\hat{\pi}_w = \Pr_{\nu}[X_0 = w_k, \dots, X_k = w_0]$ the probability that a word of positive history begins with $\phi(w)$, where $\phi(w) = w_k \dots w_1 w_0$ is the mirror of w (see Section 2.4.1). Now, for any $i \in \Sigma$, the probability $\hat{\pi}_{w \cdot i}$ is the probability that a word of positive history begins with $\phi(w \cdot i) = i \cdot \phi(w)$, namely

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

This implies, for a stationary source, the equality

$$\begin{aligned} \sum_{i \in \Sigma} \hat{\pi}_{w \cdot i} &= \Pr_{\nu}[X_1 = w_k, \dots, X_{k+1} = w_0] \\ &:= \Pr_{\nu}[X_0 = w_k, X_1 = w_{k-1} \dots, X_k = w_0] = \hat{\pi}_w. \end{aligned} \quad (2.26)$$

2.4.3 The dynamical system $\mathcal{D}$ and the transfer operator $\mathbb{H}_s$

We now assume the source $\mathcal{S}$ to be non-ambiguous, according to Definition 2.21. Then, with (2.26), the sequence $(\hat{\pi}_w)$ satisfies the set of parameterization conditions given in (2.15), and we consider the parameterization of the reverse past, associated to the sequence $(\hat{\pi}_w)$. Denote by $\tilde{T}$ the shift on $\Sigma^{\mathbb{N}}$ (here the shift towards the past of the reverse past history) and by T the shift induced by $\tilde{T}$ on $[0, 1]$ via conjugation of mappings N, M defined in section 2.3.3, namely

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

As, by definition, one has $\tilde{T}(i \cdot v) = v$, the equality $N(\tilde{T}(i \cdot v)) = N(v) = T(N(i \cdot v))$ holds; We know that the pair $([0, 1], T)$ gives rise to a complete dynamical system $\mathcal{D}$ of the interval $[0, 1]$ on the alphabet Σ .

Relation between the operator $\hat{\mathbf{P}}_s$ and the secant operator $\mathbb{H}_s$. We assume that this system is of class $\mathcal{C}^2$.

Lemma 2.26. *Consider the subset $\mathcal{F}$ of functions $X \in \mathcal{B}(\Sigma^*)$ 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 following holds:*

- (i) *The subset $\mathcal{F}$ is invariant under the action of $\hat{\mathbf{P}}_s$.*
- (ii) *The two operators, the operator $\mathbb{H}_s$ when acting on $\mathcal{C}^1([0, 1]^2)$, and the operator $\hat{\mathbf{P}}_s$ when acting on $\mathcal{F}$ are conjugate.*

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

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

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. $\square$

Expression for the Dirichlet series of interest. Finally, we have proven the following result, which constitutes the main result of this Section:

Theorem 2.27. *Consider a non-ambiguous stationary source $\mathcal{S}$, whose reverse past leads to a complete dynamical system $(\mathcal{I}, T)$ of the unit interval, of class $\mathcal{C}^2$. Consider the secant operator $\mathbb{H}_s$ of the dynamical system,*

$$\mathbb{H}_s[F](x, y) := \sum_{h \in \mathcal{H}} \left| \frac{h(x) - h(y)}{x - y} \right|^s F(h(x), h(y)), \quad (2.27)$$

whose expression involves the set $\mathcal{H}$ of inverse branches of T .

In this case, the two Dirichlet series $\Delta(s, u)$ defined in terms of the transition matrix $\mathbf{P}_s$ as in (2.16) admit alternative expressions in terms of quasi-inverses $(I - u\mathbb{H}_s)^{-1}$. Define the infinite product

$$\mathbb{Q}(s, u) := (I - u\mathbb{H}_s)^{-1} \circ \dots \circ (I - u\mathbb{H}_{s+2})^{-1} \circ \dots \quad (2.28)$$

Then, the Dirichlet series $\Delta(s, u)$ are expressed as

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) &= s \Lambda(s, u) = s (I - u\mathbb{H}_s)^{-1}[1](0, 1), \\ \Delta_D(s, u) &= \mathbb{Q}(s, u) \circ \mathbb{Q}(2, u)^{-1}[1](0, 1). \end{cases} \quad (2.29)$$

Remark. In Chapters 3 and 7, we restrict ourselves to dynamical systems of the Good Class, which lead to unambiguous sources. In contrast, we must keep the stationary hypothesis.

About the reverse past of a source $\mathcal{S}$. The conditions for the previous theorem are not completely natural since they hold on the reverse source. It would be more natural to have conditions which hold on the initial source $\mathcal{S}$. However, the notion of *natural extension* of a dynamical system allows us to return to the initial source $\mathcal{S}$.

It is not easy to get a simple definition of a natural extension of a dynamical system, and we try to define it in our setting, via the point of view of parameterization.

We begin with an unambiguous stationary source, with positive history. It is defined via its fundamental probabilities p_w which satisfy the parameterization conditions (2.15), and then the positive history is a parameterized source. With (almost) any $u \in \mathcal{I}$, we associate a word $M(u)$, and we now wish to define the reverse past of the word $M(u)$. We thus consider the “mirror parameterization” built with the probabilities $\widehat{p}_w$. As the source is stationary, these probabilities satisfy the parameterization conditions (2.15), and then the reverse past is also a parameterized source: With (almost) any $u \in \mathcal{I}$ we associate a word $\widehat{M}(u)$. The concatenation of the two

histories, the positive history from the left to the right, and the negative history, from the right to the left gives rise to a bi-infinite history.

Assume now that we begin with a dynamical source of the interval $(\mathcal{I}, T)$ (with its invariant density). Then the parameterization $M(u)$ is defined as

$$M(u) = \left(u, \tau(Tu), \tau(T^2(u)), \dots, \tau(T^k(u)), \dots \right)$$

which involves the coding τ and the shift T . However,

The reverse shift $\widehat{T}$ (that is directly defined from the parameterization $\widehat{M}$) does not define (generally speaking) a shift of class $\mathcal{C}^2$ on the unit interval

This is not always clear to obtain a good “realisation” of the natural extension via dynamical systems of the interval. Then, if the reverse source is a (stationary) dynamical system, it is not clear if the initial system is also a dynamical source in our sense. However all the properties that we ask on the probabilities of the reverse source in Chapter 3, (in particular, for periodic words) can be expressed in terms of fundamental probabilities of our initial source.

The reverse past of a stationary Markov chain. We give the example of a binary source built on a Markov chain, with the matrix $\mathbf{R}_s$ and stationary density (λ, μ) . The matrix $\mathbf{R}$ is stochastic, and the sum of each row (a, b) and (c, d) is equal to 1. The interpretation of coefficients a, b, c, d of the matrix is

$$a = p_{0|0}, \quad b = p_{1|0}, \quad c = p_{1|1}, \quad d = p_{0|1}.$$

As the density (λ, μ) is stationary, the two equalities hold $\lambda a + \mu c = \lambda$, $\lambda b + \mu d = \mu$.

Then, we consider the matrix $\widehat{\mathbf{R}}$ defined with rows

$$\left(a, \frac{\mu}{\lambda} c \right), \left(\frac{\lambda}{\mu} b, d \right).$$

It is stochastic, it has (λ, μ) as stationary density, and we now prove that this is the Markov chain of the reverse history. Define, indeed, as previously, the initial source by the sequence (X_i) and the reverse history by the sequence (Y_i) , and consider the probability of the event $\pi := \Pr(X_0 = 0, X_1 = 1)$ that equals the probability $\Pr(Y_0 = 1, Y_1 = 0)$, since the source is stationary. One has

$$\pi = \lambda \cdot p_{1|0} = \lambda \cdot b = \mu \cdot \widehat{p}_{0|1} \quad \implies \quad \widehat{p}_{0|1} = \frac{\lambda}{\mu} b.$$

Then, the matrix $\widehat{\mathbf{R}}$ is the Markov chain that defines the reverse history. Remark that the two matrices have the same characteristic polynomial, and they thus have the same spectrum.

Conclusion of this chapter.

There are two different frameworks in the thesis:

- (a) For combinatorial studies done in Chapters 5 and 6, we deal with sources given by a transition matrix $\mathbf{P}_s$. In Chapter 6, we limit ourselves to smooth sources (see Definition 6.1), which are a particular case of unambiguous sources. The sources here are not needed to be stationary.
- (b) For analytic studies, where we need a precise knowledge of the quasi-inverse $(I - u\mathbf{P}_s)^{-1}$, there are two main cases, as we already mention: the case where the pruned matrix becomes finite, or the case where it remains infinite.

- (b1) In the first case, we deal with a matrix $\mathbf{P}_s$ which acts on finite-dimensional vectors, and the functional analysis is simple. We can continue to deal with the matrix $\mathbf{P}_s$, and the source is not needed to be stationary.
- (b2) In the second case, there are again two cases, according to the digital tree:
 - For tries, we deal with a unique quasi-inverse, and we do not need the reverse past. But, when we need to conduct a precise functional analysis, and deal with tameness of sources, we change the framework and we restrict ourselves to particular initial sources, namely dynamical sources.
 - For dst, when we need to conduct a precise functional analysis, and deal with tameness of sources, we change the framework and we restrict ourselves to particular sources, which must be *stationary* and whose the reverse past defines a dynamical source.

The class of dynamical sources, and in particular, the Good Class we deal with in the future, contains all the simple sources. Then, the case (b2) is sufficient for dealing with all the simple sources (in the case of tries) and stationary simple sources (in the case of dst). Any memoryless source is stationary. Then, we only miss in the case (b2), and when we study dst's, the *non stationary Markov chains*.

Chapter 3

Tameness of Sources

Contents

3.1	Generalities.	56
3.1.1	Role of the quasi-inverse $\mathbb{T}(s, u)$	56
3.1.2	The trie operator and the dst operator.	57
3.1.3	Our needs for using the Rice methodology.	57
3.1.4	Definitions of tameness.	58
3.1.5	Plan of the chapter.	58
3.2	Periodicity and Quasi-periodicity.	59
3.2.1	Case of a memoryless source.	59
3.2.2	Case of a Markov chain.	61
3.2.3	Case of a dynamical source.	63
3.2.4	Quasi periodicity is exceptional.	64
3.2.5	An intermediary notion : p -periodicity	64
3.2.6	Contrapositive versions of the statements about periodicity.	65
3.3	Main principles for tameness.	65
3.3.1	Geometric conditions on branches. The UNI Class.	65
3.3.2	Arithmetical conditions on branches. The DIOP Classes.	66
3.3.3	Shape of tameness regions.	67
3.4	Instances of tameness.	68
3.4.1	The periodic case. Tameness in vertical strips with an infinite number of poles.	68
3.4.2	The UNI Class. Tameness in vertical strips.	68
3.4.3	The DIOP Classes. Tameness in hyperbolic regions.	69
3.4.4	Conclusion of tameness study for classical sources.	70
3.4.5	A small piece of history.	70

This Chapter studies tameness of sources. In an informal way, a function of the complex variable s is tame if it is analytic (or meromorphic) and *of polynomial growth* for $s \rightarrow \infty$. As we will see soon in Chapter 4, tameness is essential in many analytic studies, in particular for the Rice method, that will be one of the main tools of our analytic toolbox. As we mention in Chapter 2, the main Dirichlet series of interest are the mixed Dirichlet $\Delta(s, u)$ which “mix” the probabilistic properties of the source with the definition of digital trees. It appears that the

tameness of these series $\Delta(s, u)$ is “brought” by the tameness of the source, which itself relies on the tameness of the quasi-inverse $s \mapsto (I - u\mathbb{H}_s)^{-1}$ of the secant operator of the source, when u is close to 1. We are then led to conduct a precise study of the tameness of sources.

Tameness is not deeply studied in analytic combinatorics. Even for simple sources, and even for memoryless sources, this notion is often completely dismissed. It is first related to the position of the poles of the Λ series defined in Chapter 2. For the memoryless sources (p, q) with $p + q = 1$, it is related to the position of complex numbers s for which $p^s + q^s = 1$, itself related to arithmetical properties of the ratio $\log p / \log q$. And the position of these complex numbers may largely affect the asymptotics of many parameters of the source, and also parameters of structures associated with the source, as our digital trees. Tameness first intervenes as a (sufficient) *condition* for applying the Rice method, and, when applied, it intervenes in the form of the *remainder terms* of the asymptotic estimates of these parameters.

To the best of our knowledge, the first works which conduct (in the memoryless case) a detailed discussion of the position of poles are due to Fayolle *et al.* [17]. Then, always in the memoryless case, Schachinger provides a rigorous and thorough discussion of this geometry of poles [65]. Finally, the paper [22] adapts deep results described in the book of Lapidus and van Frankenhuijsen [45] and precisely relates the shape of the pole-free region to arithmetic properties of probabilities. The word *tame* was proposed by Philippe Flajolet and used for the first time in [74]. Later on, most papers which deal with probabilistic sources use similar notions and the word “tame” is now largely used, for instance in the paper [9].

However, the previous studies deal with the plain quasi-inverse $(I - \mathbb{H}_s)^{-1}$. As we wish to perform distributional studies, we are interested in the bivariate quasi-inverse $(I - u\mathbb{H}_s)^{-1}$, which is yet less studied. The paper [4] is the first (and the only?) work to deal with the bivariate quasi-inverse. This “uniform” tameness is the main subject of the present chapter and it will be central in Chapter 7, when we perform the analytic studies which lead us to the asymptotic gaussian law.

3.1 Generalities.

3.1.1 Role of the quasi-inverse $\mathbb{T}(s, u)$

There are two main cases for our studies depending on whether the source has a bounded memory or not.

(a) *Bounded memory.* The pruned graph of the source is finite, and the source is a memoryless source or a Markov chain. The matrix $\mathbf{P}_s$ (or more exactly its pruned version) can be directly studied as a function of the matrix $\mathbf{R}_s$ of the Markov chain together with the initial conditions. We do not need the source to be stationary, but we assume the Markov chain to be good, so that the matrix $\mathbf{R}_s$ satisfies the Perron Frobenius Theorem. We directly study the quasi-inverse $\mathbb{T}(s, u) = (I - u\mathbf{R}_s)^{-1}$ of the transition matrix.

(b) *Unbounded memory.* In the (general and new) case when the pruned matrix $\mathbf{P}_s$ remains infinite, the functional analysis study is more difficult, and we assume regularity properties for the source, and, more precisely for its reverse past. We then deal with the mirror operator $\hat{\mathbf{P}}_s$. Under these regularity assumptions, and provided that the source be stationary, the reverse past of the source is a good dynamical system of class $\mathcal{C}^2$, and we have proven in Theorem 2.27 that the mirror operator $\hat{\mathbf{P}}_s$ is conjugate to the secant operator $\mathbb{H}_s$ related to the dynamical system which describes the “reverse past” of the source. In this case, we study the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$ when acting on the functional space $\mathcal{C}^1(\mathcal{I}^2)$.

3.1.2 The trie operator and the dst operator.

We will deal with a generic operator which has a different definition in the two cases:

$$\mathbb{T}(s, u) = \begin{cases} (I - u\mathbf{R}_s)^{-1} & \text{(in the bounded memory case)} \\ (I - u\mathbf{H}_s)^{-1} & \text{(in the unbounded memory case)} \end{cases} \quad (3.1)$$

The operators which play the fundamental roles in the analyses of Chapter 7 are

$$\begin{cases} s\mathbb{T}(s, u) & [\text{trie}] \\ \mathbb{D}(s, u) = \mathbb{T}(s, u) \circ \mathbb{L}(s, u) & [\text{dst}] \end{cases} \quad (3.2)$$

where $\mathbb{L}(s, u) = \mathbb{Q}(s + 1, u) \circ \mathbb{Q}(2, u)^{-1}$ and $\mathbb{Q}(s, u)$ is the infinite product

$$\mathbb{Q}(s, u) := \mathbb{T}(s, u) \circ \mathbb{T}(s + 1, u) \circ \dots \circ \mathbb{T}(s + k, u) \dots, \quad (3.3)$$

3.1.3 Our needs for using the Rice methodology.

We will use the operators of (3.2) inside the Rice methodology described in Chapter 4. For this use, we need a main property for these operators when $\Re s$ is close to 1 and u close to 1, namely their *tameness*. Before defining tameness, we need to define a convenient norm for our operators when acting on $\mathcal{C}^1(\mathcal{I}^2)$:

Definition 3.1. [Norm $(1, t)$] *The norm $\|\cdot\|_{(1,t)}$ is defined on functions of the set $\mathcal{C}^1(\mathcal{I} \times \mathcal{I})$ by the equality*

$$\|F\|_{(1,t)} = \|F\|_0 + \frac{1}{|t|} \|DF\|_0, \quad (3.4)$$

where $\|\cdot\|_0$ is the sup-norm.

With this norm, we define the notion of polynomial growth for an operator $\mathbb{M}_s$ which uses the norm $\|\cdot\|_{(1,t)}$ relative to $t := \Im s$.

Definition 3.2. [Polynomial growth] *Consider an operator $\mathbb{M}_s$ which depends on the (complex) parameter s . The operator is said to be of polynomial growth in a region $\mathcal{R}$ if there exist $t_0 > 0$, $r \in \mathbb{R}^+$ and a constant $K > 0$ for which for any $s = \sigma + it \in \mathcal{R}$ with $|t| > t_0$, the norm $\|\cdot\|_{(1,t)}$ of the operator $\mathbb{M}_s$ satisfies $\|\mathbb{M}_s\|_{(1,t)} \leq K|t|^r$.*

All the operators of interest are of polynomial growth on the half-plane $\{\Re s > 1\}$ [see Chapter 7], and as we wish to use the Rice methodology, which “shifts to the left”, we focus on the *left half plane* $\{\Re s \leq 1\}$ and isolate a region $\mathcal{R}$ where the operators $s \mapsto \mathbb{T}(s, u)$, $u \mapsto \mathbb{D}(s, u)$ are analytic and of polynomial growth, for $s \in \mathcal{R}$, $|\Im s| \rightarrow \infty$, uniformly with respect to u , when u is close to 1, or at least with a controlled growth with respect to u when $u \rightarrow 1$. We will make this more precise later.

We now explain in an informal way why it is sufficient to deal with the tameness of operators $\mathbb{T}(s, u)$. This will be proven in Chapter 7, but we give the main argument now : This is due to the good properties of the factor $\mathbb{L}(s, u)$ which quantifies the “ratio” between the trie and the dst. This operator satisfies $\mathbb{L}(1, u) = I$ and Proposition 7.6 will prove that $\mathbb{L}(s, u)$ is analytic (with respect to s) and of bounded growth [i.e., of polynomial growth with $r = 0$] when $|\Im s| \rightarrow \infty$ and $\Re s \geq 1 - \delta_0 > 0$. Moreover, this bounded growth is uniform with respect to u when u belongs to the closed unity disk. It is then sufficient to deal with the operator $\mathbb{T}(s, u)$, and study its tameness when $\Re s$ is close to 1 and $|u| \leq 1$. The remainder of this chapter is devoted to this task.

3.1.4 Definitions of tameness.

We first consider the particular case when $u = 1$. This leads to the first definition which deals with the operator $\mathbb{T}(s)$.

Definition 3.3. *The operator $\mathbb{T}(s)$ is tame if there exists a region $\mathcal{R}$ [called a tameness region] which “strictly” contains the half-plane $\Re s \geq 1$, and on which the operator $\mathbb{T}(s)$ fulfills two main properties :*

- (i) *The operator $\mathbb{T}(s)$ is meromorphic on $\mathcal{R}$, with only two possibilities for its poles*
 - (ia) *it has a unique pole at $s = 1$,*
 - (ib) *or it has, for $\eta > 0$, a sequence of poles $\{s_k = 1 + k\eta; k \in \mathbb{Z}\}$*
- (ii) *It is of polynomial growth on $\mathcal{R}$.*

We now define the notion of tameness for $\mathbb{T}(s, u)$. In an informal way, it may be obtained by perturbation¹ of tameness of $\mathbb{T}(s)$.

Definition 3.4. *The operator $\mathbb{T}(s, u)$ is tame if there exist*

- (a) *a region $\mathcal{R}$ [called a tameness region] which “strictly” contains the half-plane $\Re s \geq 1$,*
- (b) *a complex neighborhood $\mathcal{U}$ of 1 and a subset $\mathcal{U}_0 \subset \mathcal{U}$ [called the tameness subset]*
- (c) *an analytic function $\sigma : \mathcal{U} \rightarrow \mathbb{C}$ with $\sigma(1) = 0$ [called the entropic function]*

such that the operator $\mathbb{T}(s, u)$ fulfills two main properties :

- (i) *For each $u \in \mathcal{U}$, the operator $s \mapsto \mathbb{T}(s, u)$ is meromorphic on $\mathcal{R}$, with only two possibilities for its poles*
 - (ia) *it has a unique pole at $s = 1 + \sigma(u)$,*
 - (ib) *or it has, for $\eta > 0$, a sequence of poles $\{s_k = 1 + \sigma(u) + k\eta; k \in \mathbb{Z}\}$*
- (ii) *For $u \in \mathcal{U}_0$, the map $s \mapsto \mathbb{T}(s, u)$ is of polynomial growth on $\mathcal{R}$, with a polynomial growth which is controlled as a function of u .*

The idea is to keep the tameness region $\mathcal{R}$ fixed. Then, the set $\mathcal{U}_0$ will depend on the possible space that $\mathcal{R}$ “gives” us to perturb. This is why we expect the shape of region $\mathcal{R}$ and the shape of set $\mathcal{U}_0$ to be closely related. We will see in this Chapter that there appears in a natural way for transfer operators $\mathbb{T}(s, u)$ of dynamical sources, with various possibilities for shapes of region $\mathcal{R}$ and subsets $\mathcal{U}_0$. We will summarize all these possibilities at the end of the Chapter.

3.1.5 Plan of the chapter.

We study the main analytical properties of the $\mathbb{T}(s, u)$ operator, in three cases, two models of simple sources (memoryless sources and good Markov chains), and the model of dynamical sources.

In Section 3.2, we focus on the case when s belongs to the vertical line $\Re s = 1$ and u is on the circle $|u| = 1$. We first describe the properties of the $\mathbb{T}(s, u)$ operator for sources with bounded memory, then, for general dynamical sources. This analysis (in these three cases) gives rise to interesting phenomena: *periodicity*², *quasi-periodicity*, and an intermediary notion, that we call *p-periodicity* which describes the behaviour of $\mathbb{T}(s, u)$ when u is a p -root of unity.

¹In the sense of Perturbation Theory (see [39]).

²We recall that the *periodicity* phenomenon (which may also occur in the Markov chain) must not be mixed up with the usual notion of aperiodicity for Markov chains which we never used under this terminology : in this thesis, a Markov chain which is irreducible and aperiodic (in the usual sense) will be said to be good.

Finally, in Section 3.3, we focus on the halfplane $\Re s < 1$. We first explain how to “guess” sufficient conditions which may lead to tameness; these natural conditions are obtained as reinforcements of non periodicity or non quasi-periodicity. We describe three conditions: a geometric condition (the Condition UNI) and two arithmetical conditions (the DIOP conditions). Finally, in Section 3.4, we recall the main results which prove that such conditions (UNI, DIOP) are sufficient to entail tameness of the operator $\mathbb{T}(s, u)$. The complex numbers s close to the vertical line $\{\Re s = 1\}$ or the complex numbers u close to 1 for which the operator $s \mapsto \mathbb{T}(s, u)$ can be proven of polynomial growth depend on the given condition. We conclude in exhibiting four types of tameness, which depend of the pairs (s, u) for which the tameness of $\mathbb{T}(s, u)$ can be obtained.

3.2 Periodicity and Quasi-periodicity.

3.2.1 Case of a memoryless source.

We begin with this simplest case.

Characterisation of the set of poles. The following result describes the first properties of the Λ series.

Lemma 3.5. *The Dirichlet series $\Lambda(s)$ of a memoryless source is meromorphic on the complex plane, is analytic on the half plane $\Re s > 1$ and has a simple pole at $s = 1$. Moreover, the set $\mathcal{Z}$ of poles is defined as*

$$\mathcal{Z} = \{s; \lambda(s) = 1\}$$

Proof. The function $s \mapsto \lambda(s)$ is analytic on the complex plane, and thus the function $s \mapsto \Lambda(s)$ is meromorphic with a set of poles $\mathcal{Z}$ defined in the Lemma. Let $\sigma := \Re s$, and assume $\sigma > 1$. Then, the inequality $|\lambda(s)| \leq \lambda(\sigma) < \lambda(1) = 1$ entails that the set $\mathcal{Z}$ is contained in the half plane $\Re s \leq 1$. $\square$

Periodicity. To the family of probabilities $\mathfrak{P} = (p_1, p_2, \dots, p_r)$, we associate the ratios

$$\alpha(i, j) := \frac{\log p_i}{\log p_j} \quad \text{for any pair } (i, j) \in \Sigma^2. \quad (3.5)$$

The following classical result proves that the position of the set $\mathcal{Z}$ with respect to the vertical line $\Re s = 1$ is related to the rationality of the ratios $\alpha(k, j)$.

Lemma 3.6. *For a memoryless source of probabilities $\mathfrak{P}$, the following conditions are equivalent:*

- (a) *The intersection $\mathcal{Z} \cap \{\Re s = 1\}$ contains a point $s \neq 1$.*
- (b) *All the ratios $\alpha(i, j)$ defined in (3.5) are rational numbers.*
- (c) *There exists $\tau > 0$ for which the equality $\mathcal{Z} \cap \{\Re s = 1\} = 1 + 2i\pi\tau\mathbb{Z}$ holds.*
- (d) *The function $\lambda(s)$ is periodic of period $2i\pi\tau$.*

A source which satisfies one of these conditions is said to be periodic.

When a memoryless source is periodic, then (e) holds

(e) *there exists an algebraic integer $a < 1$ for which all the probabilities p_i belong to the semi-group generated by a .*

Example 3.7. For any unbiased memoryless source, all the ratios $\alpha(i, j)$ are equal to 1. Such a source is periodic, and $\lambda(s)$ is periodic of period $(2\pi)/\log r$. An instance of a (non trivial) periodic memoryless source on the binary alphabet is given by $p_0 = 1/\phi, p_1 = 1/\phi^2$ where $\phi = (1 + \sqrt{5})/2$ is the golden ratio.

Proof. We will prove $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a)$

$(a) \Rightarrow (b)$. For $s = 1 + 2i\pi t$, the inequalities

$$\lambda(s) = \sum_{\ell \in \Sigma} p_\ell e^{2i\pi t \log p_\ell} \quad \text{so that} \quad |\lambda(s)| = \left| \sum_{\ell \in \Sigma} p_\ell e^{2i\pi t \log p_\ell} \right| \leq \sum_{\ell \in \Sigma} p_\ell = 1$$

always hold. Now, if the equality $\lambda(1 + 2i\pi t) = 1$ holds, this entails (by the converse of the triangular inequality) the following equalities,

$$\forall \ell, \quad e^{2i\pi t \log p_\ell} = 1, \quad \text{and then} \quad t \log p_\ell \in \mathbb{Z}. \quad (3.6)$$

Now, if $t \neq 0$, one deduces

$$\alpha_{i,j} := \frac{\log p_i}{\log p_j} \in \mathbb{Q}, \quad \forall i, j.$$

$(b) \Rightarrow (c)$. The previous proof shows that the possible solutions of the equation $\lambda(1 + 2i\pi t) = 1$ arise when

$$t \in \frac{1}{\log p_j} \mathbb{Z} \quad \forall j \in [1..r]$$

When all the real numbers $\alpha_{i,j}$ are rational, the intersection of the lattices generated by $(1/\log p_i)$ is not reduced to $\{0\}$ and has a smallest element $\tau > 0$. Then, all the solutions of the equation $\lambda(1 + 2i\pi t) = 1$ are of the form $t = \tau \mathbb{Z}$.

$(c) \Rightarrow (d)$. One always has, for any real t ,

$$\lambda(s + 2i\pi t) = \sum_{\ell} p_\ell^{s+2i\pi t} = \sum_{\ell} p_\ell^s e^{2i\pi t \log p_\ell}$$

Assume now that the equality $\lambda(1 + 2i\pi t) = 1$ holds with $t \in \tau \mathbb{Z}$. Then, the proof of $(a) \Rightarrow (b)$ shows that Relation (3.6) holds and then, for any $t \in \tau \mathbb{Z}$, one has

$$\lambda(s + 2i\pi t) = \sum_{\ell} p_\ell^s e^{2i\pi t \log p_\ell} = \sum_{\ell} p_\ell^s = \lambda(s).$$

$(d) \Rightarrow (a)$. Trivial.

$(d) \Rightarrow (e)$. Obvious if we let $a := \exp(1/|\tau|)$ □

Quasi-periodicity. For a general complex number of modulus 1, we will be also interested in the description of the set

$$\mathcal{Z} := \{s; |\lambda(s)| = 1\},$$

and its position with respect to the vertical line $\{\Re s = 1\}$. The following result provides an extension of the previous Lemma 3.6. It is related to the rationality of the ratios $\alpha(k, j, \ell)$, defined as

$$\alpha(i, j, k) := \frac{\log p_i - \log p_j}{\log p_i - \log p_k} \quad \text{for any pair } (i, j, k) \in \Sigma^3. \quad (3.7)$$

Lemma 3.8. *For a memoryless source of probabilities $\mathfrak{P}$, the following conditions are equivalent:*

- (a) *The intersection $\underline{\mathcal{Z}} \cap \{\Re s = 1\}$ contains a point $s \neq 1$.*
- (b) *All the ratios $\alpha(i, j, k)$ defined in (3.7) are rational numbers.*
- (c) *There exists $\tau > 0$ for which the equality $\underline{\mathcal{Z}} \cap \{\Re s = 1\} = 1 + 2\pi i\tau\mathbb{Z}$ holds.*

A source which satisfies one of these conditions is said to be quasi-periodic.

Proof. We will prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c)

(a) $\Rightarrow$ (b). For $s = 1 + 2i\pi t$, the inequalities

$$\lambda(s) = \sum_{\ell \in \Sigma} p_\ell e^{2i\pi t \log p_\ell} \quad \text{so that} \quad |\lambda(s)| = \left| \sum_{\ell \in \Sigma} p_\ell e^{2i\pi t \log p_\ell} \right| \leq \sum_{\ell \in \Sigma} p_\ell = 1$$

always hold. Now, if the equality $\lambda(1 + 2i\pi t) = u = e^{2i\pi\theta}$ holds, this entails (by the converse of the triangular inequality) the following equalities,

$$\forall \ell, \quad e^{2i\pi t \log p_\ell} = u = e^{2i\pi\theta}, \quad \text{and then} \quad t \log p_\ell \in \theta + \mathbb{Z}, \quad t(\log p_i - \log p_j) \in \mathbb{Z}, \quad (3.8)$$

and thus, if $t \neq 0$, one deduces

$$\alpha(i, j, k) := \frac{\log p_i - \log p_j}{\log p_i - \log p_k} \in \mathbb{Q} \quad \forall i, j, k.$$

(b) $\Rightarrow$ (c). The previous proof shows that the possible solutions of the equation $\lambda(1 + 2i\pi t) = u$ arise when

$$t \in \frac{1}{\log p_i - \log p_j} \mathbb{Z}.$$

When all the real numbers $\alpha(i, j, k)$ are rational, the intersection of all these lattices is a not reduced to $\{0\}$ and is of the form $\mathbb{Z}\tau$, with $\tau > 0$.

Fix an element of the lattice $\mathbb{Z}\tau$, of the form $t = k\tau$ for $k \in \mathbb{Z}$, and denote by θ_k the real $\theta_k \in [0, 2\pi[$ for which $k\tau \log p_1 \equiv \theta_k \pmod{1}$. Now, the pair $(k\tau, \theta_k)$ is a solution of the system

$$k\tau \log p_1 \equiv \theta \pmod{1}, \quad k\tau \log p_2 \equiv \theta \pmod{1}, \dots, \quad k\tau \log p_r \equiv \theta \pmod{1}.$$

This means that the set $\underline{\mathcal{Z}} \cap \{\Re s = 1\}$ coincides with the set $\{1 + 2i\pi k\tau, \quad k \in \mathbb{Z}\}$. □

3.2.2 Case of a Markov chain.

We now consider the case of a good Markov chain, where the expression of $\mathbb{T}(s, u) = (1 - u\mathbf{R}_s)^{-1}$ in this case involves the transition matrix of the Markov chain.

Characterization of the set of poles. There is an analog of Lemma 3.5 in the case of a good Markov chain.

Lemma 3.9. *The function $s \mapsto \mathbb{T}(s)$ of a good Markov chain is meromorphic on the complex plane, analytic on the half plane $\Re s > 1$ and has a simple pole at $s = 1$. Moreover, the set $\mathcal{Z}$ of poles is defined as*

$$\mathcal{Z} = \{s; \quad \det \mathbb{T}(s) = 0\}$$

Proof. The function $s \mapsto \mathbf{R}_s$ is analytic on the complex plane, and thus the function $s \mapsto (I - \mathbf{R}_s)^{-1}$ is meromorphic with a set of poles $\mathcal{Z}$ defined in the Lemma. Let $\sigma := \Re s$. Then, the inequality $\|\mathbf{R}_s^k\| \leq \|\mathbf{R}_\sigma^k\|$ holds and entails the inequality on the spectral radii $r(s) \leq r(\sigma)$. In the case of a good Markov chain, the spectral radius $r(\sigma)$ equals the dominant eigenvalue $\lambda(\sigma)$. Now assume that $\sigma > 1$ and prove the inequality $\lambda(\sigma) \leq \lambda(1) = 1$. As the inequality $\lambda(\sigma) \leq \lambda(1)$ holds, we assume that the equality $\lambda(\sigma) = \lambda(1)$ holds, and we look for a contradiction.

The equalities

$$\sum_j p_{i|j}^\sigma \pi_\sigma^{(j)} = \lambda(\sigma) \pi_\sigma^{(i)}, \quad \lambda(1) = 1 = \sum_i p_{i|j} = \sum_j \pi_\sigma^{(j)}$$

entail

$$\lambda(\sigma) = \sum_{i,j} p_{i|j}^\sigma \pi_\sigma^{(j)} = \sum_j \pi_\sigma^{(j)} \sum_i p_{i|j}^\sigma,$$

and thus

$$0 = \lambda(1) - \lambda(\sigma) = \sum_j \pi_\sigma^{(j)} \left[\sum_i (p_{i|j} - p_{i|j}^\sigma) \right].$$

Then all the following conditions $(\forall i \exists! j = \tau(i) \ p_{i|j} = 1)$ hold. When the Markov chain is good, there does not exist such a map $\tau : \Sigma \rightarrow \Sigma$. $\square$

Periodicity. Consider a good Markov chain, its transition matrix $\mathbf{R}$, and, for any cycle of length $k \geq 1$, of the form $\mathcal{C} := \{i_1 i_2 \dots i_k\}$, its probability $p(\mathcal{C}) := p_{i_1|i_k} p_{i_2|i_1} \dots p_{i_k|i_{k-1}}$ and its normalized probability $\pi(\mathcal{C}) = p(\mathcal{C})^{1/k}$. We also consider all the possible ratios of the form

$$\alpha(\mathcal{C}, \mathcal{K}) := \frac{\log \pi(\mathcal{C})}{\log \pi(\mathcal{K})} \quad \text{for each pair } (\mathcal{C}, \mathcal{K}) \text{ of cycles,} \quad (3.9)$$

$$\alpha(\mathcal{C}, \mathcal{K}, \mathcal{L}) := \frac{\log \pi(\mathcal{C}) - \log \pi(\mathcal{K})}{\log \pi(\mathcal{C}) - \log \pi(\mathcal{L})} \quad \text{for each triple } (\mathcal{C}, \mathcal{K}, \mathcal{L}) \text{ of cycles,} \quad (3.10)$$

These ratios play a similar role as the previous ratios $\alpha(i, j)$ or $\alpha(i, j, k)$ in the memoryless case. Indeed, the following result holds and extends the previous Lemmas 3.6 and 3.8. Its proof is omitted.

Lemma 3.10. *For a good Markov chain, with transition matrix $\mathbf{R}$, the following conditions are equivalent:*

- (a) *The intersection $\mathcal{Z} \cap \{\Re s = 1\}$ contains a point $s \neq 1$.*
- (b) *All the ratios $\alpha(\mathcal{K}, \mathcal{L})$ defined in (3.9) are rational*
- (c) *There exists $\tau > 0$ for which the equality $\mathcal{Z} \cap \{\Re s = 1\} = 1 + 2i\pi\tau\mathbb{Z}$ holds.*
- (d) *The matrix $s \mapsto \mathbf{R}_s$ is periodic of period $i\tau$.*

A Markov chain which satisfies one of these conditions is said to be periodic.

When a Markov chain is periodic, there exists an algebraic integer a and a vector of positive reals $(\nu_1, \nu_2, \dots, \nu_r)$ for which the matrix $\mathbf{R}$ is written as $\mathbf{R} = D^{-1} \mathbf{Q} D$, where D is the matrix whose diagonal is $(\nu_1, \nu_2, \dots, \nu_r)$ and all the nonzero coefficients of the matrix $\mathbf{Q}$ belong to the group generated by a .

Quasi-periodicity. We will be also interested by the possible singularities of the function $s \mapsto \mathbb{T}(s, u)$ related to the set

$$\underline{\mathcal{Z}} := \{s; \exists u, |u| = 1, \det(I - u\mathbf{R}_s) = 0\},$$

and its position with respect to the vertical line $\{\Re s = 1\}$. The following result provides an extension of the previous Lemma 3.10 in the case when u is a general complex of modulus 1. It is related to the rationality of the ratios $\alpha(\mathcal{C}, \mathcal{K}, \mathcal{L})$, defined in (3.10).

Lemma 3.11. *For a good Markov chain, with transition matrix $\mathbf{R}$, the following conditions are equivalent:*

- (a) *The intersection $\underline{\mathcal{Z}} \cap \{\Re s = 1\}$ contains a point $s \neq 1$.*
- (b) *All the ratios $\alpha(\mathcal{C}, \mathcal{K}, \mathcal{L})$ defined in (3.10) are rational*
- (c) *There exists $\tau > 0$ for which the equality $\underline{\mathcal{Z}} \cap \{\Re s = 1\} = 1 + 2i\pi\tau\mathbb{Z}$ holds.*

A Markov chain which satisfies one of these conditions is said to be quasi-periodic.

3.2.3 Case of a dynamical source.

We now consider the case of a dynamical system. In this case, $\mathbb{T}(s, u)$ is an operator which is expressed as the quasi inverse $(I - u\mathbb{H}_s)^{-1}$ of the secant transfer operator $\mathbb{H}_s$ defined in (2.10). Moreover, we limit ourselves to good dynamical sources whose definition is given in Definition 2.19. Informally speaking for good dynamical systems, the transfer operator has properties that are very similar to those of a finite matrix. We now recall the statements of Lemma 2.20 in a slightly different form, which will introduce the entropic function σ which intervenes in Definition 3.4.

Proposition 3.12. *In the case of a good dynamical source, and when (s, u) is close to $(1, 1)$ the function $s \mapsto (I - u\mathbb{H}_s)^{-1}$ is meromorphic and has a unique pole at $s = 1 + \sigma(u)$, where the function σ is defined with the Implicit Function Theorem by the conditions*

$$\sigma(1) = 0, \quad 1 - u\lambda(1 + s) = 0 \quad (3.11)$$

which involve the dominant eigenvalue $\lambda(s)$ of the operator $\mathbb{H}_s$. The function σ is called the entropic function

Now, we study the possible periodicity (or quasi-periodicity) of such a good dynamical source, and we introduce the analogs of quantities $\alpha(k, j)$ defined in (3.5) for memoryless sources or $\alpha(\mathcal{C}, \mathcal{K})$ defined in (3.9) for Markov chains. All these quantities are defined with cycles, and we are then led to study the fixed points of inverse branches $h \in \mathcal{H}^*$. First, it is clear that, for a good dynamical system [72], any inverse branch $h \in \mathcal{H}^*$ has a unique fixed point, denoted by h^* .

For an inverse branch $h \in \mathcal{H}^*$, we denote the depth of h by $p(h)$ and for $h, k, \ell \in \mathcal{H}^*$, we consider

$$\pi(h) := |h'(h^*)|^{1/p(h)}, \quad \alpha(h, k) := \frac{\log \pi(h)}{\log \pi(k)}, \quad \alpha(h, k, \ell) := \frac{\log \pi(h) - \log \pi(k)}{\log \pi(h) - \log \pi(\ell)} \quad (3.12)$$

These ratios $\alpha(h, k)$ or $\alpha(h, k, \ell)$ provide an extension of our previous quantities α that we have already defined for simple sources. The following result is an extension of the previous results described in Lemma 3.6, 3.8, 3.10 and 3.11. It relates the possible singularities of the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$ on the vertical line $\Re s = 1$ to the rationality of the ratios α 's. However, as there is an infinite number of possible ratios α , there are no longer exact equivalences.

Proposition 3.13. *Consider a dynamical system of the Good Class and its secant transfer operator $\mathbb{H}_s$, acting on the space $\mathcal{C}^1(\mathcal{I} \times \mathcal{I})$. Then, the following holds:*

- (a) *If there exists $t_0 \neq 0$ for which the spectrum $\text{Sp } \mathbb{H}_{1+it_0}$ contains an eigenvalue equal to 1, then all the ratios $\alpha(h, k)$ are rational numbers, and the set of the real numbers t for which the spectrum $\text{Sp } \mathbb{H}_{1+it}$ contains an eigenvalue equal to 1 is a lattice $\mathbb{Z} \cdot \tau$ for some $\tau > 0$. In this case, the source is said to be periodic.*
- (b) *If there exists a ratio $\alpha(h, k)$ which is not rational, then, the quasi-inverse $(I - \mathbb{H}_s)^{-1}$ is analytic on $\Re s = 1$ except at $s = 1$ where it has a simple pole.*
- (c) *If there exists $t_0 \neq 0$ for which the spectrum $\text{Sp } \mathbb{H}_{1+it_0}$ contains an eigenvalue u of modulus 1, then all the ratios $\alpha(h, k, \ell)$ are rational numbers, and the set of the real numbers t for which the spectrum $\text{Sp } \mathbb{H}_{1+it}$ contains eigenvalue u of modulus 1 is a lattice $\mathbb{Z} \cdot \tau$ for some $\tau > 0$. In this case, the source is said to be quasi-periodic.*
- (d) *If there exists a ratio $\alpha(h, k, \ell)$ which is not rational, then, the spectral radius of $\mathbb{H}_s$ is strictly less than 1 on $\{s; \Re s = 1, s \neq 1\}$ and, for any u of modulus 1, the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$ is analytic on the line $\Re s = 1$ except for $s = 1$ and $u = 1$ where it admits a simple pole.*

3.2.4 Quasi periodicity is exceptional.

We have mentioned that the periodicity phenomenon arises in a natural context for simple sources, as any unbiased memoryless source is periodic. Then a natural further question is : *Do there exist many general dynamical sources which are periodic? quasi-periodic?*

The following result [proven for instance in [1]] shows that the (quasi)-periodicity phenomenon is in a sense exceptional for general good dynamical sources: It only occurs for sources which are obtained from simple sources by some conjugation. We recall that two dynamical sources $(\mathcal{I}, T)$ and $(\mathcal{I}, U)$ are conjugate if there exists a bijection $\Phi : \mathcal{I} \rightarrow \mathcal{I}$ of class $\mathcal{C}^2$ for which $T : \Phi \circ U \circ \Phi^{-1}$. Then, It is clear that a dynamical system which is conjugate to a periodic dynamical system is itself periodic.

Proposition 3.14. *A (complete) good dynamical source may be quasi-periodic only if it is conjugate to a source with affine branches.*

We will be more interested in the contrapositive of the previous statement:

NA. Non-Affine. If a good dynamical system is not conjugate to a dynamical system with affine branches, then $\mathbb{T}(s, u)$ is analytic for any $(s, u) \neq (1, 1)$, with $\Re s = 1, |u| = 1$.

3.2.5 An intermediary notion : p -periodicity

We have shown in the present Section that the following holds:

Periodicity. All the ratios $\alpha(h, k)$ are rational for any $h, k \in \mathcal{H}^*$

$$\iff \exists s, \Re s = 1, s \neq 1 \text{ where } \mathbb{T}(s) \text{ is not analytic.}$$

Quasi-Periodicity. All the ratios $\alpha(h, k, \ell)$ are rational for any $h, k, \ell \in \mathcal{H}^*$

$$\iff \exists (s, u), \Re s = 1, |u| = 1, (s, u) \neq (1, 1) \text{ where } \mathbb{T}(s, u) \text{ is not analytic.}$$

There are particular complex numbers u which play an “intermediary” role between the particular $u = 1$ and any u of modulus 1: they are the roots ξ of unity. The situation where

the p -th roots of unity play a role is called p -periodicity and the equivalence holds, as an easy extension of our results about periodicity:

p -Periodicity. All the ratios $\alpha(h, k)$ are rational for any $h, k \in [\mathcal{H}^p]^*$

$$\iff \exists s, \Re s = 1, s \neq 1, \exists \xi, \xi^p = 1 \text{ where } \mathbb{T}(s, \xi) \text{ is not analytic.}$$

3.2.6 Contrapositive versions of the statements about periodicity.

These contrapositive versions will be more useful when we wish to strengthen the hypotheses in order to obtain (perhaps) a stronger conclusion.

$\mathcal{NP}$. Non-Periodicity. There exist $h, k \in \mathcal{H}^*$ for which $\alpha(h, k) \notin \mathbb{Q}$

$$\implies \mathbb{T}(s) \text{ is analytic for any } s, \Re s = 1, s \neq 1$$

$\mathcal{NpP}$. Non- p -Periodicity. There exist $h, k \in \mathcal{H}^p$ for which $\alpha(h, k) \notin \mathbb{Q}$

$$\implies \mathbb{T}(s, \xi) \text{ is analytic for any } s, \Re s = 1, s \neq 1 \text{ and any } \xi \text{ with } \xi^p = 1$$

$\mathcal{NQP}$. Non-Quasi-Periodicity. There exist $h, k, \ell \in \mathcal{H}^*$ for which $\alpha(h, k, \ell) \notin \mathbb{Q}$

$$\implies \mathbb{T}(s, u) \text{ is analytic for any } (s, u) \neq (1, 1), \text{ with } \Re s = 1, |u| = 1.$$

3.3 Main principles for tameness.

In the previous Section, we have studied the behaviour of $s \mapsto \mathbb{T}(s, u)$ on the halfplane $\Re s > 1$ and the vertical line $\Re s = 1$. We have thus exhibited four conditions

- one condition of geometrical type: $\mathcal{NA}$,
- three conditions of arithmetical type: $\mathcal{NP}$, $\mathcal{NpP}$, $\mathcal{NQP}$.

Each of them entails the operator $\mathbb{T}(s)$ to be analytic on the punctured line $\Re s = 1, s \neq 1$. We now focus on the *left halfplane* $\{\Re s \leq 1\}$, and we also look for tameness: this means that we wish analytic functions, with *furthermore* a polynomial growth. As Definition 3.4 describes it, the final aim is to isolate a region $\mathcal{R} \supset \{\Re s \geq 1\}$ and a subset $\mathcal{U}_0$ where the operator $s \mapsto \mathbb{T}(s, u)$ be tame on $\mathcal{R}$ for $u \in \mathcal{U}_0$. Finally, we are interested in

- the behaviour of $\mathbb{T}(s)$ on the left of the vertical line $\Re s = 1$, (not only on the vertical line)
- the tameness of $\mathbb{T}(s)$ (not only its analyticity)
- possible perturbations which would entail a good behaviour of $\mathbb{T}(s, u)$ for u close to 1, (not only at $u = 1$)

It is thus natural to define reinforcements of the previous conditions $\mathcal{NA}$, $\mathcal{NP}$, $\mathcal{NpP}$, $\mathcal{NQP}$, which would (perhaps, if we are lucky) entail the needed conclusions.

3.3.1 Geometric conditions on branches. The UNI Class.

As we are interested in systems which are *Strongly-Non-Quasiperiodic*, we are led to a reinforcement of the hypothesis of $\mathcal{NA}$ of Section 3.2.4. Note that this hypothesis is of geometric nature. The Condition UNI , proposed by Dolgopyat in [12] provides such a reinforcement, as we now explain.

One first defines a probability Pr_n on each set $\mathcal{H}^n \times \mathcal{H}^n$, in a natural way, and lets $\text{Pr}_n\{(h, k)\} := |h(\mathcal{I})| \cdot |k(\mathcal{I})|$, where $|\mathcal{J}|$ denotes the length of the interval $\mathcal{J}$. Furthermore, $\delta(h, k)$ denotes the “distance” between two inverse branches h and k of same depth, defined as

$$\delta(h, k) = \inf_{x \in \mathcal{I}} |\Psi'_{h,k}(x)| \quad \text{with} \quad \Psi_{h,k}(x) = \log \left| \frac{h'(x)}{k'(x)} \right|. \quad (3.13)$$

The distance $\delta(h, k)$ measures the difference between the “shape” of the two branches h, k .

The **UNI** Condition, stated as follows (see [12]), is a *geometric condition* which expresses that the probability that two inverse branches have almost the same “shape” is very small:

Definition 3.15. [Condition **UNI**]. A good dynamical system $(\mathcal{I}, T)$ with contraction ratio $\rho < 1$ satisfies the **UNI** condition if its set $\mathcal{H}^*$ of inverse branches satisfies the following

For any $\hat{\rho} \in]\rho, 1[$, and for any integer n , one has $\Pr_n[\delta \leq \hat{\rho}^n] \ll \hat{\rho}^n$.

For a source with affine branches, the “distance” δ is always zero, and the probabilities $\Pr_n[\delta \leq \hat{\rho}^n]$ are all equal to 1. Such a source never satisfies the Condition **UNI**. Conversely, a good dynamical source of the **UNI** Class cannot be conjugate to a source with affine branches, as it is proven by Baladi and Vallée [1].

Then, the condition **UNI** provides a re-inforcement (of geometrical type) of the hypothesis

The system is not conjugate to a dynamical system with affine branches

We can expect a re-inforcement of the conclusion of Assertion $\mathcal{NA}$ which will lead (if we are lucky) to tameness for $\mathbb{T}(s, u)$. We will see that it will be indeed the case.

3.3.2 Arithmetical conditions on branches. The DIOP Classes.

All the statements of Section 3.2.6 deal with real numbers $\alpha(h, k)$ or $\alpha(h, k, \ell)$ which must be *not rational*. In the same vein as previously, we look for a reinforcement of these hypotheses, and we wish to deal with real numbers which “strongly differ” from rational numbers. What does it mean exactly (in a more formal setting...)? We are then led to approximability of real numbers by rational numbers, and, more precisely, to *diophantine* numbers.

Irrationality exponent and Diophantine numbers. The *irrationality exponent* of an irrational number was introduced by Liouville. The irrationality exponent of the irrational number x is defined by

$$\mu(x) := \sup \left\{ \nu, \left| x - \frac{p}{q} \right| \leq \frac{1}{q^{2+\nu}} \text{ for an infinite number of pairs } (p, q) \right\}.$$

The irrationality exponent of the irrational x is then a measure of its approximability by rational numbers.

The approximability of an irrational number x is closely related to properties of its continued fraction expansion, since truncations of this expansion give rise to the rational numbers that provide the best rational approximations of the irrational x . When all the quotients that occur in the continued fraction expansion of x are bounded, the irrational x cannot be well approximable by rationals.

An irrational number x is *diophantine* if its irrationality exponent is finite. Then, a diophantine irrational number is not too well approximated by rational numbers: it can be viewed (in an informal way...) as an irrational number which “strongly differs” from a rational number.

The DIOP classes. Then, the reinforcements that we look for the previous hypotheses will deal with *diophantine* numbers. This natural idea is due to Dolgopyat [13], and leads to define two classes of sources.

The following condition, denoted by **DIOP2**,

There exist two branches $h, k \in \mathcal{H}^*$ for which the ratio $\alpha(h, k)$ is diophantine.

provides a reinforcement of the notion of *Non-Periodicity*

There exist two branches $h, k \in \mathcal{H}^*$ for which the ratio $\alpha(h, k)$ is not rational.

This is a good candidate for the notion of *Strongly-Non-Periodicity*. We can expect in this case a “good behaviour” for $\mathbb{T}(s)$ for any $\Re s = 1, s \neq 1$.

In the same vein, the following condition, denoted by DIOP3,

There exist three branches $h, k, \ell \in \mathcal{H}^*$ for which the ratio $\alpha(h, k, \ell)$ is diophantine
provides a reinforcement of *Non-Quasi-Periodicity*

There exist three branches $h, k, \ell \in \mathcal{H}^*$ for which the ratio $\alpha(h, k, \ell)$ is not rational.

This is a good candidate for the notion of *Strongly-Non-Quasi-Periodicity*. We can expect a “good behaviour” for $\mathbb{T}(s, u)$ for any $\Re s = 1, |u| = 1$ and $(s, u) \neq (1, 1)$

In fact, the DIOP2 condition is a good candidate for the notion of *Strongly-Non-Periodicity* for a sequence formed by the multiples m_q of some integer m , as we now explain. It is clear that the product of a diophantine number by a rational number is itself diophantine (with the same irrationality exponent). Furthermore, the equality

$$\frac{c}{d} \alpha(h, k) = \alpha(h^c, k^d)$$

holds. Then, from a pair (h, k) with $h \in \mathcal{H}^a$ and $k \in \mathcal{H}^b$ which leads to a diophantine number $\alpha(h, k)$, we let $m := \text{lcm}(a, b)$, with $m = ca = db$ and we first consider the branches h^c, k^d which both belong to $\mathcal{H}^m$, and for which $\alpha(h^c, k^d)$ is diophantine. We then build a sequence of pairs (h_ℓ, k_ℓ) , with h_ℓ and k_ℓ both in $\mathcal{H}^{m\ell}$ for any $\ell \geq 1$. Then, the DIOP2 condition is a good candidate for the notion of *Strongly-non-p-Periodicity* when the integer p is any multiple of the integer m . We can expect for DIOP2 a “good behaviour” for $\mathbb{T}(s, \xi)$ for any $\Re s = 1, s \neq 1$ and any root ξ of unity that satisfies $\xi^{m\ell} = 1$ for some fixed m and any $\ell \geq 1$.

We will see in the sequel that these reinforcements will be well-adapted to our purposes: These *arithmetical* conditions will indeed lead to tameness, as we see in the following of the chapter.

3.3.3 Shape of tameness regions.

As we will see soon, regions with an hyperbolic shape or vertical strips arise in a natural way as possible tameness regions for interesting subclasses of sources. This is why the following definition will be important in the sequel. It describes two cases of possible tameness regions $\mathcal{R}$ which will occur for a classical source:

Definition 3.16. [Shape of regions] *A region $\mathcal{R} \supset \{\Re s \geq 1\}$ has a*

- (a) *S-shape* (shorthand for Strip shape) *if $\mathcal{R}$ is a vertical strip $\Re(s) > 1 - \delta$ for some $\delta > 0$.*
- (b) *H-shape* (shorthand for Hyperbolic shape) *if $\mathcal{R}$ is an hyperbolic region $\mathcal{R}$, defined as, for some $A, B, \rho > 0$*

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

When they exist, δ is the width, ρ is the hyperbolicity exponent.

A vertical strip can be viewed as a region with a zero hyperbolicity exponent. We are interested by tameness regions which are the largest possible. Then, it is natural to define the hyperbolicity exponent of the source $\mathcal{S}$ as the *infimum of all the hyperbolicity exponents* of tameness regions of the source $\mathcal{S}$. For instance, if the source admits as tameness region a vertical strip, then the hyperbolicity exponent of the source equals 0. There also exist some sources for which the singularities of the Λ function come close to the vertical line $\Re s = 1$ very fast, with an exponential speed. Such sources have an hyperbolicity exponent equal to ∞ . This leads to the following definition:

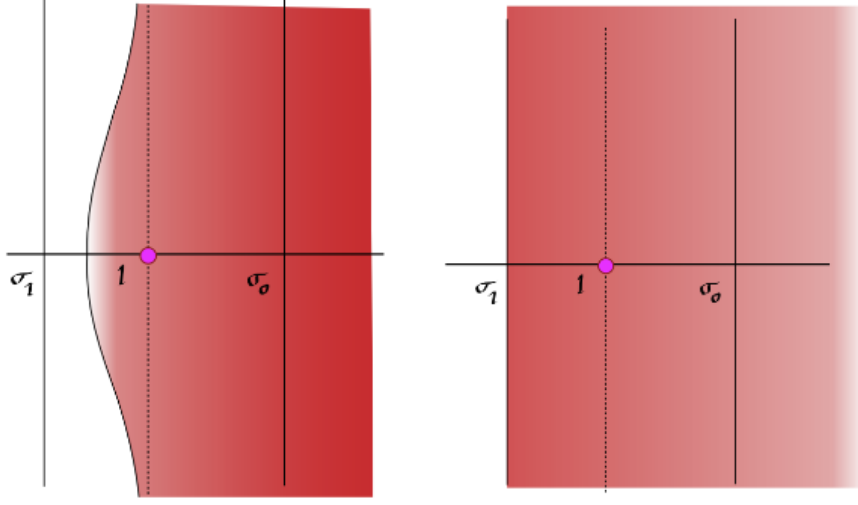


Figure 3.1: Two possible tameness regions $\mathcal{R}$. On the left, the case of the H -shape. On the right, the case of a S -shape.

3.4 Instances of tameness.

3.4.1 The periodic case. Tameness in vertical strips with an infinite number of poles.

In the case of a periodic source, which is conjugate to a simple source, the function $s \mapsto \lambda(s)$ is periodic of period $i\eta$, and there is a vertical strip on the left of the vertical line $\Re s = 1$ where the Λ function is analytic and of polynomial growth. There exists in this case a tameness region of the source which is a vertical strip. Then:

Proposition 3.17. *Consider a periodic source, with period $i\eta$, and its entropic function σ defined in (3.11). There exist a complex neighborhood $\mathcal{U}$ of $u = 1$ and a vertical strip $\mathcal{R} := \{\Re s > 1 - \delta\}$ such that the operator $\mathbb{T}(s, u)$ fulfills two main properties :*

- (i) *For each $u \in \mathcal{U}$, the operator $s \mapsto \mathbb{T}(s, u)$ is meromorphic on $\mathcal{R}$, with a unique family of poles at $s_k = 1 + \sigma(u) + ik\eta$*
- (ii) *For $u \in \mathcal{U}$, the map $s \mapsto \mathbb{T}(s, u)$ is of polynomial growth on $\mathcal{R}$, uniformly with respect to u .*

3.4.2 The UNI Class. Tameness in vertical strips.

We have seen that a good dynamical source of the UNI Class cannot be conjugate to a source with affine branches, as it is proven by Baladi and Vallée [1]. Then, the condition UNI excludes all the simple sources, which cannot be S -tame. The strength of the Condition UNI is due to the fact that this condition is sufficient to imply strong tameness:

Theorem 3.18. [Dolgopyat, Baladi–Vallée, Cesaratto–Vallée] *Consider a dynamical source of the Good Class, with UNI type, and its entropic function σ defined in (3.11). There exist a complex neighborhood $\mathcal{U}$ of $u = 1$ and a vertical strip $\mathcal{R} := \{s; \Re s > 1 - \delta\}$ (with $\delta > 0$) such that the operator $\mathbb{T}(s, u)$ fulfills two main properties:*

- (i) *For each $u \in \mathcal{U}$, the operator $s \mapsto \mathbb{T}(s, u)$ is meromorphic on $\mathcal{R}$, with only a pole at $s = 1 + \sigma(u)$*
- (ii) *For $u \in \mathcal{U}$, the map $s \mapsto \mathbb{T}(s, u)$ is of polynomial growth on $\mathcal{R}$, uniformly with respect to u .*

3.4.3 The DIOP Classes. Tameness in hyperbolic regions.

Tameness of simple sources. It is possible to define the irrationality exponent of a finite family of numbers, provided that they are not all rational. The irrationality $\mu(\mathcal{S})$ of a non-periodic simple source $\mathcal{S}$, is then defined as the irrationality exponent of the set

$$\{\alpha(\mathcal{C}, \mathcal{K}); \quad \mathcal{K}, \mathcal{C} \text{ cycles of length } \leq r\}.$$

The source is diophantine if the irrationality exponent is finite. For a memoryless source over an alphabet of size r , the irrationality exponent satisfies almost everywhere the equality

$$\mu(\mathfrak{P}) + 1 = \frac{1}{r - 1}$$

(“everywhere” means here: when the probability family $\mathfrak{P}$ is randomly chosen in the subset $\{(p_1, p_2, \dots, p_r) : p_j > 0, p_1 + p_2 + \dots + p_r = 1\}$ with respect to the Lebesgue measure)

Theorem 3.19. [Roux-Vallée] *For a simple non-periodic source, there exists an exact relation between the two exponents –the irrationality exponent μ and the hyperbolicity exponent ρ , namely the equality $\rho = 2\mu + 2$. A simple source is never S -tame. A diophantine non-periodic source is H -tame.*

The remark above entails that the hyperbolicity exponent of a non-periodic memoryless source over an alphabet of size r is “almost everywhere” equal to $2/(r - 1)$. The hyperbolicity exponent of a binary source is “almost everywhere” equal to 2.

Tameness of a dynamical source in hyperbolic regions. The following result relates the arithmetic properties of the probabilities of the source to the geometry of the tameness region. This is the main contribution of Roux’ PhD thesis [60].

Theorem 3.20. [Dologopyat, Naud, Melbourne, Roux–Vallée] *Consider a dynamical source of the Good Class and DIOP type, and its entropic function σ defined in (3.11). There exist an hyperbolic region $\mathcal{R}$ and a complex neighborhood $\mathcal{U}$ of $u = 1$ such that the operator $\mathbb{T}(s, u)$ fulfills two main properties :*

- (i) *For each $u \in \mathcal{U}$, the operator $s \mapsto \mathbb{T}(s, u)$ is meromorphic on $\mathcal{R}$, with a unique pole at $s = 1 + \sigma(u)$,*
- (ii) *Consider the following subsets of $\mathcal{U}$,*

$$\begin{aligned} \mathcal{T} &:= \{u \in \mathcal{U}; u = e^{i\theta}, \theta \in \mathbb{R}\}, \\ \mathcal{T}_m &:= \{u \in \mathcal{U}; u = \exp[2i\pi a/(mb)], (a, b) \in \mathbb{Z} \times \mathbb{N}\}. \end{aligned} \tag{3.14}$$

- (b3) *For a DIOP3 source, for any $u \in \mathcal{T}$, the operator $\mathbb{T}(s, u)$ is of polynomial growth in $\mathcal{R}$, uniformly with respect to $u \in \mathcal{T}$.*
- (b2) *For a DIOP2 source, there exists an integer m and a real $r \geq 0$, such that for any $u = \exp[2i\pi a/(mb)] \in \mathcal{T}_m$, the operator $\mathbb{T}(s, u)$ is of polynomial growth in $\mathcal{R}$, and satisfies $\|\mathbb{T}(s, u)\|_{(1,t)} \leq Kb|t|^r$.*

In the case of a general source, the optimality of the tameness region is not proven, and there is only an upper bound on the hyperbolicity exponent. We cannot exclude that there may exist a vertical strip as a tameness region, for which the hyperbolicity exponent equals 0. This happens when the UNI condition is also fulfilled.

3.4.4 Conclusion of tameness study for classical sources.

We now return to the Definition 3.4, which depends on

- the shape of the region $\mathcal{R}$,
- the number of poles in $\mathcal{R}$
- the shape of subset $\mathcal{U}_0$.

We have found natural instances of classical sources which lead to various types of tameness, that are defined in the table of Figure 3.2:

Definition of tameness type	$\mathcal{R}$	Number of poles	$\mathcal{U}_0 \subset \mathcal{U}$	Examples of occurrences
P -tame	A vertical strip	∞	$\mathcal{U}$	Periodic Sources
S -tame	A vertical strip	1	$\mathcal{U}$	UNI Sources
Strongly H -tame	A hyperbolic region	1	$\mathcal{T}$ defined in (3.14)	DIOP3 Sources
Weakly H -tame	A hyperbolic region	1	$\mathcal{T}_m$ defined in (3.14)	DIOP2 Sources

Figure 3.2: Various types of tameness defined by the shape of the region $\mathcal{R}$, the number of poles in $\mathcal{R}$ and the shape of the subset $\mathcal{U}_0$ of the complex neighborhood of $\mathcal{U}$

Moreover, the following is true:

- simple sources are never S -tame, but they may be H -tame or P -tame, according to arithmetic properties of their probabilities.
- dynamical sources may be P -tame only if they are “conjugate” to simple sources.

The various possibilities are recalled in Figure 3.3.

Tameness type	Simple sources	Dynamical sources not conjugate to simple sources	Arithmetic conditions on ratios α	Geometric conditions on branches
P -tame	Possible	Impossible	$\alpha(h, k) \in \mathbb{Q}$	—
Quasi-periodic	Possible	Impossible	$\alpha(h, k, \ell) \in \mathbb{Q}$	—
S -tame	Impossible	Possible	—	UNI
Strongly H -tame	Possible	Possible	DIOP3	—
Weakly H -tame	Possible	Possible	DIOP2	—

Figure 3.3: Tameness properties for classical sources, and sufficient conditions under which these tameness properties hold.

3.4.5 A small piece of history.

The UNI and DIOP conditions are introduced by Dolgopyat in the papers [12], [13]. He proves that, under these conditions, and in the case of a finite alphabet, the quasi inverse of the *plain* transfer operator is analytic and of polynomial growth in a region on the left of the line $\Re s = 1$. When the UNI condition holds, this region is a vertical strip. When the DIOP condition holds, this region is of hyperbolic type. There are extensions of these previous results to the quasi-inverse of the *secant* operator, which are proven to hold for the UNI condition by Cesaratto and Vallée [4], and for the DIOP condition by Roux and Vallée [61].

Conclusion of the Chapter

We have exhibited various tameness properties of the operator $\mathbb{T}(s, u)$ which will play a central role in the analytic studies performed in Chapter 7. This Chapter 7 indeed relies in a central way on the Rice methodology, where tameness is essential. We now describe in the next Chapter 4 the main tools we will use in our analysis, and, in particular, the Rice methodology.

Chapter 4

Main tools from analysis of algorithms

Contents

4.1	Bernoulli and Poisson models.	74
4.1.1	Bernoulli model.	74
4.1.2	Poisson model.	75
4.1.3	Relation between the two models. The Poisson Generating Function	75
4.1.4	Interest of the Poisson model.	75
4.1.5	Binomial recurrence.	76
4.1.6	Return from the Poisson model to the Bernoulli model	77
4.2	Mellin Transform	77
4.2.1	Basic Properties	77
4.2.2	Functional properties.	79
4.3	The Poisson-Mellin-Newton-Rice cycle	80
4.3.1	The Valuation-Degree Condition	80
4.3.2	Existence and expression of the analytic lifting $\varpi(s)$	80
4.4	Rice's methodology	81
4.4.1	First step: An integral form.	81
4.4.2	Second step: shifting to the left.	83
4.4.3	Tameness of $\varpi(s)$.	83
4.4.4	Asymptotic estimates for B_n .	84
4.5	Proof of the remainder estimates in the Rice method.	86
4.5.1	The statement	86
4.5.2	Estimates near the real axis.	86
4.5.3	Far from the real axis.	87
4.5.4	Proof of Proposition 4.16.	88
4.6	Depoissonization	89
4.6.1	A general depoissonization result.	89
4.6.2	The $\mathcal{JS}$ conditions.	90
4.6.3	Analytic extension of the sequence B_n .	90
4.7	Laplace Transform	91

We present the main analytic tools that will be used in the thesis. They are classical and largely used in analytic combinatorics, in particular in analysis of digital trees (tries and digital search trees). They are described in [24, 67, 71] for instance. However, we sometimes combine these tools in a non classical way, as we will see later.

In this chapter, we first define in Section 4.1 the probabilistic models which we are going to deal with. The initial model is the Bernoulli model where the number of words is fixed. As we will explain later, the Poisson model –where the number of words follows a Poisson law– is often easier to deal with, and this is why the first steps of our algebraic analyses will be performed inside this probabilistic model [see Chapters 5 and 6]. Then, we wish to return to the (natural) Bernoulli model.

The following of this chapter is devoted to this “return” from the Poisson model to the Bernoulli model. There are two main possible paths for this return: the Depoissonization methods and the Rice method. Here, in this thesis, we choose the last one, and this is why we focus in this Chapter on the Rice method. The Rice method starts with a binomial recurrence and transforms it into an integral along a vertical line of the complex plane. It deals with an analytic lifting of a sequence, and the Mellin transform provides conditions of existence for this analytic lifting. This is why we first recall some basic facts on the Mellin transform in Section 4.2 that are used next in the cycle “Poisson-Mellin-Newton-Rice” which proves the existence of this lifting. This cycle gathers all the main tools of this chapter and is central in our work : it is described in Section 4.3. Then, we describe the Rice method in Section 4.4 and we finally explain how the Rice method provides remainder estimates in Section 4.5. In the last section 4.7, we give some very basic properties of Laplace transform which will be used in Chapter 6.

4.1 Bernoulli and Poisson models.

Consider the set $\mathcal{S}^*$ of all the infinite words of the source. We wish to study a random variable $R : \mathcal{S}^* \rightarrow \mathbb{R}$. This means that the variable R is defined on any sequence $\mathcal{X} := (X_1, X_2, \dots, X_n)$ obtained by n uniform and independent drawings in the source $\mathcal{S}$ of infinite words. This drawing can be made via the mapping M defined in chapter 2: we draw n uniform real numbers $x_1, x_2, \dots, x_n$ from the unit interval $\mathcal{I} = [0, 1]$ and we let $X_i := M(x_i)$. Remark that all our parameters of interest (profile, depth) are defined on trees that are themselves built on sequence of words $\mathcal{X}$. Then, these parameters depend (in an indirect way) on the sequence $\mathcal{X}$ and they are random variables of the previous type.

4.1.1 Bernoulli model.

Such a variable R can be described by the sequence (R_n) of variables where R_n is the restriction of R to the set $\mathcal{S}^n$. In this model, called the *Bernoulli model*, and denoted by $(\mathcal{B}_n, \mathcal{S})$, we deal with a finite sequence $\mathcal{X}$ (of fixed cardinality n) of infinite words independently produced by the same source $\mathcal{S}$, namely $\mathcal{X} \in \mathcal{S}^n$. We are interested in the behaviour of $R(\mathcal{X})$ where $\mathcal{X}$ is randomly chosen in $\mathcal{S}^n$. We then study the random variable $R_n := R|_{\mathcal{S}^n}$ via its expectation $B_n := \mathbb{E}[R_n]$ (also denoted by $\mathbb{E}_n[R]$). It proves useful to consider the exponential generating function $B(z)$ of the sequence (B_n) ,

$$B(z) = \sum_{n \geq 0} B_n \frac{z^n}{n!}. \quad (4.1)$$

4.1.2 Poisson model.

Rather than fixing the cardinality n of the sequence $\mathcal{X} \in \mathcal{S}^*$, it is often more convenient to consider that the sequence $\mathcal{X} \in \mathcal{S}^*$ has a variable number N of elements that obeys a Poisson law of parameter z , defined by its distribution

$$\Pr[N = k] = e^{-z} \frac{z^k}{k!}. \quad (4.2)$$

Such a random variable N satisfies $\mathbb{E}[N] = \text{Var}[N] = z$: thus, in this model, N is narrowly concentrated near its mean z with a high probability so that the rate z plays a role much similar to the cardinality of $\mathcal{X}$. This model is called the Poisson model of rate z and is denoted by $(\mathcal{P}_z, \mathcal{S})$. It is then composed with two main steps:

- (a) The number N of words is drawn according to the Poisson law
- (b) Then, the N words are independently drawn from the source $\mathcal{S}$.

A variable $R : \mathcal{S}^* \rightarrow \mathbb{R}$ can be studied in this probabilistic model $(\mathcal{P}_z, \mathcal{S})$, and we denote by $\mathbb{E}_{[z]}[R]$ its expectation in this model.

4.1.3 Relation between the two models. The Poisson Generating Function

Dealing with conditional probabilities with respect to the events $[N = n]$, we obtain

$$\mathbb{E}_{[z]}[R] = \sum_{n \geq 0} \mathbb{E}_{[z]}[R|N = n] \cdot \Pr_{[z]}[N = n].$$

By definition of the models, the equality $\mathbb{E}_{[z]}[R|N = n] = \mathbb{E}[R_n]$ holds. Then, using the definition of the Poisson law entails

$$\mathbb{E}_{[z]}[R] = \sum_{n \geq 0} B_n \left(e^{-z} \frac{z^n}{n!} \right) = e^{-z} \sum_{n \geq 0} B_n \frac{z^n}{n!}$$

We are led to the following definition:

Definition 4.1. [Poisson transform] *The Poisson transform of a sequence (B_n) is defined by*

$$P(z) = e^{-z} \sum_{n \geq 0} B_n \frac{z^n}{n!}. \quad (4.3)$$

Lemma 4.2. *Consider a random variable R defined on the set $\mathcal{S}^*$. The expectation $\mathbb{E}_{[z]}[R]$ in the Poisson model is the Poisson transform $P(z)$ of the sequence $(\mathbb{E}_n[R])$ of its expectations in the Bernoulli model.*

4.1.4 Interest of the Poisson model.

In the Poisson model, the computations are often easy and “transparent”. We consider the following two examples, which are central in the sequel of the thesis :

Lemma 4.3. *Consider a source $\mathcal{S}$ and the Poisson model $(\mathcal{P}_z, \mathcal{S})$, where the number N of words of the source follows a Poisson law of parameter z . The following holds:*

- (a) *If the source $\mathcal{S}$ admits fundamental probabilities p_w (defined in Chapter 2), the number N_w of words which begin with the prefix w follows a Poisson law of parameter $p_w z$.*
- (b) *Consider a random variable $R := R_N$ defined on $\mathcal{S}^*$ with integer values and denote by $P(z)$ its expectation in the Poisson model $(\mathcal{P}_z, \mathcal{S})$. Then the expectation in the Poisson model $(\mathcal{P}_z, \mathcal{S})$ of the variable $S := R_{N+1}$ associated to the variable $N + 1$ is $P'(z) + P(z)$.*

Proof.

(a) In the Bernoulli model where N is fixed and equal to n , the variable N_w follows a binomial law, and

$$\Pr_n[N_w = k] = \binom{n}{k} p_w^k (1 - p_w)^{n-k}.$$

As the equality $\Pr_{[z]}[N_w = k | N = n] = \Pr_n[N_w = k]$ holds, we obtain, using the distribution of the variable N ,

$$\begin{aligned} \Pr_{[z]}[N_w = k] &= \sum_{n \geq k} \Pr_{[z]}[N_w = k | N = n] \cdot \Pr_{[z]}[N = n] \\ &= \sum_{n \geq k} \binom{n}{k} p_w^k (1 - p_w)^{n-k} \cdot \left(e^{-z} \frac{z^n}{n!} \right). \end{aligned} \quad (4.4)$$

The equalities

$$\binom{n}{k} \frac{z^n}{n!} = \frac{z^k}{k!} \cdot \frac{z^{n-k}}{(n-k)!}, \quad e^{-z} = e^{-zp_w} \cdot e^{-z(1-p_w)}$$

then entail the result, namely

$$\Pr_{[z]}[N_w = k] = e^{-zp_w} \frac{(zp_w)^k}{k!}.$$

(b) Denote by B_n the expectation of R when $N = n$. The exponential generating function C of the expectations of the variable $S := R_{N+1}$ equals

$$C(z) = \sum_{n \geq 0} B_{n+1} \frac{z^n}{n!} = \sum_{n \geq 1} B_n \frac{z^{n-1}}{(n-1)!}.$$

It thus satisfies $C(z) = B'(z)$. Then the Poisson generating function $Q(z)$ of the variable S is related to the Poisson generating function $P(z)$ of the variable R , via the equality

$$Q(z) = e^{-z} C(z) = e^{-z} \frac{d}{dz} (e^z P(z)) = P(z) + P'(z),$$

which completes the proof. $\square$

4.1.5 Binomial recurrence.

The following result relates the sequence (B_n) and the sequence formed with the coefficients of $P(z)$.

Lemma 4.4. *Consider a Poisson generating function $P(z)$ and its series expansion given by coefficients (P_n) under the form*

$$P(z) := e^{-z} \sum_{n \geq 0} B_n \frac{z^n}{n!} = \sum_{n \geq 0} (-1)^n P_n \frac{z^n}{n!}.$$

Then, the following binomial recurrences hold between the sequences (B_n) and (P_n)

$$(i) \quad B_n = \sum_{k=0}^n \binom{n}{k} (-1)^k P_k, \quad (ii) \quad P_n = \sum_{k=0}^n \binom{n}{k} (-1)^k B_k. \quad (4.5)$$

Proof. To prove (i), we observe that $B(z) = e^z P(z)$ and thus

$$B(z) = \left(\sum_{n \geq 0} \frac{z^n}{n!} \right) \left(\sum_{n \geq 0} (-1)^n P_n \frac{z^n}{n!} \right) = \sum_{n \geq 0} \frac{z^n}{n!} \left(\sum_{k=0}^n \binom{n}{k} (-1)^k P_k \right)$$

This entails the first relation. Moreover, the relation between the two sequences is involutive, and the second result follows. $\square$

4.1.6 Return from the Poisson model to the Bernoulli model

Assume now that the expectation of the random variable R is known in the Poisson model. Is it easy to return to the Bernoulli model $(\mathcal{B}_n, \mathcal{S})$ and obtain the expectation of R in the Bernoulli model? There are two possibilities, and the diagram of Table 4.1 is very useful for understanding the two points of view.

(a) [The left path.] We deal with the sequence (P_n) of the coefficients of $P(z)$ defined as

$$P_n := (-1)^n [z^n] P(z), \quad \text{with the expansion} \quad P(z) := \sum_{n \geq 0} (-1)^n P_n \frac{z^n}{n!}.$$

This method can be used when there exists an analytic lifting $\varpi(s)$ for the sequence (P_n) which moreover satisfies *tameness* properties. Then the binomial recurrence is transferred into an relation which expressed the sequence B_n as an integral along a vertical line which involves the analytic lifting. This is the Rice method.

(b) [The right path.] We directly deal with the series $P(z)$, and its asymptotics as $z \rightarrow \infty$, in particular in *cones*. Then, it is possible to *depoissonize* and prove, under some natural conditions, that the two sequences B_n and $P(n)$ behave in the same asymptotic way.

In this thesis, we decided to use the Rice methodology. However, we will explain the main principles of the Depoissonization method in Section 4.6, at the end of this Chapter. Both paths strongly rely on the Mellin transform. However, some easy properties of this transform are needed for the Rice method, whereas deeper properties of the Mellin transform are used in the Depoissonization process.

4.2 Mellin Transform

The Mellin transform is a very useful tool in analytic combinatorics and analysis of algorithms. For Rice's path, we only need basic facts about Mellin transform that we now describe as follow.

4.2.1 Basic Properties

The definition domain of a Mellin transform will be a vertical strip. We thus introduce the notation $< a, b >$ for an *open* strip of the complex plane, namely the subset of complex numbers $s := \sigma + it$ for which $\sigma \in]a, b[$. Such a strip is represented in Figure 4.6.

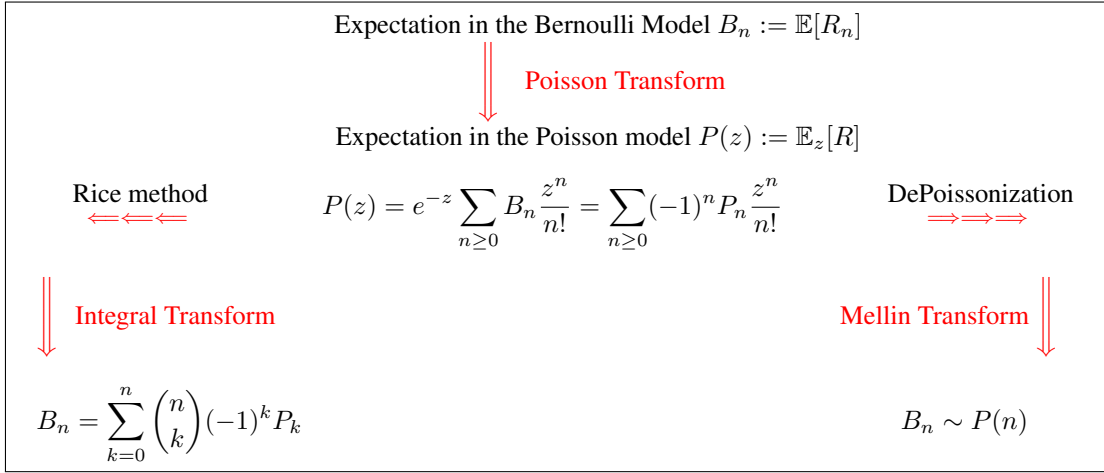


Table 4.1: This figure shows how to study a random variable R in the Bernoulli model. We first study it in the Poisson model $\mathcal{P}_z$ via the series $P(z)$. Then, we wish to return to the Bernoulli model, and there are two possible ways: the right way use the Depoissonization scheme, and the left-way uses the Rice methodology.

Definition 4.5. Let $f(x)$ be a real function defined on $(0, +\infty)$ and Lebesgue integrable. Then its Mellin transform is a complex valued function that is defined by

$$\mathcal{M}[f(x); s] := f^*(s) = \int_0^\infty f(x) x^{s-1} dx \quad (4.6)$$

The largest open strip $< a, b >$ in which the integral converges is called the fundamental strip. The function f^* is analytic inside $< a, b >$.

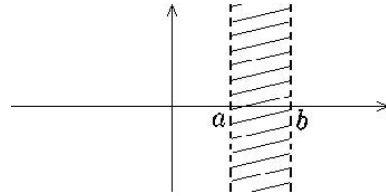


Figure 4.1: Fundamental strip $< a, b >$.

Lemma 4.6. If $f(x) = O(x^u)$ as $x \rightarrow 0^+$ and $f(x) = O(x^v)$ as $x \rightarrow +\infty$, then the Mellin transform $f^*(s)$ exists in the fundamental strip $< -u, -v >$.

Proof. We write with $\sigma := \Re s$

$$\begin{aligned} \left| \int_0^\infty f(x) x^{s-1} dx \right| &\leq \int_0^1 |f(x)| x^{\sigma-1} dx + \int_1^\infty |f(x)| x^{\sigma-1} dx \\ &\leq c_1 \int_0^1 x^{\sigma+u-1} dx + c_2 \int_1^\infty x^{\sigma+v-1} dx, \end{aligned}$$

where c_1 and c_2 are constants. The first integral exists for $\sigma > -u$ and the second for $\sigma < -v$. This proves the lemma. $\square$

Let $\alpha \in \mathbb{R}$. The power function $f : x \mapsto x^\alpha$ has the same order at $x = 0$ and $x = \infty$. Then, its fundamental strip is empty, and the Mellin function does not exist. However, there are many instances of function f for which the Mellin transform f^* exist.

Example 4.7. [The Gamma function.] Let $f(x) = e^{-x}$. Then, with the two estimates

$$f(x) = O(1) = O(x^0) \quad \text{as } x \rightarrow 0, \quad \forall M > 0 \quad f(x) = O(x^{-M}) \quad \text{as } x \rightarrow \infty,$$

Lemma 4.6 entails that $f^*(s)$ exists in the fundamental strip $< 0, +\infty >$. The Mellin transform of the function $x \mapsto e^x$ is the Gamma function defined as

$$\Gamma(s) := \int_0^{+\infty} e^{-x} x^{s-1} dx$$

Here is a table which describes the Mellin transforms of variants of the function $x \mapsto e^{-x}$. This shows three different functions that have the same Mellin transform, with different fundamental strips.

Function	Mellin Transform	Fund. Strip
e^{-x}	$\Gamma(s)$	$< 0, +\infty >$
$e^{-x} - 1$	$\Gamma(s)$	$< -1, 0 >$
$e^{-x} - 1 + x$	$\Gamma(s)$	$< -2, -1 >$

Table 4.2: Mellin transforms of variants of the function $x \mapsto e^{-x}$.

4.2.2 Functional properties.

There are many interesting functional properties for the Mellin transform. Here, we consider two important rules, relative to the change of scale and the derivation.

Lemma 4.8. Let f be a function whose Mellin transform $s \mapsto f^*(s)$ exists in the fundamental strip $< a, b >$. Then, the following rules hold for the function $f_\mu : x \mapsto f(\mu x)$ relative to the change of scale and for the derivation $f \rightarrow f'$.

$$\begin{aligned} (a) \quad & f^*(\mu s) = \mu^{-s} f^*(s) \quad \text{for } s \in < a, b > \\ (b) \quad & (f')^*(s) = -(s-1) f^*(s-1) \quad \text{for } s \in < a'-1, b'-1 > \end{aligned} \quad (4.7)$$

where the fundamental strip $< a', b' >$ is the intersection of the fundamental strip of f and the fundamental strip of $x \mapsto x f'(x)$.

Proof.

(a) The function $f_\mu : x \mapsto f(\mu x)$ satisfies $f_\mu(x) = \Theta f(x)$ when $x \rightarrow 0$ or when $x \rightarrow \infty$. Then the fundamental strip for f_μ is the same as for f . For any $s \in < a, b >$, the change of variable $y = \mu x$ gives the result.

(b) This rule is better understood when one considers the operator

$$\Delta := x \cdot \frac{d}{dx}$$

Consider the case where the fundamental strips of f and Δf have a non-empty intersection $< a', b' >$. This is the case in particular when the derivative of the asymptotic form of f coincides with the asymptotic form of the derivative, and in this case, the two fundamental strips are the same. For $s \in < a', b' >$ and if f is of class $\mathcal{C}^1$, integration by parts yields

$$\int_0^\infty f'(x) x^s dx = \int_0^\infty \Delta f(x) x^{s-1} dx = [f(x) x^s]_0^\infty - s \int_0^\infty f(x) x^{s-1} dx$$

and the term $[f(x) x^s]_0^\infty$ equals 0 inside the strip $< a', b' >$. $\square$

4.3 The Poisson-Mellin-Newton-Rice cycle

We now show how to obtain an analytic lifting of the sequence (P_k) , via the Mellin transform P^* of the Poisson generating function $P(z)$.

4.3.1 The Valuation-Degree Condition

This will be an important condition for obtaining such an analytic lifting.

Definition 4.9. For a sequence B of general term B_k , with $B_k \in \mathbb{C}$,

- (i) the valuation $\text{val}(B)$ is the smallest index of non-zero elements of B .
- (ii) the degree $\text{deg}(B)$ is the infimum of all c such that $B_k = O(n^c)$.
- (iii) the Valuation-Degree Condition (VLC) is the strict inequality $\text{val}(B) > \text{deg}(B)$.

The “good” case for finding a simple lifting $\varpi(s)$ of the sequence P_n arises when the Valuation-Degree Condition $\text{val}(B) > \text{deg}(B)$ holds, as we explain. This is not a quite restrictive hypothesis, since if the inequality $\text{deg}(B) \geq \text{val}(B)$ holds, we can subtract to the sequence B_k its leading terms so that the new sequence satisfies the inequality $\text{val}(B) > \text{deg}(B)$. In the case when the inequality $\text{val}(B) > \text{deg}(B)$ holds, we use the Mellin transform $P^*(s)$ of the Poisson generating function $P(z)$ in order to build the analytic lifting $\varpi(s)$ of the sequence (P_n) .

4.3.2 Existence and expression of the analytic lifting $\varpi(s)$

Proposition 4.10. Consider a sequence B_k which satisfies the Valuation-Degree Condition $\text{val}(B) > \text{deg}(B)$, and let $\text{val}(B) := 1 + k_0$, $\text{deg}(B) = c$. Denote by $P(z)$ its Poisson generating function, and by P_k the coefficients of $P(z)$,

$$P(z) := e^{-z} \sum_{k \geq 1+k_0} B_k \frac{z^k}{k!} = \sum_{k \geq 1+k_0} (-1)^k P_k \frac{z^k}{k!},$$

so that the binomial relations hold between the two sequences (B_k) and (P_k) , namely

$$B_n = \sum_{k=1+k_0}^n \binom{n}{k} (-1)^k P_k, \quad P_n = \sum_{k=1+k_0}^n \binom{n}{k} (-1)^k B_k. \quad (4.8)$$

Then, the sequence P_k admits an analytic lifting $\varpi(s)$ on the halfplane $\Re s > c$ which involves the Mellin transform $P^*(s)$ under the form

$$\varpi(s) = \frac{P^*(-s)}{\Gamma(-s)}. \quad (4.9)$$

Proof. There are three main steps.

Step 1. We prove that the Mellin transform $P^*(s)$ of the Poisson generating function $P(z)$ exists in the fundamental strip $\langle -1 - k_0, -c \rangle$ and is well defined by the following formal exchange of integration and summation, which is justified in the fundamental strip $\langle -1 - k_0, -c \rangle$,

$$\begin{aligned} P^*(s) &= \sum_{k \geq 1+k_0} \frac{B_k}{k!} \int_0^\infty e^{-z} z^k z^{s-1} dz = \sum_{k \geq 1+k_0} \frac{B_k}{k!} \Gamma(k+s) \\ &= \Gamma(s) \left[\sum_{k=1+k_0}^\infty B_k \frac{s(s+1) \dots (s+k-1)}{k!} \right]. \end{aligned}$$

Indeed, each $\Gamma(s+n)$ is well defined for $n \geq 1+k_0$ as soon as $\Re s > -1 - k_0$. Furthermore, the following estimate that holds uniformly with respect to s in a compact subset of $\mathbb{C} \setminus \mathbb{Z}_{\leq 1}$,

$$\frac{s(s+1) \dots (s+n-1)}{n!} = \frac{n^{s-1}}{\Gamma(s)} \left[1 + O\left(\frac{1}{n}\right) \right] \quad (n \rightarrow \infty). \quad (4.10)$$

proves that the last series is uniformly convergent for $\Re s < -c$.

Step 2. If we now let

$$\varpi(s) := \frac{P^*(-s)}{\Gamma(-s)} = \sum_{k=1+k_0}^\infty (-1)^k B_k \frac{s(s-1) \dots (s-k+1)}{k!}, \quad (4.11)$$

the right-side is expressed as a Newton interpolation series which converges on the half-plane $\Re s > c$ and defines there an analytic continuation $\varpi(s)$ of the ratio $P^*(-s)/\Gamma(-s)$.

Step 3. With Equations (4.8) and (4.11), the function $\varpi(s)$ satisfies $\varpi(n) = P_n$ for $n \geq 1+k_0$. $\square$

Then, the cycle “Poisson–Mellin–Newton–Rice” provides the analytic lifting $\varpi(s)$ which will be used in the Rice method. With the expression given by (4.9) the singularities of $\varpi(s)$ are often apparent. But this cycle “Poisson–Mellin–Newton–Rice” *does not prove a priori* that the analytic lifting $\varpi(s)$ is of polynomial growth. This property is also essential for using the Rice methodology, as we will see in the next two Sections, and this property *has to be proven by other means*. We will apply this proposition with $k_0 = 1$ and $c = 1$ (i.e., $\text{val}(B) = 2$, $\text{deg}(B) = 1$) in Lemmas 5.1 of Chapter 5 and 6.2 of Chapter 6.

4.4 Rice’s methodology

The Rice Formula introduced by Nörlund [53, 54], transforms a binomial sum into an integral in the complex plane.

4.4.1 First step: An integral form.

We recall that we are interested in the binomial sum of the form

$$B_n := \sum_{k=1+k_0}^n (-1)^k \binom{n}{k} P_k. \quad (4.12)$$

Definition 4.11. [Polynomial growth] A function $s \mapsto \varpi(s)$ defined in an unbounded domain Ω is said to be of polynomial growth if there exists r for which the estimate $|\varpi(s)| = O(|s|^r)$ holds as $s \rightarrow \infty$ on Ω . When Ω is included in a vertical strip $\{s; \Re s \in [a, b]\}$, this means: there exists r for which the equality $|\varpi(s)| = O(|\Im s|^r)$ holds as $s \rightarrow \infty$ on Ω .

Proposition 4.12. [Rice's Integral]. Assume that the sequence P_n in (4.12) admits a lifting $\varpi(s)$ which is analytic with polynomial growth in the half-plane $\Re s > c$ for some $c \in]k_0, 1 + k_0[$. Then, for n large enough and $d \in]c, 1 + k_0[$, the sequence B_n admits an integral representation of the form

$$B_n = -\frac{1}{2i\pi} \int_{d-i\infty}^{d+i\infty} \varpi(s) L_n(s) ds \quad \text{with} \quad L_n(s) = \frac{(-1)^n n!}{s(s-1)(s-2)\dots(s-n)}. \quad (4.13)$$

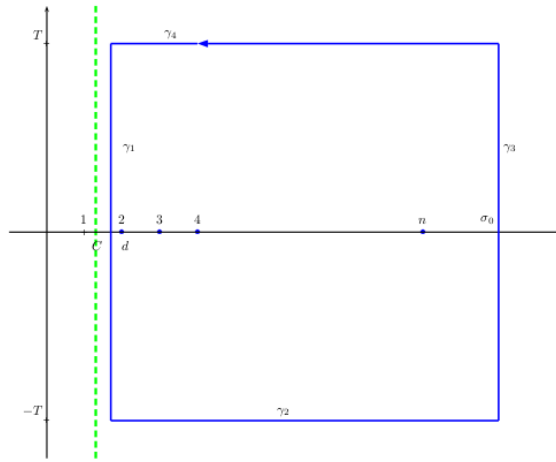


Figure 4.2: The picture presents a contour of the integral for $\sigma_0 = 1$.

The function $L_n(s)$ is called the Rice kernel of order n .

Proof. First, we consider the rectangle $\mathcal{A}_M$ described in Figure 4.2, delimited by the contour ρ_M defined by the two vertical lines $\Re s = d$ (with $d \in]c, 1 + k_0[$), $\Re s = n + M$ and two horizontal lines $\Im s = \pm M$. If the contour ρ_M is taken counterclockwise, then the Residue Theorem applies and entails the equality

$$\begin{aligned} \frac{1}{2i\pi} \int_{\rho_M} L_n(s) \cdot \varpi(s) ds &= \sum_{k=1+\sigma_0}^n \text{Res}[L_n(s)\varpi(s); s=k] \\ &= (-1)^n \sum_{k=1+k_0}^n (-1)^k \binom{n}{k} \varpi(k) \\ &= (-1)^n \sum_{k=1+k_0}^n (-1)^k \binom{n}{k} P_k = B_n \end{aligned} \quad (4.14)$$

Next, the integral on the curve ρ_M is the sum of four integrals. Let now M tend to ∞ . The integrals on the right, top and bottom lines tend to 0, due to the polynomial growth of the function $\varpi(s)$. The integral on the left becomes

$$-\int_{d-i\infty}^{d+i\infty} L_n(s) \cdot \varpi(s) ds,$$

and we have proven (4.13). For details on the proof, we may refer to [53, 54] or [24]. $\square$

4.4.2 Second step: shifting to the left.

Then, along general principles in analytic combinatorics as explained in [24, 25], the integration line can be pushed to the left, as soon as $L_n(s) \cdot \varpi(s)$ (closely related to $\varpi(s)$) has good analytic properties.

Proposition 4.13. *Assume that the lifting $\varpi(s)$ of the sequence P_n is meromorphic in a region $\mathcal{R}$ on the left of $\Re s = 1 + k_0$ and of polynomial growth there (for $|\Im s| \rightarrow \infty$). Then*

$$B_n = - \left[\sum_k \text{Res}[L_n(s) \varpi(s); s = s_k] + \frac{1}{2i\pi} \int_{\mathcal{C}} L_n(s) \varpi(s) ds \right], \quad (4.15)$$

where $\mathcal{C}$ is a curve (oriented from the south to the north) of class $\mathcal{C}^1$ included in $\mathcal{R}$ and the sum is extended to all poles s_k of $L_n(s)$ inside the domain $\mathcal{D}$ delimited by the vertical line $\Re s = d$ and the curve $\mathcal{C}$.

Proof. The proof is similar to Lemma 4.12. We apply the Residue Theorem to the function $L_n(s) \varpi(s)$ inside the region $\mathcal{D}_M$ which is the intersection of the domain $\mathcal{D}$ with the horizontal strip $|\Im s| \leq M$, and denote $\mathcal{C}_M$ the curve (taken counterclockwise) which borders the region $\mathcal{D}_M$. Since $\varpi(s)$ is meromorphic, this is the same for $L_n(s)$ and

$$\frac{1}{2i\pi} \int_{\mathcal{C}_M} L_n(s) \varpi(s) ds = \sum_{s_k \in \mathcal{D}_M} \text{Res}[L_n(s) \varpi(s) : s = s_k]$$

where the sum is extended to all poles s_k of $L_n(s) \varpi(s)$ inside the domain $\mathcal{D}_M$. Now, when $M \rightarrow \infty$, the integrals on the two horizontal segments tend to 0, since $\varpi(s)$ is of polynomial growth, and

$$\begin{aligned} \lim_{M \rightarrow \infty} \int_{\mathcal{C}_M} L_n(s) \varpi(s) ds &= \int_{d-i\infty}^{d+i\infty} L_n(s) \varpi(s) ds - \int_{\mathcal{C}_2} L_n(s) \varpi(s) ds \\ &= 2i\pi \sum_{s_k \in \mathcal{D}} \text{Res}[L_n(s) \varpi(s) : s = s_k]. \end{aligned}$$

We have then proven the result. $\square$

The dominant singularities of $L_n(s) \varpi(s)$ provide the asymptotic behaviour of B_n , and the remainder integral is estimated using the polynomial growth of $L_n(s)$ when $|\Im(s)| \rightarrow \infty$.

4.4.3 Tameness of $\varpi(s)$.

For this shifting to the left, in the previous section we need, a region $\mathcal{R}$ on the left of $\Re s = 1 + k_0$, where $\varpi(s)$ is of polynomial growth (for $|\Im s| \rightarrow \infty$) and meromorphic. We need also a good knowledge of its poles. These properties are described by the tameness of the function $\varpi(s)$ at the point σ_0 which is the (possible) rightmost singularity of $\varpi(s)$.

Definition 4.14. *A function $\varpi(s)$ is tame at σ_0 if one of the three following properties holds:*

(a) [*S*-shape] (shorthand for Strip shape) *there exists a vertical strip $\Re(s) > \sigma_0 - \delta$ for some $\delta > 0$ where $\varpi(s)$ is meromorphic, has a sole pole (of order $\ell_0 > 0$) at $s = \sigma_0$ and is of polynomial growth as $|\Im s| \rightarrow +\infty$.*

(b) [*H*-shape] (shorthand for Hyperbolic shape) *there exists a hyperbolic region $\mathcal{R}$, defined as, for some $A, B, \beta_0 > 0$*

$$\mathcal{R} := \{s = \sigma + it; |t| \geq B, \sigma > \sigma_0 - \frac{A}{t^{\beta_0}}\} \cup \{s = \sigma + it; \sigma > \sigma_0 - \frac{A}{B^{\beta_0}}, |t| \leq B\},$$

where $\varpi(s)$ is meromorphic, with a sole pole (of order $\ell_0 > 0$) at $s = \sigma_0$ and is of polynomial growth in $\mathcal{R}$ as $|\Im s| \rightarrow +\infty$.

(c) [*P*-shape] (shorthand for Periodic shape) *there exists a vertical strip $\Re(s) > \sigma_0 - \delta$ for some $\delta > 0$ where $\varpi(s)$ is meromorphic, has only a pole (of order $\ell_0 > 0$) at $s = \sigma_0$ and a family (s_k) (for $k \in \mathbb{Z} \setminus \{0\}$) of simple poles at points $s_k = \sigma_0 + 2ki\pi t$ with $t \neq 0$, and is of polynomial growth as $|\Im s| \rightarrow +\infty$ ¹.*

There are four parameters relative to the tameness: the real σ_0 is the position, the integer ℓ_0 is the order, and, when they exist, the real δ is the abscissa, and the real ρ is the exponent.

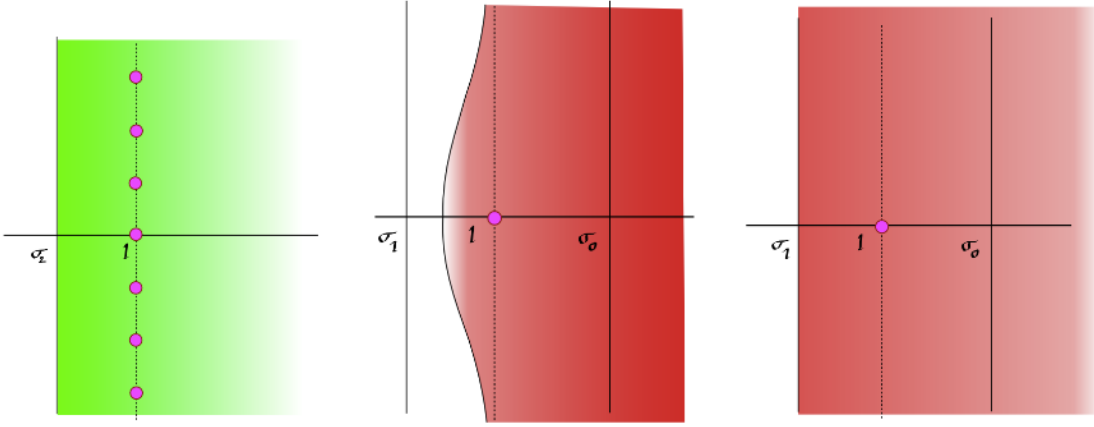


Figure 4.3: Three possible domains where the function $\varpi(s)$ is analytic and of polynomial growth : Periodic shape (on the left), Hyperbolic shape (at the middle) and Strip-shape (on the right).

This definition arises in a natural way, since we will see that the analytic lifting $\varpi(s)$ will be closely related to the Dirichlet series $s \mapsto \Delta(s, u)$ introduced in Chapter 2, which inherits tameness properties of the source, defined in Chapter 3.

4.4.4 Asymptotic estimates for B_n .

The tameness of $\varpi(s)$ entails the tameness of

$$L_n(s)\varpi(s) \quad \text{with} \quad L_n(s) := \frac{(-1)^n n!}{s(s-1)\dots(s-n)}$$

and the tameness parameters for $L_n(s)\varpi(s)$ are closely related to those of $\varpi(s)$. However, there are two cases for the order, according to whether the position σ_0 be an integer or not. The tameness order m_0 for $L_n(s)\varpi(s)$ at σ_0 equals the tameness order ℓ_0 when σ_0 is not an integer and $\ell_0 + 1$ if σ_0 is an integer. With the tameness properties of $L_n(s)\varpi(s)$, the shifting

¹More precisely, this means that $\varpi(s)$ is of polynomial growth on a family of horizontal lines $t = t_k$ with $t_k \rightarrow \infty$, and on vertical lines $\Re(s) = \sigma_0 - \delta'$ with some $\delta' < \delta$.

to the left inside the tameness region is possible and leads to the following result. Note that, from Proposition 4.13, the dominant part of the asymptotics comes from considering poles in the region $\mathcal{R}$ (residue calculus) and the error term comes from the evaluation of the integral on the curve $\mathcal{C}_2$ defined in Proposition 4.13.

Proposition 4.15. *The following holds for the sequence (B_n) , when it is related to $\varpi(s)$ by the Rice formula (4.13). If $\varpi(s)$ is tame at $s = \sigma_0$, with an order denoted by m_0 , the integer which equals ℓ_0 if σ_0 is not an integer and $\ell_0 + 1$ if σ_0 is an integer. Then there exists a polynomial Q of degree $m_0 - 1 \geq 0$ such the following asymptotics hold, depending on the tameness shape:*

(a) *With a S-shape and width δ_0 , for any $\delta < \delta_0$, one has, for $n \rightarrow \infty$,*

$$(-1)^{n+1} B_n = n^{\sigma_0} Q(\log n) + O(n^{\sigma_0 - \delta}).$$

(b) *With a H-shape and exponent β_0 , then, for any β with $\beta < 1/(\beta_0 + 1)$, one has, for $n \rightarrow \infty$,*

$$(-1)^{n+1} B_n = n^{\sigma_0} Q(\log n) + O\left(n^{\sigma_0} \cdot \exp[-(\log n)^\beta]\right)$$

(c) *With a P-shape and width δ_0 , then, for any $\delta < \delta_0$, one has, for $n \rightarrow \infty$,*

$$(-1)^{n+1} B_n = n^{\sigma_0} (Q(\log n) + \Phi(n)) + O(n^{\sigma_0 - \delta})$$

where $n^{\sigma_0} \cdot \Phi(n)$ is the part of the expansion brought by the family of the non real poles of $G(s)$ located on the vertical line $\Re s = \sigma_0$.

Proof. If $\varpi(s)$ is of polynomial growth in a region $\mathcal{R}$, the line of integration $\Re s = d$ can be moved to the left until a curve ρ , which lies inside the region $\mathcal{R}$, provided residues of poles inside $\mathcal{D}$ be taken into account. If $L_n(s)\varpi(s)$ has a pole of order $m_0 > 0$ at σ_0 , it admits near this pole the singular expression

$$L_n(s)\varpi(s) = \sum_{k=1}^{m_0} \frac{a_k}{(s - \sigma_0)^k}$$

and this pole contributes with a quantity of the form

$$n^{\sigma_0} \left[\sum_{k=0}^{m_0-1} (-1)^k \frac{a_k}{k!} \log^k n \right]$$

In the cases (a) or (c), the curve ρ can be chosen as a vertical line of equation $\Re s = \alpha$ with $\alpha = \sigma_0 - \beta$. In the case (b), the curve ρ can be chosen as an hyperbolic curve of the form

$$\rho := \left\{ s = \sigma + it, |t| \geq B, \sigma = \sigma_0 - \frac{A}{|t|^{\beta_0}} \right\} \cup \left\{ s = \sigma + it, \sigma = \sigma_0 - \frac{A}{B^{\beta_0}}, |t| \leq B \right\}$$

for some strictly positive constants (A, B, β) .

The remainder of the proof (made in the next Section) is devoted to the computation of the integral

$$\int_{\rho} L_n(s)\varpi(s) ds,$$

and proves the following : if $\varpi(s)$ is of polynomial growth on the curve ρ as $|s| \rightarrow \infty$, this integral is of order $n^{\sigma_0} O(n^{-\delta})$ in the cases (a) and (c) of Proposition 4.15. It is of order $n^{\sigma_0} O(\exp[-(\log n)^\beta])$ with $\beta < 1/(1 + \beta_0)$ in case (b). $\square$

4.5 Proof of the remainder estimates in the Rice method.

4.5.1 The statement

Proposition 4.16. Associate to a function $\varpi(s)$ the function $L_n(s) \cdot \varpi(s)$ with

$$L_n(s) := \frac{(-1)^n n!}{s(s-1) \dots (s-n)}.$$

(i) Consider the vertical line $\Re s = \alpha$, and assume that $\varpi(s)$ be continuous on $\Re s = \alpha$ and be of at most polynomial growth there: $\varpi(s) = O(s^r)$ as $|s| \rightarrow \infty$ on $\Re s = \alpha$. Then the integral on the vertical $\Re s = \alpha$ of $L_n(s)\varpi(s)$ admits the following estimate, as $n \rightarrow \infty$:

$$\int_{\Re s = \alpha} L_n(s)\varpi(s)ds = O(n^\alpha).$$

(ii) Consider a curve ρ of hyperbolic type, namely of the form:

$$\rho := \{s = \sigma + it, |t| \geq B, \sigma = \sigma_0 - \frac{A}{|t|^{\beta_0}}\} \cup \{s = \sigma + it, \sigma = \sigma_0 - \frac{A}{B^{\beta_0}}, |t| \leq B\}$$

for some strictly positive constants (A, B, β_0) , and assume that $\varpi(s)$ be continuous on ρ and be of at most polynomial growth there $\varpi(s) = O(|s|^r)$ as $|s| \rightarrow \infty$. Then the integral of $L_n(s)\varpi(s)$ on the curve ρ admits the following estimate, as $n \rightarrow \infty$,

$$\int_{\rho} L_n(s)\varpi(s)ds = n^{\sigma_0} \cdot O(\exp[-(\log n)^\beta]), \quad \text{with } \beta < \frac{1}{1 + \beta_0}.$$

For the proof, we only need to consider the upper half-plane. We use $T = \sqrt{n}$ as a cut-off point and decompose each of the curves –the vertical line or the hyperbolic curve ρ – into two parts. The first result provides estimates when s is near the real axis ($|\Im s| \leq T$) and the second lemma deals with the case when s is far from the real axis ($|\Im s| \geq T$).

4.5.2 Estimates near the real axis.

Lemma 4.17. For s outside of a fixed sector containing the negative real axis in its interior, and under the condition $|s| \leq \sqrt{n}$, one has, as $n \rightarrow \infty$:

$$L_n(s) = \frac{n!(-1)^n}{s(s-1) \dots (s-n)} = -n^s \Gamma(-s) \left(1 + O\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{s^2}{n}\right) \right). \quad (4.16)$$

Also, for any s fixed with $s \notin \mathbb{N}$, one has

$$L_n(s) = -n^s \Gamma(-s) \left(1 + O\left(\frac{1}{n}\right) \right). \quad (4.17)$$

Proof. One has

$$\frac{n!(-1)^n}{s(s-1) \dots (s-n)} = -\frac{n!}{-s(-s+1) \dots (-s+n)} = -\frac{\Gamma(n+1)\Gamma(-s)}{\Gamma(n-s+1)}.$$

Stirling's formula holds in the complex plane, provided a sector around the negative real axis is avoided. Under this condition, one has

$$\Gamma(w+1) = w^w e^{-w} \sqrt{2\pi w} \left(1 + O\left(\frac{1}{w}\right) \right), \quad |w| \rightarrow +\infty. \quad (4.18)$$

With the Stirling formula: :

$$\begin{aligned} \frac{\Gamma(n+1)}{\Gamma(n-s+1)} &= \frac{n^n e^{-n} \sqrt{2\pi n}}{(n-s)^{n-s} e^{s-n} \sqrt{2\pi(n-s)}} \left(1 + O\left(\frac{1}{n}\right)\right) \\ &= \exp[n \log n - (n-s) \log(n-s) - s] \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= \exp[s \log n - (n-s) \log(1-s/n) - s] \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \end{aligned}$$

In the region under consideration, we have $s/n = O(1/\sqrt{n})$, which is a small quantity, so that $\log(1+s/n) = s/n + O(s^2/n^2)$. Consequently,

$$\begin{aligned} \frac{\Gamma(n+1)}{\Gamma(n-s+1)} &= n^s \exp\left[O\left(\frac{s^2}{n}\right)\right] \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= n^s \left(1 + O\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{s^2}{n}\right)\right), \end{aligned}$$

and the estimate (4.16) results. The proof of (4.17) is similar, even simpler, via the relation $s/n = O(1/n)$. $\square$

4.5.3 Far from the real axis.

Lemma 4.18. Fix any number $m > 0$. Then, there exists a computable constant $K_m > 0$ such that for n large enough, $s = b + it$, b fixed and $t \geq \sqrt{n}$, one has

$$|L_n(s)| \leq \frac{K_m}{t^m} e^{-B\sqrt{n}},$$

with $B = \log(\sqrt{2})$.

Proof. The proof is done for $b = 0$, but it extends to any value b fixed. Choose an integer $m > 0$ and set $A = \lfloor \sqrt{n} \rfloor$. We write

$$\left| \frac{n!}{s(s-1)(s-2)\dots(s-n)} \right| = \frac{1}{|s|} \prod_{a=1}^m \left| \frac{a}{a-s} \right| \prod_{a=m+1}^{m+A} \left| \frac{a}{a-s} \right| \prod_{a=m+A+1}^n \left| \frac{a}{a-s} \right|$$

The first product has a trivial bound

$$\prod_{a=1}^m \left| \frac{a}{a-s} \right| < \frac{m!}{t^m} \quad (4.19)$$

For the second product, the complex s is close to the imaginary axis when $n \rightarrow \infty$. The triangle $(a, 0, s)$ is approximately a right triangle. The angle β at a satisfies, for n large enough,

$$\tan(\beta) \sim \frac{|s|}{|a|} \geq 1 \quad \text{and thus} \quad \left| \frac{a}{a-s} \right| = \cos(\beta) < \cos\left(\frac{\pi}{4}\right) = \left(\frac{1}{\sqrt{2}}\right)^A.$$

resulting in

$$\prod_{a=m+1}^{m+A} \left| \frac{a}{s-a} \right| < \left(\frac{1}{\sqrt{2}}\right)^A. \quad (4.20)$$

For the third product, we plainly use the triangle inequality, which gives $|a/(a-s)| < 1$ and

$$\prod_{a=m+A+1}^n \left| \frac{a}{a-s} \right| < 1. \quad (4.21)$$

Collecting (4.19), (4.20), (4.21), we have:

$$\left| \frac{n!}{s(s-1)(s-2)\dots(s-n)} \right| < \frac{m!}{t^m} \left(\frac{1}{\sqrt{2}} \right)^A = \frac{m!}{t^m} e^{-B\sqrt{n}}$$

Then, $K_m = m!$ and $B = \log(\sqrt{2})$. $\square$

4.5.4 Proof of Proposition 4.16.

We only need to consider the upper half-plane. We use $T = \sqrt{n}$ as a cut-off point and decompose each positive part $\tilde{\rho}$ of the curve –the vertical line or the hyperbolic curve ρ – into two parts.

Case of a vertical line. We use the decomposition

$$\int_{\tilde{\rho}} L_n(s) \varpi(s) ds = \int_{s=\alpha}^{\alpha+iT} L_n(s) \varpi(s) ds + \int_{s=\alpha+iT}^{\alpha+i\infty} L_n(s) \varpi(s) ds.$$

Near the real axis, namely for $s \in [\alpha, \alpha + iT]$, we apply Lemma 4.17:

$$\int_{s=\alpha}^{\alpha+iT} L_n(s) \varpi(s) ds = \int_{s=\alpha}^{\alpha+iT} n^s \Gamma(-s) \varpi(s) (1 + O(n^{-1})) ds \quad (4.22)$$

As the fast decay of $\Gamma(s)$ compensates more for the polynomial growth of $\varpi(s)$ and $|n^s| = n^\alpha$, the integral is $O(n^\alpha)$.

Far from the real axis, namely for $s \in [\alpha + iT, \alpha + \infty t]$, we apply Lemma 4.18:

$$\int_{s=\alpha+iT}^{\alpha+i\infty} |L_n(s) \varpi(s)| ds < K_m e^{-B\sqrt{n}} \int_{t=T}^{\infty} \frac{t^r}{t^m} dt = O(e^{-B\sqrt{n}}) \quad (4.23)$$

for n large enough, provided m has been chosen such that $m > r + 2$. The combination of Equations (4.22), (4.23) yields the claimed estimate in the case of a vertical line.

Case of an hyperbolic curve. Consider now the case of an hyperbolic curve, and consider the two parts of the curve $\tilde{\rho}$: the curve ρ^- (near the real axis) and the curve ρ^+ (near imaginary infinity).

$$\int_{\tilde{\rho}} L_n(s) \varpi(s) ds = \int_{\rho^+} L_n(s) \varpi(s) ds + \int_{\rho^-} L_n(s) \varpi(s) ds. \quad (4.24)$$

In the case of the curve ρ^+ , which can be compared to a vertical line, we apply Lemma 4.17 and

$$\left| \int_{\rho^+} L_n(s) \varpi(s) ds \right| < K_m \int_T^\infty O(t^r) \cdot O(t^{-m}) \cdot e^{-B\sqrt{n}} dt = O(e^{-B\sqrt{n}}), \quad (4.25)$$

for n large enough, provided m has been chosen such that $m > r + 2$.

Now, near the real axis, Lemma 4.18 gives

$$\int_{\rho^-} L_n(s) \varpi(s) ds = \left(\int_{\rho^-} n^s \Gamma(-s) \varpi(s) ds \right) (1 + O(n^{-1})). \quad (4.26)$$

Letting $s := \sigma + it$, and $B := \log n$, we use the following estimates

$$|n^s| = n^\sigma = n^{\sigma_0} \exp[-A B t^{-\beta_0}], \quad |\varpi(s)\Gamma(-s)| \leq \exp[-Kt],$$

(for some $K > 0$). The first one is due to the definition of the curve whereas the second one uses the fast decay of $\Gamma(-s)$ which more than compensates for the polynomial growth of $\varpi(s)$. If we let $B := \log n$, the modulus of the integrand is at most

$$|L_n(s)\varpi(s)| \leq n^{\sigma_0} \exp[-Kt - ABt^{-\beta_0}].$$

When n (and then B) is fixed, the minimum of the function $t \mapsto Kt + A B t^{-\beta_0}$ is reached for $t^{\beta_0+1} = \beta_0 B/K$. Then the maximum of $|L_n(s)\varpi(s)|$ is of order $n^{\sigma_0} \exp[-(\log n)^\beta]$ with $\beta < 1/(1 + \beta_0)$. Using the same principles as in the Laplace method, we obtain the estimate

$$\int_{\rho^-} L_n(s)\varpi(s)ds = n^{\sigma_0} O(\exp[-(\log n)^\beta]) \quad \text{with } \beta < 1/(1 + \beta_0).$$

this yields the claimed estimate in the case of a hyperbolic curve.

4.6 Depoissonization

We wish to describe more precisely the right path mentioned in Section 4.1.6. We recall the general framework: We wish to study a variable R and we denote by B_n its expectation in the Bernoulli model $\mathcal{B}_n$. When we work in the Poisson model, of rate z , the exponential generating function $B(z)$ of the sequence (B_n) ,

$$B(z) = \sum_{n \geq 0} B_n \frac{z^n}{n!}. \quad (4.27)$$

coincides with the expectation of the variable R in the Poisson model up to a factor of e^{-z} .

The Depoissonization method is used when one has a good knowledge of the Poisson transform $P(z)$ itself (whereas the Rice methodology will be used when one has a good knowledge of its coefficients $\varphi(n)$). As previously, we wish to return to the Bernoulli model, and we can expect that $P(n)$ which is the expectation of the Poisson model (when the cardinality N follows a Poisson law of rate $z = n$) is close to B_n (which is the expectation in the Bernoulli model when N is fixed and equal to n).

This return to the Bernoulli model needs a good behaviour of $P(z)$ with respect to cones. For $\theta < \pi/2$, the cone S_θ is the set of complex numbers z whose argument $\arg z$ satisfies the inequality $|\arg z| \leq \theta$.

4.6.1 A general depoissonization result.

We use the following theorem of Jacquet and Szpankowski. This new result which is very helpful and greatly simplifies the previous depoissonization results due to the same authors, described in [34] and [35].

Theorem 4.19 (Jacquet and Szpankowski, 2014). *Let $P(z)$ be the Poisson transform of a sequence B_n , namely*

$$P(z) = e^{-z} \sum_{n \geq 0} B_n \frac{z^n}{n!}.$$

This series is assumed to be absolutely convergent for $z \in \mathbb{C}$. Assume moreover that the sequence B_n admits an analytic extension $z \mapsto T(z)$ in a cone S_α , which is of polynomial growth in this cone S_α when $|z| \rightarrow \infty$, i.e., there exists γ such that, for $l|z| \rightarrow \infty$

$$|\arg(z)| \leq \alpha \implies |T(z)| = O(|z|^\gamma).$$

Then, there exists a cone S_β where the following estimate holds for $|z| \rightarrow \infty$,

$$P(z) = T(z) + \frac{z}{2}T''(z) + O(z^{\gamma-2}).$$

4.6.2 The $\mathcal{JS}$ conditions.

This theorem is based itself on two previous results of the same authors. The first one describes two sufficient conditions (i) and (ii) sometimes called the $\mathcal{JS}$ conditions under which one can relate the asymptotic behaviours of P and T . The proof of these theorems makes a strong use of the Mellin transform.

Theorem 4.20 (Jacquet and Szpankowski, 1998). *Let $P(z)$ be the Poisson transform of a sequence B_n , namely*

$$P(z) = e^{-z}B(z), \quad \text{with} \quad B(z) = \sum_{n \geq 0} B_n \frac{z^n}{n!},$$

that is assumed to be an entire function of z . It is also assumed that there exists $\theta \in]0, \pi/2[$ for which the two following conditions simultaneously hold in a linear cone S_θ , for some real numbers $a, b, r > 0, \beta$, and $\delta < 1$:

- (i) For $z \in S_\theta$, $|z| > r \implies |P(z)| \leq b|z|^\beta$.
- (ii) For $z \notin S_\theta$, $|z| > r \implies |P(z)e^z| = |B(z)| \leq a \exp(\delta|z|)$.

Then, one has $B_n \sim P(n)$ for $n \rightarrow \infty$.

4.6.3 Analytic extension of the sequence B_n .

We now explain how it is possible to obtain sufficient conditions under which Conditions (i) and (ii) are fulfilled.

- Condition (i) is proven to hold with the use of Mellin transform. We study the asymptotics of $P(n)$ (for $n \rightarrow \infty$) with the Mellin transform P^* of the function $z \mapsto P(z)$.
- Condition (ii) is proven to hold as soon as the coefficients B_n satisfy the following: There exists a function $z \mapsto T(z)$ such that $T(n) = B_n$ which exists in a linear cone, is analytic there and is of polynomial growth when $|z| \rightarrow \infty$.

The following Theorem makes precise the previous sentence:

Theorem 4.21 (Jacquet and Szpankowski, 1999). *Let $T(z)$ be an analytic continuation of a sequence B_n which is $O(|z|^\gamma)$ in a linear cone. Then, for some θ_0 , and for all linear cones S_θ with $\theta < \theta_0$, there exist $\delta < 1$ and $a > 0$ such that the exponential generating function $B(z)$ of $T(n)$ satisfies*

$$z \notin S_\theta \implies |B(z)| \leq a \exp(\delta|z|).$$

Then, with the two previous theorems 4.20 and 4.21, the proof of Theorem 4.19 is obtained.

4.7 Laplace Transform

We use the Laplace transform in Chapter 6, and we recall here some simple properties which will be of interest for us. We follow here the handbook [3] Ch. 15, Sec. 15.2.

Definition 4.22. Let f be a function defined on $]0, +\infty[$. The Laplace transform of f (if it exists) is denoted by $\mathcal{L}[f]$, and defined by

$$\mathcal{L}[f](t) := \int_0^{\infty} e^{-tx} f(x) dx := F(t) \quad (4.28)$$

where t is a complex number.

Lemma 4.23. When the function f has the Laplace transform F , then the properties of Table 4.3 hold.

Given function g	its Laplace transform G
$f(\mu x)$	$\mu^{-1} F(t \mu^{-1})$
1	$1/t$
e^{-x}	$1/(t+1)$
x	$1/t^2$
$e^{-x\mu} - 1 + x\mu$	$\mu t^{-2} (1 + t\mu^{-1})^{-1}$

Table 4.3: Basic Properties of Laplace transform

Proof. The first three rules are clear. Rule 4 is obtained with integration by part. Rule 5 is deduced with the first four ones. Indeed, with $f(x) = e^{-x} - 1 + x$, one has with Rules 2, 3, 4

$$F(t) := \int_0^{+\infty} e^{-tx} [e^{-x} - 1 + x] = \frac{1}{1+t} - \frac{1}{t} + \frac{1}{t^2} = \frac{1}{t^2(1+t)} \quad (4.29)$$

With Rule 1, we deduce the result. $\square$

Rule 5 will be used in Chapter 6, in the proof of Proposition 6.4.

Conclusion of the Chapter.

We provide main tools which are needed in the following Chapters. We consider two probabilistic models, the Bernoulli and Poisson, where we are interested in the former model but it is easier to deal with the later model. To return from the Poisson to Bernoulli, there are two possibilities: either by Rice or DePoissonization. In this thesis, we will use the Rice's method which involved the "Poisson-Mellin-Newton-Rice cycle". We also state few DePoissonization's theorems.

The general approach which is described in this Chapter will be instantiated in the following three chapters. We now return to our problem : the study of the depth of digital trees built on a general source. In the next two chapters (Chapters 5 and 6), we perform a combinatorial and

algebraic analysis, and we are interested in obtaining an exact expression for some generating functions. The application of the method described here gives rise to exact expressions which are written as binomial sums and involve an analytic lifting, which is exactly “our” $\Delta(s, u)$ that has been already mentioned in Chapter 2 and defined in (2.16). Then, in the last Chapter 7, we perform the second part of the analysis, namely the analytic and asymptotic analysis, and we apply the second step of the Rice method in order to obtain asymptotic estimates that will be used in the probabilistic study. The tameness of the function $s \mapsto \Delta(s, u)$ will be essential and it strongly relies on the tameness of the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$, as it was defined in Definition 3.4 of Chapter 3.

Chapter 5

Profile and depth for simple sources – Algebraic analyses

Contents

5.1	Main objects of the algebraic study	94
5.1.1	Parameters of a digital tree.	94
5.1.2	Generating functions.	94
5.1.3	Outline of the method.	95
5.2	Functional equations for Poisson generating functions	96
5.2.1	Case of memoryless sources	96
5.2.2	Case of Markov chains	98
5.3	Analytic liftings via Mellin transforms of Poisson generating functions.	99
5.3.1	Memoryless case.	99
5.3.2	Markov Chains case.	100
5.4	Exact expressions of the expectations in the Bernoulli model	103
5.4.1	Expressions as binomial sums	103
5.4.2	Integral expressions for the probability generating functions.	103
5.4.3	An additive expression for $\Delta_D(s, u)$ in the unbiased memoryless case.	104
5.4.4	An exact expression for the profile in the unbiased memoryless case.	105

This is the first chapter of the thesis that is precisely devoted to the analysis of the depth of digital trees on a general source, and we will perform here the first part of the analysis, namely the algebraic part. The originality of our work relies on the fact that we deal with a *general* source. However, here, in this chapter, we only consider the case where the digital trees are built on simple sources. Of course, the whole analysis is now classical and very well-known¹, but we revisit this analysis with the architecture of the previous chapter, which gives rise to a simple process, completely transparent and with (very) few computations. Moreover, the analyses for the two digital trees (trie, dst) are completely parallel, and are performed together.

¹ At least the algebraic analysis, because, as we already mentioned, the analytic part of the analysis is not complete, even in the case of simple sources.

5.1 Main objects of the algebraic study

5.1.1 Parameters of a digital tree.

We first recall the main parameters of a digital tree, which have been previously defined in Chapter 1 (See Definition 1.5). In particular, the notion of full nodes is important to unify the two digital structures (trie and dst).

We are mainly interested in the profile and the (typical) depth. The profile of a digital tree (trie or dst) is defined by the sequence $b_{N,k}$ of random variables which count the number of full nodes at level k in a digital tree of size N . We consider the random variable

$$b_N(u) = \sum_{k \geq 0} b_{N,k} u^k, \quad (5.1)$$

and we are interested in its behaviour in the Bernoulli model $\mathcal{B}_n$ where the size N is fixed and equal to n . In particular, the expectation $B_n(u)$ of the variable $b_n(u)$ in the Bernoulli model is central in our study because it is closely related to the probability generating function to the depth D_n . We now recall this relation, already explained in Chapter 1.

For a tree of size n (i.e., with n full nodes), denote $d_{n,i}$ as the level of the full node containing (a part of) the i -th key. We observe the relation

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

where $\mathbb{I}[\cdot]$ denotes the Iverson bracket. Then, if $B_{n,k} := \mathbb{E}[b_{n,k}]$ is the average profile, the following equality holds,

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

The typical depth (called in the thesis 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, and it is closely related to the average profile.

5.1.2 Generating functions.

Consider the expectation $B_{n,k}$ of the variable $b_{n,k}$ in the Bernoulli model $\mathcal{B}_n$, and introduce the Poisson generating functions relative to the profile

$$P(z, u) = e^{-z} \sum_{n \geq 0, k \geq 0} B_{n,k} u^k \frac{z^n}{n!} = e^{-z} \sum_{n \geq 0} B_n(u) \frac{z^n}{n!}. \quad (5.2)$$

Then, from the previous Chapter 4, the series $P(z, u)$ is just the expectation of the variable $b_N(u)$ defined in (5.1) in the Poisson model $\mathcal{P}_z$, when N follows a Poisson law with parameter z , whereas $B_n(u)$ is the expectation of $b_N(u)$ in the Bernoulli model $\mathcal{B}_n$, when N is fixed and equal to n .

We have the following relations

$$B_n(u) = n\mathbb{E}[u^{D_n}], \quad B_n(1) = \sum_{k \geq 0} B_{n,k} = n, \quad B'_n(1) = n\mathbb{E}[D_n], \quad B(z, 1) = z. \quad (5.3)$$

We will deal with “underlined” versions of the generating functions, defined as

$$\underline{P}(z, u) := \frac{P(z, u) - z}{u - 1} = e^{-z} \sum_{n \geq 0} \underline{B}_n(u) \frac{z^n}{n!}, \quad (5.4)$$

with

$$\underline{B}_n(u) = \frac{B_n(u) - B_n(1)}{u - 1} = \frac{B_n(u) - n}{u - 1} = n \frac{\mathbb{E}[u^{D_n} - 1]}{u - 1}, \quad (5.5)$$

and (mainly implicitly) with its coefficients $\underline{P}_n(u)$ (which are polynomials with respect to u) defined as

$$\underline{P}(z, u) = \sum_{n \geq 0} (-1)^n \underline{P}_n(u) \frac{z^n}{n!}, \quad \underline{B}_n(u) = \sum_{k=2}^n (-1)^k \binom{n}{k} \underline{P}_k(u) \quad (5.6)$$

Valuation-Degree condition. We are interested in the underlined sequence $\underline{B}_n(u)$, because it fulfills the Valuation-Degree Condition defined in Chapter 4 (see Definition 4.9).

Lemma 5.1. *The two sequences $\underline{B}_n(u)$ (relative to tries and dst) satisfy the Valuation-Degree Condition for $|u| \leq 1$, with a valuation equal to 2 and a degree equal to 1. Furthermore, in the case of dst, the leading coefficient $\underline{P}_2(u) = \underline{B}_2(u)$ of $\underline{P}(z, u)$ equals 1.*

Proof. For the two digital trees (tries or dst), one has $B_0(u) = 0$ and $B_1(u) = 1$. Then the sequence $\underline{B}_n(u)$ are of valuation at least 2. Furthermore, for the dst, one has $B_2(u) = 1 + u$, so that $\underline{B}_2(u) = (1 + u - 2)/(u - 1) = 1$. More precisely, one has, in the trie case,

$$b_{1,0} = 1, \quad b_{N,0} = 0 \quad \text{for } N \neq 1, \quad b_{0,k} = 0 \quad \text{for any } k \geq 0 \quad \text{and} \quad b_{1,k} = 0 \quad \text{for } k \geq 1. \quad (5.7)$$

□

5.1.3 Outline of the method.

We recall that our final goal (in the algebraic step) is to obtain an expression for the probability generating function $\mathbb{E}[e^{uD_n}]$, itself closely related to the expectation $B_n(u)$ (in the Bernoulli model $\mathcal{B}_n$) of the variable $b_{N,u}$ defined in (5.1).

There are three main steps in the algebraic part of the analysis.

Step 1. The recursive definition of the digital tree is first and *directly* translated into a functional equation for the (normalized) Poisson generating functions $\underline{P}(z, u)$. We operate this transfer *in a direct way*, whereas usual approach first write ... a recurrence in ... the Bernoulli model.

Step 2. Our next aim is the study the Mellin transform $s \mapsto Z(s, u)$ of the function $z \mapsto \underline{P}(z, u)$ [defined in (5.6)]. Indeed, as the *Valuation-Degree Condition* is fulfilled, we know with the principles of the *Poisson-Mellin-Rice-Newton Cycle* recalled in Proposition 4.10 that the quotient

$$\Delta(s, u) := \frac{Z(-s, u)}{\Gamma(-s)}$$

will provide an analytic lifting of the sequence $n \mapsto \underline{P}_n(u)$ which is needed to apply the Rice methodology. We then transfer the functional equation which is satisfied by $\underline{P}(z, u)$ into a functional equation which is satisfied by $\Delta(s, u)$. This functional equation is easy to solve and we obtain an exact expression for the analytic lifting $\Delta(s, u)$.

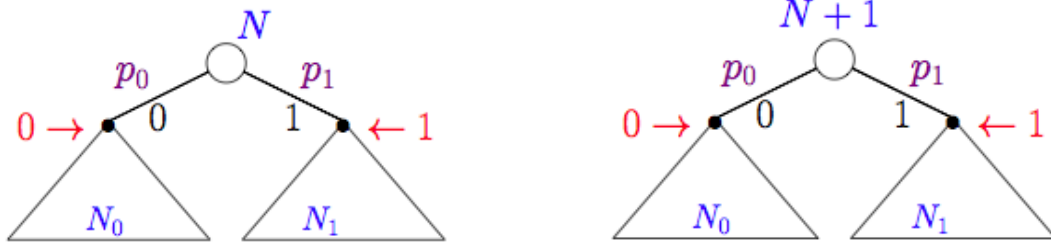


Figure 5.1: An explanation of the recurrences for tries and dst, with their similarities and their differences.

Step 3. We then deduce an exact expression for $\underline{B}_n(u)$, as a binomial sum of the form

$$\underline{B}_n(u) := \sum_{k=2}^n (-1)^k \binom{n}{k} \Delta(k, u)$$

which involves the values at integer $s = k$ of the analytic lifting $\Delta(s, u)$.

We also obtain an alternative expression as a Rice integral of the form

$$\underline{B}_n(u) = \frac{1}{2i\pi} \int_{d-i\infty}^{d+i\infty} \Delta(s, u) L_n(s) ds \quad \text{with} \quad L_n(s) := \frac{(-1)^n n!}{s(s-1)\dots(s-n)}$$

(with $d > 1$). This last expression will be the beginning step for the analytical part of our analysis, that will be performed in the following of this thesis. The analytic study is related to the (uniform) tameness of $s \mapsto \Delta(s, u)$

Now, in this chapter, we focus on the three algebraic steps. We summarize some original features of our work: we directly deal with the Poisson model, we deal with underlined versions $\underline{B}_n(u)$ for which the Valuation-Degree Condition is satisfied, and we explicitly use the cycle Poisson-Mellin-Rice-Newton. We also perform the two analyses (tries, dst) in a strongly parallel way.

5.2 Functional equations for Poisson generating functions

5.2.1 Case of memoryless sources

Proposition 5.2. [Memoryless] *Consider a memoryless source with probability (p_i) . Then the Poisson normalized bivariate generating function $\underline{P}(z, u)$ of the profile satisfies*

$$\begin{cases} \underline{P}(z, u) = z(1 - e^{-z}) + u \sum_{i \in \Sigma} \underline{P}(p_i z, u) & \text{[for trie]} \\ \frac{d}{dz} \underline{P}(z, u) + \underline{P}(z, u) = z + u \sum_{i \in \Sigma} \underline{P}(p_i z, u). & \text{[for dst]} \end{cases} \quad (5.8)$$

Proof. We recall that we directly deal with the Poisson model, first for tries, then for dst. Figure 5.1 can help to better understand the recurrences.

Case of trie. For a trie of size N , the sequence $b_{N,k}$ of the profile satisfies the basic recurrence,

$$b_{N,k} = \sum_{j \in \Sigma} b_{N_j, k-1} \quad \text{for } N \geq 2, k \geq 1, \quad (5.9)$$

where N_j is the number of nodes in the j -th subtree (This is also the number of words which begin with the symbol j). Taking into account the initial conditions described in (5.7), the recurrence formula for the variables $b_N(u)$ deals with the Iverson's bracket² of the boolean function ($N = 1$) and is:

$$b_N(u) = \llbracket N = 1 \rrbracket (u - 1) + u \sum_{j \in \Sigma} b_{N_j}(u),$$

Now, in the Poisson model $(\mathcal{P}_z, \mathcal{S})$, and with Lemma 4.3, each variable N_j follows a Poisson law of parameter $p_j z$, and

$$\mathbb{E}_z \llbracket N = 1 \rrbracket = \Pr_z[N = 1] = ze^{-z}.$$

This entails the relation

$$P(z, u) = ze^{-z}(1 - u) + u \sum_{j \in \Sigma} P(p_j z, u),$$

and finally the expression for the underlined series $\underline{P}(z, u)$ in the case of the trie.

Case of dst. For a dst of size N , the sequence $b_{N,k}$ of the profile satisfies the basic recurrence

$$b_{N+1,k} = \sum_{j \in \Sigma} b_{N_j, k-1} \quad \text{for } N \geq 0, k \geq 1, \quad (5.10)$$

where N_j is the number of nodes in the j -th subtree. This entails the following recurrence for the variables $b_N(u)$

$$b_{N+1}(u) = 1 + u \sum_{j \in \Sigma} b_{N_j}(u)$$

Now, in the model $(\mathcal{P}_z, \mathcal{S})$, Lemma 4.3 entails that each N_j follows a Poisson law of parameter $p_j z$, and the expectation in this Poisson model of the right-side is

$$1 + u \sum_{j \in \Sigma} P(p_j z, u)$$

Furthermore, Lemma 4.3 of the previous Chapter also shows that the expectation of the variable $b_{N+1}(u)$ is equal to

$$\frac{d}{dz} P(z, u) + P(z, u).$$

This entails the relation

$$\frac{d}{dz} P(z, u) + P(z, u) = 1 + u \sum_{j \in \Sigma} P(p_j z, u),$$

and finally the expression for the underlined series $\underline{P}(z, u)$ in the case of the dst. $\square$

Remark. The difference between the two recurrences (5.9) and (5.10) is due to the fact that the root of a dst contains a word, in contrast to the trie. This translation of 1 in (5.10) entails the existence of the derivative in the functional equation. This explains why the study of the dst will be more involved.

²For a boolean random variable X , the Iverson bracket $\llbracket X \rrbracket$ is a random variable of values in $\{0, 1\}$ which equals 1 if X is true and 0 if X is false.

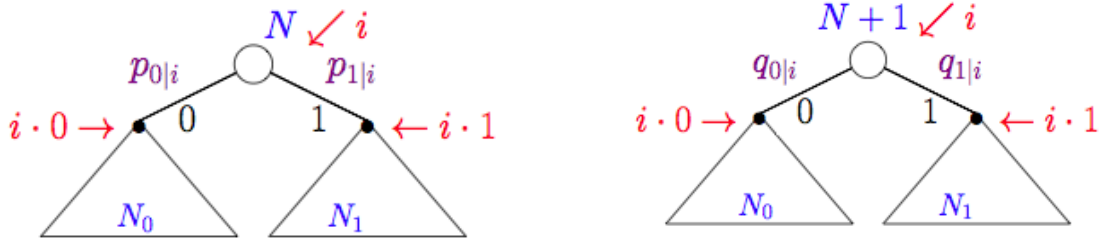


Figure 5.2: An explanation of the recurrences for tries and dst, with their similarities and their differences.

5.2.2 Case of Markov chains

Proposition 5.3. [Markov chains] *Consider a Markov chain with the transition matrix $\mathbf{R} := (p_{j|i})$ and an initial distribution $V = (v_i)$ also denoted by $(p_{i|\epsilon})$. For $i \in \Sigma \cup \{\epsilon\}$, consider its shifted source $\mathcal{S}_{(i)}$. Then the Poisson underlined generating functions $\underline{P}^{(i)}(z, u)$ relative to the source $\mathcal{S}_{(i)}$ satisfy a system of functional equations:*

$$\left\{ \begin{array}{ll} \underline{P}^{(i)}(z, u) = z(1 - e^{-z}) + u \sum_{j \in \Sigma} \underline{P}^{(j)}(p_{j|i}z, u) & \text{[for trie]} \\ \frac{d}{dz} \underline{P}^{(i)}(z, u) + \underline{P}^{(i)}(z, u) = z + u \sum_{j \in \Sigma} \underline{P}^{(j)}(p_{j|i}z, u) & \text{[for dst]} \end{array} \right. \quad (5.11)$$

Proof. In the same vein as previously, Figure 5.2 can help to better understand the recurrences.

For a Markov chain of order 1, denote by $\mathcal{S}_{(i)}$ the shifted source, in the sense of Chapter 2, namely the source formed with the words which begin with symbol i , from which i is removed, and by $\mathcal{S} = \mathcal{S}_{(\epsilon)}$ the initial source. We consider all the random variables that have been previously defined, but we now use the upper index (i) to refer to the source $\mathcal{S}_{(i)}$. Now, the sequences $b_{N,k}^{(i)}$ satisfy the following recurrences

$$\left\{ \begin{array}{ll} b_{N,k}^{(i)} = \sum_{j \in \Sigma} b_{N_j, k-1}^{(i,j)} & \text{for } N \geq 2, k \geq 1 \quad \text{[for trie]} \\ b_{N+1,k}^{(i)} = \sum_{j \in \Sigma} b_{N_j, k-1}^{(i,j)}, & \text{for } N \geq 0, k \geq 1 \quad \text{[for dst]} \end{array} \right. \quad (5.12)$$

which involve the sources $\mathcal{S}_{(i,j)}$. As the source $\mathcal{S}$ is a Markov chain of order 1, then two sources $\mathcal{S}_{(i,j)}$ and $\mathcal{S}_{(j)}$ are equivalent (in the sense of Chapter 2) for any $i \in \Sigma \cup \{\epsilon\}$. Furthermore, as previously, the variable N_j denotes the number of words which begin by j . But, now, the variable N_j has a distribution which depends on the initial source $\mathcal{S}_{(i)}$; namely, it follows a Poisson distribution of parameter $p_{j|i}z$. $\square$

Remark. In the memoryless case, the system of functional equations for dst involve three operations:

- (i) the differentiation d/dz with respect to z ,
- (ii) the change of variable $z \mapsto qz$ and

(iii) the conditional on the source $\mathcal{S}_{(i)}$.

In contrast, no differentiation occurs in the functional equations for tries which are thus simpler to deal with.

5.3 Analytic liftings via Mellin transforms of Poisson generating functions.

We now use the Mellin transform to obtain the analytic liftings $s \mapsto \Delta(s, u)$ of the sequence $n \mapsto \underline{P}_n(u)$ which are needed to use the Rice Formula.

We have already remarked that the Valuation-Degree Condition is fulfilled, and we apply the principles of the cycle Poisson-Mellin-Rice-Newton, described in Proposition 4.10. Then, these analytic liftings are given by the Mellin transform: if $Z(s, u)$ is the Mellin transform of the function $z \mapsto \underline{P}(z, u)$, we know that the analytic liftings of the sequence $n \mapsto \underline{P}_n(u)$ are equal to the ratios

$$\frac{Z(-s, u)}{\Gamma(-s)}$$

We will provide explicit expressions for this ratio in the two cases of interest (trie or dst).

5.3.1 Memoryless case.

Proposition 5.4. *Consider a memoryless source with probability (p_i) . Then the analytic liftings $s \mapsto \Delta(s, u)$ of the sequence $n \mapsto \underline{P}_n(u)$ defined in (5.6) are defined for $\Re s > 1$ and $|u| \leq 1$. They are expressed as quasi-inverses,*

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) = s(1 - u\lambda(s))^{-1} := s\Lambda(s, u) & \text{[for trie]} \\ \Delta_D(s, u) = Q(s, u)Q(2, u)^{-1} & \text{[for dst]} \end{cases}. \quad (5.13)$$

which involve the function $\lambda(s)$ and the infinite product $Q(s, u)$ defined as

$$\lambda(s) = \sum_{i \in \Sigma} p_i^s, \quad Q(s, u) := \prod_{j \geq 0} \frac{1}{1 - u\lambda(s + j)} \quad (5.14)$$

Let $p := \max p_i$. The infinite product $Q(s, u)$ defined in (5.14) is normally convergent on any subset of the form $\{(s, u); \Re s \geq 1 + a, |u| \leq (1/2)p^{-a}\}$ for any $a > 0$.

Proof. There are three main steps in the proof: we first deal with $Z(s, u)$, and we obtain in each case a functional equation, that we transfer as a functional equation on $\Delta(s, u)$. Then, we solve it, and we obtain the final expression for $\Delta(s, u)$.

Step 1. With Lemma 5.1, the Mellin transforms $Z(s, u)$ of the functions $z \mapsto \underline{P}(z, u)$ exist in the fundamental strip $\mathcal{B} := < -2, -1 >$ and satisfy the functional equations

$$\begin{cases} Z(s, u) = -\Gamma(s + 1) + u \sum_{j \in \Sigma} p_j^{-s} Z(s, u) = -\Gamma(s + 1) + u\lambda(-s)Z(s, u) & \text{[for trie]} \\ Z(s, u) - (s - 1)Z(s - 1, u) = u \sum_{j \in \Sigma} p_j^{-s} Z(s, u) = u\lambda(-s)Z(s, u) & \text{[for dst]} \end{cases}. \quad (5.15)$$

We use here the following two properties of the Mellin Transform given in Lemma 4.7,

$$\mathcal{M}[f(ax); s] = a^{-s} \mathcal{M}[f(x); s], \quad \mathcal{M}[f'(x); s] = -(s-1) \mathcal{M}[f(x); s-1].$$

Step 2. We now return to $\Delta(s, u)$ and use the identity $(s-1)\Gamma(s-1) = \Gamma(s)$. For the trie, we can directly solve the equation and obtain an expression which involves the Λ -series of Chapter 2, namely

$$\Delta(s, u) - u\lambda(s)\Delta(s, u) = -s \quad \text{or} \quad \Delta(s, u) = s(1 - u\lambda(s))^{-1} := s\Lambda(s, u).$$

For the dst, the analytic lifting satisfies the relations:

$$\Delta(s+1, u) = \Delta(s, u) - u\lambda(s)\Delta(s, u) \quad \text{or} \quad \Delta(s, u) = (1 - u\lambda(s))^{-1} \Delta(s+1, u)$$

and thus, for any $k \geq 1$,

$$\Delta(s, u) = (1 - u\lambda(s))^{-1} (1 - u\lambda(s+1))^{-1} \dots (1 - u\lambda(s+k-1))^{-1} \Delta(s+k, u)$$

The infinite product $Q(s, u)$ defined in (5.14) is convergent, as we will show in the following Lemma 5.5. Then, the sequence $k \mapsto \Delta(s+k, u)$ admits a limit $a(u)$ for $k \rightarrow \infty$ and the equality $\Delta(s, u) = Q(s, u)a(u)$ holds. It remains to evaluate $a(u)$.

Step 3. We now use Lemma 5.1, together with the fact that $\Delta(s, u)$ extends the sequence $\underline{P}_n(u)$. This entails the equality $\Delta(2, u) = 1$, and then the equality $a(u) = Q(2, u)^{-1}$. $\square$

Lemma 5.5. *The infinite product $Q(s, u)$ defined in (5.14) is normally convergent on the subset $\{(s, u); \Re s \geq 1+a, |u| \leq (1/2)p^{-a}\}$.*

Proof. The real $p := \max(p_i)$ satisfies $p < 1$, and we consider $\sigma := \Re s \geq 1+a$. Then, the inequality $\lambda(\sigma) \leq p^a$ holds, and one has

$$|\lambda(s)| \leq \lambda(\sigma) \leq p^a, \quad \text{and for any } k \geq 0, \quad |u\lambda(s+k)| \leq \lambda(\sigma)|u|p^k < |u|p^{a+k}.$$

Consider u and a that satisfy $|u|p^a \leq 1/2$. Thus, the inverses $(1 - u\lambda(s+k))^{-1}$ are well-defined for any $k \geq 0$, and

$$|(1 - u\lambda(s+k))^{-1}| \leq \frac{1}{1 - |u|p^{a+k}} = 1 + p^k \frac{|u|p^a}{1 - |u|p^{a+k}} \leq 1 + p^k,$$

Since the series of general term p^k is convergent, the infinite product $Q(s, u)$ is normally convergent and defines an analytic function on the domain $\{(s, u); \Re s \geq 1+a, |u| \leq (1/2)p^{-a}\}$ $\square$

5.3.2 Markov Chains case.

Now, we adopt a matricial approach, and we consider vectorial analytic liftings $s \mapsto \Delta(s, u)$ whose components are the analytic liftings $\Delta_{(i)}(s, u)$ of the functions $z \mapsto \underline{P}_n^{(i)}(u)$ (for $i \in \Sigma$). Remark that the vectorial analytic liftings $s \mapsto \Delta(s, u)$ of the vectorial functions $z \mapsto \underline{P}_n(u)$ whose components are the functions $z \mapsto \underline{P}_n(u)$.

Proposition 5.6. *Consider a good Markov chain with transition matrix $\mathbf{R} := (p_{j|i})$.*

Then the vectorial analytic liftings $s \mapsto \Delta(s, u)$ are defined for $\Re s > 1$ and $|u| \leq 1$ and are expressed as quasi-inverses which involve the matrix $\mathbf{R}_s := (p_{j|i}^s)$,

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) &= s(1 - u \mathbf{R}_s)^{-1} \mathbf{1}, & \text{[for trie]} \\ \Delta_D(s, u) &= \mathbf{Q}(s, u) \mathbf{Q}(2, u)^{-1} \mathbf{1}, & \text{[for dst]} \end{cases} \quad (5.16)$$

where the vector $\mathbf{1}$ has all its components equal to 1 and $\mathbf{Q}(s, u)$ is an infinite product of matrices defined as

$$\mathbf{Q}(s, u) := (1 - u\mathbf{R}_s)^{-1} \cdot (1 - u\mathbf{R}_{s+1})^{-1} \cdots (1 - u\mathbf{R}_{s+k})^{-1} \cdots \quad (5.17)$$

Denote by $\lambda(s)$ the dominant eigenvalue of $\mathbf{R}_s$, and let $\rho := \exp[-\lambda'(1)]$. Then, the infinite matrix product $\mathbf{Q}(s, u)$ defined in (5.17) is normally convergent on any subset of the form $\{(s, u); \Re s \geq 1 + a, |u| \leq (1/2)p^{-a}\}$ for any $a > 0$.

Proof. We adapt the previous three steps to the present case.

Step 1. For any $|u| \leq 1$ and $s \in \mathcal{B} :=]-2, -1[$, the Mellin transforms of conditional normalized profile $Z^{(i)}(s, u) = \mathcal{M}[z \mapsto \underline{P}^{(i)}(z, u); s]$ satisfy the equations

$$\begin{cases} Z^{(i)}(s, u) = -\Gamma(s+1) + u \sum_{j \in \Sigma} p_{j|i}^{-s} Z^{(j)}(s, u) & \text{[for trie]} \\ Z^{(i)}(s, u) - (s-1)Z^{(i)}(s-1, u) = u \sum_{j \in \Sigma} p_{j|i}^{-s} Z^{(j)}(s, u) & \text{[for dst]} \end{cases} \quad (5.18)$$

As previously, we use the following two properties of the Mellin Transform defined in (4.7)

$$\mathcal{M}[f(ax); s] = a^{-s} M[f(x); s] \quad \mathcal{M}[f'(x); s] = -(s-1)M[f(x); s-1].$$

Step 2. We now return to $\Delta^{(i)}(s, u)$, and we consider the vectorial functions $\Delta(s, u)$ whose components are the functions $\Delta^{(i)}(s, u)$ (for $i \in \Sigma$), each of them being defined on the subset $\{\Re s > 1, |u| \leq 1\}$. The previous system of equations is transferred into a vectorial equation which involves the matrix $\mathbf{R}_s$, and this gives, for each type of tree,

$$\begin{cases} \Delta(s, u) = s\mathbf{1} + u\mathbf{R}_s\Delta(s, u) & \text{or} & \Delta(s, u) = s(I - u\mathbf{R}_s)^{-1}\mathbf{1} \\ \Delta(s+1, u) = \Delta(s, u) - u\mathbf{R}_s\Delta(s, u) & \text{or} & \Delta(s, u) = (I - u\mathbf{R}_s)^{-1}\Delta(s+1, u) \end{cases},$$

where the vector $\mathbf{1}$ has all its components equal to 1. For tries, the expression given in the statement of Proposition is obtained. For the dst, the vectorial analytic lifting $\Delta(s, u)$ satisfies the relations, for any $k \geq 1$,

$$\Delta(s, u) = (1 - u\mathbf{R}_s)^{-1}(1 - u\mathbf{R}_{s+1})^{-1} \cdots (1 - u\mathbf{R}_{s+k-1})^{-1}\Delta(s+k, u)$$

The infinite product $\mathbf{Q}(s, u)$ defined in (5.17) is convergent, as it is proven in the following Lemma 5.7. Then, the sequence $k \mapsto \Delta(s+k, u)$ admits a limit $\mathbf{a}(u)$ for $k \rightarrow \infty$ and the equality $\Delta(s, u) = \mathbf{Q}(s, u)\mathbf{a}(u)$ holds. It remains to evaluate $\mathbf{a}(u)$.

Step 3. We now use Lemma 5.1, together with the fact that $\Delta(s, u)$ extends the sequence $\underline{P}_n(u)$. This entails the equality $\Delta(2, u) = \mathbf{1}$, and then the equality $\mathbf{a}(u) = \mathbf{Q}(2, u)^{-1}\mathbf{1}$. $\square$

Lemma 5.7. *Consider a Markov chain, moreover p -smooth. Then, the infinite matrix product $\mathbf{Q}(s, u)$ defined in (5.17) is normally convergent on any subset of the form $\{(s, u); \Re s \geq 1 + a, |u| \leq (1/2)p^{-a}\}$ for any $a > 0$.*

Proof. Define, for $\sigma \in \mathbb{R}$,

$$\mu(\sigma) := \sup \left\{ \sum_{i \in \Sigma} q_{i|w}^\sigma; w \in \Sigma^\star \right\}. \quad (5.19)$$

For $\sigma := \Re s \geq 1 + a$, the norm $\|\mathbf{P}_s\|$ satisfies,

$$\|\mathbf{P}_s\| \leq \mu(\sigma) \leq p^a, \quad \text{and, more generally,} \quad \|u\mathbf{P}_s\| \leq |u|p^a.$$

Assume now that the inequality $|u|p^a < 1/2$ holds. Then, one has $\|u\mathbf{P}_{s+k}\| \leq |u|p^a p^k < 1$, for any $k \geq 0$. The quasi-inverses $(I - u\mathbf{P}_{s+k})^{-1}$ are well-defined for any $k \geq 0$, and their norms satisfy

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

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 $\{\Re s \geq 1 + a\} \times \{|u| \leq (1/2)p^{-a}\}$. $\square$

We are now ready to perform the last step.

Proposition 5.8. *Consider a Markov chain with a transition matrix $\mathbf{R} := (p_{j|i})$ and an initial distribution (v_i) . For $\Re s > 1$ and $|u| \leq 1$, the initial liftings $\Delta(s, u) := \Delta^{(\epsilon)}(s, u)$ are expressed in terms of the matrix $\mathbf{R}_s$ and vector $V_s := (v_i^s)$ as*

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) &= s(1 + u {}^t V_s \cdot (I - u\mathbf{R}_s)^{-1} \cdot \mathbf{1}) = s\Lambda(s, u) & \text{[for trie]} \\ \Delta_D(s, u) &= 1 + u(A(s, u) - A(2, u)) & \text{[for dst]} \end{cases}$$

where $A(s, u) := \sum_{j \geq 0} {}^t V_{s+j} \mathbf{Q}(s+j, u) \mathbf{Q}(2, u)^{-1} \mathbf{1}$ (5.20)

is a series which involves the infinite product $\mathbf{Q}(s, u)$ defined in (5.17). The series $A(s, u)$ is convergent for $\Re s > 1$ and $|u| \leq 1$.

Proof. [For dst.] When iterating for $k \geq 1$ the equation

$$\Delta^{(\epsilon)}(s, u) - \Delta^{(\epsilon)}(s+1, u) = u {}^t V_s \Delta(s, u)$$

which involves the vectorial lifting $\Delta(s, u)$ defined in Equation (5.16), we obtain

$$\Delta^{(\epsilon)}(s, u) - \Delta^{(\epsilon)}(s+k, u) = u \sum_{j=0}^{k-1} {}^t V_{s+j} \Delta(s+j, u).$$

Now, for $k \rightarrow \infty$, the series in the right-side is exactly the series $uA(s, u)$ defined in (5.20) and, as previously, $\Delta^{(\epsilon)}(s+k, u)$ has a limit when $k \rightarrow \infty$ denoted by $a_{(\epsilon)}(u)$. The equality

$$\Delta^{(\epsilon)}(s, u) = a_{(\epsilon)}(u) + uA(s, u)$$

together with the equality $\Delta^{(\epsilon)}(2, u) = 1$ now entail the equality

$$\Delta^{(\epsilon)}(s, u) = 1 + u(A(s, u) - A(2, u)).$$

[For tries.] The identity of tries follows from the relation

$$\Delta^{(\epsilon)}(s, u) = s + u {}^t V_s \Delta(s, u) = s + u {}^t V_s s(I - u\mathbf{R}_s)^{-1} \mathbf{1}$$

$\square$

5.4 Exact expressions of the expectations in the Bernoulli model

We now provide exact expressions for the probability generating functions of profile and typical depth. We are also ready to use the Rice methodology as soon as the analytic liftings will be proven to be tame.

5.4.1 Expressions as binomial sums

Theorem 5.9. *For any simple source, and any of the two digital trees, there are exact expressions of the probability generating function $B_n(u)$ of the profile and the probability generating function $G_n(u)$ of the depth, with binomial sums of the form*

$$n \frac{G_n(u) - 1}{u - 1} = \frac{B_n(u) - n}{u - 1} = \sum_{k=2}^n (-1)^k \binom{n}{k} \Delta(k, u), \quad (5.21)$$

that involve the values at integers $s = k$ of the functions $s \mapsto \Delta(s, u)$ defined in Proposition 5.4 [for memoryless source] or in Proposition 5.6 [for Markov chains].

(a) In the memoryless case, the sequence $\Delta(k, u)$ is expressed with $\lambda(s) := \sum p_i^s$ as

$$\Delta(k, u) = \begin{cases} \Delta_T(k, u) &= \frac{k}{1 - u\lambda(k)} & \text{[for trie]} \\ \Delta_D(k, u) &= \prod_{j=2}^{k-1} (1 - u\lambda(j)) \quad (\text{for } k \geq 3), \quad \Delta(2, u) = 1 & \text{[for dst]} \end{cases}.$$

(b) In the Markov chains case, the sequence $\Delta(k, u)$ is expressed with the matrix $\mathbf{R}_s$ and vector $V_s = (v_i^s)$ as

$$\begin{cases} \Delta_T(k, u) &= k (1 + u {}^t V_k \cdot (1 - u \mathbf{R}_k)^{-1} \cdot \mathbf{1}) \\ \Delta_D(k, u) &= 1 + u \sum_{j=2}^{k-1} {}^t V_j \cdot \left[\prod_{i=1}^{j-2} (1 - u \mathbf{R}_i) \right] \cdot \mathbf{1} \quad (\text{for } k \geq 3), \quad \Delta(2, u) = 1 \end{cases}.$$

5.4.2 Integral expressions for the probability generating functions.

We have proven in Lemmas 5.5 and 5.7 that the Dirichlet series are bounded on subsets of the form $\{(s, u); \Re s \geq 1 + a, |u| \leq (1/2)\rho^{-a}\}$ with any $a > 0$. Then, we may apply the first step of the Rice method, and we obtain an integral form for the generating functions.

Theorem 5.10. *For any simple source, and any of the two digital trees, there are integral expressions of the probability generating function $B_n(u)$ of the profile and the probability generating function $G_n(u)$ of the depth,*

$$n \frac{G_n(u) - 1}{u - 1} = \frac{B_n(u) - n}{u - 1} = \frac{1}{2i\pi} \int_{d-i\infty}^{d+i\infty} \Delta(s, u) L_n(s) ds$$

(with $d > 1$) that involve the Rice kernel $L_n(s)$ and the functions $s \mapsto \Delta(s, u)$.

(a) In the memoryless case, the liftings $\Delta(s, u)$ are expressed in Proposition 5.4 and involve $\lambda(s) := \sum p_i^s$.

(b) In the Markov chains case, the liftings $\Delta(s, u)$ are expressed in Proposition 5.6 and involve the matrix $\mathbf{R}_s$ and vector $V_s = (v_i^s)$.

5.4.3 An additive expression for $\Delta_D(s, u)$ in the unbiased memoryless case.

In this case, there exists an alternative form for the Dirichlet series $\Delta_D(s, u)$ of dst, with an additive form, which is obtained via classical formula of q -analogs.

Proposition 5.11. *In the case of an unbiased memoryless source, the mixed Dirichlet series $\Delta_D(s, u)$ admits an additive expression*

$$\Delta_D(s, u) = \sum_{k \geq 0} u^k \delta_k(s) \quad \text{with} \quad \delta_k(s) = \sum_{i=0}^k a_i(s) b_{k-i}.$$

The sequences $a_\ell(s), b_\ell$ involve the finite products Q_i defined as

$$Q_i := (1 - r^{-1})(1 - r^{-2}) \dots (1 - r^{-i}) \quad \text{for } i \geq 1 \quad \text{and} \quad Q_0 = 1, \quad (5.22)$$

under the form

$$a_k(s) := \frac{1}{Q_k} r^{(1-s)k}, \quad b_k := \frac{(-1)^k}{Q_k} r^{-k(k+1)/2}.$$

Proof. In the unbiased memoryless case, the probabilities p_i are all equal to $q = 1/r$ (where r is the size of the alphabet) and the function $\lambda(s)$ is equal to $\lambda(s) = r \cdot r^{-s} = r^{1-s}$. Then, the following equalities hold

$$\lambda(s+j) = r^{1-s} \cdot r^{-j}, \quad 1 - u\lambda(s+j) = 1 - vq^j \quad \text{with} \quad q = r^{-1}, \quad v := ur^{1-s}$$

In this case, we can use two formulae that come from the theory of q -analogs, which transform infinite products into series,

$$\frac{1}{(1-v)(1-vq)(1-vq^2) \dots (1-vq^j) \dots} = \sum_{k \geq 0} \frac{v^k}{(1-q)(1-q^2) \dots (1-q^k)}, \quad (5.23)$$

$$(1-v)(1-vq)(1-vq^2) \dots (1-vq^j) \dots = \sum_{k \geq 0} \frac{(-1)^k v^k q^{k(k-1)/2}}{(1-q)(1-q^2) \dots (1-q^k)}. \quad (5.24)$$

The first one (5.23) will be used to transform $Q(s, u)$ and the second (5.24) will be used to transform $Q(2, u)^{-1}$. Using the sequence Q_i defined in (5.22) and with the first formula (5.23), we obtain

$$Q(s, u) = \prod_{j \geq 0} \frac{1}{1 - u\lambda(s+j)} = \sum_{k \geq 0} u^k \frac{r^{(1-s)k}}{(1-r^{-1})(1-r^{-2}) \dots (1-r^{-k})},$$

and $Q(s, u)$ is written as

$$Q(s, u) = \sum_{k \geq 0} a_k u^k, \quad \text{with} \quad a_k(s) := \frac{1}{Q_k} r^{(1-s)k}.$$

With the second formula 5.24,

$$Q(2, u)^{-1} = \prod_{j \geq 0} (1 - u\lambda(2+j)) = \sum_{k \geq 0} u^k \frac{(-1)^k}{Q_k} r^{-k(k+1)/2},$$

and $Q(2, u)^{-1}$ is written as

$$Q(2, u)^{-1} = \sum_{k \geq 0} b_k u^k \quad \text{with} \quad b_k := \frac{(-1)^k}{Q_k} r^{-k(k+1)/2}$$

Finally, the Dirichlet series $\Delta(s, u)$ is now written as a series

$$\Delta(s, u) = Q(s, u)Q(2, u)^{-1} = \sum_{k \geq 0} u^k \delta_k(s) \quad \text{with} \quad \delta_k(s) := \sum_{i=0}^k a_i(s) b_{k-i}.$$

□

5.4.4 An exact expression for the profile in the unbiased memoryless case.

This gives rise to an exact formula for the profile, already obtained by Louchard and Szpankowski in [48]. In the case of tries, the profile formula is

$$B_{n,k} = r^k \sum_{\ell=2}^n (-1)^\ell \ell \binom{n}{\ell} \frac{1}{r^{\ell k}} (r^{\ell-1} - 1)$$

Proposition 5.12. *For an unbiased memoryless source, there exists an exact expression for the dst profile which involves the quantities defined in Proposition 5.11 under the form*

$$B_{n,k} = \sum_{\ell=2}^n (-1)^\ell \binom{n}{\ell} (\delta_{k-1}(\ell) - \delta_k(\ell)). \quad (5.25)$$

Proof. Clear from Proposition 5.11 and (5.21) □

Remark. Such a formula is not known in the case when the dst is built on a simple source which is not a memoryless unbiased source. We return to this problem in the next chapter.

Conclusion of the Chapter.

We have revisited here the well-known approach which performs the combinatorial steps for the analysis of the profile and the depth in digital trees when they are built on simple sources. We better understand the structure of the approach and the main properties of the trees and of the source that are important in such an analysis. We are now ready to extend the approach when the digital trees are built on a general source.

Chapter 6

Profile and depth for a general source – Algebraic Analysis

Contents

6.1	General strategy	108
6.1.1	A sequence of shifted sources.	108
6.1.2	A sequence of generating functions.	108
6.1.3	General strategy for the algebraic study.	109
6.2	The basic recurrence and the system of functional equations.	110
6.3	The expression of the analytic lifting as a product.	112
6.4	Expression of the analytic lifting as a series.	114
6.5	Final expressions in the Bernoulli model.	117
6.5.1	Expression as binomial sums.	117
6.5.2	Expression as Rice's integrals.	118
6.5.3	Expression of the sequence $B_{n,k}$	118
6.5.4	Towards to the next chapter	119

We now consider the original framework of this thesis, namely the study of digital trees (tries or dst) when they are built on words which are independently emitted by a *general source*. We first use the (recursive) definition of the digital trees, as given in Chapter 1, and already used in a strong way in Chapter 5. But the new actor is now the general source, and we consider here the modelisation of a source via its generalized transition matrix $\mathbf{P}$, as we define it in Section 2.3.1. We are led to assume a supplementary hypothesis for the source $\mathcal{S}$ and its matrix $\mathbf{P}$, namely its smoothness¹:

Definition 6.1. *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 $(i, w) \in \Sigma \times \Sigma^*$.*

We will see in this Chapter that the approach described in the previous Chapter admits a clean extension to any general (smooth) source. This is why the present Chapter has exactly the same

¹This hypothesis is probably too strong, and it is probably sufficient to assume that the matrix $\mathbf{P}$ be irreducible and aperiodic. But, we are not specialists for Markov chain with an infinite number of states, and, anyway, for the analytical part of the study, we will be led to stronger hypotheses.

structure as Chapter 5. We begin the study in the Poisson model, where the recursive definition of digital trees gives rise to a system of functional equations satisfied by the Poisson generating functions. Then, we wish to return to the Bernoulli model, with the Rice methodology. We first need to get a nice expression of the analytic lifting. As previously, the system in the Poisson model transfers as a system satisfied by the analytic liftings, and we obtain a close expression for these analytic liftings $\Delta(s, u)$ in a multiplicative way.

However, we are also interested in expressions of another type, with an additive structure. This additive point of view will be of particular interest if we wish to obtain an exact expression of the sequence $B_{n,k}$. In the last chapter, we have remarked that such an exact expression is not known, except in a very particular case when the source is an unbiased memoryless source. And, in this case, we use a formula that comes from the theory of q -analogs, and seems to be completely *ad hoc* for this case. In order to obtain the expression of $\Delta(s, u)$ as a series, we will use a tool that has not yet appeared in this thesis, the Laplace transform, just described at the end of Chapter 4, but not yet used. The formula which will be obtained can be viewed as an extension of the formula (5.23), and seems not known, even for a biased memoryless source.

This chapter is also based on the same original features that have been already used in the previous Chapter: we *directly* deal with the Poisson model, we deal with underlined versions $\underline{B}_n(u)$ for which the *Valuation-Degree* Condition is satisfied, and we explicitly use the *Poisson-Mellin-Rice-Newton* cycle. We also perform the two analyses (tries, dst) in a *strongly parallel* way.

6.1 General strategy

6.1.1 A sequence of shifted sources.

We consider a given general source $\mathcal{S}$ (only assumed to be smooth). We are interested in the depth of digital trees (tries or dst) built on the source $\mathcal{S}$, and we also deal with the profile of these digital trees. Thus, we introduce the random variables

$$b_N(u) := \sum_{k \geq 0} u^k b_{N,k},$$

where the variable $b_{N,k}$ is the number of full nodes at level k in a digital tree (trie or dst) of size N built on the source $\mathcal{S}$.

The main idea here is to consider *together* all the shifted sources $\mathcal{S}_{(w)}$ and the random variables

$$b_N^{(w)}(u) := \sum_{k \geq 0} u^k b_{N,k}^{(w)}$$

where the variable $b_{N,k}^{(w)}$ is the number of full nodes at level k in a digital tree (trie or dst) of size N built on the source $\mathcal{S}_{(w)}$. We will study the system of these shifted sources. Then, we will return to our initial source.

6.1.2 A sequence of generating functions.

We wish to study the profile $b_N^{(w)}(u)$ in the Bernoulli model $(\mathcal{B}_n, \mathcal{S}_{(w)})$, but we know that the study is easier in the Poisson model. We then introduce the Poisson generating functions

$$P^{(w)}(z, u) := e^z \sum_{n \geq 0} B_n^{(w)}(u) \frac{z^n}{n!},$$

which are the expectations of the variable $b_N^{(w)}(u)$ in the Poisson model $(\mathcal{P}_z, \mathcal{S}_{(w)})$. As in the previous chapter, we deal with the underlined versions,

$$\underline{P}^{(w)}(z, u) := \frac{P^{(w)}(z, u) - z}{u - 1} = e^{-z} \sum_{n \geq 0} \underline{B}_n^{(w)}(u) \frac{z^n}{n!} \quad (6.1)$$

with

$$\underline{B}_n^{(w)}(u) = \frac{B_n^{(w)}(u) - B_n(1)}{u - 1} = \frac{B_n^{(w)}(u) - n}{u - 1} \quad (6.2)$$

The coefficients $\underline{P}_n(u)$ of the underlined Poisson generating function are defined as

$$\underline{P}^{(w)}(z, u) = \sum_{n \geq 0} (-1)^n \underline{P}_n^{(w)}(u) \frac{z^n}{n!}, \quad \underline{B}_n^{(w)}(u) = \sum_{k=2}^n (-1)^k \binom{n}{k} \underline{P}_k^{(w)}(u) \quad (6.3)$$

As previously, it will be easy to return to the probability generating function of the depth (underlined and non underlined), due to the equalities

$$\underline{G}_n^{(w)}(u) := \frac{G_n^{(w)}(u) - 1}{u - 1} = \frac{1}{n} \frac{B_n^{(w)}(u) - n}{u - 1} = \frac{1}{n} \underline{B}_n^{(w)}(u).$$

Lemma 6.2. *For any $w \in \Sigma^*$, the two sequences $\underline{B}_n^{(w)}(u)$ (relative to tries and dst) satisfy the Valuation-Degree Condition for $|u| \leq 1$, with a valuation equal to 2 and a degree equal to 1. Furthermore, for dst's, the leading coefficient $\underline{P}_2^{(w)}(u) = \underline{B}_2^{(w)}(u)$ of $\underline{P}^{(w)}(z, u)$ equals 1.*

We wish to obtain an alternative expression of the generating functions $\underline{P}^{(w)}(z, u)$ of the profile. We first derive in Proposition 6.3 a system of functional equations. Next, we perform an algebraic study, which first provides in Proposition 6.6 an exact expression for the probability generating function of the depth and introduces a central object in our study, the mixed Dirichlet series $\Delta(s, u)$ (which mixes properties of the source and properties of the tree) for which we obtain an alternative expression in Proposition 6.4. Then, it will be possible, in the next chapter, to perform the asymptotic analysis.

6.1.3 General strategy for the algebraic study.

Our method is composed of five main steps, each dedicated to the use of one of the three main tools.

- (a) We deal with the Poisson generating functions $\underline{P}^{(w)}(z, u)$ and we first obtain a system of functional equations [Proposition 6.3]. This system of functional equations involves, in the case of dst, the differentiation d/dz with respect to z and this creates one of the main difficulties in the dst analysis. There are two main transforms with which the differentiation “disappears”: the Laplace transform and the Mellin transform. The Mellin transform was used in the previous chapter, but not the Laplace transform.
- (b) We use the cycle Poisson-Mellin-Newton-Rice, and we obtain a first expression of the analytic lifting $\Delta(s, u)$ as an infinite product [Proposition 6.4]. This is an analysis which extends the analysis of the previous chapter.
- (c) We also use the Laplace transform and we obtain a second expression of the analytic lifting $\Delta(s, u)$ as a series [Proposition 6.6]. This is a new point of view which was not explicitly used before in such contexts.

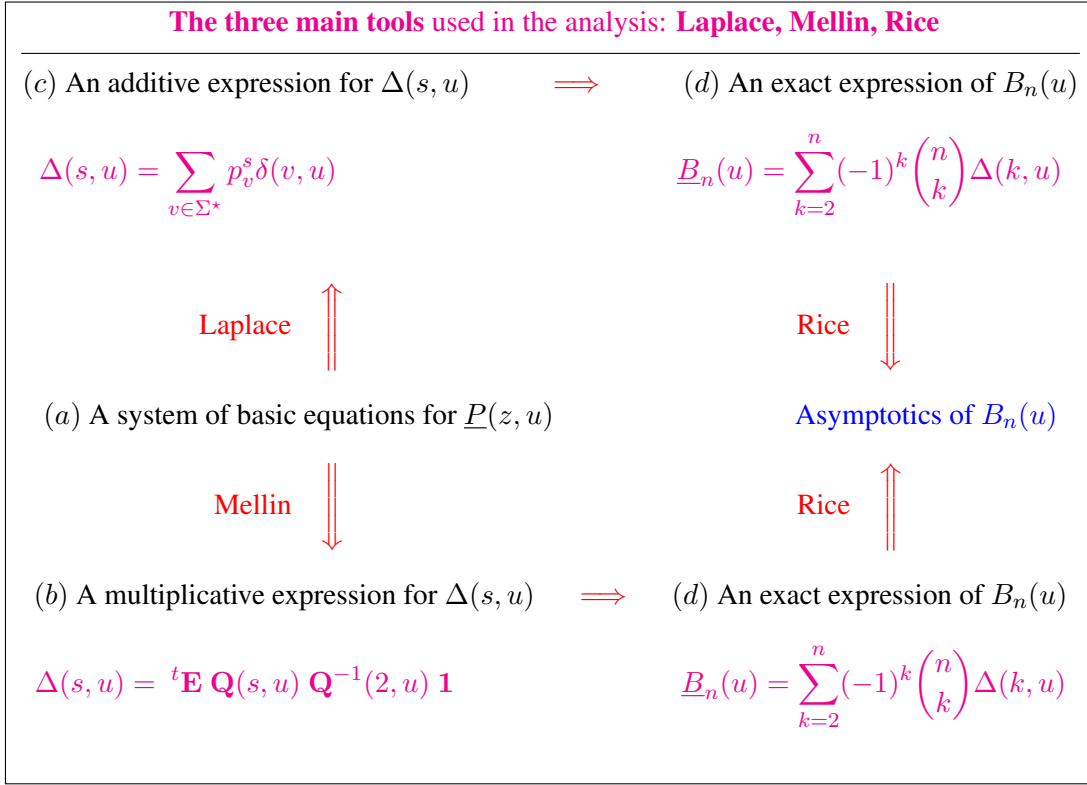


Figure 6.1: Strategy for the algebraic study.

- (d) We obtain exact expressions for the probability generating functions ($B_n(u)$ for the profile, and $G_n(u)$ for the depth) under the form of binomial sums which involve the sequence $\Delta(k, u)$.
- (e) As the analytic lifting $s \mapsto \Delta(s, u)$ is of polynomial growth² on the halfplane $\Re s \geq 1 + a$ with any $a > 0$, we perform the first step of the Rice method (shifting to the right) and the conclusion of our algebraic analysis is the expression of the probability generating functions as an integral over a vertical line of the complex plane.

In the next Chapter, we aim to perform the shifting to the left of this integral– and it would be possible if we have a precise knowledge of the operator $\mathbf{P}_s$. This does not seem possible for a general source. This is why we consider in the next chapter dynamical sources, study the transfer operator $\mathbb{H}_s$ instead of the transition matrix $\mathbf{P}_s$, and notably their tameness, as it is defined in Chapter 3.

Now, in this chapter, we focus on the five algebraic steps described in Figure 6.1

6.2 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 $\mathcal{S}$.

²It is actually of bounded growth

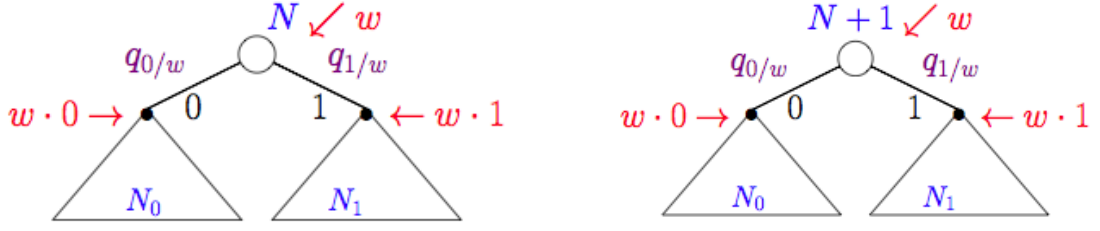


Figure 6.2: An explanation of the recurrences for tries and dst, with their similarities and their differences.

Proposition 6.3. Consider a source $\mathcal{S}$ and its shifted sources $\mathcal{S}_{(w)}$. Then, the underlined Poisson generating functions $\underline{P}^{(w)}(z, u)$ of the profiles relative to the shifted sources $\mathcal{S}_{(w)}$ are solutions of the system of functional equations

$$\begin{cases} \underline{P}^{(w)}(z, u) = z(1 - e^{-z}) + u \sum_{i \in \Sigma} \underline{P}^{(w \cdot i)}(q_{i|w} z, u) & \text{[for trie]} \\ \underline{P}^{(w)}(z, u) + \frac{d}{dz} \underline{P}^{(w)}(z, u) = z + u \sum_{i \in \Sigma} \underline{P}^{(w \cdot i)}(q_{i|w} z, u) & \text{[for dst]} \end{cases} \quad (6.4)$$

Proof. We directly deal with the Poisson model, first for tries, then for dst. Figure 6.2 may help to understand the recurrences, in each of the two cases.

Case of trie. For a trie of size N , the sequence $b_{N,k}^{(w)}$ of the profile satisfies the basic recurrence, which is explained in Figure 6.2 (on the left).

$$b_{N,k}^{(w)} = \sum_{j \in \Sigma} b_{N_j, k-1}^{(w \cdot j)} \quad \text{for } N \geq 2, k \geq 1, \quad (6.5)$$

where N_j is the number of nodes in the j -th subtree (This is also the number of words of the sequence which begin with the symbol j). Then the recurrence formula for the variables $b_N(u)$ is

$$b_N^{(w)}(u) = \llbracket N = 1 \rrbracket (u - 1) + u \sum_{j \in \Sigma} b_{N_j}^{(w \cdot j)}(u),$$

where $\llbracket \cdot \rrbracket$ is the Iverson's bracket. Now, in the Poisson model $(\mathcal{P}_z, \mathcal{S}_{(w)})$, Lemma 4.3 entails that each variable N_j follows a Poisson law of parameter $q_{j|w} z$, and

$$\mathbb{E}_z \llbracket N = 1 \rrbracket = \Pr_z[N = 1] = ze^{-z}.$$

This entails the relation

$$P^{(w)}(z, u) = ze^{-z}(1 - u) + u \sum_{j \in \Sigma} P^{(w \cdot j)}(q_{j|w} z, u),$$

and finally the expression for the underlined series $\underline{P}^{(w)}(z, u)$ in the case of the trie.

Case of dst. For a dst of size N , the sequence $b_{N,k}$ of the profile satisfies the basic recurrence which is explained in Figure 6.2 (on the right).

$$b_{N+1,k}^{(w)} = \sum_{j \in \Sigma} b_{N_j, k-1}^{(w \cdot j)} \quad \text{for } N \geq 0, k \geq 1, \quad (6.6)$$

where N_j is the number of nodes in the j -th subtree. This entails the following recurrence for the variables $b_N(u)$

$$b_{N+1}^{(w)}(u) = 1 + u \sum_{j \in \Sigma} b_{N_j}^{(w \cdot j)}(u)$$

Now, in the Poisson model $(\mathcal{P}_z, \mathcal{S}_{(w)})$, with Lemma 4.3, each N_j follows a Poisson law of parameter $q_{j|w} z$, and the expectation in the Poisson model $(\mathcal{P}_z, \mathcal{S}_{(w)})$ of the right-side is

$$1 + u \sum_{j \in \Sigma} P^{(w)}(q_{j|w} z, u)$$

Furthermore, Lemma 4.3 entails that the expectation of the variable $b_{N+1}(u)$ in the model $(\mathcal{P}_z, \mathcal{S}_{(w)})$ is equal to

$$\frac{d}{dz} P^{(w)}(z, u) + P^{(w)}(z, u).$$

This entails the relation

$$\frac{d}{dz} P^{(w)}(z, u) + P^{(w)}(z, u) = 1 + u \sum_{j \in \Sigma} P^{(w \cdot j)}(q_{j|w} z, u),$$

and finally the expression for the underlined series $\underline{P}^{(w)}(z, u)$ in the case of the dst. $\square$

Remark. As in the previous chapter, the system of functional equations for dst involve three operations:

- (i) the differentiation d/dz with respect to z ,
- (ii) the change of variable $z \mapsto qz$ and
- (iii) the conditional on the source $\mathcal{S}_{(w)}$.

In contrast, there is no derivation occurs in the functional equations for tries which are thus simpler to deal with.

6.3 The expression of the analytic lifting as a product.

We extend here the approach of [37]. We have already remarked in Lemma 6.2 that the Valuation-Degree Condition is fulfilled for any source; and in particular, any shifted source $\mathcal{S}_{(w)}$, and we apply the principles of the cycle Poisson-Mellin-Rice-Newton, described in Proposition 4.10. Then, the analytic liftings are given by the Mellin transform: if $Z^{(w)}(s, u)$ is the Mellin transform of the function $z \mapsto \underline{P}^{(w)}(z, u)$, we know that the analytic liftings of the sequence $n \mapsto \underline{P}_n^{(w)}(u)$ are equal to the ratios

$$\frac{Z^{(w)}(-s, u)}{\Gamma(-s)}.$$

Proposition 6.4. Consider a p -smooth source $\mathcal{S}$ and its generalized matrix $\mathbf{P}_s$. Then, the analytic lifting $\Delta(s, u)$ of the sequence $\underline{P}_n(u)$ admits the following expression

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) &= s {}^t\mathbf{E} (I - u\mathbf{P}_s)^{-1} \mathbf{1}, & \text{[for trie]} \\ \Delta_D(s, u) &= {}^t\mathbf{E} \mathbf{Q}(s, u) \cdot \mathbf{Q}(2, u)^{-1} \mathbf{1} & \text{[for dst]} \end{cases}. \quad (6.7)$$

where $\mathbf{1}$ is the vector (indexed with Σ^*) whose all the components equal 1, $\mathbf{E}$ is the vector (indexed with Σ^*) whose all the components equal 0 except the one relative to ϵ which equals 1. Moreover, $\mathbf{Q}(s, u)$ is an infinite product of operators

$$\mathbf{Q}(s, u) := (I - u\mathbf{P}_s)^{-1} \cdot \dots \cdot (I - u\mathbf{P}_{s+k})^{-1} \cdot \dots \quad (6.8)$$

For a p smooth source, the infinite product $\mathbf{Q}(s, u)$, is normally convergent on the Cartesian product $\{\Re s \geq 1 + a\} \times \{|u| \leq (1/2)p^{-a}\}$.

Proof. There are three main steps in the proof: we first deal with $Z^{(w)}(s, u)$ and we obtain a system of functional equations that we transfer into a system of functional equations for $\Delta^{(w)}(s, u)$. Then, we solve it and we obtain the final expression for $\Delta(s, u)$.

Step 1. With Lemma 6.2 and Proposition 4.10, the Mellin transforms $Z^{(w)}(s, u)$ of the functions $z \mapsto \underline{P}^{(w)}(z, u)$ exist in the fundamental strip $\mathcal{B} := \langle -2, -1 \rangle$ and satisfy the functional equations

$$\begin{cases} Z^{(w)}(s, u) = -\Gamma(s+1) + u \sum_{j \in \Sigma} q_{j|w}^{-s} Z^{(w \cdot j)}(s, u) & \text{[for trie]} \\ Z^{(w)}(s, u) - (s-1)Z^{(w)}(s-1, u) = u \sum_{j \in \Sigma} q_{j|w}^{-s} Z^{(w \cdot j)}(s, u) & \text{[for dst]} \end{cases} \quad (6.9)$$

We have used here the two properties of the Mellin Transform described in Lemma 4.8

$$\mathcal{M}[f(ax); s] = a^{-s} \mathcal{M}[f(x); s], \quad \mathcal{M}[f'(x); s] = -(s-1) \mathcal{M}[f(x); s-1]$$

Step 2. We now return to $\Delta^{(w)}(s, u)$ and we consider the vectorial functions $\Delta(s, u)$ whose components are the functions $\Delta^{(w)}(s, u)$ (for $w \in \Sigma^*$), each of them being defined on the subset $\{\Re s > 1, |u| \leq 1\}$. The previous system of equations is transferred into a vectorial equation which involves the generalized transition matrix $\mathbf{P}_s$ and we obtain:

$$\begin{cases} \Delta(s, u) = s\mathbf{1} + u\mathbf{P}_s\Delta(s, u) & \text{or} \quad \Delta(s, u) = s(I - u\mathbf{P}_s)^{-1}\mathbf{1} \\ \Delta(s+1, u) = \Delta(s, u) - u\mathbf{P}_s\Delta(s, u) & \text{or} \quad \Delta(s, u) = (I - u\mathbf{P}_s)^{-1}\Delta(s+1, u) \end{cases}.$$

where the vector $\mathbf{1}$ has all its components equal to 1. For tries, the expression given in the statement of Proposition is obtained. For the dst, the vectorial analytic lifting $\Delta(s, u)$ satisfies the relations, for any $k \geq 1$,

$$\Delta(s, u) = (1 - u\mathbf{P}_s)^{-1}(1 - u\mathbf{P}_{s+1})^{-1} \dots (1 - u\mathbf{P}_{s+k-1})^{-1}\Delta(s+k, u)$$

The infinite product $\mathbf{Q}(s, u)$ defined in (6.8) is convergent, as it is proven in the following Lemma 6.5. Then, the sequence $k \mapsto \Delta(s+k, u)$ admits a limit $\mathbf{a}(u)$ for $k \rightarrow \infty$ and the equality $\Delta(s, u) = \mathbf{Q}(s, u)\mathbf{a}(u)$ holds. It remains to evaluate $\mathbf{a}(u)$.

Step 3. We now use Lemma 5.1, together with the fact that $\Delta^{(w)}(s, u)$ extends the sequence $\underline{P}_n^{(w)}(u)$. This entails the equality $\Delta(2, u) = \mathbf{1}$, and then the equality $\mathbf{a}(u) = \mathbf{Q}(2, u)^{-1}\mathbf{1}$. We now focus on the first component $\Delta(s, u) := \Delta_{(\epsilon)}(s, u)$ of the vector $\Delta(s, u)$, and we derive an exact expression for the analytic lifting of the sequence $\underline{P}_n(u)$ relative to the initial source $\mathcal{S} := \mathcal{S}_{(\epsilon)}$ $\square$

We now prove that the infinite product is convergent.

Lemma 6.5. Denote by $\mathcal{B}(\Sigma^*)$ the Banach space formed with the bounded complex functions $\Sigma^* \rightarrow \mathbb{C}$, endowed with the norm $\|X\| := \sup |X_{(w)}|$. Then, for a p smooth source, the infinite products $\mathbf{Q}(s, u)$, is normally convergent on the Cartesian product $\{\Re s \geq 1 + a\} \times \{|u| \leq (1/2)p^{-a}\}$.

Proof. Define, for $\sigma \in \mathbb{R}$,

$$\mu(\sigma) := \sup \left\{ \sum_{i \in \Sigma} q_i^\sigma; w \in \Sigma^* \right\}. \quad (6.10)$$

For $\sigma := \Re s \geq 1 + a$, the norm $\|\mathbf{P}_s\|$ satisfies,

$$\|\mathbf{P}_s\| \leq \mu(\sigma) \leq p^a, \quad \text{and, more generally,} \quad \|u\mathbf{P}_s\| \leq |u|p^a.$$

Assume now that the inequality $|u|p^a < 1$ holds. Then, for any $k \geq 0$, one has $\|u\mathbf{P}_{s+k}\| \leq |u|p^a p^k < p^k$. Then the quasi-inverses $(I - u\mathbf{P}_{s+k})^{-1}$ are well-defined for any $k \geq 0$, and their norms satisfy

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

We now consider the case when $|u|p^a < 1/2$, then $|u|p^a p^k < 1/2$ for any $k \geq 0$, and

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

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 $\{\Re s \geq 1 + a\} \times \{|u| \leq (1/2)p^{-a}\}$. $\square$

6.4 Expression of the analytic lifting as a series.

The use of the Laplace transform is not very usual in the digital trees analyses. There are yet some instances, in particular the papers [21] or [41].

Proposition 6.6. Let $\mathcal{S}$ be a p -smooth source with fundamental probabilities p_w . In the case of the dst, the analytic lifting $\Delta_{(w)}(s, u)$ of the modified sequence $\underline{P}_n(u)$ admits an alternative form as a series

$$\Delta(s, u) := \sum_{v \in \Sigma^*} \delta(v, u) p_v^s, \quad \text{with} \quad \delta(v, u) := \frac{1}{p_v} \sum_{w \geq v} u^{|w|} p_w \prod_{\substack{\alpha \in [\epsilon, w] \\ \alpha \neq v}} \frac{1}{1 - p_v p_\alpha^{-1}}, \quad (6.11)$$

which exists for $\Re s > 1$ and $|u| \leq 1$. This is a power series with respect to u , and

$$\Delta_k(s) := [u^k] \Delta(s, u) = \sum_{w \in \Sigma^k} p_w \sum_{v \leq w} p_v^{s-1} \prod_{\substack{\alpha \in [\epsilon, w] \\ \alpha \neq v}} \frac{1}{1 - p_v p_\alpha^{-1}},$$

Proof. We consider the Laplace transform $\mathcal{L}$, which transforms $\underline{P}^{(w)}(z, u)$ into $C^{(w)}(t, u)$, defined as

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

With Proposition 6.3 and main properties of the Laplace transform recalled in Lemma 4.3, the generating functions $\widehat{C}^{(w)}(t, u) := t^2 C^{(w)}(t, u)$ satisfy the system of functional equations

$$(t+1)\widehat{C}^{(w)}(t, u) = 1 + u \sum_{i \in \Sigma} \frac{1}{q_{i|w}} \widehat{C}^{(w \cdot i)}\left(\frac{t}{q_{i|w}}, u\right). \quad (6.12)$$

We first focus on $\widehat{C}(t, u)$ relative to the initial source $\mathcal{S} := \mathcal{S}_{(\epsilon)}$. Iterating Relation (6.12), and using the multiplicative property of conditional probabilities, one obtains

$$\widehat{C}(t, u) = \sum_{w \in \Sigma^*} u^{|w|} p_w \prod_{v \leq w} \frac{1}{1 + t p_v^{-1}} \quad (6.13)$$

We recall that we will transform products of rational fractions into sums, and we simply use decomposition into partial fractions. We obtain

$$\prod_{v \leq w} \frac{1}{1 + t p_v^{-1}} = \sum_{v \leq w} \frac{r(v, w)}{1 + t p_v^{-1}}, \quad \text{with } r(v, w) := \prod_{\alpha \in [\epsilon, w] \setminus \{v\}} \frac{1}{1 - p_v p_\alpha^{-1}}. \quad (6.14)$$

We set

$$\delta(v, u) := \frac{1}{p_v} \sum_{w \geq v} r(v, w) u^{|w|} p_w \quad (6.15)$$

We will see later in Lemma 6.7 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 $\widehat{C}(t, u)$,

$$\widehat{C}(t, u) = \sum_{v \in \Sigma^*} \delta(v, u) \frac{p_v}{1 + t p_v^{-1}}, \quad (6.16)$$

and we now return to $C(t, u)$,

$$C(t, u) = \frac{1}{t^2} \widehat{C}(t, u) = \sum_{v \in \Sigma^*} \delta(v, u) \frac{1}{t^2} \frac{p_v}{1 + t p_v^{-1}}. \quad (6.17)$$

We now remark the equality (Lemma 4.3)

$$\mathcal{L}[e^{-x p_v} - 1 + x p_v](t) = \frac{p_v}{t^2(1 + t p_v^{-1})},$$

We now apply the inverse Laplace transform on both sides of (6.16) to recover the Poisson generating function $\underline{P}(z, u)$ relative to the source $\mathcal{S} = \mathcal{S}_\epsilon$,

$$\underline{P}(z, u) = \sum_{v \in \Sigma^*} \delta(v, u) [e^{-z p_v} - 1 + z p_v]. \quad (6.18)$$

This generating function is a harmonic sum, and then its Mellin transform equals

$$Z(s, u) = \Gamma(s) \sum_{v \in \Sigma^*} p_v^{-s} \delta(v, u), \quad \text{and then} \quad \Delta(s, u) = \sum_{v \in \Sigma^*} p_v^s \delta(v, u)$$

which provides the result. $\square$

Remark. We remark the following relations

$$\text{for } \alpha < v, \quad p_v p_\alpha^{-1} = p_{v|\alpha}, \quad \text{for } \alpha > v, \quad p_v p_\alpha^{-1} = p_{\alpha|v}^{-1}$$

Then, $r(v, w)$ is a product of quasi-inverses of conditional probabilities and decomposes into two factors, according to the position of prefix α with respect to v , and

$$r(v, w) := \beta(v) \cdot \gamma(v, w) \quad \text{with} \quad \beta(v) := \left(\prod_{\alpha \in [\epsilon, v[} \frac{1}{1 - p_{v|\alpha}} \right), \quad (6.19)$$

$$\gamma(v, w) := (-1)^{|w|-|v|} \left(\prod_{\alpha \in]v, w]} \frac{p_{\alpha|v}}{1 - p_{\alpha|v}} \right) \quad (\text{for } v < w), \quad \text{and} \quad \gamma(w, w) = 1.$$

We then prove that the previous series is convergent

Lemma 6.7. *If the source $\mathcal{S}$ is p -smooth, the Dirichlet series defined in (6.11) is absolutely convergent for $\Re s > 1$ and $|u| \leq 1$.*

Proof. Define $\mu(\sigma)$ as in (6.10), and remind that $\mu(\sigma) \leq p^a$ for $\sigma \geq 1 + a$. The term $\delta(v, u)$ decomposes into two factors, $\delta(v, u) = \beta(v) \cdot \gamma(v, u)$, with

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

where

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

Using the two inequalities

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

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

$$\left| \sum_{v \in \Sigma^*} \beta(v) p_v^s \right| \leq \sum_{v \in \Sigma^*} \beta(v) p_v^\sigma \leq \sum_{k \geq 0} \mu(\sigma)^k \prod_{i=1}^k \frac{1}{1 - p^i}$$

and, with the classical equality already used in Chapter 5, as (5.23)

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

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

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

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

$$\gamma(\mathcal{S}, u) = \sum_{k \geq 0} \gamma_k(\mathcal{S}, u), \quad \text{with} \quad \gamma_k(\mathcal{S}, u) = \sum_{w \in \Sigma^k} (-u)^{|w|} p_w \prod_{v \in \mathcal{P}_w} \frac{p_v}{1 - p_v}.$$

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

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

Consider $\theta < 1$. As soon as p satisfies $p^{k+1} \leq \theta/(1 + \theta)$, the quotient $p_{w \cdot i}/(1 - p_{w \cdot i})$ is less than θ and $|\gamma_{k+1}(\mathcal{S}, u)| \leq \theta |u| \gamma_k(\mathcal{S}, u)$. $\square$

This ends the proof of Proposition 6.6.

6.5 Final expressions in the Bernoulli model.

We summarize the results about the probability generating function of the profile that we have obtained during the algebraic analysis. They provide exact expressions of the probability generating function $B_n(u)$ of the profile and the probability generating function $G_n(u)$ for the depth. They hold for any *smooth* source.

6.5.1 Expression as binomial sums.

Theorem 6.8. *Consider any general smooth source, with a generalized transition matrix $\mathbf{P}_s$, and fundamental probabilities (p_w) , and a digital tree (trie or dst) built with n independently drawn from the source. For each type of digital trees, the probability generating functions, $B_n(u)$ for the profile and $G_n(u)$ for the depth, are expressed as binomial sums,*

$$B_n(u) = nG_n(u), \quad G_n(u) = 1 + \frac{1}{n}(u-1) \sum_{\ell=2}^n (-1)^\ell \binom{n}{\ell} \Delta(\ell, u)$$

which involve the sequence $\Delta(\ell, u)$. The sequence $\Delta(\ell, u)$ depends on the type of tree, and has two possible definitions

(a) *The first ones explicitly involve the matrix $\mathbf{P}_s$:*

$$\Delta(\ell, u) = \begin{cases} \Delta_T(\ell, u) = \ell {}^t \mathbf{E} (I - u \mathbf{P}_\ell)^{-1} \mathbf{1} & \text{[for trie]} \\ \Delta_D(\ell, u) = {}^t \mathbf{E} (I - u \mathbf{P}_{\ell-1}) \dots (I - u \mathbf{P}_2) \mathbf{1}, \quad \Delta(2, u) = 1 & \text{[for dst]} \end{cases}.$$

where $\mathbf{1}$ is the vector (indexed with Σ^*) whose all the components equal 1, $\mathbf{E}$ is the vector (indexed with Σ^*) whose all the components equal 0 except the one relative to ϵ which equals 1.

(b) *The second ones explicitly involve the fundamental probabilities:*

$$\begin{cases} \Delta_T(\ell, u) = \ell \sum_{w \in \Sigma^*} u^{|w|} p_w^\ell, \\ \Delta_D(\ell, u) = \sum_{v \in \Sigma^*} \delta(v, u) p_v^\ell, \quad \text{with} \quad \delta(v, u) := \frac{1}{p_v} \sum_{w \geq v} u^{|w|} p_w \prod_{\substack{\alpha \in [\epsilon, w] \\ \alpha \neq v}} \frac{1}{1 - p_v p_\alpha^{-1}}. \end{cases}$$

6.5.2 Expression as Rice's integrals.

We have proven in Lemmas 6.7 and 6.5 that $s \mapsto \Delta(s, u)$ is of bounded growth on any halfplane of the form $\Re s \geq 1 + a$ with $a > 0$ and $|u| \leq (1/2)p^{-a}$. Then, we may perform the first step of the Rice methodology and transform the binomial sum into an integral over a vertical line.

Theorem 6.9. *Consider any general smooth source, with a generalized transition matrix $\mathbf{P}_s$, and fundamental probabilities (p_w) , and a digital tree (trie or dst) built with n independently drawn from the source. For each type of digital trees, the probability generating functions, $B_n(u)$ for the profile and $G_n(u)$ for the depth, are expressed as Rice's integral. On the closed disk $|u| \leq 1$, they admit the following integral expressions (with $d > 1$)*

$$B_n(u) = nG_n(u), \quad G_n(u) = 1 + \frac{1}{n}(1-u) \frac{1}{2i\pi} \int_{d-i\infty}^{d+i\infty} \Delta(s, u) L_n(s) ds \quad (6.20)$$

(where $L_n(s)$ is the Rice Kernel defined in (4.13)) that involve the functions $s \mapsto \Delta(s, u)$, defined as a series or as an infinite product which exist for $\Re s > 1$ and $|u| \leq 1$.

(i) As an infinite product which involves the generalized transition matrix $\mathbf{P}_s$

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) &= s {}^t\mathbf{E} (I - u\mathbf{P}_s)^{-1} \mathbf{1} & \text{[for trie]} \\ \Delta_D(s, u) &= {}^t\mathbf{E} \mathbf{Q}(s, u) \cdot \mathbf{Q}(2, u)^{-1} \mathbf{1} & \text{[for dst]} \end{cases} \quad (6.21)$$

where $\mathbf{1}$ is the vector (indexed with Σ^*) whose all the components equal 1, $\mathbf{E}$ is the vector (indexed with Σ^*) whose all the components equal 0 except the one relative to ϵ which equals 1 and the infinite product

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

(ii) As a series which involves the fundamental probabilities (p_w)

$$\begin{cases} \Delta_T(s, u) &= s \sum_{w \in \Sigma^*} u^{|w|} p_w^s, \\ \Delta_D(s, u) &= \sum_{v \in \Sigma^*} \delta(v, u) p_v^s, \quad \text{with} \quad \delta(v, u) := \frac{1}{p_v} \sum_{w \geq v} u^{|w|} p_w \prod_{\substack{\alpha \in [\epsilon, w] \\ \alpha \neq v}} \frac{1}{1 - p_v p_\alpha^{-1}}. \end{cases}$$

6.5.3 Expression of the sequence $B_{n,k}$.

The additive expression of $\Delta(s, u)$ as a power series with respect to u is particularly useful when we wish extract the coefficients $[u^k] \Delta(s, u)$. One then obtains an expression for the sequence $B_{n,k}$.

Theorem 6.10. *There is an exact expression for the profile of digital trees, in term of the coefficient $\delta_k(s) := [u^k] \Delta(s, u)$ as*

$$B_{n,k} = \sum_{\ell=2}^n (-1)^\ell \binom{n}{\ell} (\delta_{k-1}(\ell) - \delta_k(\ell)).$$

with

$$\delta_k(s) = \begin{cases} s \sum_{w \in \Sigma^k} p_w^s & \text{[for trie]} \\ \sum_{w \in \Sigma^k} p_w \sum_{v \leq w} p_v^{s-1} r(v, w), & \text{[for dst]} \end{cases},$$

where, in the case of *dst*, the coefficient $r(v, w)$ is

$$r(v, w) := \prod_{\substack{\alpha \in [\epsilon, w] \\ \alpha \neq v}} \frac{1}{1 - p_v p_\alpha^{-1}} = (-1)^{|w| - |v|} \left(\prod_{\alpha \in [\epsilon, v[} \frac{1}{1 - p_v p_\alpha} \right) \cdot \left(\prod_{\alpha \in]v, w]} \frac{p_{\alpha|v}}{1 - p_{\alpha|v}} \right).$$

Remark. Using the decomposition of $r(v, w)$ as a product, that was described in (6.19) and recalled above, the difference $\delta_k(s) - \delta_{k-1}(s)$ can (perhaps) be simplified, and this (perhaps) gives rise to a formula for the profile which (perhaps) may be studied with the Rice Formula.

6.5.4 Towards to the next chapter

We have now obtained exact expressions of the probability generating functions of interest $G_n(u)$ (for the depth) and $B_n(u)$ for the profile. We have performed the first step of the Rice methodology. We are now ready for the last step, where we wish to obtain asymptotic estimates for $G_n(u)$ when $n \rightarrow \infty$ and u close to 1. This will be essential to derive asymptotic estimates for the mean $\mathbb{E}[D_n]$ and the variance $\text{Var}[D_n]$, and exhibit a Gaussian asymptotic law for the depth.

We choose the Rice method for obtaining such estimates for $G_n(u)$. Then, we start with the expression (6.20), and we wish to shift the integral to the left. This would be possible if we have a good knowledge of the maps $s \mapsto \Delta(s, u)$ on the left of the vertical line $\Re s = 1$. This does not seem possible to directly deal with the operator $\mathbf{P}_s$ when acting on bounded sequences. This is why we have explained in Chapter 2 how to restrict the class of general sources to the class of stationary sources whose reverse past is a dynamical source. Then, in this case, as described at the end of Chapter 2, we can “replace” the matrix $\mathbf{P}_s$ by the (secant) transfer operator $\mathbb{H}_s$ of the dynamical source, and we start the analysis of the next chapter; with the following result.

Theorem 6.11. *Consider any stationary source $\mathcal{S}$, whose reverse past is a dynamical source, with a secant transfer operator $\mathbb{H}_s$. Consider a digital tree (trie or *dst*) built with n independently drawn from the source $\mathcal{S}$. For each type of digital trees, the probability generating functions, $G_n(u)$ of the depth is expressed as a binomial sum,*

$$G_n(u) = 1 + \frac{1}{n}(u - 1) \sum_{\ell=2}^n (-1)^\ell \binom{n}{\ell} \Delta(\ell, u) \quad (6.22)$$

which involves the function $\Delta(s, u)$. The function $\Delta(s, u)$ depends on the type of tree,

$$\Delta(s, u) = \begin{cases} \Delta_T(s, u) = s(I - u\mathbb{H}_s)^{-1}[1](0, 1) & \text{[for trie]} \\ \Delta_D(s, u) = (I - u\mathbb{H}_s)^{-1} \circ \mathbb{Q}(s + 1, u) \circ \mathbb{Q}(2, u)^{-1}[1](0, 1) & \text{[for dst]} \end{cases}.$$

and (possibly) involves the infinite product

$$\mathbb{Q}(s + 1, u) = (I - u\mathbb{H}_{s+1})^{-1}(I - u\mathbb{H}_{s+2})^{-1} \cdots (I - u\mathbb{H}_{s+k})^{-1} \cdots$$

It remains to apply the Rice method to the previous binomial recurrence. We do not perform here the first step of the Rice method, because we wish to obtain for our next analyses an integral expression which holds on a complex neighborhood of 1. Then, the expression (6.22) will be our starting point for the next Chapter.

Chapter 7

Distribution of the depth

Contents

7.1	Statements of our main results.	122
7.1.1	Asymptotic estimates for the mean and the variance.	122
7.1.2	Asymptotic Gaussian law.	123
7.2	Probabilistic theorems for the study of asymptotic distribution.	123
7.2.1	Goncharov Theorem	123
7.2.2	Speed of convergence towards the normal law	124
7.2.3	Quasi-Power Theorem	124
7.2.4	Our study	124
7.3	Beginning the study	125
7.3.1	The mixed Dirichlet series.	125
7.3.2	Tameness of the mixed Dirichlet series on the right of $\Re s = 1$	125
7.3.3	An integral expression for the generating function $G_n(u)$	126
7.4	Applying the Rice method.	127
7.4.1	Properties of the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$ for (s, u) close to $(1, 1)$	127
7.4.2	Main functions of the analysis.	128
7.4.3	Various types of tameness.	128
7.5	Three estimates.	129
7.5.1	An expression for the moment generating function of the depth D_n	129
7.5.2	An expression for the cumulant generating function	131
7.5.3	An expression for the normalized characteristic function.	133
7.6	Proofs of the main results	134
7.6.1	Asymptotic estimates for the mean value $\mathbb{E}[D_n]$ and the variance $\mathbb{V}[D_n]$	134
7.6.2	Asymptotic Gaussian law for the depth	134
7.6.3	Speed of convergence.	135
7.6.4	Expression of the subdominant constants.	137
7.6.5	Conclusion of the Chapter.	138

We now present the final results of the thesis. The following two main results, stated as Theorem 7.1 and Theorem 7.2, constitute with the last Theorems of the previous Chapter, namely Theorem 6.11 and Theorem 6.9, the main results of the Thesis. These results are very general as they refer to the two possible digital trees and various sources which intervene via their type of tameness.

The method we choose for obtaining the distributional analysis of the depth is the Rice method, described in Chapter 4. In the previous chapter, we exhibit the two analytic liftings $\Delta(s, u)$ which will intervene in the Rice method. They can be viewed as mixed Dirichlet series, which “mix” the behaviour of the source with the structure of the digital tree whose type is denoted by $X \in \{T, D\}$ (T for trie, D for digital search tree). The source intervenes with its secant transfer operator $\mathbb{H}_s$, defined at the end of Chapter 2, mainly via its quasi-inverse $\mathbb{T}(s, u) := (I - u\mathbb{H}_s)^{-1}$. We know that the Rice method relies on the tameness of such an analytic lifting $\Delta(s, u)$. We already mentioned at the beginning of Chapter 3 (and we prove this fact in the present chapter) that the tameness of $\Delta(s, u)$ only relies on the tameness of the operator $\mathbb{T}(s, u)$, as it was defined in Definition 3.4. We have exhibited at the end of Chapter 3 three types of tameness which appear in a natural way for classical sources (see the Table in Figure 7.1). The possible type of tameness is denoted by $Y \in \{S, P, H\}$.

As it was described in Chapter 4, the Rice method provides estimates with a main term which is brought by the residues of $\Delta(s, u)$, whereas the remainder term depends on tameness properties of $\Delta(s, u)$. Then, the remainder terms will depend on the tameness type of the source, while the main terms depend both on the source, and on the tree type X . More precisely, the main terms depend strongly on the source and it just “adjusted” by the tree type X .

In Section 7.1, we state the two main results, Theorem 7.1 and Theorem 7.2. Then, Section 7.2 describes three classical results in probability theory that are used in the Chapter. They mainly deal with the moment generating functions M_n , the cumulant generating functions C_n and the normalized characteristic functions $\widehat{M}_n$. Section 7.3 begins the study, provides a first expression for these generating functions, mainly based on the results of the previous Chapter 6, and states their basic properties (namely their analyticity). Then, we wish to apply the Rice method, and we need a good knowledge on $\Delta(s, u)$, about its residues and its tameness; we thus define the main functions which will intervene in our further computations. Section 7.4 applies the Rice method in the present context; it first provides an estimate for the moment generating function M_n which is further transferred to the other two generating functions C_n and $\widehat{M}_n$. Section 7.6 uses these estimates together with the three classical theorems of Section 7.2 and concludes the proofs for Theorems 7.1 and 7.2.

7.1 Statements of our main results.

7.1.1 Asymptotic estimates for the mean and the variance.

Theorem 7.1. *Consider a stationary source $\mathcal{S}$, whose reverse past is a dynamical source which is assumed to be tame with a type $Y \in \{S, P, H\}$. Consider any of the two types of digital trees –trie (type T) or dst (type D)– built on a random sequence of n words independently drawn from the source. Then, the mean and the variance of their depth D_n admit the following asymptotic expansions, for any $X \in \{T, D\}$ and $Y \in \{S, P, H\}$,*

$$\begin{aligned}\mathbb{E}[D_n] &= \mu \log n + \mu_X + R_Y(n) \\ \text{Var}[D_n] &= \nu \log n + \nu_X + R_Y(n)\end{aligned}\tag{7.1}$$

[Dominant constants]. The constants μ, ν are expressed with the dominant eigenvalue $\lambda(s)$ of the source, as

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

The only case where $\nu = 0$ arises for an unbiased memoryless source.

[Sub-dominant constants]. The constants μ_X, ν_X depend both on the source and on the type $X \in \{T, D\}$ of digital tree and the inequality $\mu_T > \mu_D$ holds. There exist explicit (but involved) expressions for these constants.

[Remainder terms]. The type of functions $R_Y(n)$ only depends on the type Y of source tameness. The functions $R_Y(n)$ admit the general form:

$$R_Y(n) = \begin{cases} O(n^{-\delta}) & \text{if the source is } S\text{-tame} \\ O(\exp[-(\log n)^\rho]) & \text{if the source is } H\text{-tame} \\ \Pi(n) + O(n^{-\delta}) & \text{if the source is } P\text{-tame} \end{cases} \quad (7.2)$$

Here, δ is the width, ρ is related to hyperbolic exponent ρ_0 via the relation $\rho < 1/(1 + \rho_0)$ and $\Pi(n)$ is a periodic function of $\log n$, whose period η is the period of the source.

7.1.2 Asymptotic Gaussian law.

Theorem 7.2. Consider a stationary source $\mathcal{S}$, whose reverse past is a dynamical source which is assumed to be tame of any type. Assume moreover $\mathcal{S}$ not to be conjugated to an unbiased memoryless source. Consider a digital tree (trie or dst) built on n words independently drawn from the source. Then,

(a) the depth D_n of the digital tree asymptotically follows a Gaussian law

$$\frac{D_n - \mathbb{E}[D_n]}{\sqrt{\text{Var}[D_n]}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (7.3)$$

(b) Moreover, the speed of convergence towards the Gaussian law is of order $(\log n)^{-1/2}$.

7.2 Probabilistic theorems for the study of asymptotic distribution.

We present three main theorems, that are classic results in Probability Theory, and are useful for obtaining asymptotic distributions, in particular asymptotic Gaussian laws. The Goncharov Theorem or equivalently the Lévy's continuity theorem, which can be founded in Durrett [15] or Feller [18], is a classical tool for proving the asymptotic Gaussian law, whereas the Berry-Esseen inequality is the standard tool for studying the speed of convergence. We also describe the Quasi-Powers Theorem, due to Hwang. When it can be applied (and this will not be always the case here), this is a “turn-key” theorem, very easy to use.

7.2.1 Goncharov Theorem

Theorem 7.3. [Goncharov] Consider a sequence of random variables D_n , and consider the probability generating function $G_n(u) := \mathbb{E}[u^{D_n}]$ of the variable D_n . Denote by μ_n the mean value $\mathbb{E}[D_n]$ of D_n and by ν_n the standard deviation of D_n and consider the characteristic function $\widehat{M}_n(i\tau)$ of the variable $\widehat{D}_n := (D_n - \mu_n)/\nu_n$, namely

$$\widehat{M}_n(i\tau) := \exp \left[-i\tau \frac{\mu_n}{\nu_n} \right] G_n(e^{i\tau/\nu_n}) = \exp \left[-i\tau \frac{\mu_n}{\nu_n} \right] M_n \left(\frac{i\tau}{\nu_n} \right). \quad (7.4)$$

Then, if for any real τ , the sequence $\widehat{M}_n(i\tau)$ tends to $e^{-\tau^2/2}$, then the variables D_n asymptotically follow a Gaussian law.

7.2.2 Speed of convergence towards the normal law

Theorem 7.4. [Berry–Esseen inequality] Let F_n be the distribution function of a random variable X_n , and $M_n(i\tau) := \mathbb{E}[\exp(i\tau X_n)]$ its characteristic function. Denote Φ the distribution function of the Gaussian law. Then, for any $T > 0$, the following inequality holds

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq b \int_{-T}^{+T} \frac{1}{|\tau|} |M_n(i\tau) - e^{-\tau^2/2}| d\tau + \frac{r(b)}{T}, \quad (7.5)$$

where b satisfies $b \geq 1/(2\pi)$ and $r(b)$ is a constant function of b .

7.2.3 Quasi-Power Theorem

The Quasi-Powers theorem, due to Hwang, uses the moment generating function $M_n(w) := G_n(e^w)$, with strong analyticity conditions on a neighborhood of 0. It *directly* provides asymptotic expressions for the mean and the variance, and *exhibits the cancellation phenomenon* in the computation of the variance. It also provides the *speed* of convergence towards the Gaussian law. The Theorem may be applied under three main conditions:

- (a) the moment generating functions $M_n(w)$ are analytic on a neighborhood $\mathcal{W}$ of 0,
- (b) they are written as a *product* between a main term and a remainder term,
- (c) the remainder term is written as $1 + \epsilon_n(w)$, where $\epsilon_n(w)$ is an *analytic* function which satisfies $\epsilon_n(0) = 0$ and is *uniformly* bounded by a sequence κ_n on $\mathcal{W}$. Then, the two first derivatives of ϵ_n at zero can be easily bounded with the Cauchy formula.

Finally, Hwang deals with the cumulant generating function, and obtains the asymptotic estimates for the mean and variance, which are the first two coefficients of the series expansion of the cumulant at $w = 0$. He proves the following result:

Theorem 7.5. [Quasi-Powers Theorem (Hwang, 1994)]. Consider a sequence of random variables D_n and their moment generating functions $M_n(w) := G_n(e^w)$. Suppose that the functions $M_n(w)$ are analytic in a complex neighborhood $\mathcal{W}$ of zero, and satisfy

$$M_n(w) = G_n(e^w) = \exp[\beta_n U(w) + V(w)] (1 + \epsilon_n(w)), \quad (7.6)$$

where the function $\epsilon_n(w)$ is bounded by a sequence κ_n , uniformly on $\mathcal{W}$. Moreover, $U(w)$ and $V(w)$ are analytic on $\mathcal{W}$, the sequence β_n tends to ∞ , and the sequence κ_n tends to zero. Then, the mean and the variance satisfy

$$\begin{aligned} \mathbb{E}_n[D_n] &= U'(0) \beta_n + V'(0) + O(\kappa_n), \\ \text{Var}_n[D_n] &= U''(0) \beta_n + V''(0) + O(\kappa_n) \end{aligned} \quad (7.7)$$

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

7.2.4 Our study

Here, our moment generating functions are always analytic in a neighborhood $\mathcal{W}$ of 0, and Condition (a) is always satisfied. In the case of S -tameness or P -tameness, conditions (b) and (c) are also fulfilled, and the Quasi-powers Theorem may be applied: this is the easy case. The

more difficult cases arise in the case of H tameness, since the moment generating function is no longer written as a product. In the case of strong H -tameness, we may perform a scale change, in order to get such a factorization, but only when w is imaginary. Then, the Cauchy formula cannot be applied, and we must get a direct proof of the boundedness of the derivatives of $\epsilon_n(w)$ at $w = 0$. The case of the weak H -tameness is still more involved, since the estimates are not obtained for any imaginary $i\tau$, but only on a dense subset, and these estimates are non uniform...

7.3 Beginning the study

We set up the scenery. We recall the expression of the analytic liftings (obtained in the previous chapter), prove their tameness “on the right”, and perform the first step of the Rice method (“on the right”).

7.3.1 The mixed Dirichlet series.

In the previous chapter, we exhibit the liftings $\Delta(s, u)$ which are expressed with the secant operator $\mathbb{H}_s$ of the form

$$\Delta(s, u) := \begin{cases} \Delta_T(s, u) &= s(I - u\mathbb{H}_s)^{-1}[1](0, 1), \\ \Delta_D(s, u) &= \mathbb{Q}(s, u) \circ \mathbb{Q}(2, u)^{-1}[1](0, 1). \end{cases} \quad (7.8)$$

Here, $\mathbb{Q}(s, u)$ is the infinite product $\mathbb{Q}(s, u) = (I - u\mathbb{H}_s)^{-1} \circ (I - u\mathbb{H}_{s+1})^{-1} \circ \dots$.

Therefore, the operators involved are all related to the quasi inverse $\mathbb{T}(s, u) = (I - u\mathbb{H}_s)^{-1}$, as

$$\Delta(s, u) := \begin{cases} \Delta_T(s, u) &= s\mathbb{T}(s, u)[1](0, 1), \\ \Delta_D(s, u) &= \mathbb{T}(s, u) \circ \mathbb{L}(s, u)[1](0, 1). \end{cases} \quad (7.9)$$

where $\mathbb{L}(s, u)$ is an infinite product defined as a function of the infinite product $\mathbb{Q}(s, u)$,

$$\mathbb{L}(s, u) := \mathbb{Q}(s + 1, u) \mathbb{Q}(2, u)^{-1}. \quad (7.10)$$

As we will see, the operator $\mathbb{L}(s, u)$ plays a secondary role whereas the operator $\mathbb{T}(s, u)$ plays the principal role.

7.3.2 Tameness of the mixed Dirichlet series on the right of $\Re s = 1$.

We now consider the case of a dynamical source of the Good Class, and the situation on the right of the vertical line $\Re s = 1$.

Proposition 7.6. [Tameness of $\mathbb{L}(s, u)$] *Consider a dynamical source of the Good Class, its secant transfer operator $\mathbb{H}_s$, and the norm $\|\cdot\|_{(1,t)}$ defined in (3.4). There exist a complex neighborhood $\mathcal{U}$ of $u = 1$ and a real $a \in]0, 1]$ such that the following holds :*

- (i) *For any $b > a$, there exists a constant $K > 0$ such that, for any $u \in \mathcal{U}$ and any s with $\Re s \in]1 + a, 1 + b[$, the norms $\|\cdot\|_{(1,t)}$ of the operators $\mathbb{H}_s^k$, $(I - u\mathbb{H}_s)^{-1}$ and $I - u\mathbb{H}_s$ satisfy*

$$\|\mathbb{H}_s^k\|_{(1,t)} \leq K(2 + \sigma)\rho^{k(\sigma-1)}, \quad \|(I - u\mathbb{H}_s)^{-1}\|_{(1,t)} \leq 1 + 2K(2 + \sigma)u\rho^{\sigma-1}$$

$$\|I - u\mathbb{H}_s\|_{(1,t)} \leq 1 + K(2 + \sigma)u\rho^{\sigma-1}$$

The operators $(I - u\mathbb{H}_s)^{-1}$ and $I - u\mathbb{H}_s$ are analytic as functions of (s, u) on $\{\Re s > 1 + a\} \times \mathcal{U}$.

- (ii) The infinite products $\mathbb{Q}(s+1, u)$, $\mathbb{Q}^{-1}(s+1, u)$ are normally convergent on $\Re s > a$. They define analytic functions of (s, u) on $\{\Re s > a\} \times \mathcal{U}$.
- (iii) The function $(s, u) \mapsto \mathbb{L}(s, u)$ is analytic on $\{\Re s > a\} \times \mathcal{U}$. The norm $\|\cdot\|_{(1,t)}$ of the operator $\mathbb{L}(s, u)$ is uniformly bounded on $\{\Re s > a\} \times \mathcal{U}$.
- (iv) The mixed Dirichlet series $(s, u) \mapsto \Delta(s, u)$ are analytic on $\{\Re s > 1+a\} \times \mathcal{U}$. The modulus $|\Delta(s, u)|$ is uniformly bounded on $\{\Re s > 1+a\} \times \mathcal{U}$.

Proof. Assertion (i). The first inequality is of the type Lasota Yorke and uses the norm $\|\cdot\|_{(1,t)}$ defined in (3.4) of Chapter 3. Denote by $a_k(\sigma)$ the norm-sup of the function $\mathbb{H}_\sigma^k[1]$, and consider the function $G := \mathbb{H}_s^k[F]$. Its norm $\|\cdot\|_{(1,t)}$ satisfies

$$\|G\|_{(1,t)} \leq a_k(\sigma) \left[\|F\|_0 + \frac{1}{|t|} \left(L|s| \|F\|_0 + \rho^k \|F\|_1 \right) \right],$$

where L is the distortion constant and ρ the contraction. With the inequalities $a_k(\sigma) \leq L\rho^{k(\sigma-1)}$, and $|s| \leq \sigma + |t|$, the first inequality is proven. We now choose $a \in]0, 1]$ and the neighborhood $\mathcal{U}$ of $u = 1$ so that $|u|\rho^a \leq 1/2$ for any $u \in \mathcal{U}$. The second and the third ones are easily deduced from the first one and the previous inequality $|u|\rho^a \leq 1/2$.

Assertion (ii) Consider s with $\Re s > a$. Then the norm $\|\cdot\|_{(1,t)}$ of the infinite product $\mathbb{Q}(s+1, u)$ satisfies

$$\|\mathbb{Q}(s+1, u)\|_{(1,t)} \leq \prod_{k=1}^{\infty} \|(I - u\mathbb{H}_{s+k})^{-1}\|_{(1,t)} \leq \prod_{k=0}^{\infty} \left[1 + 2K(1 + \sigma + k)u\rho^{\sigma+k} \right],$$

and the last infinite product is convergent. $\square$

7.3.3 An integral expression for the generating function $G_n(u)$.

With the previous result, we perform the first step of the Rice methodology, and this provides first important results about the sequences G_n, M_n, C_n of various generating functions: these sequences are analytic on a neighborhood of $w = 0$.

Proposition 7.7. *The following holds :*

- (a) There exist a neighborhood $\mathcal{W}$ of $w = 0$ and a real $a > 0$ for which the moment generating function $M_n(w) := G_n(e^w) := \mathbb{E}[\exp(wD_n)]$ of the depth D_n of a digital tree (trie or dst) admits an integral representation of Rice type along the vertical line $\Re s = d$ (with $d > 1+a$), with

$$n\underline{G}_n(e^w) := n \frac{G_n(e^w) - 1}{e^w - 1} = -\frac{1}{2i\pi} \int_{d-i\infty}^{d+i\infty} \Delta(s, e^w) L_n(s) ds \quad (7.11)$$

which involves the Rice kernel

$$L_n(s) := \frac{(-1)^n n!}{s(s-1) \dots (s-n)}$$

and the mixed Dirichlet series $s \mapsto \Delta(s, u)$ defined in (7.8).

- (b) *There exists a complex neighborhood $\mathcal{W}$ of 0 on which the sequence $M_n(w) := G_n(e^w)$ of the moment generating functions and the sequence $C_n(w) := \log M_n(w)$ of the cumulant generating functions are analytic. In particular, the two sequences $C_n(0)$ and $C_n''(0)$ are well defined.*

Proof. For Assertion (a), we apply the first step of Rice methodology described in Proposition 4.12. This is possible thanks to the previous Proposition 7.6 which entails the polynomial growth of $\Delta(s, u)$ on the right of $\Re s = 1$. This proves Assertion (a).

For Assertion (b), we use the fact that the series $s \mapsto \Delta(s, e^w)$ is bounded on the vertical line, uniformly with respect to $w \in \mathcal{W}$. Then, with Assertion (i) of Proposition 4.16, the integral of the right member of (7.11) defines a sequence of analytic functions $n\underline{G}_n(e^w)$ uniformly bounded on a neighborhood $\mathcal{W}$ of 0. The expression for the continuant, as

$$C_n(w) := \log [1 + (e^w - 1)\underline{G}_n(w)]$$

defines thus a sequence of analytic functions on $\mathcal{W}$. □

7.4 Applying the Rice method.

Now, we wish to shift “to the left”, and use the second step of the Rice method. We know that we will need a good knowledge of $\Delta(s, u)$ on the left, (namely its residues, and its tameness) and the present section gathers the “kit” which will be needed in our further study.

7.4.1 Properties of the quasi-inverse $(I - u\mathbb{H}_s)^{-1}$ for (s, u) close to $(1, 1)$

We now recall an important property of transfer operators $\mathbb{H}_s$ relative to dynamical sources of the Good Class. See [4] for a more precise statement.

Proposition 7.8. *Consider a dynamical source of the Good Class, and the dominant eigenvalue $\lambda(s)$ of the secant operator $\mathbb{H}_s$. Consider the function U defined from the dominant eigenvalue $\lambda(s)$ via the implicit equation*

$$e^w \lambda(1 + U(w)) = 1 \quad \text{with} \quad U(0) = 0. \quad (7.12)$$

- (a) *The operator $s \mapsto \mathbb{T}(s, e^w) = (I - e^w \mathbb{H}_s)^{-1}$ has a simple pole at $s = 1 + U(w)$, and its residue there, denoted by $\mathbb{A}(w)$, involves spectral objects of $\mathbb{H}_s$ at $s = 1 + U(w)$, namely, the dominant eigenvalue $\lambda(s)$ and the dominant projector $\mathbb{G}_s$,*

$$\mathbb{A}(w) := \text{Res} [s \mapsto \mathbb{T}(s, e^w); 1 + U(w)] = \frac{-1}{e^w \lambda'(1 + U(w))} \mathbb{G}_{1+U(w)}. \quad (7.13)$$

- (b) *The function $s \mapsto \Delta(s, e^w)$ has a simple pole at $s = 1 + U(w)$ and the residue $A(w) := \text{Res}[\Delta(s, e^w); s = 1 + U(w)]$ depends on the type of the digital tree, it is denoted by $A_T(w)$ or $A_D(w)$ according to the type of tree, and*

$$A(w) := \begin{cases} A_T(w) &= (1 + U(w)) \cdot \mathbb{A}(w)[1](0, 1) \\ A_D(w) &= \mathbb{A}(w)[\mathbb{L}(1 + U(w), e^w)](0, 1) \end{cases}. \quad (7.14)$$

At $w = 0$, they both satisfy $A_T(0) = A_D(0) = -1/\lambda'(1) = 1/h(S)$, where $h(S)$ is the entropy of the source.

7.4.2 Main functions of the analysis.

In the following, the function U plays a fundamental role, it satisfies $U(0) = 0$ and its first two derivatives are expressed with the dominant eigenvalue $\lambda(s)$ as

$$U'(0) = -\frac{1}{\lambda'(1)}, \quad U''(0) = \frac{\lambda'(1)^2 - \lambda''(1)}{\lambda'(1)^3}. \quad (7.15)$$

We also use the function

$$V(w) := (e^w - 1)\Gamma(-1 - U(w))A(w), \quad (7.16)$$

which is analytic in a neighborhood of 0 and satisfies $V(0) = 1$ (this is due to the fact that $t\Gamma(-1 - t)$ is analytic in the neighborhood of 0 and equals 1 for $t = 0$). We also deal with the sequence F_n related to the Rice kernel that can be estimated with Lemma 4.17

$$F_n(t) := -\frac{L_n(1+t)}{n^t\Gamma(-1-t)}, \quad F_n(t) = \exp \left[O \left(\frac{t^2}{n} \right) \right]. \quad (7.17)$$

Moreover, if the source is periodic of period $i\eta$, we introduce the series Π_n , defined with the sequence $U_k(w) = U(w) + ik\eta$

$$\Pi_n(w) := \sum_{k \in \mathbb{Z}^*} \pi_{n,k}(w) n^{ik\eta} \quad \text{with} \quad \pi_{n,k}(w) := \Gamma(-1 - U_k(w))F_n(U_k(w)). \quad (7.18)$$

Remark that the function V depends both on the source and the tree (via the function A), whereas the sequence Π_n only depend on the source, and the sequence F_n is universal.

7.4.3 Various types of tameness.

We copy here the Table which already appeared in Figure 3.2 of Chapter 3. It must be understood with Definition 3.4 of the same Chapter 3.

Definition of tameness type	$\mathcal{R}$	Number of poles	$\mathcal{U}_0 \subset \mathcal{U}$	Examples of occurrences
P -tame	A vertical strip	∞	$\mathcal{U}$	Periodic Sources
S -tame	A vertical strip	1	$\mathcal{U}$	UNI Sources
Strongly H -tame	A hyperbolic region	1	$\mathcal{T}$ defined in (3.14)	DIOP3 Sources
Weakly H -tame	A hyperbolic region	1	$\mathcal{T}_m$ defined in (3.14)	DIOP2 Sources

Figure 7.1: Various types of tameness defined by the shape of the region $\mathcal{R}$, the number of poles in $\mathcal{R}$ and the shape of the subset $\mathcal{U}_0$ of the complex neighborhood of $\mathcal{U}$

We recall that in the case of weak H -tameness, the set $\mathcal{T}_m$ gathers complex numbers of modulus 1 whose argument belongs to the set

$$\mathcal{Q}_m := \left\{ \frac{2\pi a}{mk}; \quad a \in \mathbb{Z}, k \geq 1 \right\}, \quad (7.19)$$

and in this case the bound depends linearly of the denominator k .

We now apply Proposition 4.16, together with Definition 3.4:

Proposition 7.9. *The order of the remainder term $R_Y(n)$ associated with the tameness type Y is described in the table of Figure 7.2. In the case of weakly H -tameness, the estimates hold when the denominator of the argument of u is polynomial with respect to $\log n$.*

Definition of tameness type	$\mathcal{R}$ and its parameter	$\mathcal{U}_0 \subset \mathcal{U}$	Order of the remainder term	Uniformity w.r.t $u \in \mathcal{U}_0$
P -tame	Vertical strip (δ)	$\mathcal{U}$	$n^{-\delta}$	Yes
S -tame	Vertical strip (δ)	$\mathcal{U}$	$n^{-\delta}$	Yes
Strongly H -tame	Hyperbolic region (ρ_0)	$\mathcal{T}$ def. in (3.14)	$\exp[-(\log n)^\rho]$	Yes
Weakly H -tame	Hyperbolic region (ρ_0)	$\mathcal{T}_m$ def. in (3.14)	$\exp[-(\log n)^\rho]$	No

Figure 7.2: Various types of remainders associated with various types of tameness. For a vertical strip, δ is the width; for a hyperbolic region, ρ_0 is the exponent, and the exponent ρ satisfies $\rho < 1/(1 + \rho_0)$. The last column answers the question “Is the remainder uniform with respect to $u \in \mathcal{U}_0$ ”? The answer is always yes *except* in the case of weakly H -tameness, where the remainder depends in a linear way on the denominator of the argument of u_0 .

7.5 Three estimates.

We are now ready to apply the second step of the Rice method: this entails an estimate for the sequence M_n of moment generating functions, that we transfer to the sequence C_n of cumulants, then to the sequence $\widehat{M}_n$ of normalized characteristic functions.

7.5.1 An expression for the moment generating function of the depth D_n .

Applying the second step of the Rice methodology (shifting to the left) leads to a first estimate for the moment generating functions $M_n(w)$ of the depth D_n .

Proposition 7.10. *Consider a digital tree (trie or dst) built on n words independently drawn from the source of the Good Class . Consider the sequence F_n relative to the Rice Kernel described in (7.17). Associate with the pair (source, digital tree) the functions defined in Section 7.4.2, namely the functions U, A, V , and the sequence Π_n when the source is periodic. Then, the following holds:*

- (a) *There exists a neighborhood $\mathcal{W}$ of 0 and a sequence of analytic functions R_n defined in $\mathcal{W}$ for which the moment generating function $M_n(w)$ of the depth decomposes as*

$$M_n(w) := \mathbb{E}[\exp(wD_n)] = P_n(w) + (1 - e^w)Q_n(w) + (1 - e^w)R_n(w), \quad (7.20)$$

where the term Q_n only appears in the case when the source is periodic. Moreover

$$\begin{aligned} P_n(w) &= n^{U(w)} D_n(w), \quad \text{with } D_n(w) = V(w) F_n(U(w)) \\ Q_n(w) &= n^{U(w)} A(w) \Pi_n(w) \end{aligned}$$

- (b) *Consider now a complex $w_0 \in \mathcal{W}$ for which the operator $s \mapsto \mathbb{T}(s, u)$ is tame at $u = e^{w_0}$, on a region $\mathcal{R}$ on the left of the vertical line $\Re s = 1$. Then*

- (b1) *At $w = w_0$, the equality holds*

$$R_n(w_0) = \frac{1}{2i\pi} \int_{\rho} \Delta(s, e^{w_0}) L_n(s) ds, \quad (7.21)$$

with a contour ρ which borders the region $\mathcal{R}$ on the left.

(b2) The function R_n admits at w_0 the following estimate $R_n(w_0) = O(R_Y(n))$ which depends on the type of tameness,

$$R_Y(n) = \begin{cases} n^{-\delta} & \text{for } Y = P \text{ or } Y = S \\ \exp[-(\log n)^\rho] & \text{for } Y = H \end{cases}. \quad (7.22)$$

(b3) The same estimates hold for all the sequences formed with the derivatives $R_n^{(k)}(w_0)$ for any order k , the equality

$$R_n^{(k)}(w_0) = \frac{1}{2i\pi} \int_{\rho} \Delta^{(k)}(s, e^{w_0}) L_n(s) ds, \quad (7.23)$$

and, in the non periodic case; the two first derivatives of M_n satisfies at $w = w_0$.

$$\begin{aligned} M_n'(w) &= n^{U(w)} [D_n(w)U'(w) \log n + D_n'(w)] + O(R_Y(n)) \\ M_n''(w) &= n^{U(w)} D_n(w) [U''(w) \log n + (U'(w))^2 \log n] \\ &\quad + n^{U(w)} [2D_n'(w)U'(w) \log n + D_n''(w)] + O(R_Y(n)) \end{aligned}$$

The same type of formula exists in the periodic case.

Proof.

Assertions (a), (b1), (b2). Consider some complex w close to 0 for which the operator $s \mapsto \mathbb{T}(s, u)$ is tame at $u = e^w$. Then, with Proposition 7.6, the function $s \mapsto \Delta(s, u)$ is itself tame at $u = e^w$, and we use the Rice methodology: Then, the line of integration $\Re(s) = d$ can be moved to the left in (4.13), until a curve $\mathcal{C} \subset \mathcal{R}$, with residues taken into account. The three cases of tameness lead to various possibilities for the curve $\mathcal{C}$ and the residues.

(i) In the non periodic case, there are two residues $s = 1 + U(w)$ and $s = 1$ to be taken into account, and

$$\begin{aligned} n[G_n(e^w) - 1] &= -(e^w - 1) \operatorname{Res} [\Delta(s, e^w) L_n(s); s = 1 + U(w)] \\ &\quad - (e^w - 1) \operatorname{Res} [\Delta(s, e^w) L_n(s); s = 1] \\ &\quad - (e^w - 1) \frac{1}{2i\pi} \int_{\mathcal{C}} \Delta(s, e^w) L_n(s) ds \end{aligned}$$

The second residue of the above equation at $s = 1$ is equal to $-n/(1 - e^w)$, and it remains

$$M_n(w) := G_n(e^w) = \frac{1}{n}(1 - e^w) \operatorname{Res} [\Delta(s, e^w) \cdot L_n(s); s = 1 + U(w)] \quad (7.24)$$

$$+ (1 - e^w) \frac{1}{n} \frac{1}{2i\pi} \int_{\mathcal{C}} \Delta(s, e^w) L_n(s) ds. \quad (7.25)$$

With Lemma 4.17, and Definition of functions A given in (7.14), V in (7.16) and F_n given in (7.17), the first residue is equal to,

$$\begin{aligned} \operatorname{Res} [\Delta(s, e^w) \cdot L_n(s); s = 1 + U(w)] &= \frac{1}{n} L_n(1 + U(w)) A(w) \\ &= -\frac{n^{U(w)+1}}{n} A(w) \Gamma(-1 - U(w)) F_n(U(w)) \end{aligned}$$

And thus, the equality holds :

$$(1 - e^w) \operatorname{Res} [\Delta(s, e^w) \cdot L_n(s); s = 1 + U(w)] = n^{U(w)} V(w) F_n(U(w))$$

(ii) In the periodic case, the curve $\mathcal{C}$ is a vertical line of equation $\Re s = 1 - \delta$, and there are two kinds of residue. As previously, the residue at $s = 1$, but also the family of residues at points $1 + U_k(w)$ (for $k \in \mathbb{Z}$) where $U_k(w) = U(w) + ik\eta$, and η is the period. It then remains

$$\begin{aligned} M_n(w) := G_n(e^w) &= \frac{1}{n}(1 - e^w) \sum_{k \in \mathbb{Z}} \text{Res} [\Delta(s, e^w) \cdot L_n(s); s = 1 + U_k(w)] \\ &\quad + \frac{1}{n}(1 - e^w) \frac{1}{2i\pi} \int_{\Re s = 1 - \delta} \Delta(s, e^w) L_n(s) ds. \end{aligned}$$

With Lemma 4.17, and definition of functions A given in (7.14), F_n given in (7.17), and Π_n in (7.18), the sum of residues is equal to

$$\begin{aligned} &\frac{1}{n}(1 - e^w) \sum_{k \in \mathbb{Z}^*} \text{Res} [\Delta(s, e^w) \cdot L_n(s); s = 1 + U_k(w)] \\ &= \frac{1}{n}(1 - e^w) A(w) \sum_{k \neq 0} L_n(1 + U_k(w)) \\ &= n^{U(w)}(e^w - 1) A(w) \sum_{k \neq 0} n^{ik\eta} \Gamma(-1 - U_k(w)) F_n(U_k(w)) \\ &= (e^w - 1) A(w) \Pi_n(w) \end{aligned}$$

Now Proposition 7.9 provides estimates for the remainder integrals, according to the type of tameness. This ends the proof for Assertions (a), (b1) and (b2).

Assertion (b3) Denote by $\mathbb{T}^{(k)}(s, e^w)$ and (resp.) by $\Delta^{(k)}(s, e^w)$ the k -th derivatives of $w \mapsto \mathbb{T}(s, e^w)$ (resp.) $w \mapsto \Delta(s, e^w)$ with respect to w . The k -th derivative of a quasi-inverse is expressed with a polynomial of quasi-inverses; for instance for $k = 1$ and $k = 2$,

$$\begin{aligned} \mathbb{T}^{(1)}(s, e^w) &= e^w \mathbb{H}_s \circ \mathbb{T}(s, e^w)^2 \\ \mathbb{T}^{(2)}(s, e^w) &= e^w \mathbb{H}_s \circ \mathbb{T}(s, e^w)^2 + 2e^{2w} \mathbb{H}_s^2 \circ \mathbb{T}(s, e^w)^3 \end{aligned}$$

If now the operator $s \mapsto \mathbb{T}(s, e^w)$ is of polynomial growth on a region $\mathcal{R}$ at $w = w_0$, then, for any $k \geq 1$, its derivative $\mathbb{T}^{(k)}(s, e^w)$ is also of polynomial growth on the same region $\mathcal{R}$. And it is also true for the derivatives $\Delta^{(k)}(s, e^w)$. Then, we take the derivative “under the integral”; and (7.23) holds.

On the other hand, we take the derivatives of the whole decomposition in (7.20), at $w = w_0$, and we obtain the two estimates for the derivatives $M'_n(w_0)$ and $M''_n(w_0)$ of Assertion (b3). $\square$

7.5.2 An expression for the cumulant generating function

The cumulant generating functions are very useful to obtain asymptotic expansions for the mean value $\mathbb{E}[D_n]$ and the variance $\mathbb{V}[D_n]$. They directly prove without (almost any) computations the “cancellation” phenomenon that occurs in the leading term of the variance.

Proposition 7.11. *Consider the functions of the decomposition (7.20). Then, the cumulant generating functions $C_n(w) := \log M_n(w)$ admit the following convenient additive decomposition, which involves analytic functions on a neighborhood of 0,*

- (a) *In case of S -tameness or P -tameness, when w belongs to a complex neighborhood of 0, the convenient decomposition is:*

$$C_n(w) = \log P_n(w) + \log \left[1 + (1 - e^w) \frac{Q_n(w)}{P_n(w)} \right] + \log \left[1 + (1 - e^w) \frac{R_n(w)}{P_n(w) + Q_n(w)} \right].$$

Note that the factor Q_n (related to the possible periodicity of the source) is zero in the case of S -tameness.

- (b) In case of strong H -tameness, we let $a_n = \Theta(\log n)^{-1/2}$. Then, the convenient decomposition is

$$C_n(i\tau a_n) = \log P_n(i\tau a_n) + \log \left[1 + (1 - e^{i\tau a_n}) \frac{R_n(i\tau a_n)}{P_n(i\tau a_n)} \right].$$

- (c) In the non periodic case, at $w = 0$, one has, with the remainder term R_Y specific to each tameness $Y \in \{S, H\}$,

$$\begin{cases} C'_n(0) &= U'(0) \log n &+ D'_n(0) &+ O(R_Y(n)) \\ C''_n(0) &= U''(0) \log n &+ D''_n(0) - D'_n(0)^2 &+ O(R_Y(n)) \end{cases}.$$

There are also the same type of formulae in the periodic case.

Proof. We use the decomposition of $M_n(w)$ described in (7.20). As the two sequences

$$\frac{Q_n(w)}{P_n(w)}, \quad \frac{R_n(w)}{P_n(w) + Q_n(w)}$$

are uniformly bounded on a neighborhood of 0, the functions

$$\log \left[1 + (1 - e^w) \frac{Q_n(w)}{P_n(w)} \right], \quad \log \left[1 + (1 - e^w) \frac{R_n(w)}{P_n(w) + Q_n(w)} \right]$$

are analytic on a neighborhood $\mathcal{W}_1$ of zero, and we deduce Assertion (a).

Assertion (c) The first two derivatives of C_n at $w = 0$ are expressed with the first two derivatives of M_n at $w = 0$ as $C'_n(0) = M'_n(0)$, $C''_n(0) = M''_n(0) - M'_n(0)^2$. Moreover, for any type of tameness, the operator $\mathbb{T}(s, e^w)$ is tame at $w = 0$, and we can always use the computations of Assertion (b3) of the previous proposition, together with the equality $D_n(0) = 1$. This leads to the result (in the non periodic case), after some nice cancellations.

Assertion (b). In the case of H -tameness, the first term $P_n(i\tau)$ is not always main term on a neighborhood $\mathcal{T}$ of zero. In the case of strong H -tameness, we operate a scale change, and we consider a variable of the form $i\tau a_n$, with $a_n = \Theta(\log n)^{-\gamma}$ with γ to be chosen soon. Then

$$n^{U(w)} = \exp \left[i\tau a_n \log n U'(0) - \frac{1}{2} \tau^2 a_n^2 \log n U''(0) + \tau^3 O(a_n^3 \log n) \right].$$

Choose now $\gamma > 1/3$ and $\gamma > (1 - \rho)/2$ where ρ is the hyperbolic exponent [for instance $\gamma = 1/2$ is a good choice]. Then the term $P_n(i\tau a_n)$ has an order strictly larger than $R_n(i\tau a_n)$ on a neighborhood $\mathcal{W}$ of 0. Furthermore the functions

$$\tau \mapsto \log \left[1 + (1 - e^{i\tau a_n}) \frac{R_n(i\tau a_n)}{P_n(i\tau a_n)} \right]$$

are analytic on a neighborhood $\mathcal{W}_1$ of zero. □

7.5.3 An expression for the normalized characteristic function.

We now describe estimates for the normalized characteristic functions which intervene in the statement of Goncharov Theorem.

Proposition 7.12. *Consider a digital tree (trie or dst) built on n words independently drawn from the same good dynamical source, and consider the function U defined by (7.12). Denote its first two derivatives as $\mu := U'(0)$ and $\nu = U''(0)$. Assume the following*

(a) *The expectation $\mu_n := \mathbb{E}[D_n]$ and the variance $\nu_n^2 := \mathbb{V}[D_n]$ satisfy*

$$\mu_n - \mu \log n = O(1), \quad \nu_n^2 - \nu \log n = O(1).$$

(b) *There exists a real τ for which the operator $s \mapsto \mathbb{T}(s, u)$ is tame at $u = e^{i\tau/\nu_n}$.*

Then, the normalized characteristic function defined as

$$\widehat{M}_n(i\tau) := \exp \left[-i\tau \frac{\mu_n}{\nu_n} \right] G_n \left(e^{i\tau/\nu_n} \right) = \exp \left[-i\tau \frac{\mu_n}{\nu_n} \right] M_n \left(\frac{i\tau}{\nu_n} \right) \quad (7.26)$$

admits the following expansion, as soon as $i\tau/\nu_n$ belongs to $\mathcal{W}$,

$$\widehat{M}_n(i\tau) = e^{-\tau^2/2} \exp[f_n(\tau)] \left[1 + O \left(\frac{\tau}{\nu_n} \right) \right] + O \left(\frac{\tau}{\nu_n} \right) \widehat{R}_n, \quad (7.27)$$

which involves a function f_n that satisfies

$$|f_n(\tau)| \leq \alpha \frac{|\tau|}{\nu_n} + \beta \frac{|\tau|^3}{\nu_n} \quad \text{for some constants } \alpha, \beta, \quad (7.28)$$

and the remainder term $\widehat{R}_n(\tau)$ depends on the type of tameness and satisfies

$$\widehat{R}_n = \begin{cases} O(n^{-\delta}) & \text{for } P \text{ or } S\text{-tameness} \\ O(\exp[-(\log n)^\rho]) & \text{for } H\text{-tameness} \end{cases}. \quad (7.29)$$

The last estimate holds in the case of weak H -tameness when τ/ν_n belongs to the tameness set $\mathcal{Q}_m$ with a denominator k_n polynomial with respect to $\log n$.

Proof. The proof follows the ideas given in the proof of Hwang. With hypothesis (b), Proposition 7.10 applies and provides a decomposition of $M_n(w)$ given in (7.20). We first isolate in the sum $P_n + Q_n$ the factor $n^{U(w)}$ and we let $w := i\tau/\nu_n$. Then, we first focus on the product

$$B_n(\tau) := \exp \left[-i\tau \frac{\mu_n}{\nu_n} \right] \cdot \exp \left[U \left(\frac{i\tau}{\nu_n} \right) \cdot \log n \right].$$

One has, when $i\tau/\nu_n$ belongs to $\mathcal{W}$,

$$\begin{aligned} B_n(\tau) &= \exp \left[-i\tau \frac{\mu_n}{\nu_n} + \left(i\mu \frac{\tau}{\nu_n} - \frac{\nu}{2} \frac{\tau^2}{\nu_n^2} + O \left(\frac{\tau^3}{\nu_n^3} \right) \right) \log n \right] \\ &= e^{-\tau^2/2} \exp \left[i(\mu \log n - \mu_n) \frac{\tau}{\nu_n} + (\nu_n^2 - \nu \log n) \frac{\tau^2}{2\nu_n^2} + O \left(\frac{\tau^3}{\nu_n^3} \log n \right) \right]. \end{aligned}$$

The expansion involves the two differences $\mu \log n - \mu_n$ and $\nu_n^2 - \nu \log n$ which are assumed to be $O(1)$ thanks to Hypothesis (a). Then

$$B_n(\tau) = e^{-\tau^2/2} \exp[f_n(\tau)]$$

where $f_n(\tau)$ satisfies the inequality in (7.28).

We now return to the normalized characteristic function $\widehat{M}_n(i\tau)$. The term which involves the remainder term has the same form as previously, except in the case of weak H -tameness where it is multiplied by a factor of the form $O(\log n)^\gamma$, which may be “absorbed” by the term of the form $\exp[-(\log n)^\beta]$: it becomes a remainder term $\widehat{R}_n$ of the same form as the previous one with a smaller β .

$$\begin{aligned}\widehat{M}_n(\tau) &= B_n(\tau) \cdot \left[1 + O\left(\frac{\tau}{\nu_n}\right) \right] + O\left(\frac{\tau}{\nu_n}\right) \cdot \widehat{R}_n \\ &= e^{-\tau^2/2} \exp[f_n(\tau)] \left[1 + O\left(\frac{\tau}{\nu_n}\right) \right] + O\left(\frac{\tau}{\nu_n}\right) \cdot \widehat{R}_n.\end{aligned}$$

□

7.6 Proofs of the main results

7.6.1 Asymptotic estimates for the mean value $\mathbb{E}[D_n]$ and the variance $\mathbb{V}[D_n]$.

We are now ready to prove Theorem 7.1.

Proof of Theorem 7.1. This is just an application of Proposition 7.11.

Assertion (c) is sufficient to conclude in all the cases, except perhaps in the periodic case (where we do not perform the computation for the derivatives of C_n).

However, we have stated Assertions (a) and (b) because they lead to the estimates without any computation. In the case of S -tameness or P -tameness, Assertion (a) is used as in the Quasi-Power Theorem, in conjunction with the Cauchy Theorem.

We can also use Assertion (b) to prove the Theorem in the case of strong H -tameness. We cannot use the Cauchy theorem, but we can directly deal with the derivatives of the remainder terms at $\tau = 0$, using Assertion (b3) of the previous proposition. All the first derivatives contain a factor a_n , and the second derivatives contain a factor a_n^2 . This leads to the result, too.

The only case when we actually need the computation of Assertion (c) is the weak H -tameness.

□

Proposition 7.13. [Log-convexity of the function λ .] *The function $s \mapsto \log \lambda(s)$ is always convex and is always strictly convex except if the source is an unbiased memoryless source.*

Proof. This is a very classical result whose proof can be found in [72] for instance.

□

We do not make precise here the expression for the subdominant constants. This will be done in Section 7.6.4 and gives rise to Figure 7.3.

7.6.2 Asymptotic Gaussian law for the depth

We are now ready to prove Assertion (a) of Theorem 7.2. This is an application of Proposition 7.12 together with the Goncharov Theorem.

Proof of Assertion (a) of Theorem 7.2. We separate two cases, according to the type of uniform tameness: the first case deals with S and P shapes, together with the strong H -tameness. Here, for any fixed τ , there exists an integer n_0 (depending on τ) for which, for $n \geq n_0$, the operator

$u \mapsto \mathbb{T}(s, u)$ is tame at $u = e^{i\tau/\nu_n}$. Then, we directly apply Proposition 7.12 for τ and any $n \geq n_0$. In the expression of (7.27), each factor of the first term,

$$\exp[f_n(\tau)], \quad 1 + O\left(\frac{\tau}{\nu_n}\right)$$

tends to 1, and the last term tends to 0. Then, we may apply Goncharov's Theorem, and the asymptotic normality is proven in this case.

It remains the case of weakly H -tameness, that is more involved. Then, the tameness hypothesis holds only for particular numbers τ , which belong to the tameness set $\mathcal{Q}_m$ already defined in (7.19)

$$\mathcal{Q}_m := \left\{ \frac{2\pi a}{mk}; \quad a \in \mathbb{Z}, k \geq 1 \right\}, \quad (7.30)$$

relative to a given integer m . The idea is to associate with each real τ a sequence τ_n which satisfies the following three conditions:

- (a) Each number τ_n/ν_n belongs to $\mathcal{Q}_m$
- (b) The difference $e^{-\tau_n^2/2} - e^{-\tau^2/2}$ tends to zero.
- (c) The difference $\widehat{M}_n(\tau_n) - \widehat{M}_n(\tau)$ tends to zero

These three conditions may be fulfilled together: We consider indeed a sequence k_n of integers [to be further chosen] and we associate the rational of denominator mk_n closest to $\tau/(2\pi\nu_n)$, that is denoted it by $\tau_n/(2\pi\nu_n)$. Then

$$\frac{\tau_n}{\nu_n} \in \mathcal{Q}_m, \quad \left| \frac{\tau}{2\pi\nu_n} - \frac{\tau_n}{2\pi\nu_n} \right| \leq \frac{1}{mk_n},$$

Then, with the Mean Value Theorem,

$$\left| e^{-\tau_n^2/2} - e^{-\tau^2/2} \right| = O\left(\frac{\nu_n}{k_n}\right), \quad \left| \widehat{M}_n(\tau) - \widehat{M}_n(\tau_n) \right| = O\left(\frac{\mu_n}{k_n}\right).$$

(The last inequality holds as the derivative of $\widehat{M}_n(\tau)$ is $O(\mu_n)$ the product of two functions of modulus 1, whose derivative is bounded by μ_n).

These two sequences tend to 0 as soon as k_n is chosen as $\Theta(\log n)^\gamma$ with $\gamma > 1$. Furthermore, since τ_n/ν_n belongs to $\mathcal{Q}_m$ and as soon as τ_n/ν_n belongs to $\mathcal{W}$, Proposition 7.12 can be applied with a remainder $\widehat{R}_n$ replaced by $\widetilde{R}_n := k_n \widehat{R}_n = O(\log n)^\gamma R_n$ which also tends to 0, and finally the sequence $\widehat{M}_n(\tau_n) - e^{-\tau_n^2/2}$ tends to zero. Now, the triangular inequality

$$|\widehat{M}_n(\tau) - e^{-\tau^2/2}| \leq |\widehat{M}_n(\tau) - \widehat{M}_n(\tau_n)| + |\widehat{M}_n(\tau_n) - e^{-\tau_n^2/2}| + |e^{-\tau_n^2/2} - e^{-\tau^2/2}|$$

together with the Goncharov Theorem leads to the asymptotic normality. $\square$

7.6.3 Speed of convergence.

It remains to prove Assertion (b) of Theorem 7.2. This is an application of Proposition 7.12 together with the Berry-Esseen Inequality.

Proof of Assertion (b) of Theorem 7.2. We will apply the Berry Esseen inequality with $T := T_n = c\nu_n$ with c small enough. We will use a refinement of the previous ideas, and as previously, we consider two cases; the easy case gathers all the tameness types except the weak H -tameness,

and the second case deals with weak H -tameness. In both cases, we consider the Berry-Essen function defined as

$$H_n(\tau) = \frac{1}{\tau} \left[\widehat{M}_n(i\tau) - e^{-\tau^2/2} \right].$$

Easy case. Proposition 7.12 together with (7.27), provides the following decomposition for the Berry-Esseen function $|H_n(\tau)|$,

$$\frac{1}{\tau} \left[\widehat{M}_n(i\tau) - e^{-\tau^2/2} \right] = \frac{1}{\tau} \left[B_n(\tau) - e^{-\tau^2/2} \right] + O\left(\frac{1}{\nu_n}\right) \cdot B_n(\tau) + O\left(\frac{1}{\nu_n}\right) \cdot R_n, \quad (7.31)$$

where $B_n(\tau)$ is written as the product $e^{-\tau^2/2} \exp[f_n(\tau)]$ with

$$f_n(\tau) \leq \frac{\alpha|\tau| + \beta|\tau^3|}{\nu_n}.$$

The difference $|B_n(\tau) - e^{-\tau^2/2}|$ equals the product $e^{-\tau^2/2} |\exp[f_n] - 1|$, and we apply the inequality $|e^x - 1| \leq x e^x$ to $x = f_n$. This gives

$$|\exp[f_n(\tau)] - 1| \leq \frac{\alpha|\tau| + \beta|\tau^3|}{\nu_n} \exp\left[\frac{\alpha|\tau| + \beta|\tau^3|}{\nu_n}\right].$$

Remark that, for $|\tau| \leq T_n := c\nu_n$ with c small enough, there exists $\gamma > 0$ for which

$$|f_n(\tau)| \leq \gamma + \frac{\tau^2}{4}, \quad \text{and thus} \quad |\exp[f_n(\tau)] - 1| \leq \frac{\alpha|\tau| + \beta|\tau^3|}{\nu_n} \exp\left[\gamma + \frac{\tau^2}{4}\right]. \quad (7.32)$$

Then, the functions involved in the first two terms of (7.31) satisfies on the interval $[-T_n, T_n]$,

$$\frac{1}{|\tau|} |B_n(\tau) - e^{-\tau^2/2}| \leq \frac{\alpha + \beta\tau^2}{\nu_n} \exp\left[\gamma - \frac{\tau^2}{4}\right], \quad |B_n(\tau)| \leq \exp\left[\gamma - \frac{\tau^2}{4}\right]$$

We thus exhibit a function

$$H^+(\tau) := (1 + \tau^2) e^{-\tau^2/4}, \quad (7.33)$$

which provides on the interval $[-T_n, +T_n]$ an upper bound for H_n , of the form

$$H_n(\tau) \leq \frac{K}{\nu_n} [H^+(\tau) + R_n]. \quad (7.34)$$

The function H^+ is integrable on $]-\infty, +\infty[$, and thus its integral on the interval $[-T_n, +T_n]$ is $O(1)$. Finally one has

$$\int_{-T_n}^{+T_n} H_n(\tau) d\tau = O\left(\frac{1}{\nu_n}\right) + O(R_n) = O\left(\frac{1}{\nu_n}\right),$$

since $O(R_n)$ is of order smaller than $O(1/\nu_n)$. Using the inequality of Berry-Esseen, this proves the asymptotic Gaussian law with a speed of convergence of order $O(1/\nu_n) = O((\log n)^{-1/2})$. This ends the proof in the easy case.

Case of the weak H -tameness. We use the same idea as in the previous proof of Assertion (a) : we consider an integer $k_n \Theta(\log n)^\gamma$ with γ to be chosen later. We associate with each real τ the rational of denominator mk_n closest to $\tau/(2\pi\nu_n)$, that is denoted it by $\tau_n/(2\pi\nu_n)$. Then the real $\tau_n/(2\pi\nu_n)$ belongs to the tameness set $\mathcal{Q}_m$. We then “replace” the function $H_n(\tau)$ by the staircase function which equals to $H_n(\tau_n)$ on the interval of length $2\pi\nu_n/(mk_n) = O(\nu_n/k_n)$ to which τ belongs.

The estimate of the integral of the function H_n over the interval $[-T_n, +T_n]$, denoted as $I[H_n, T_n]$, is obtained with the corresponding estimate of the integral $I[H^+, T_n]$ where H^+ is the upperbound previously obtained and defined in (7.33). There are three steps:

- (a) We compare $I[H_n, T_n]$ with its Riemann sum $S[H_n, T_n, s_n]$ of step $s_n = O(\nu_n/k_n)$. The error is of order $\epsilon_n^{(1)} := T_n \cdot s_n \cdot \sup |H'_n| = O(\nu_n^2 \mu_n^2 / k_n)$, since the derivative of H_n is $O(\mu_n^2)$
- (b) We get an upper bound of the Riemann sum $S[H_n, T_n, s_n]$ of H_n with the Riemann sum $S[H^+, T_n, s_n]$ of the function H^+ with applying our previous proof at each point τ_n . This is possible since the tameness holds at $u = e^{i\tau_n/\nu_n}$. This provides for $S[H_n, T_n, s_n]$ a bound of order $O(1/\nu_n) \cdot S[H^+, T_n, s_n] + O(R_n/\nu_n)$
- (c) We compare the Riemann sum $S[H^+, T_n, s_n]$ of H^+ of step $s_n = O(\nu_n/k_n)$ to the integral $I[H^+, T_n]$. The error is of order $\epsilon_n^{(2)} := T_n \cdot s_n \cdot \sup |(H^+)'| = O(\nu_n^2/k_n)$ since the derivative of H^+ is $O(1)$

We summarize

$$\begin{aligned}
 I[H_n, T_n] &\leq S[H_n, T_n, s_n] + \epsilon_n^{(1)} \\
 &\leq \frac{K}{\nu_n} S[H^+, T_n, s_n] + LR_n + \epsilon_n^{(1)} \\
 &\leq \frac{K}{\nu_n} [I[H^+, T_n] + \epsilon_n^{(2)}] + LR_n + \epsilon_n^{(1)}
 \end{aligned}$$

This proves that the integral of H_n over the interval $[-T_n, +T_n]$ is of order

$$I[H_n, T_n] = O\left(\frac{1}{\nu_n}\right) + O\left(\frac{\nu_n}{k_n}\right) + O\left(\frac{\nu_n^6}{k_n}\right).$$

If we choose $k_n = \Theta(\log n)^{7/2}$, the integral $I[H_n, T_n]$ is of order $O(1/\nu_n)$, and with the Berry-Esseen inequality, this proves a speed of convergence of order $O(1/\nu_n) = O(\log n)^{-1/2}$ $\square$

7.6.4 Expression of the subdominant constants.

A general form for the constants. The constants of the analysis are brought by the derivatives at zero of the three functions U , $\log V$ and $\log F_n$, whose expression is provided in (7.12), (7.16) and (7.17). There are also periodic terms, but we focus here on the non periodic case. They are described in Figure 7.3

All the constants of the analysis will involve the following basic constants a, b, c, d, e . The constants a and b

$$a = -\lambda'(1), \quad b = \lambda''(1)$$

are brought by the derivatives of U . The constants c, d, e are brought by the derivatives of $\log V$, and involve the derivative of the dominant projector $s \mapsto \mathbb{G}_s$ at $s = 1$, together with the two derivatives of $(s, u) \mapsto \mathbb{L}(s, u)$ at $(s, u) = (1, 1)$,

$$\mathbb{L}^{[1]} := \frac{d}{du} \mathbb{L}(s, u)|_{(1,1)}, \quad \mathbb{L}^{[2]} := \frac{d^2}{ds du} \mathbb{L}(s, u)|_{(1,1)}$$

under the form $c = \mathbb{G}'_1[1](0, 1)$, $d = \mathbb{G}_1 \circ \mathbb{L}^{[1]}[1](0, 1)$, $e = \mathbb{G}_1 \circ \mathbb{L}^{[2]}[1](0, 1)$. The Euler constant γ is brought by the derivatives of the function $\log F_n$ at 0.

(a) The dominant constants are

$$\mu = \frac{1}{a}, \quad \nu = \frac{b - a^2}{a^3}$$

(b) The subdominant constants for the mean are

$$\mu_T = \frac{b}{2a^2} + \frac{(\gamma - 1)}{a} + \frac{c}{a} \quad \mu_D = \mu_T + \frac{d}{a}$$

The inequality $\mu_T > \mu_D$ holds.

(c) The subdominant constants ν_T, ν_D of the variance are both of the form

$$\nu_X = 2(d_X + (\gamma - 1)c_X) + \mu_X - \mu_X^2$$

and involve constants c_X, d_X

$$\begin{aligned} c_T &= \frac{b}{a^3} - \frac{1}{a} + \frac{c}{a^2} & c_D &= c_T + \frac{d}{a^2} \\ d_T &= \frac{b^2}{4a^2} - \frac{b}{a^2} + \frac{bc}{a^3} & d_D &= d_T + \frac{a^2 - b}{a^3}d + \frac{d}{a^2} + \frac{e}{a} \end{aligned}$$

Figure 7.3: Here are the expressions of the constants which appear in Theorem 7.1 in the non periodic case.

Particular case of a memoryless source. In this case, all the constants are expressed with the function $\lambda(s)$ and its derivatives at integer values of s . The basic constants are:

$$\begin{aligned} a &= -\sum p_i \log p_i & b &= \sum p_i \log^2 p_i \\ c &= 0, & d &= -\sum_{k \geq 2} \frac{b_k}{1 - a_k} & e &= -\sum_{k \geq 2} \frac{b_k - da_k(1 - a_k)}{(1 - a_k)^2} \end{aligned}$$

and they themselves involve the two sequences

$$a_k = \lambda(k) = \sum p_i^k \quad b_k = -\lambda'(k) = -\sum p_i^k \log p_i$$

7.6.5 Conclusion of the Chapter.

We obtained very precise results for the limiting distribution, in all the cases: the two digital trees, the three types of tame sources. The speed of convergence towards the Gaussian law is proven to be optimal in all the cases, always of order $O(\log n)^{-1/2}$. In contrast, the remainder terms in the asymptotic estimates depend on the type of tameness. The study is classical in the case of S and P tameness, where we can use directly the Quasi-Powers Theorem, but it is more involved (and less classical) in the case of H -tameness (which is in a sense the general case for

simple sources). Then our results also apply to the case of simple sources where they provide completely proven estimates.

In fact, the proofs of the chapter can be adapted to a general framework which may arise in many instances of distributional analyses where the Rice method has to be applied, and when the lifting $\Delta(s, u)$ is built on a quasi-inverse of the form $(I - u\mathbb{H}_s)^{-1}$ which is analytic with respect to u when u is close to 1 and $\Re s$ close to 1, has nice properties when (s, u) is close to $(1, 1)$ (this is needed for studying the residues), has a nice behaviour on the left of $\Re s = 1$ when $\Im s \rightarrow \infty$ and u close to 1 (with many variations about this notion)...

Conclusion and Open problems

In this thesis, we perform the analysis of the depth of two digital trees –the trie and the digital search tree– when the words are independently emitted by a general source. This result extends, and sometimes makes more precise, the previous analyses that have been performed for simple sources.

We propose a new modelling for a general source that is based on a generalized transition matrix which extends the transition matrix of a Markov chain. We precisely define the notion of tameness for a source, we show how it intervenes in our problem, and we give three main instances for such a definition, which arise in a natural way for sources, even for simple sources.

We explain the similarities (and the differences) between the probabilistic behaviors of the two digital structures, with the similarities of their Dirichlet series. We obtain asymptotic estimates for the mean and the variance of the depth, and make precise the form of the remainder terms, in relation with the type of tameness of the source. We compare the two constant terms in the mean (for trie and dst) and we prove that the constant term of the dst is strictly smaller than the constant term of the trie: this quantifies the better efficiency of the dst comparing to the trie. We exhibit an asymptotic gaussian law for the depth, with an optimal speed of convergence.

We also obtain an exact expression of the average profile for a general source, whereas it was previously already known only for a unbiased memoryless source.

Open problems.

The present work also asks the following questions, which remain unsolved at the end of the thesis, and appear to be very interesting (even though some of them seem quite challenging).

Importance of the stationarity hypothesis for the source. For tries, the analysis does not assume the source to be stationary, because the “reverse past” is not needed in Chapter 2. We may deal with the initial source provided it defines a good dynamical source. However, in the case of dst, the analysis was only performed when the source is stationary or simple. In the case when the source has an unbounded memory, we need to deal with the “past”, and the stationary condition arises for building the “reverse past” of the source. We continue the analysis with this reverse past. This restriction does not seem completely natural, and we do not actually understand the dissymmetry between tries and dst. We thus ask the question: What happens if the source is not stationary? Is it possible to perform the analysis for a non-stationary source?

A combinatorial interpretation for our extension of q -analog formula. We have obtained a general formula which generalizes an initial formula that comes from the domain of q -analog formulae. The initial formula of q analogs can be viewed as a particular case of “our” formula when the source is an unbiased memoryless. Of course, the formula can be explained from a syntactic point of view by an expansion of an infinite product as a series (as the initial formula

does). “Our” formula does not seem to be known in combinatorics¹ and a “direct” proof does not seem to be clear inside the domain. This formula actually comes from the dst structure, and was obtained with thinking about this structure (in fact its mixed Dirichlet series) with two different points of view (with the Mellin transform at one hand and the Laplace transform at the other hand). Are there other instances of formulae which are explained with a Laplace/Mellin comparison? Is the dst the “good” combinatorial structure which explains this type of formulae?

Beginning the study of the dst profile. We have obtained an exact (but involved) formula for the average profile $B_{n,k}$ of the dst which extends the previous expression only obtained for unbiased memoryless sources; in this last case, this formula is related to the q analog formula (see [14, 42, 48], and see above) analyze asymptotically the average profile for a symmetric digital search tree under various ranges of level. We derive this formula from the additive expression of the Δ -series. There exist studies (for instance the paper [14]) which perform a precise analysis of the average dst profile for a general memoryless source without the explicit formula (since it was yet unknown...). The authors directly deal with recurrences. As our formula is quite involved, we are not sure at all if it could be a starting point of a precise asymptotic analysis. Is it the case? Is it possible to (at least) begin with this formula the study of the dst profile ?

Exact order of the remainder terms. We have obtained remainder terms in the asymptotic estimates of the mean and the variance of the depth, in the case when the source is tame. We use tameness of the source, and we shift to the left *inside* the tameness region: we do not take into account the *next* singularities of the Δ function that we would meet on the left. Of course, we may adopt another strategy, and decide to take into account the next poles (and their residues) at last for simple sources when we have some knowledge about them. This is the approach which is often used, in particular by Jacquet and Szpankowski in many of their papers. This leads to estimate a sum over residues of the same type as (7.18). Consider the particular case $w = 0$. The poles, located at s_k are now no longer regularly distributed on the line $\Re s = 1$, they are of the form $s_k = 1 + \sigma_k + it_k$ and *generally speaking* the series which represents the sum of residues is (approximately) of the form

$$\Psi(n) := \sum_{k \in \mathbb{Z}^*} \Gamma(-1 - s_k - it_k) n^{-s_k - it_k}$$

The paper [22] studies this function $\Psi(n)$ when the source is memoryless and H -tame of exponent ρ_0 . The paper proves that the *series of the moduli* is of exact order $\Theta(\exp[(\log n)^\rho])$, with $\rho \leq 1/(1 + \rho_0)$, that is the same order as our remainder. It is claimed in many papers that the function $\Psi(n)$ is of smaller order, but it seems very difficult to study it more precisely. Thus, we do not claim that we attain the right order in the remainder terms, but we do not see how to attain a smaller order.

Towards the analysis of the Lempel-Ziv algorithms for a general source. We wished to perform the analysis of compression algorithms of Lempel-Ziv type. The complexity of this algorithm strongly depends on the probabilistic behaviour of the *phrase length*, which is related to the typical depth of the dst. However, as the paper [37] shows it, the probabilistic models are not the same, as the dst is built on words that are independently drawn from the same source, whereas the Lempel-Ziv algorithm deals with a *unique* word, and all the phrases that are built from this unique word. The paper [37] explains how to deal with intermediate models, but this

¹Many thanks to Jean-Gabriel Luque for interesting discussions on the subject...

study is convenient for simple sources (with bounded memory), in which one phrase depends only on a bounded number of previous phrases. This is no longer the case for general sources. Then we tried to deal with other types of shifted sources. Previously, the source $\mathcal{S}_{(w)}$ was the source formed with the words which begin with w (from which the prefix w is removed). Now we have to deal with the source of the *first return to w* . This means that the convenient shifted source is $S_{<w>}$ which gather the words which begin with the prefix $<w>$ defined as the first return to w : The prefix $<w>$ contains two occurrences of w , and only two, and the second occurrence is at the end of $<w>$. However, the transition matrix between these LZ-shifted sources is not easy to build... And we did not succeed with this modelling of the Lempel-Ziv algorithm....

Appendix A

Analytic and meromorphic functions

We review here two important theorems in complex analysis: Cauchy's integral formula and Cauchy's residue theorem. We follow here the books of Flajolet and Sedgewick [24]; and Szpankowski[71].

A.1 Analytic functions

Analytic functions can be characterized by one of the three equivalent ways: by convergent series, by differentiability property, or by integrals vanishing on cycles.

Convergent Series. Let $f(z)$ be a function of a complex variable z . We say that $f(z)$ is analytic at point $z = a$ if it has a convergent series representation in a neighborhood of $z = a$, that is,

$$f(z) = \sum_{n \geq 0} f_n(z - a)^n, \quad z \in B(a, r),$$

where $B(a, r)$ is a ball with center a and radius $r > 0$. From this definition, if $f(z)$ is analytic at $z = a$ then there is a disk called the *disk of convergence* such that the series representing $f(z)$ is convergent inside this disk and divergent outside the disk. The radius of the disk is called *radius of convergence*.

Holomorphic function. A function $f(z)$ is holomorphic at a point $z = a$ if it has a derivative at $z = a$ defined as

$$\frac{df(z)}{dz} \Big|_{z=a} = \lim_{z \rightarrow a} \frac{f(z) - f(a)}{z - a}.$$

Thus a holomorphic function is analytic since it has a local convergent series representation.

Integrals. if $f(z)$ is an analytic function on an open simply connected set, then

$$\oint f(z) := \int_C f(z) = 0$$

along any closed path C inside this set.

A.2 Meromorphic Functions and Residues.

A quotient of two analytic functions gives a *meromorphic function* that is analytic everywhere but a set of points called *poles*, where the dominator vanishes. More formally, a meromorphic function $f(z)$ can be represented in a neighborhood of $z = a$ with $z \neq a$ by the Laurent series as:

$$f(z) = \sum_{n \geq -M} f_n(z-a)^n$$

for some integer M . We say that $f(z)$ has a *pole* of order M at $z = a$ if $f_{-M} \neq 0$.

Definition A.1. [Residue.] *The residue of $f(z)$ at a point a is the coefficient at $(z-a)^{-1}$ in the Laurent expansion of $f(z)$ around a , and it is denoted as*

$$\text{Res}[f(z); z = a] := f_{-1} = \lim_{z \rightarrow a} (z-a)f(z).$$

Lemma A.2. [Residues of meromorphic functions]

(a) *Suppose f, g are holomorphic in a region containing z_0 which is a zero of g ($g(z_0) = 0$) of multiplicity 1 and $f(z_0) \neq 0$, $g'(z_0) \neq 0$. Then f/g has a pole of order 1 at z_0 and the residue*

$$\text{Res} \left[\frac{f(z)}{g(z)}; z = z_0 \right] = \frac{f(z_0)}{g'(z_0)}$$

(b) *Suppose z_0 is a pole of f of order n . Then*

$$\text{Res}[f(z); z = z_0] = \frac{1}{(n-1)!} \lim_{z \rightarrow z_0} \frac{d^{n-1}}{dz^{n-1}} [(z-z_0)^n f(z)].$$

Cauchy Residue Theorem. Residues are very important in evaluating contour integrals as shown by the following theorem.

Theorem A.3. [Cauchy Residue Theorem] *If $f(z)$ is analytic within and on the boundary of a simple closed curve C except at a finite number of poles $a_1, a_2, \dots, a_N$ inside of C having residues $\text{Res}[f(z); z = a_j]$, $j = 1, \dots, N$, then*

$$\frac{1}{2\pi i} \int_C f(z) dz = \sum_{j=1}^N \text{Res}[f(z); z = a_j], \quad (\text{A.1})$$

where the curve C is traversed counterclockwise.

Proof. We can give the proof when f has only one pole at $z = a$. Since $f(z)$ is meromorphic, it has the Laurent expansion of the form $\sum_{n \geq -1} f_n(z-a)^n$. Then the integral over C ,

$$\int_C f(z) dz = \sum_{n \geq 0} f_n \int_C (z-a)^n dz + f_{-1} \int_C \frac{dz}{z-a} = 2\pi i \text{Res}[f(z), z = a].$$

This follows the fact that the first integral is zero, $f_{-1} = \text{Res}[f(z), z = a]$ and the last integral equal $2\pi i$. In fact, by the change of variable $z-a = e^{2\pi i t}$,

$$\int_C \frac{dz}{z-a} = 2\pi i \int_0^1 dt = 2\pi i$$

□

Cauchy Coefficient Formula. The residue theorem can be used to prove the next very crucial result.

Theorem A.4. [Cauchy Coefficient Formula] *Let $f(z)$ be analytic inside a simple connected region with C being a closed curve oriented counterclockwise that encircles the origin $z = 0$. Then, for any integer n*

$$f_n := [z^n]f(z) = \frac{1}{2\pi i} \oint_C f(z) \frac{dz}{z^{n+1}}.$$

Moreover, the following holds

$$f^{(k)}(z) = \frac{k!}{2\pi i} \oint_C \frac{f(w)dw}{(z-w)^{k+1}}$$

where $f^{(k)}(z)$ is the k th derivative of $f(z)$.

Proof.

$$f_n := [z^n]f(z) = \text{Res}[f(z)z^{-n-1}, z = 0] = \frac{1}{2\pi i} \oint_C f(z) \frac{dz}{z^{n+1}}$$

□

A.3 Gamma function.

We review here some properties of *Euler's Gamma function*. The Gamma function was first introduced by Leonhard Euler to generalize the factorial to non integer values.

Definition A.5. *The Gamma function is defined as*

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \quad \Re(z) > 0. \quad (\text{A.2})$$

We see that the above integral can generalize $n!$. By integration by part, one obtains

$$\Gamma(z+1) = - \int_0^\infty t^z d(e^{-t}) = z\Gamma(z) \quad (\text{A.3})$$

Observe that $\Gamma(1) = 1$ and for n natural $\Gamma(n) = n!$. Remark further that the equality $s\Gamma(s) = \Gamma(s+1)$ allows us to evaluate not only in the right half complex plane but also in every other point of complex plane. For example we can write $-\frac{1}{2}\Gamma(-\frac{1}{2}) = \Gamma(\frac{1}{2}) = \sqrt{\pi}$, so $\Gamma(-\frac{1}{2}) = -2\sqrt{\pi}$. However, the problem occurs with nonpositive integers. For instance, the statement $0\Gamma(0) = \Gamma(1)$ doesn't allow us to evaluate the Gamma function at 0. In fact, there is a pole of the Gamma function at 0 and more generally at every negative integer.

Analytic continuation. The Gamma function can be extended analytically to the whole complex plane. We first extend the definition (A.3) to $-1 < \Re(z) < 0$ by writing

$$\Gamma(z) = \frac{\Gamma(z+1)}{z}, \quad -1 < \Re(z) < 0.$$

$\Gamma(z+1)$ is well defined in the region $-1 < \Re(z) < 0$ since $\Re(z+1) > 0$. At $z = 0$ there is a pole whose residue is easy to evaluate; that is,

$$\text{Res}[\Gamma(z); z = 0] = \lim_{z \rightarrow 0} z\Gamma(z) = 1.$$

Now we can further extend to $-2 < \Re(z) < -1$ by applying (A.3) twice to get

$$\Gamma(z) = \frac{\Gamma(z+2)}{z(z+1)}, \quad -2 < \Re(z) < -1.$$

Observe that

$$\text{Res}[\Gamma(z); z = -1] = \lim_{z \rightarrow -1} (z+1)\Gamma(z) = -1.$$

In general, by assuming that $\Gamma(z)$ is defined up to the strip $-n < \Re(z) < -n+1$, then the extension to $-n-1 < \Re(z) < -n$ is obtained by

$$\Gamma(z) = \frac{\Gamma(z+n+1)}{z(z+1) \cdots (z+n-1)(z+n)}.$$

The residue at $z = -n$ becomes

$$\text{Res}[\Gamma(z); z = -n] = \lim_{z \rightarrow -n} (z+n)\Gamma(z) = \frac{(-1)^n}{n!}. \quad (\text{A.4})$$

for all $z = 0, -1, -2, \dots$. Therefore, the Gamma function is analytic everywhere except at $z = 0, -1, -2, \dots$.

A.4 Dirichlet series

A.4.1 Classical Dirichlet series.

Associate with the sequence (a_n) of complex numbers the series of the form

$$A(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}. \quad (\text{A.5})$$

We wish to know for which complex numbers s the series $A(s)$ is absolutely convergent? Observe the equality

$$\left| \frac{a_n}{n^s} \right| = \frac{|a_n|}{n^{\Re(s)}}.$$

This leads us to associate with $A(s)$ another Dirichlet $\underline{A}(s)$ defined for real numbers σ as

$$\underline{A}(\sigma) = \sum_{n=1}^{\infty} \frac{|a_n|}{n^{\sigma}}.$$

Therefore, if the series $\underline{A}$ converges for $\sigma = r$, it also converges for $\sigma \geq r$. Denote by r_a be a lower bound of such real numbers r , then the series $\underline{A}$ converges for $\Re(s) > r_a$ and diverges for $\Re(s) < r_a$. The real number r_a is called the *abscissa* of absolute convergence.

Theorem A.6. *Suppose that the Dirichlet series $\underline{A}(\sigma)$ is not always convergent. Then there exists a real number r_a called the abscissa of absolute convergence for which the series $\underline{A}(\sigma)$ is convergent for $\sigma > r_a$ and divergent for $\Re(s) < r_a$.*

We also consider the abscissa of simple convergence. Consider the set of real numbers σ for which the series $A(\sigma + it)$ is convergent for at least one real t . The abscissa of simple convergence r_c of the Dirichlet series $A(s)$ is the infimum of such real numbers σ . In other words, if the abscissa r_c is different from $\pm\infty$, this is the largest real σ for which the Dirichlet

series $A(s)$ is divergent on every point of the halfplane $\Re s < \sigma$. Of course, for a Dirichlet series with positive numbers, the two abscissae of convergence –simple and absolute– are equal.

A particular case of the Dirichlet series is the Riemann zeta function which is defined as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

It converges absolutely for $\Re(s) > 1$, and for $s = 1$ it diverges, so the series of the absolute diverges for $\Re(s) < 1$. In fact, Dirichlet series converges absolutely for all $\Re(s) > 1$ as long as $|a_n|$ is bounded for all $n \geq 1$

A.4.2 Our Dirichlet series

The Λ series, defined as

$$\Lambda(s) = \sum_{w \in \Sigma^*} p_w^s$$

is a generalization of the classical Dirichlet series with positive coefficients. And it is possible to define the abscissa of absolute convergence, as the infimum of all the real numbers σ for which the series $\Lambda(\sigma)$ is convergent. We know that the series is divergent at $s = 1$. Then the abscissa of absolute convergence $r_a(\Lambda)$ satisfies $r_a(\Lambda) \geq 1$. And on any halfplane of the form $\Re s \geq \sigma > r_a$, the series is normally convergent and defines an analytic function of s .

We are interested in the thesis (particularly in Chapter 3.) in the case when the Dirichlet series $\Lambda(s)$ has an abscissa of absolute convergence equal to 1. And the tameness studies in which domains on the left of the vertical line $\Re s = 1$, the series $\Lambda(s)$ is convergent, without being absolutely convergent.

Bibliography

- [1] BALADI, V., AND VALLÉE, B. Euclidean algorithms are Gaussian. *Journal of Number Theory* 110 (2005), 331–386.
- [2] BOURDON, J. Size and path length of Patricia tries: Dynamical sources context. *Random Struct. Algorithms* 19, 3-4 (2001), 289–315.
- [3] BRONSHTEIN, I., SEMENDYAYEV, K., MUSIOL, G., AND MUEHLIG, H. *Handbook of Mathematics*. Springer, 2004.
- [4] CESARATTO, E., AND VALLÉE, B. Gaussian distribution of trie depth for strongly tame sources. *Combinatorics, Probability and Computing* 24, Special Issue 01 (January 2015), 54–103.
- [5] CLÉMENT, J. *Arbres Digitaux et Sources Dynamiques*. PhD thesis, Université de Caen, 2000.
- [6] CLÉMENT, J., FILL, J. A., NGUYEN THI, T., AND VALLÉE, B. Towards a realistic analysis of the QuickSelect algorithm. *Theory of Computing Systems Special issue for STACS 2013* (To appear).
- [7] 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.
- [8] 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. In *Proceedings of STACS* (2013), pp. 598–609.
- [9] CLÉMENT, J., NGUYEN THI, T. H., AND VALLÉE, B. Towards a realistic analysis of some popular sorting algorithms. *Combinatorics, Probability and Computing* 24, Special Issue 01 (January 2015), 104–144.
- [10] DE LA BRIANDAIS, R. File searching using variable length keys. In *Papers Presented at the the March 3-5, 1959, Western Joint Computer Conference* (New York, NY, USA, 1959), IRE-AIEE-ACM '59 (Western), ACM, pp. 295–298.
- [11] DEVROYE, L. A note on the average depth of tries. *Computing* 28, 4 (1982), 367–371.
- [12] DOLGOPYAT, D. On decay of correlations in Anosov flows. *Ann. of Math.* 147, 2 (1998), 357–390.
- [13] DOLGOPYAT, D. Prevalence of rapid mixing in hyperbolic flows. *Ergod. Th. & Dynam. Sys.* 18 (1998), 1097–1114.

- [14] DRMOTA, M., AND SZPANKOWSKI, W. The Expected Profile of Digital Search Trees. *J. Combinatorial Theory, Ser. A*, 118 (2011), 1939–1965.
- [15] DURRETT, R. *Probability: Theory and Examples*, second ed. Wadsworth, 1996.
- [16] FAGIN, R., NIEVERGELT, J., PIPPENGER, N., AND STRONG, H. R. Extendible hashing—a fast access method for dynamic files. *ACM Trans. Database Syst.* 4, 3 (Sept. 1979), 315–344.
- [17] FAYOLLE, G., FLAJOLET, P., AND HOFRI, M. On a functional equation arising in the analysis of a protocol for a multiaccess broadcast channel. *Advances in Applied Probability* 18 (1986), 441–472.
- [18] FELLER, W. *An Introduction to Probability Theory and its Applications*, vol. II. John Wiley & Sons, 1971.
- [19] FLAJOLET, P. The ubiquitous digital tree. In *STACS 2006, 23rd Annual Symposium on Theoretical Aspects of Computer Science, Marseille, France, February 23-25, 2006, Proceedings* (2006), vol. 3884 of *Lecture Notes in Computer Science*, Springer, pp. 1–22.
- [20] FLAJOLET, P. A journey between Rice, Mellin and Poisson. *Personal communication* (2008).
- [21] FLAJOLET, P., AND RICHMOND, B. Generalized Digital Trees and their difference differential equations. *Random Structures and Algorithms* vol 3 (1992).
- [22] FLAJOLET, P., ROUX, M., AND VALLÉE, B. Digital trees and memoryless sources: from arithmetics to analysis. In *Proceedings of AofA’10, DMTCS* (2010), proc AM, pp. 231–258.
- [23] FLAJOLET, P., AND SEDGEWICK, R. Digital Search Trees revisited. *Siam J. Comput* Volume 15, 3 (August 1986), 748 – 767.
- [24] FLAJOLET, P., AND SEDGEWICK, R. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [25] 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.
- [26] FUCHS, M., LEE, C.-K., AND PRODINGER, H. Approximate Counting via the Poisson-Laplace-Mellin Method. *DMTCS Proceedings*, 01 (2012), 13–28.
- [27] GONNET, G. H., AND BAEZA-YATES, R. *Handbook of Algorithms and Data Structures: in Pascal and C*, second ed. Addison-Wesley, 1991.
- [28] HUBALEK, F. On the variance of the internal path length of generalized digital trees—the mellin convolution approach. *Theoretical computer science* 242, 1 (2000), 143–168.
- [29] HUN, K., AND VALLÉE, B. Typical Depth of a Digital Search Tree built on a general source. In *Proceedings of ANALCO’14* (2014), SIAM, pp. 1–15.
- [30] HWANG, H. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics* 19 (1998), 329–343.

- [31] JACQUET, P., AND RÉGNIER, M. Limiting Distributions for Trie Parameters. In *Lecture Notes in Computer Science*. 214, pp. 196–210, 1986.
- [32] JACQUET, P., AND RÉGNIER, M. Trie partitioning process: Limiting distributions. In *Lecture Notes in Computer Science, Lecture Notes in Computer Science, Lecture Notes in Computer Science* (1986), N.-Y. Springer Verlag, Ed., vol. 214, pp. 196–210.
- [33] JACQUET, P., AND SZPANKOWSKI, W. Analysis of digital tries with Markovian Dependency. *IEEE Trans. Information Theory* 37 (1991), 1470–1475.
- [34] JACQUET, P., AND SZPANKOWSKI, W. Analytical de-Poissonization and its applications. *Theoretical Computer Science* 201, 1-2 (1998), 1–62.
- [35] JACQUET, P., AND SZPANKOWSKI, W. Entropy computations for discrete distributions: towards analytic information theory. *IEEE International Symposium on Information Theory* (1998).
- [36] JACQUET, P., AND SZPANKOWSKI, W. Limiting distribution of lempel ziv’78 redundancy. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on* (July 2011), pp. 1509–1513.
- [37] JACQUET, P., SZPANKOWSKI, W., AND TANG, J. Average Profile of the Lempel-Ziv Parsing Scheme for a Markovian Source. *Algorithmica*, 31 (2001), 318–360.
- [38] JR., E. G. C., AND EVE, J. File structures using hashing functions. *Commun. ACM* 13, 7 (1970), 427–432.
- [39] KATO, T. *Perturbation Theory for Linear Operators*. Springer-Verlag, 1980.
- [40] KIRSCHENHOFER, P., AND PRODINGER, H. Some further results on digital search trees. In *Automata, Languages and Programming*, L. Kott, Ed., vol. 226 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 1986, pp. 177–185.
- [41] KIRSCHENHOFER, P., PRODINGER, H., AND SZPANKOWSKI, W. Digital Search Trees Again Revisited: The Internal Path Length Perspective. *SIAM Journal on Computing* 23, 3 (1994), 598–616.
- [42] KNESSL, C., AND SZPANKOWSKI, W. On the average profile of symmetric digital search trees. *Online Journal of Analytic Combinatorics* 4 (2009).
- [43] KNUTH, D. E. *The Art of Computer Programming: Sorting and Searching.*, third ed., vol. 3. Addison-Wesley, 1998.
- [44] KONHEIM, A. G., AND NEWMAN, D. J. A note on growing binary trees. *Discrete Math.* 4, 1 (Jan. 1973), 57–63.
- [45] LAPIDUS, M., AND VAN FRANKENHUIJSEN, M. *Fractal Geometry, Complex Dimensions and Zeta Functions: Geometry and Spectra of Fractal Strings*. Springer, 2006.
- [46] LECKEY, K., NEININGER, R., AND SZPANKOWSKI, W. Towards more realistic probabilistic models for data structures: the external path length in tries under the Markov model. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms* (2013), SIAM, pp. 877–886.

- [47] LOUCHARD, G. Exact and asymptotic distributions in digital and binary search trees. *RAIRO - Theoretical Informatics and Applications - Informatique Théorique et Applications* 21, 4 (1987), 479–495.
- [48] LOUCHARD, G., AND SZPANKOWSKI, W. Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm. *IEEE Transactions on Information Theory* 41, 2 (March 1995), 478–488.
- [49] LOUCHARD, G., SZPANKOWSKI, W., AND TANG, J. Average profile of the generalized digital search tree and the generalized Lempel-Ziv algorithm,. *SIAM J. Computing* 28 (1999), 935–954.
- [50] MAHMOUD, H. *Evolution of Random Search Trees*. John Wiley, 1992.
- [51] MENDELSON, H. Analysis of extendible hashing. *Software Engineering, IEEE Transactions on*, 6 (1982), 611–619.
- [52] NGUYEN THI, T. *Towards a realistic analysis of sorting and searching algorithms*. PhD thesis, Université de Caen, 2014.
- [53] NÖRLUND, N. E. Leçons sur les équations linéaires aux différences finies. In *Collection de monographies sur la théorie des fonctions*. Gauthier-Villars, Paris, 1929.
- [54] NÖRLUND, N. E. *Vorlesungen über Differenzenrechnung*. Chelsea Publishing Company, New York, 1954.
- [55] PARK, G., HWANG, H.-K., NICODÈME, P., AND SZPANKOWSKI, W. Profile of Tries. *Siam J. Comput.* Vol 38 (5) (2009).
- [56] PITTEL, B. Paths in a random digital tree: Limiting distributions. *Advances in Applied Probability* (1986), 139–155.
- [57] PRODINGER, H. Approximate counting via Euler transform. *Mathematica Slovaca* 44, 5 (1994), 569–574.
- [58] PU, I. M. *Fundamental of Data Compression*. Elsevier, 2006.
- [59] RÉGNIER, M., AND JACQUET, P. New results on the size of tries. *Information Theory, IEEE Transactions on* 35, 1 (Jan 1989), 203–205.
- [60] ROUX, M. *Séries de Dirichlet, Théorie de l’information, et Analyse d’algorithmes*,. PhD thesis, University of Caen, 2011.
- [61] ROUX, M., AND VALLÉE, B. Information theory: Sources, Dirichlet series, and realistic analyses of data structures. In *Proceedings 8th International Conference Words 2011* (2011), vol. 63 of *EPTCS*, pp. 199–214.
- [62] RUELLE, D. Dynamical zeta functions and transfer operators. *Notices Amer. Math. Soc* 69 (2002), 887–895.
- [63] SALOMON, D. *A concise Introduction fo Data Compression*. Springer, 2008.
- [64] SAYOOD, K. *Introduction to Data Compression*, second ed. ACADEMIC PRESS, 2000.

- [65] SCHACHINGER, W. Limiting distributions for the costs of partial match retrievals in multidimensional tries. *Random Structures & Algorithms* 17, 3-4 (2000), 428–459.
- [66] SEDGEWICK, R. *Algorithms in C: Fundamentals, Data Structures, Sorting, Searching*, third ed ed. Addison-Wesley, 1998.
- [67] SEDGEWICK, R., AND FLAJOLET, P. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, 1996.
- [68] STORER, J. A. *Data Compression: methods and theory*. Computer science press, 1988.
- [69] SZPANKOWSKI, W. Some Results on V-ary Asymmetric tries. *Journal of Algorithms*, 9 (1988), pp. 224–244.
- [70] SZPANKOWSKI, W. A Characterization of Digital Search Trees From the Successful Search Viewpoint. *Theoretical Computer Science*, 85 (1991), 117–134.
- [71] SZPANKOWSKI, W. *Average-Case Analysis of Algorithms on Sequences*. John Wiley, 2001.
- [72] VALLÉE, B. Dynamical sources in information theory: Fundamental intervals and word prefixes. *Algorithmica* 29, 1/2 (2001), 262–306.
- [73] VALLÉE, B. Théorie de l’information: Modèles, Algorithmes, Analyse. In *Notes du Mini-Cours, EJC du GDR IM, Chambéry, France*. EJCIM, 2010.
- [74] VALLÉE, B., CLÉMENT, J., FILL, J. A., AND FLAJOLET, P. The number of symbol comparisons in Quicksort and Quickselect. In *ICALP (1)* (2009), pp. 750–763.
- [75] YAO, A. C. A note on the analysis of extendible hashing. 84–86.
- [76] ZIV, J., AND LEMPEL, A. A universal algorithm for sequential data compression. *IEEE Transactions on information theory* 23, 3 (1977), 337–343.
- [77] ZIV, J., AND LEMPEL, A. Compression of individual sequences via variable-rate coding. *Information Theory, IEEE Transactions on* 24, 5 (1978), 530–536.

Résumé. Cette thèse effectue des analyses probabilistes de la profondeur des arbres digitaux [tries et arbres digitaux de recherche(dst)] quand ils sont construits sur des mots émis par une source générale. Cette étude est liée à des algorithmes de compression de type Lempel-Ziv qui sont basés sur l'utilisation d'arbres digitaux (tries or dst). La complexité des algorithmes qui utilisent ces structures de données sont liés à la forme de ces arbres, et nous sommes ici intéressés par le comportement probabiliste d'un paramètre important, la profondeur typique ou la profondeur. Nous introduisons un nouveau point de vue sur les sources générales et nous nous concentrons alors sur le modèle des sources dynamiques. La source intervient dans l'analyse par sa "tameness", et nous définissons des notions précises de "tameness" qui sont nouvelles. La thèse utilise des méthodes de combinatoire analytique et nous introduisons des fonctions génératrices (de type Dirichlet), qui caractérisent le comportement de l'arbre (trie ou dst) quand il est construit sur la source. Comme la source est une source dynamique, nous effectuons une analyse dynamique, qui mélange de manière originale des méthodes de la combinatoire analytique et des méthodes de la théorie des systèmes dynamiques (i.e. les opérateurs de transfert et leurs propriétés spectrales). Nous utilisons également de nombreux objets et méthodes de la combinatoire analytique classique, comme les transformées de Poisson, Laplace, et Mellin, que nous mélangeons d'une nouvelle façon. Nous fournissons également un point de vue unifié sur l'analyse des deux types d'arbres digitaux, alors que les analyses classiques sont dédiées à un type précis d'arbres. Enfin, nous montrons que, pour les deux types d'arbres digitaux, pour une large classe de sources, la profondeur typique suit asymptotiquement une loi gaussienne, avec une vitesse de convergence optimale.

Mots clés: *Combinatoire analytic, Arbres digitaux, Profondeur d'un arbre digital, Sources, Systèmes dynamiques, "Tameness" d'un source.*

Abstract. This thesis performs probabilistic analyses of the depth of digital trees [tries and digital search trees (dst)] when they are built on words emitted by a general source. This study is related to compression algorithms of Lempel-Ziv type which are based on the use of digital trees (tries or dst). The complexity of algorithms which use these data structures are related to the shape of these trees, and we are here interested by the probabilistic behaviour of an important parameter, the typical depth, or depth. We introduce a new point of view on general sources, and we then focus on the model of dynamical sources. The source intervenes in the analysis via its tameness, and we define precise notions of tameness which are new. The thesis deals with methods in analytic combinatorics, and we introduce (Dirichlet) generating functions, which characterize the behaviour of the tree (trie or dst) when it is built on the source. As the source is a dynamical source, we perform a dynamical analysis, which mixes in an original setting methods from analytic combinatorics and methods from dynamical system theory (namely transfer operators, and their spectral properties). We also use many objects and methods from classical analytic combinatorics, as Poisson, Laplace, and Mellin transforms, that we mix in a new way. We also provide an unified point of view on the analysis of the two types of digital trees (tries and dst), whereas the classical analyses are dedicated to one of the precise types of trees. Finally, we prove that, for the two types of digital trees, for a large class of sources, the typical depth follows an asymptotic gaussian law, with an optimal speed of convergence.

Key words: *Analytic combinatorics, Digital trees, Depth of a digital tree, Sources, Dynamical systems, Tameness of a source*
