On the Stack-Size of General Tries

Jérémie BOURDON[(1)], Markus NEBEL[(2)], Brigitte VALLÉE[(1)]

[(1)] GREYC, Université de Caen, F-14032 Caen (France)
[(2)] FB Informatik, Johann Wolfgang Goethe-Universität,
D-60054 Frankfurt a. M. (Germany)

Contact author Jérémie BOURDON
e-mail: bourdon@info.unicaen.fr, fax=+33 (0)2 31 56 73 30

Abstract

Digital trees or tries are a general purpose flexible data structure that implements dictionaries built on words. The present paper is focussed on the average-case analysis of an important parameter of this tree-structure, i.e., the stack-size. The stack-size of a tree is the memory needed by a storage-optimal preorder traversal. The analysis is carried out under a general model in which words are produced by a source (in the information-theoretic sense) that emits symbols. Under some natural assumptions that encompass all commonly used data models (and more), we obtain a precise average-case and probabilistic analysis of stack-size. Furthermore, we study the dependency between the stack-size and the ordering on symbols in the alphabet: we establish that, when the source emits independent symbols, the optimal ordering arises when the most probable symbol is the last one in this order.

Keywords: Average-case Analysis of data-structures, Information Theory, Trie, Mellin analysis.

AMS Subject Classification: 68P05, 68W40, 94A15.

1 Introduction

Digital trees or tries are a versatile data structure that implements dictionary operations on sets of words (namely insert, delete and query), as well as set-theoretic operations like set union or set intersection. The idea can be traced back to Fredkin [8] who coined the name *trie* as a hybrid between "tree" and "retrieval". As an abstract structure, tries are based on a splitting according to symbols encountered in words: if X is a set of words, and $\Sigma = \{a_1, \ldots, a_r\}$ is the alphabet, then the trie associated with X is defined recursively by the rule:

$$\mathrm{Tr}X = \langle \mathrm{Tr}(X \setminus a_1), \ldots, \mathrm{Tr}(X \setminus a_r) \rangle,$$

where $X \setminus \alpha$ is the subset which collects all the suffixes of those words of X that begin with symbol α; recursion is halted as soon as X contains less than two elements. The advantage of the trie is that it only keeps the minimal prefix set of symbols that is necessary to distinguish all the elements of X. Figure 1 (page 4) shows an example of a trie built on a set of 7 words on the alphabet $\{a, b, c\}$.

Evidently, the tree $\mathrm{Tr}X$ supports the search for any word w in the set X by following an access path dictated by the successive symbols of w. By similar means, the trie implements insertions and deletions, so that it is a fully dynamic dictionary data type. In addition, tries efficiently support set-theoretic operations like union and intersection [27], as well as partial match queries or interval search [7, 23], and suitable adaptions make them a method of choice for complex text processing tasks [9, Ch. 7]. These various applications justify to consider the trie structure as one of the central general-purpose data structures of computer science [9, 15, 17, 24]. The cost of the main operations is measured by three parameters of the tree structure —height, number of internal nodes and external path length— that have been already precisely analyzed. Knuth's book, vol. 3, [15] contains the first analyses of parameters of tries, though these are restricted to additive parameters (number of internal nodes and external path length) in an essential way. The first works regarding trie height are due to Yao [30], Régnier [21, 22], Flajolet [5], and Szpankowski [25, 26]. When it appeared in 1992, Mahmoud's book [17] gave a general synthesis on trie analyses and the current state of knowledge.

This paper is devoted to the average–case analysis of another important parameter, namely the stack-size. When choosing an order on the symbols of the alphabet, a preorder traversal of the trie TrX gives the list of the words of X in the lexicographical order. When the traversal is implemented in a recursive way, the height of the trie exactly measures the recursion-depth needed. However, very often, the recursion is removed, or the technique known as end-recursion removal is used for saving recursive calls [24]. In this case, the last subtrie (i.e., the subtrie that is relative to the last symbol in the alphabet order) is not "put on the stack". Then, the amount of memory that is needed is no more measured by the height. Another parameter that is called the stack-size is now convenient: it is a kind of "biased" height where any edge whose label is the last symbol has zero cost. This last symbol is in a sense "excluded".

For tries, this parameter of interest has not been extensively studied. Recently, Nebel in [18] and [19] has adopted a combinatorial point of view where all possible shapes of binary tries of a given size are taken with equal probability. Nebel's model is of interest from the standpoint of combinatorial analysis. For other classes of trees, there is a rich literature related to the stack-size and similar notions of height that starts with the historical paper of De Bruin, Knuth and Rice [3] where the average height of planted plane trees was considered (note, that by the rotation correspondence this parameter equals the stack-size of binary trees). Further results can be found in [14] and [13] and in the references given there.

A parameter similar to the stack-size is the so called Horton-Strahler number of a tree. It specifies the recursion-depth needed for a traversal when end-recursion removal is applied and the subtrees are visited in no fixed order; the order is chosen such that the recursion depth is minimal. Other applications of this parameter are related to geology, molecular biology, synthetic images of trees and channel networks (see [29] for details). For tries in the Bernoulli model, the same parameter has been studied by Devroye and Kruszewski [4] and Nebel [20] provides results in a combinatorial model.

The present paper considers the natural model for characterizing the performance of the trie data structure. First of all, we specify how the words used to generate the (random) trie are produced. We consider here an alphabet $\Sigma := \{a_1, a_2, \ldots a_r\}$ of cardinality r (possibly infinite) and a mechanism $\mathcal{S}$ (called a source in information theory contexts) that produces infinite sequences of symbols of Σ. Such an infinite sequence is called an (infinite) word and the set of words is denoted by Σ^∞. Here, the source is quite general: it may emit independent symbols, and we call it a memoryless source; it may be a Markov chain, where each symbol may only depend on a fixed number of previous symbols. However, the dependence between emitted symbols may be even more general and involve unbounded part of past history. A class of such sources that come from dynamical systems is introduced in [28], and we refer to them as dynamical sources. Our probabilistic model is then the so-called Bernoulli model of size n denoted by $\mathcal{B}(n, \mathcal{S})$: it considers all possible sets X of fixed cardinality n formed with independent source words. We aim to analyze the probabilistic behavior of the stack-size $s(X)$ of trie TrX when the cardinality n becomes large.

In this paper, we explain why the stack-size and the height have very similar behavior. In particular, we show that the stack-size is nothing else than a modified height. More precisely, we prove that the stack-size is exactly the height relative to an alphabet where one of the symbols is "excluded". In a recent paper [1], Clément, Flajolet and Vallée proved the following results on the analysis of trie height:

For "perfect" sources $\mathcal{S}$ in the Bernoulli model $\mathcal{B}(n, \mathcal{S})$, the trie height has an expectation of order $\log n$ and its probability distribution is asymptotically of the double exponential type. More precisely,

$$E[h_n] = \frac{2}{|\log c(\mathcal{S})|}\log n + Q_{\mathcal{S}}(\log n) + \left[\frac{\gamma + \log \rho(\mathcal{S})}{|\log c(\mathcal{S})|} + \frac{1}{2}\right] + o(1),$$

$$\lim_{n\to\infty} \sup_{k\geq 0} \left|\Pr\{h_n \leq k\} - \exp[-\rho(\mathcal{S})\ c(\mathcal{S})^k\, n^2]\right| = 0.$$

Here, γ is the Euler constant; furthermore, $c(\mathcal{S})$ and $\rho(\mathcal{S})$ are two positive constants, and $Q_{\mathcal{S}}(u)$ is a periodic function of very small amplitude; all depend on the source $\mathcal{S}$.

In this paper, we show how the same methods can be adapted to the study of this modified height. Those are our results concerning stack size:

For "perfect" sources $\mathcal{S}$ in the Bernoulli model $\mathcal{B}(n,\mathcal{S})$, the m–stack–size relative to the excluded symbol m has an expectation of order $\log n$ *and its probability distribution is asymptotically of the double exponential type. More precisely,*

$$E[s_n^{[m]}] = \frac{2}{|\log c_m(\mathcal{S})|}\log n + Q_{m,\mathcal{S}}(\log n) + \left[\frac{\gamma + \log\rho_m(\mathcal{S})}{|\log c_m(\mathcal{S})|} + \frac{1}{2}\right] + o(1),$$

$$\lim_{n\to\infty} \sup_{k\geq 0} \left|\Pr\{s_n^{[m]} \leq k\} - \exp[-\rho_m(\mathcal{S})\ c_m(\mathcal{S})^k\, n^2]\right| = 0.$$

Here, γ is the Euler constant; furthermore, $c_m(\mathcal{S})$ and $\rho_m(\mathcal{S})$ are two positive constants, $Q_{m,\mathcal{S}}(u)$ is a periodic function of very small amplitude; all depend on the source $\mathcal{S}$ and the "excluded" symbol m.

The constants that intervene in the analysis of the stack-size or the height for classical sources are explicit. This is the case for all memoryless sources, and, in particular, for sources which emit independent and equiprobable symbols. In the latter case, all the choices of "excluded" symbol entail the same constants $\tilde{c}(\mathcal{S}), \tilde{\rho}(\mathcal{S})$ for trie stack-size to be compared to the constants $c(\mathcal{S})$ and $\rho(\mathcal{S})$ that intervene in the height

$$\tilde{c}(\mathcal{S}) = \frac{1}{r+1}, \qquad c(\mathcal{S}) = \frac{1}{r}, \qquad \tilde{\rho}(\mathcal{S}) = \rho(\mathcal{S}) = \frac{1}{2}.$$

Thus, for the usual unbiased binary trie, one obtains for the expected stack-size

$$E[s_n] = \frac{2}{\log 3}\log n + \tilde{Q}(\log n) + \left[\frac{\gamma - \log 2}{\log 3} + \frac{1}{2}\right] + o(1),$$

to be compared to the height

$$E[h_n] = \frac{2}{\log 2}\log n + Q(\log n) + \left[\frac{\gamma}{\log 2} - \frac{1}{2}\right] + o(1).$$

When the memoryless source emits r symbols $a_1, a_2, \ldots, a_r$ with respective probabilities $p(a_1), p(a_2), \ldots p(a_r)$ (r possibly infinite), then the constants $c_m(\mathcal{S}), \rho_m(\mathcal{S})$ of trie stack-size are explicit too. We compare those constants to the constants $c(\mathcal{S})$ and $\rho(\mathcal{S})$ that intervene in the height,

$$c(\mathcal{S}) = \sum_{j=1}^{r} p(a_j)^2, \qquad c_m(\mathcal{S}) = \frac{1}{1-p(m)^2}\sum_{j|a_j\neq m} p(a_j)^2 = 1 - \frac{1-c(\mathcal{S})}{1-p(m)^2} \qquad \rho_m(\mathcal{S}) = \rho(\mathcal{S}) = \frac{1}{2}.$$

These expressions prove the following (intuitive) fact : for a fixed memoryless source, the constant $|\log c_m(\mathcal{S})|$ is an increasing function of the "excluded probability" $p(m)$, so that the optimal choice for the excluded symbol arises when this symbol is the most probable one. This intuitive fact seems to be true for more general sources, and we state as a conjecture.

The continued fraction expansion is an important example of a dynamical source with unbounded memory. The height of tries has been studied in this context in [1] and involves the constant $c(\mathcal{S}) \doteq 0.19945881818343767$, sometimes known as "Vallée's constant". This constant also intervenes in various two-dimensional generalizations of the Euclidean algorithm [2]. Numerical investigations provide approximate values of the sequence of constants $(c_m(\mathcal{S}))_{m\geq 1}$. This sequence appears to be increasing and tends to $c(\mathcal{S})$ as $m\to\infty$. Here are the four first values for $c_m(\mathcal{S})$,

$$c_1 \sim 0.055, \quad c_2 \sim 0.173, \quad c_3 \sim 0.191, \quad c_4 \sim 0.196,$$

Plan of the paper. We first recall some basic facts about the trie structure and its main parameters; we introduce the parameter of interest, and describe the two probabilistic models (the Poisson model and the Bernoulli model) that are used later. Then, we show that the series of prefix probabilities play a fundamental rôle and we introduce the notion of perfect sources that provides a general framework where the series of prefix probabilities is well–behaved. We analyze the stack-size in that context, first in the Poisson model, then in the Bernoulli model.

2 A first approach for analyzing the trie stack-size.

2.1. Trie structure. We recall our framework: we consider here an alphabet $\Sigma := \{a_1, a_2, \dots a_r\}$ of cardinality r (finite or denumerable) and a source $\mathcal{S}$ which could be of a quite general type. We deal with the problem of comparing n infinite words independently produced by the same general source by comparing their prefixes. The classical underlying structure is a tree, called a trie [15, 17].
With any finite set X of infinite words produced by the same source, we associate a *trie*, $\mathrm{Tr}X$, defined by the following recursive rules:

(R_0) If $X=\emptyset$, then $\mathrm{Tr}X$ is the empty tree,

(R_1) If $X=\{x\}$ has a cardinality equal to 1, then $\mathrm{Tr}X$ consists of a single *leaf node* represented as $\square$ that contains x,

(R_2) If X has a cardinality at least 2, then $\mathrm{Tr}X$ is an *internal node* represented generically by 'o' to which are attached r subtrees,

$$\mathrm{Tr}X = \langle o, \mathrm{Tr}X(a_1), \mathrm{Tr}X(a_2), \dots, \mathrm{Tr}X(a_r)\rangle,$$

where $X(a_j)$ collects all the suffixes of those words in X that begin with a first symbol equal to a_j. The edge that attaches the subtrie $\mathrm{Tr}X(a_j)$ is labelled by the symbol a_j.

Such a tree structure underlies the classical radix sorting methods. It can be built by following the recursive rules R_0, R_1, R_2. Its internal nodes are closely linked to prefixes of words of X. More precisely, each internal node of $\mathrm{Tr}X$ corresponds to a prefix w shared by at least two words of X. In the sequel, the probability p_w that an infinite source word begins with prefix w plays an important rôle. Figure 1 shows an example of a trie built on a set of 7 words on the alphabet $\{a,b,c\}$. The prefixes used in the trie are $\{a, b, ca, cb, cca, ccb, ccc\}$

2.2. Height and stack-size. The depth of a node v in a trie is the number of edges that connect v with the root: it is also the length of the prefix that labels the path from the root to v. The height of the trie is the maximum level of any leaf. It is a measure of the depth needed for a recursive preorder traversal of $\mathrm{Tr}X$ and is recursively defined as follows

$$h(\mathrm{Tr}X) := \begin{cases} 0 & \text{for } |X|\leq 1, \\ 1+\max\{\, h(\mathrm{Tr}X(a_1)), h(\mathrm{Tr}X(a_2)), \dots, h(\mathrm{Tr}X(a_{r-1})), h(\mathrm{Tr}X(a_r))\,\} & \text{otherwise.}\end{cases}$$

The height of the trie of Figure 1 equals 3. The trie $\mathrm{Tr}X$ has height at most k provided that there exists no word $w\in\Sigma^k$ that is the common prefix of two words of X.

We now choose an order on the alphabet, so that the trie becomes a planar tree. When the alphabet is finite, the last symbol with respect to this order, denoted by m for $m = a_r$, plays a special rôle in a preorder

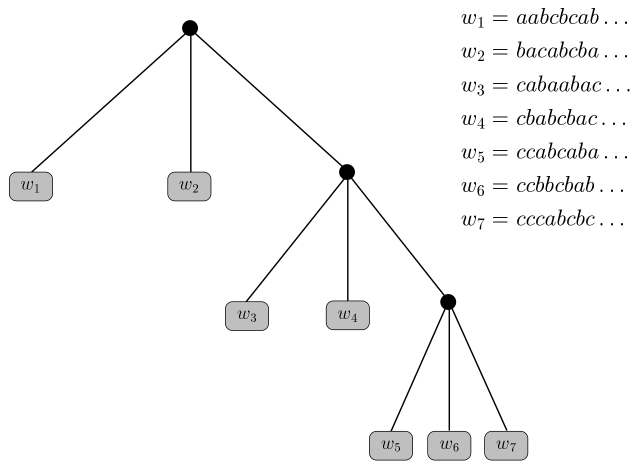


Figure 1: An example of ternary trie built on the set $\{w_1, \dots, w_7\}$.

traversal where the last recursive call is removed, since, in this case, the corresponding subtrie is not put on the stack. According to this observation, the stack-size is recursively defined as follows

$$s(\mathrm{Tr}X) := \begin{cases} 0 & \text{for } |X| \le 1, \\ \max\{1 + \max\{\, s(\mathrm{Tr}X(a_1)), s(\mathrm{Tr}X(a_2)), \dots, s(\mathrm{Tr}X(a_{r-1}))\,\}\ ,\ s(\mathrm{Tr}X(a_r))\,\} & \text{otherwise.} \end{cases}$$

More generally, in any alphabet (finite or infinite), we can "exclude" any symbol m, and we denote the stack-size relative to the last excluded symbol m by $s^{[m]}$. Then

$$s^{[m]}(\mathrm{Tr}X) = \begin{cases} 0 & \text{for } |X| \le 1, \\ \max\{1 + \max_{j|a_j \neq m}\{\, s(\mathrm{Tr}X(a_j))\,\}\ ,\ s(\mathrm{Tr}X_m)\,\} & \text{otherwise.} \end{cases}$$

For the trie of Figure 1, one has $s^{[a]} = 3, s^{[b]} = 3, s^{[c]} = 1$. We denote by $|w|_m$ the number of symbols of w that are distinct from m, and we call it the m–length of prefix w. For instance, one has $|ca|_c = 1, |ca|_b = 2$. In the same vein, the m–level of a node is the number of edges whose label is distinct from m. Then, the stack-size of the trie relative to the choice of symbol m as the last symbol is the maximum m–level of any leaf. Leaves relative to a prefix w which ends with symbol m do not intervene in the stack-size, because there exists a sister–node related to a longer m–level; more precisely, a leaf that is labelled by a prefix of the form $w = xm$ has a sister-node (internal or external) of the form $w' = xn$ with $n \neq m$. Then $|w'|_m = |w|_m + 1$. Thus, the leaves that are "useful" for evaluating the stack-size are related to finite prefixes that do not end with symbol m. For instance, the leaf w_4 (whose label is the prefix cb of b-level 1) is not useful for the b-stack-size of Figure 1, since its sister-nodes of label cc or ca are of b–level 2.

2.3. Change of the alphabet. As already seen, a useful prefix is a finite word that ends with a symbol distinct from m. Such a word is an element of the set

$$\Gamma_m^* \qquad \text{with} \qquad \Gamma_m := \{m\}^*\ [\Sigma \setminus \{m\}]. \tag{1}$$

Therefore, it is convenient to deal with a different alphabet. Each word of the (infinite) set $\{m\}^*\ [\Sigma \setminus \{m\}]$ is coded by a new symbol b_i for $i \in \{1, \dots, r-1\} \times \mathbb{N}$. More precisely, the symbols $a_1, \dots, a_r$ ($a_j \neq m$) of $\Sigma \setminus \{m\}$ are re-labelled as $\{b_{1,0}, b_{2,0}, \dots b_{r-1,0}\}$ and then the word $m^\ell b_{j,0}$ is re-labelled as $b_{j,\ell}$. The new alphabet $\{b_i, i \in \{1 \dots, r-1\} \times \mathbb{N}\}$ (now depending on the symbol m) is denoted by Γ_m. The useful prefixes of m–depth k are exactly the words of Γ_m^k. Now, the trie $\mathrm{Tr}X$ has a stack-size of at most k provided that there exists no word $w \in \Gamma_m^k$ which is the common prefix of two words of X.

The new coding changes the initial source $\mathcal{S}$ into a new source which depends on the excluded symbol m and is denoted by $\mathcal{S}_m$.

2.4. Bernoulli and Poisson models. The purpose of average–case analysis of data structures is to characterize the mean value of their parameters under a well-defined probabilistic model that describes the initial distribution of its inputs. In the present paper, we adopt the following general model : we work with a finite set X of infinite words independently produced by the same source. The cardinality n of the set X is usually fixed and the probabilistic model is then called the Bernoulli model of size n and denoted by $\mathcal{B}(n)$. However, rather than fixing the cardinality n of the set X, it proves technically convenient to assume that the set X has a variable number N of elements that obeys a Poisson law of parameter z,

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

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

2.5. Analyses of height and stack-size in the Poisson model. Consider a random trie $\mathrm{Tr}X$ produced by a source under the Poisson model $\mathcal{P}(z)$. Recall that the trie $\mathrm{Tr}X$ has a height at most k provided that there exists no word $w \in \Sigma^k$ that is a common prefix of two words of X. In a similar vein, the trie $\mathrm{Tr}X$ has a stack-size at most k provided that there exists no word $w \in \Gamma_m^k$ that is a common prefix of two words of X. Each set Σ^k or Γ_m^k is a set of independent prefixes, so that the independence property holds. Moreover, the number of words that begin with a finite prefix w is itself distributed as a Poisson variable of rate zp_w (where p_w denotes the probability that a source word begins with prefix w). Thus, the probabilities of these two events (a trie with a height of at most k or a stack-size of at most k) can now be expressed as a function of the prefix probabilities p_w. The following representations for the distribution of the height and the m-stack-size in the Poisson model

$$H_k(z) = \prod_{w\in\Sigma^k} (1+zp_w)\, e^{-zp_w} = e^{-z} \prod_{w\in\Sigma^k} (1+zp_w), \quad S_k^{[m]}(z) = \prod_{w\in\Gamma_m^k} (1+zp_w)\, e^{-zp_w} = e^{-z} \prod_{w\in\Gamma_m^k} (1+zp_w), \tag{2}$$

respectively yield the expected height and the expected stack-size

$$H(z) = \sum_{k=0}^{\infty} [1 - H_k(z)], \qquad S^{[m]}(z) = \sum_{k=0}^{\infty} \left[1 - S_k^{[m]}(z)\right]. \tag{3}$$

2.6. Mellin analysis. The analysis of height and stack-size in the Poisson model $\mathcal{P}(z)$ are based on estimates of the individual probabilities $H_k(z)$ and $S_k^{[m]}(z)$ followed by a Mellin analysis. First, taking logarithms, one gets

$$\log H_k(z) = \sum_{w\in\Sigma^k} \big[- zp_w + \log(1+zp_w)\big], \qquad \log S_k^{[m]}(z) = \sum_{w\in\Gamma_m^k} \big[- zp_w + \log(1+zp_w)\big]. \tag{4}$$

The inequalities $\quad |\log H_k(z) + \frac{z^2}{2}\sum_{w\in\Sigma^k} p_w^2| \le \frac{z^3}{2}\sum_{w\in\Sigma^k} p_w^3, \quad |\log S_k^{[m]}(z) + \frac{z^2}{2}\sum_{w\in\Gamma_m^k} p_w^2| \le \frac{z^3}{2}\sum_{w\in\Gamma_m^k} p_w^3 \quad (5)$

show that the main terms to be analyzed are so-called harmonic sums of the form

$$G(z) = \sum_{w \in \mathcal{U}} g(z\, p_w), \qquad \text{for some set } \mathcal{U}. \tag{6}$$

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

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

Since the Mellin transform of $x \mapsto g(\mu x)$ is μ^{-s} times the transform $g^*(s)$ of g, the Mellin transform of G defined in (6) is

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

Our analysis will involve the so–called series of prefix probabilities of depth k or of m–depth k

$$\Lambda_k(s) := \sum_{w \in \Sigma^k} p_w^s, \qquad \Lambda_k^{[m]}(s) := \sum_{w \in \Gamma_m^k} p_w^s. \tag{7}$$

The largest open strip $< \alpha, \beta >$ where the integral converges is called *fundamental strip*. There is a general phenomenon which makes the Mellin transform quite useful: Poles of the Mellin transform are in direct correspondence with terms in the asymptotic expansion of the original function at ∞. For the asymptotic evaluation of a harmonic sum $G(x)$ this principle applies provided the Dirichlet series $\Delta_{\mathcal{U}}(s)$ and the transform $g^*(s)$ are each analytically continuable and of controlled growth. Then, as $x \to \infty$ the asymptotic expansion of $G(x)$ is closely related to the sum of residues right to the fundamental strip,

$$G(x) \sim -\sum \text{Res } (g^*(s) \Delta_{\mathcal{U}}(s) x^- s).$$

For details on the methodology we refer to [6].

2.7. Depoissonization techniques. The same quantities (distribution and expectation of the height and the stack-size) under the Bernoulli model $\mathcal{B}(n)$ are easily obtained from their corresponding expressions in the Poisson model $\mathcal{P}(z)$. Let $h_{k,n}$ and $s_{k,n}^{[m]}$ denote the respective probabilities that a trie has a height of at most k or has a stack-size of at most k in the model $\mathcal{B}(n)$. Then, the associated exponential generating functions satisfy

$$\sum_n h_{k,n} \frac{z^n}{n!} = e^z H_k(z) = \prod_{w \in \Sigma^k} (1 + zp_w), \qquad \sum_n s_{k,n}^{[m]} \frac{z^n}{n!} = e^z S_k^{[m]}(z) = e^z \prod_{w \in \Gamma_m^k} (1 + zp_w), \tag{8}$$

and the corresponding expectations are

$$E[h_n] = n!\, [z^n] \sum_{k=0}^{\infty} \Big[e^z - \prod_{w \in \Sigma^k} (1 + zp_w) \Big], \qquad E[s_n^{[m]}] = n!\, [z^n] \sum_{k=0}^{\infty} \Big[e^z - \prod_{w \in \Gamma_m^k} (1 + zp_w) \Big].$$

3 Perfect sources.

3.1. Definition of perfect sources. A trie built on words emitted by a source $\mathcal{S}$ will have a good performance if the words have themselves good splitting properties. The probability that two independent words coincide until depth k can be expressed as $\Lambda_k(2)$ where $\Lambda_k(s)$ is defined in (7). More generally, the probability that ℓ independent words coincide until depth k can be expressed as $\Lambda_k(\ell)$. The splitting process will be efficient if, for each fixed ℓ, this probability is a fast decreasing function of k ($k \to \infty$). It is exponentially decreasing for usual sources. A source which implies such a behavior will be called *perfect*.

Definition 1. *A source $\mathcal{S}$ is said to be perfect if the series $\Lambda_k(s)$ of prefix probabilities of depth k satisfies the following quasi-power property,*

$$\Lambda_k(s) := \sum_{w \in \Sigma^k} p_w^s = A(s)\, \lambda(s)^k\ [1 + O(\sigma(s)^k)] \qquad \text{for } k \to \infty, \text{ and } s \geq 1,$$

where the function $\lambda : s \to \lambda(s)$ is log*–concave and the quantity $\sigma(s)$ satisfies $|\sigma(s)| < 1$ for $s \geq 1$.*
A source $\mathcal{S}$ is said to be m–perfect if the source $\mathcal{S}_m$ is perfect, i.e., the series $\Lambda_k^{[m]}(s)$ satisfies the following quasi-power property,

$$\Lambda_k^{[m]}(s) := \sum_{w \in \Gamma_m^k} p_w^s = B_m(s)\ \mu_m(s)^k\ [1 + O(\sigma_m(s)^k)] \qquad \text{for } k \to \infty, \text{ and } s \geq 1,$$

where the function $\mu_m : s \to \mu_m(s)$ is log*–concave and the quantity $\sigma_m(s)$ satisfies $|\sigma_m(s)| < 1$ for $s \geq 1$.*
A source that is both perfect and m–perfect for all symbols m is said to be totally perfect. The quantities that intervene in the quasi-power property are called the parameters of the source.

We shall see in the sequel that the log–concavity property is useful for comparing the two terms $z^2\Lambda_k(2)$ and $z^3\Lambda_k(3)$ for suitable values $k = k(z)$.

3.2. Perfection of classical sources. We prove now that all classical sources are totally perfect.

Proposition 1. *All the memoryless sources and all the ergodic Markov chains are totally perfect.*

Proof.
(*a*) *Memoryless sources.* If $\mathcal{S}$ is a memoryless source which emits symbols $a_1, a_2, \ldots, a_r$ with respective probabilities $p(a_1), p(a_2), \ldots p(a_r)$, the modified source $\mathcal{S}_m$ is a memoryless source too. It emits symbols b_i, $i \in \{1, \ldots, r-1\} \times \mathbb{N}$, and the respective probabilities satisfy $p_{(b_{(i,k)})} = p_{(b_{(i,0)})} \times p_{(m)}^k$, so that

$$\Lambda_k(s) = \lambda(s)^k, \quad \Lambda_k^{[m]}(s) = \mu_m(s)^k, \tag{9}$$

$$\text{with} \quad \lambda(s) = \sum_{\ell=1}^{r} p(a_\ell)^s \quad \text{and} \quad \mu_m(s) = \frac{1}{1 - p(m)^s} \sum_{\ell \mid a_\ell \neq m} p(a_\ell)^s = 1 - \frac{1 - \lambda(s)}{1 - p(m)^s}.$$

Then $A(s) = B_m(s) = 1$. Moreover, it is clear that both functions $\lambda(s)$ and $\mu_m(s)$ are log-concave.

(*b*) *Markov chains.* Let $\mathcal{S}$ be a source associated to a Markov chain with initial probabilities $(\pi_1, \pi_2, \ldots, \pi_r)$. For some s, we consider the r vector $\pi(s) := (\pi_1^s, \pi_2^s, \ldots, \pi_r^s)$ and the $r \times r$ matrix $\mathbf{P}_s := (p_{i|j}^s)$. Then the series of prefix probabilities $\Lambda_k(s)$ can be expressed with the k-th power of matrix $\mathbf{P}_s$, as

$$\Lambda_k(s) = {}^t e\ \mathbf{P}_s^k\, \pi(s) \qquad \text{with} \qquad e = (1, 1 \ldots 1).$$

When the Markov chain is ergodic, the matrix $\mathbf{P}_s$ has good dominant spectral properties: by classical Perron-Frobenius theorem, for real values of parameter s, it possesses a unique dominant eigenvalue $\lambda(s)$, so that the quasi-power property of Definition holds. Moreover, the function λ is log–concave.
We suppose that the excluded symbol is $m := a_\ell$, and we consider two matrices built from matrix $\mathbf{P}_s$: the $r \times r$ matrix $\mathbf{M}_s$ where all the rows of $\mathbf{P}_s$ except the ℓ-th one are replaced by zero rows, and the $r \times r$ matrix $\mathbf{N}_s$ where the ℓ-th row of $\mathbf{P}_s$ is replaced by a zero row. The matrix $\mathbf{L}_s := \mathbf{N}_s\,(I - \mathbf{M}_s)^{-1}$ replaces the matrix $\mathbf{P}_s$ when changing the alphabet, and finally

$$\Lambda_k^{[m]}(s) = {}^t e\, \mathbf{L}_s^k\, \pi(s) \qquad \text{with} \qquad e = (1, 1 \ldots 1).$$

When the Markov chain is ergodic, the matrix $\mathbf{L}_s$ has good dominant spectral properties: for real values of parameter s, it possesses a unique dominant eigenvalue $\mu(s)$, so that the quasi-power property of Definition 3.1. holds. Moreover, the function μ is log–concave. $\square$

3.3. Dynamical sources. Dynamical sources, introduced in [28], encompass and generalize the two previous models of sources. They are associated with expanding maps of the interval. We refer to [28] for more details. We just recall the definition of such sources and the main properties.

Definition 2. *A dynamical source $\mathcal{S}$ is defined by four elements:*

- (a) *an alphabet Σ finite or denumerable,*
- (b) *a topological partition of $\mathcal{I} :=]0,1[$ with disjoint open intervals $\mathcal{I}_a, a \in \Sigma$,*
- (c) *an encoding mapping σ which is constant and equal to a on each $\mathcal{I}_a$,*
- (d) *a shift mapping T whose restriction to $\mathcal{I}_a$ is a real analytic bijection from $\mathcal{I}_a$ to $\mathcal{I}$. Let h_a be the local inverse of T restricted to $\mathcal{I}_a$ and $\mathcal{H}$ be the set $\mathcal{H} := \{h_a, a \in \Sigma\}$. There exists a common complex neighbourhood $\mathcal{V}$ of $\bar{\mathcal{I}}$ on which the set $\mathcal{H}$ satisfies the following:*
  - (d_1) *the mappings h_a extend to holomorphic maps on $\mathcal{V}$, mapping $\mathcal{V}$ strictly inside $\mathcal{V}$,*
  - (d_2) *the mappings $|h'_a|$ extend to holomorphic maps $\tilde{h}_a$ on $\mathcal{V}$ and there exists $\delta_a < 1$ for which $0 < \tilde{h}_a(z) \leq \delta_a$ for $z \in \mathcal{V}$,*
  - (d_3) *there exists $\gamma < 1$ for which the series $\sum_{a\in\Sigma} \delta_a^s$ converges on $\Re(s) > \gamma$.*

The word $M(x)$ of Σ^∞ emitted by the source is then formed with the sequence of symbols $\sigma T^j(x)$

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

If the unit interval is endowed with a real analytic density f, the source is called a *Probabilistic Dynamical Source* and is denoted by $(\mathcal{S}, F)$, where F is the distribution function associated to the density f. All the infinite words that begin with the same prefix $w := m_1 \ldots m_k$ correspond to real numbers x that lie in a same "fundamental" interval $\mathcal{I}_w =]h_w(0), h_w(1)[$ with $h_w := h_{m_1} \circ h_{m_2} \circ \cdots \circ h_{m_k}$, generated by iterations of the shift T. The probability p_w that a word begins with prefix w is then the measure of this interval $\mathcal{I}_w$, i.e.,

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

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

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

and the restriction of T on $\mathcal{I}_a$ is the affine mapping defined by $T(q_a) = 0$ and $T(q_{a+1}) = 1$.

(Markov chains). Any Markov chain with a finite alphabet can be associed to a dynamical system. We consider $r+1$ copies of $\mathcal{I}$, where r is the cardinality of the alphabet. Each copy $\mathcal{I}_\ell :=]\ell, \ell+1[$ memorizes the last emitted symbol a_ℓ (eventually, $\ell = 0$). Denoting by Φ_m the translation $\Phi_m(x) := x + m$, we then define, for $1 \leq i \leq r$ and $0 \leq j \leq r$

$$\mathcal{I}_{i,j} := \Phi_j(]q_{i|j}; q_{i+1|j}[), \text{ where } q_{i|j} := \sum_{a<i} p_{a|j},$$

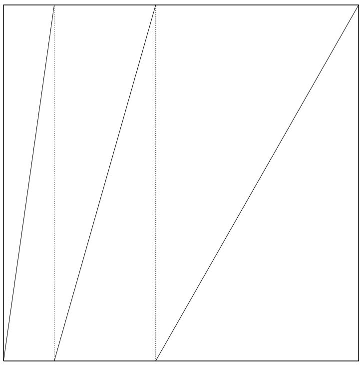
and the restriction $T_{i,j}$ of T on $\mathcal{I}_{i,j}$ that is the affine mapping defined implicitly by $\Phi_i \circ T_{i,j} \circ \Phi_j^{-1}(q_{i|j}) = 0$ and $\Phi_i \circ T_{i,j} \circ \Phi_j^{-1}(q_{i+1|j}) = 1$. The system associated to partition $\mathcal{I}_{i,j}$ of $]0, r+1[$ and to branches $T_{i,j}$ is a "general" dynamical system.

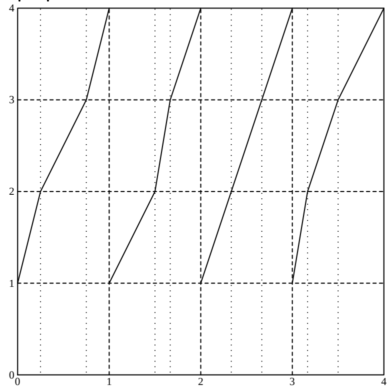
(Continued Fraction). The continued fraction transformation is an example of a source with unbounded memory (in a sense, it is the derivative $T'(x)$ that keeps memory of the previous history.) The alphabet is

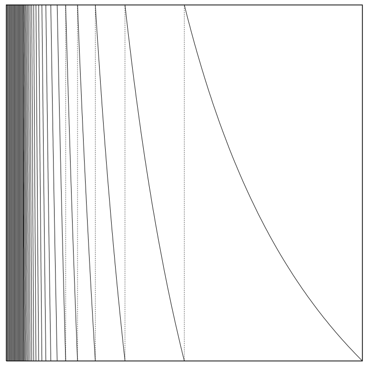
then $\mathbb{N}^*$, the topological partition is defined by $\mathcal{I}_a :=]1/(a+1), 1/a[$ and the restriction of T on $\mathcal{I}_a$ is the decreasing linear fractional transformation $T(x) := (1/x) - a$. In other words,

$$T(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor, \qquad \text{and } \sigma(x) = \lfloor \frac{1}{x} \rfloor.$$

The following figures present three examples of sources represented by their shift function T: namely, a memoryless source of probabilities $(\frac{1}{7}, \frac{2}{7}, \frac{4}{7})$, a Markov chain of length 3 and the continued fraction expansion.







Generating operators. There is a direct relationship between the dynamics of the source $\mathcal{S}$, the Dirichlet series and the spectral properties of an operator closely related to the way the shift T transforms probability distributions. The basic ingredient, well-developped in dynamical system theory is the "classical" Ruelle operator,

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

which depends on a parameter s and is defined through the analytic extensions $\tilde{h}$ of $|h'|$. This operator can be viewed as a density transformer since if X is a random variable with density f, then the density of $T(X)$ is $\mathcal{G}_1[f]$. The dynamics of the process is a priori described by $s = 1$, but many other properties appear to be dependent upon complex values of s other than 1.

However, the classical Ruelle operator cannot generate at the same time both ends of the fundamental intervals, so that it is not adequate in information theory context. In [28], Vallée has introduced a new tool, the "generalized" Ruelle operator, that involves secants of inverse branches

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

instead of tangents $|h'(x)|$ of inverse branches.

Each symbol a produced by source $\mathcal{S}$ is "generated" by a Ruelle operator $\mathbf{G}_{s,[a]}$

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

that involves the analytic extension $\widetilde{H}_a$ of the secant H_a of the inverse branch h_a. These operators act on functions F of two (complex) variables. A finite word $w := m_1 m_2 \ldots m_k$ is generated by the composition operator $\mathbf{G}_{s,[w]} := \mathbf{G}_{s,[m_k]} \circ \ldots \circ \mathbf{G}_{s,[m_1]}$. All possible prefixes of length k are thus generated by the k–th power of the Ruelle operator relative to source $\mathcal{S}$,

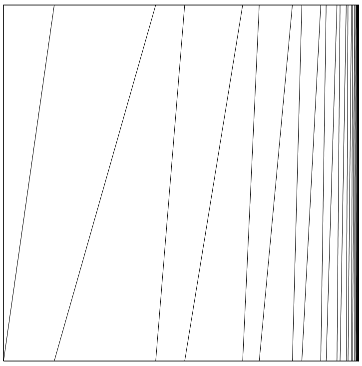
$$\mathbf{G}_s := \sum_{a \in \Sigma} \mathbf{G}_{s,[a]}. \tag{10}$$

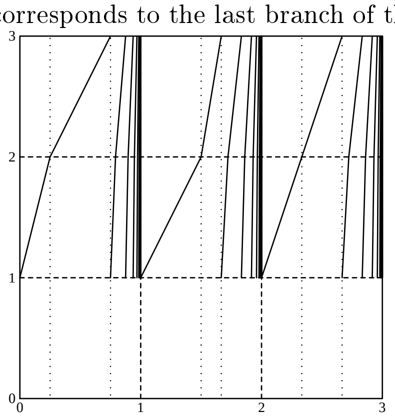
Then, the series $\Lambda_k(s)$ of prefix probabilities can be expressed by means of the k-th power of $\mathbf{G}_s$

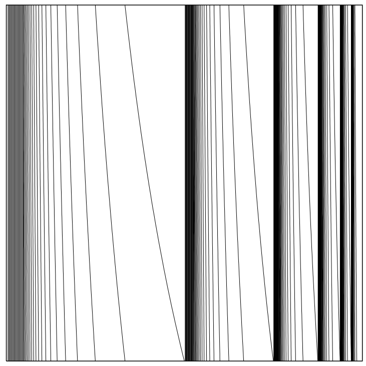
$$\Lambda_k(s) = \mathbf{G}_s^k[Q^s](0, 1), \text{ where } Q(x, y) := \left| \frac{F(x) - F(y)}{x - y} \right|$$

is the secant of the distribution F.

Change of the source. The following figures present the induced sources $\mathcal{S}_m$ associated to the previous examples when m is the letter that corresponds to the last branch of the initial source.







The change of the source (and thus the change of the alphabet) also translates into a change of the Ruelle operator. The relation (1) on Γ_m gives an expression of this operator by means of $\mathbf{G}_s$ and $\mathbf{G}_{s,[m]}$,

$$\mathbf{L}_{s,[m]} := (\mathbf{G}_s - \mathbf{G}_{s,[m]}) \circ (I - \mathbf{G}_{s,[m]})^{-1} \tag{11}$$

so that all possible prefixes of Γ_m^k are thus generated by the k–th power of the Ruelle operator $\mathbf{L}_{s,[m]}$ and the series $\Lambda_k^{[m]}(s)$ can be expressed by means of the k-th power of $\mathbf{L}_{s,[m]}$

$$\Lambda_k^{[m]}(s) = \mathbf{L}_{s,[m]}^k[Q^s](0,1), \text{ where } Q(x,y) := \left|\frac{F(x) - F(y)}{x - y}\right|$$

is the secant of the distribution F.

3.4. Perfection of dynamical sources. We prove now that all dynamical sources are totally perfect.

Proposition. *All the dynamical sources are totally perfect. In that cases, the ratios that intervene in the quasi-power properties are dominant eigenvalues of suitable Ruelle operators.*

Proof. If condition (d) holds, we can prove the following: for $\Re(s) > \gamma$, the Ruelle operator $\mathbf{G}_s$ acts on the Banach space $B_\infty(\mathcal{V})$ formed with all functions F that are holomorphic in the domain $\mathcal{V} \times \mathcal{V}$ and are continous on the closure $\bar{\mathcal{V}} \times \bar{\mathcal{V}}$, endowed with the sup-norm. It is compact, even more nuclear in the sense of Grothendieck [10, 11]. Furthermore, for real values of parameter s, it has positive properties that entail (via Theorems of Perron-Frobenius style due to Krasnoselvski [16]) the existence of dominant spectral objects: there exists a unique dominant eigenvalue $\lambda(s)$ positive, analytic for $s > \gamma$, a dominant eigenfunction denoted by Ψ_s, and a dominant projector E_s. Under normalization condition $E_s[\Psi_s] = 1$, these last two objects are unique too. Then, the compacity entails a spectral gap between the dominant eigenvalue and the remainder of the spectrum, that separates the operator $\mathbf{G}_s$ in two parts: the "part" relative to the dominant eigenvalue and the "part" relative to the remainder of the spectrum. This gives access to the quasi-power property

$$\Lambda_k(s) = A(s)\lambda(s)^k[1 + O(\sigma(s)^k)],$$

where $A(s)$ is a positive constant and $\sigma(s)$ is a constant satisfying $R(s)/\lambda(s) < \sigma(s) < 1$ that depends on the modulus $R(s)$ of subdominant eigenvalues of $\mathbf{G}_s$. The Log-concavity of $\lambda(s)$ comes from a maximum property of the operator $\mathbf{G}_s$. Finally, the source $\mathcal{S}_m$ satisfies the same properties and is also perfect. ◻

4 Analysis of the trie stack-size for perfect source.

We state now the main result of the paper:

Theorem. *Let $\mathcal{S}$ be a m–perfect source of parameters (B_m, μ_m). Then, in the Bernoulli model $\mathcal{B}(n, \mathcal{S})$, the m–stack–size relative to the excluded symbol m has an expectation of order $\log n$. Its probability distribution is asymptotically of the double exponential type. More precisely,*

$$E[s_n^{[m]}] = \frac{2}{|\log \mu_m(2)|} \log n + Q_{m,\mathcal{S}}(\log n) + \left[\frac{\gamma + \log B_m(2) - \log 2}{|\log \mu_m(2)|} + \frac{1}{2}\right] + o(1),$$

$$\lim_{n\to\infty} \sup_{k\geq 0} \left| \Pr\{s_n^{[m]} \leq k\} - \exp[-\frac{B_m(2)}{2}\, \mu_m(2)^k\, n^2] \right| = 0.$$

Here, γ is the Euler constant and $Q_{m,\mathcal{S}}(u)$ is a periodic function of very small amplitude that depends on the source $\mathcal{S}$ and the "excluded" symbol m.

Proof. There are three main steps in the proof. First, we work in the Poisson model, and we analyze the distribution and the expectation, then we translate the results into the Bernoulli model. In the proof, the excluded symbol m is fixed and we therefore omit its notation.

(1) Probability distribution of the stack-size under the Poisson model. Here, we compare three sequences

$$S_k(z) := S_k^{[m]}(z) = \prod_{w\in\Gamma^k} (1 + zp_w)\, e^{-zp_w}, \qquad \widetilde{S}_k(z) := \exp\big[-\frac{z^2}{2} \sum_{w\in\Gamma^k} p_w^2\big], \qquad \widehat{S}_k(z) := \exp[-\frac{B(2)}{2} z^2\, \mu(2)^k],$$

and we prove that

$$\sum_{k\geq 0} \big[S_k(z) - \widetilde{S}_k(z)\big] = o(1) \qquad \text{and} \qquad \sum_{k\geq 0} \big[\widetilde{S}_k(z) - \widehat{S}_k(z)\big] = o(1), \tag{12}$$

so that the individual probabilities $S_k(z)$ admit a double exponential approximation $\widehat{S}_k(z)$.

(a) First, we compare $S_k(z)$ with $\widetilde{S}_k(z)$; we prove that the term $z^3 \sum p_w^3$ involved in the relation (5) is really an error term when using the log-concavity of μ that guarantuees the existence of a number d in the interval $]3/\log\mu(3), 2/\log\mu(2)[$. Then, if we set $\kappa(z) := \lfloor d \log z \rfloor$, there exist positive numbers ε, ε' and $\nu \in [0; 1[$ such that the following three properties hold:

$$(C_1) : z^2 \mu(2)^{\kappa(z)} \geq z^{\varepsilon}, \quad (C_2) : z^3 \mu(3)^{\kappa(z)} \leq z^{-\varepsilon}, \quad (C_3) : (\forall w, |w|_m \geq \kappa(z))\ zp_w \leq \nu.$$

The sum (12) in then split into two parts according to the integer $\kappa(z)$. For the indices $k \leq \kappa(z)$, each term of the sum is small. The property (C_3) of index $\kappa(z)$, the Quasi-Power Property at $s = 2$, and finally property (C_1) of index $\kappa(z)$ imply that

$$\sum_{k\leq\kappa(z)} \big[\widetilde{S}_k(z) - S_k(z)\big] \leq z^3 \kappa(z) \exp\big[-d_0 z^2 \mu(2)^{\kappa(z)}\big] \leq z^3 \kappa(z) \exp(-d_0 z^{\varepsilon}) = o(1).$$

The second part of the sum relative to $k > \kappa(z)$ is also $o(1)$. Here, although the terms of the sum are not small, their differences can be bounded by means of the Quasi-Power Property at $s = 3$ in conjunction with the property (C_2) of index $\kappa(z)$,

$$\sum_{k>\kappa(z)} \big[S_k(z) - \widetilde{S}_k(z)\big] \leq d_1 z^3 \frac{\mu(3)^{\kappa(z)}}{1 - \mu(3)} \leq z^{-\varepsilon'} = o(1).$$

(b) Now, the Quasi-Power Property at $s = 2$ implies that $\widehat{S}_k(z)$ well approximates to $\widetilde{S}_k(z)$ and

$$\sum_{k\geq 0} [S_k(z) - \widehat{S}_k(z)] = o(1).$$

This provides the asymptotic probability distribution of the stack-size $S_k(z)$ in the Poisson model of parameter z

$$\lim_{z\to\infty} \sup_{k\geq 0} \left|\Pr\{S_k(z) \leq k\} - \exp[-\,\frac{B(2)}{2}\,\mu(2)^k\, z^2]\right| = 0.$$

(2) Expected stack-size under the Poisson model. The harmonic sum that approximates the average stack-size in a Poisson model

$$\widehat{S}(z) := \sum_{k=0}^{\infty}\big(1 - \exp[-\,\frac{B(2)}{2}\,\mu(2)^k\, z^2]\big), \qquad \text{has Mellin transform} \qquad \widehat{S}^*(s) = -\frac{1}{2}\,(\frac{B(2)}{2})^{-s/2}\,\frac{\Gamma(s/2)}{1-\mu(2)^{s/2}}.$$

The fundamental strip is $\langle -2, 0\rangle$, with the singular expansion at $s=0$ being

$$\widehat{S}^*(s) \asymp \frac{2}{|\log\mu(2)|}\frac{1}{s^2} - [\frac{\gamma + \log B(2) - \log 2}{|\log \mu(2)|} + \frac{1}{2}]\frac{1}{s}, \quad (s=0).$$

There are also regularly spaced poles on the line $\Re(s) = 0$ that entail periodic fluctuations. This gives the expected stack-size under the Poisson model.

(3) Analysis of the stack-size under the Bernoulli model. The depoissonization is then a refinement of the method used for the Poisson model. Indeed, the Cauchy integral formula, in conjunction with the "$\exp - \log$" transformation applied to $e^z S_k(z)$, gives an approximation for the probabilities $s_{k,n}$

$$s_{k,n} \approx \frac{n!}{2i\pi}\int_{\gamma} \exp\big(-\frac{z^2}{2}\sum_{w\in\Gamma^k} p_w^2\big)\,\frac{e^z\,dz}{z^{n+1}}.$$

Here, γ is a simple closed contour encircling positively the origin. This integral can be estimated by a saddle point method under a refinement of the three conditions (C_1), (C_2) and (C_3) and with a similar splitting of the index k. The probabilities $s_{k,n}$ are then well approximated by

$$s_{k,n} \approx \exp\big[-\frac{B(2)}{2}\,\mu(2)^k\, n^2\big].$$

Finally, all the estimates for the stack-size in the Poisson model remain valid in the Bernoulli model. $\square$

The optimal choice of the excluded symbol. The constant $c_m(\mathcal{S}) = \mu_m(2)$ is always less than 1. Then, the formula for the expected stack-size shows that the stack-size is an increasing function of constant $\mu_m(2)$. When $\mathcal{S}$ is a fixed memoryless source, the constant $\mu_m(2)$ has a simple expression (9) which involves the probabilities of the symbols, and the optimal choice for the excluded symbol m arises when it is the most probable one.

The same fact can be conjectured for all perfect sources.

Conjecture. *For all totally perfect sources, the optimal choice for the excluded symbol arises when the excluded symbol is the most probable.*

We do not know how to prove the conjecture. We just now give some elements that show that the conjecture is plausible. First, we have to define the probability of symbol m. Consider two Ruelle operators associated to the Ruelle operator $\mathbf{G}_s$ defined in (10); for the first one, denoted by $\widetilde{\mathbf{G}}_{s,u}$, the symbol m is marked by variable u; for the second one, denoted by $\widehat{\mathbf{G}}_{s,u}$, all the symbols except m are marked by u,

$$\widetilde{\mathbf{G}}_{s,u} := u\,\mathbf{G}_{s,[m]} + \sum_{a\neq m}\mathbf{G}_{s,[a]}, \qquad \widehat{\mathbf{G}}_{s,u} := \mathbf{G}_{s,[m]} + u\sum_{a\neq m}\mathbf{G}_{s,[a]}.$$

The probability $p(m)$ of symbol m can be expressed with the dominant eigenvalue $\tilde{\lambda}(s,u)$ of $\widetilde{\mathbf{G}}_{s,u}$, or with the dominant eigenvalue $\widehat{\lambda}(s,u)$ of $\widehat{\mathbf{G}}_{s,u}$

$$p(m) := \frac{d}{du}\tilde{\lambda}(1,u)|_{u=1} = 1 - \frac{d}{du}\hat{\lambda}(1,u)|_{u=1}. \tag{13}$$

On the other hand, the relation $(I - \widehat{\mathbf{G}}_{s,u})^{-1} = (I - u\mathbf{L}_{s,[m]})^{-1}\,(I - \mathbf{G}_{s,[m]})^{-1}$ that involves $\mathbf{L}_{s,[m]}$ defined in (11) exactly expressed the decomposition $\Sigma^* = \Gamma_m^*\ \{m\}^*$. With (13), it entails the equality

$$|\mu'_m(1)| \ = \frac{|\lambda'(1)|}{1 - p(m)}$$

that proves that the largest value of $|\mu'_m(1)|$ is obtained for the most probable symbol m. Furthermore, for two symbols m and n for which $p(m) > p(n)$, the functions $\nu_m := |\log \mu_m|$ and $\nu_n := |\log \mu_n|$ satisfy the following

$$\nu_m(1) = 0, \qquad \nu_n(1) = 0, \qquad \nu'_m(1) > \nu'_n(1),$$

so that, on a neighbourhood of $s = 1$, we have $\nu_m(s) > \nu_n(s)$, and we wish to prove the inequality $\nu_m(2) > \nu_n(2)$.

5 Some experimentations.

In this section, we present some experimental results which we compare to the theoretic formulae of the paper.

5.1. Memoryless sources. The first set of experimentations is for memoryless sources.

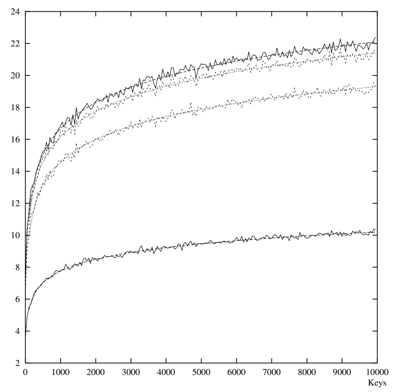


Figure 2: A plot of the formula for the average height and stack-size of ternary tries with different choices for the excluded symbol compared to simulation results (averaged on 20 entries).

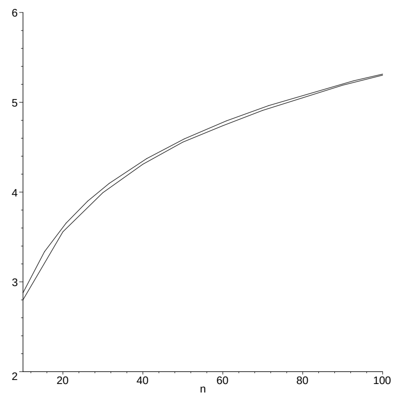


Figure 3: A plot of the formula (lowest curve of Figure 2) compared to simulation results.

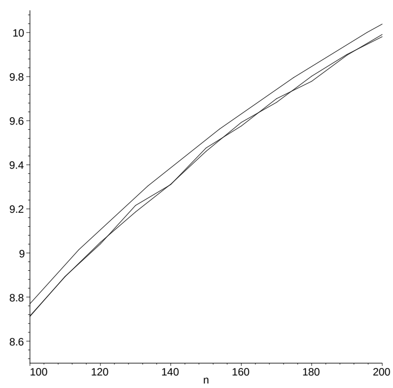


Figure 4: A plot of the formula for unbiased binary tries compared to simulation results for the unbiased binary source and for 24 bit integers.

Figure 2 shows a plot of the formula of the average stack-size of ternary tries which are relative to a memoryless source of probabilities $1/7, 2/7, 4/7$. The lowest curve corresponds to the case when the most probable symbol is excluded and the highest corresponds to the height. Figure 3 proves the good quality of our asymptotics since there is almost no difference between the simulations and the prediction by means of the formula (even for tries of smaller sizes).
Finally, Figure 4 shows the average stack-size of unbiased binary tries. The highest of the three curves is related to the formula; the other two curves result from two types of simulations. The first type of simulations is made for unbiased binary sources, while the second type uses random integer data that usually are assumed to fit the assumptions of unbiased binary source model.

5.2. Continued fraction source. The second set of experimentations concerns tries built with the continued fractions expansion of uniformly distributed reals of $[0;1]$.

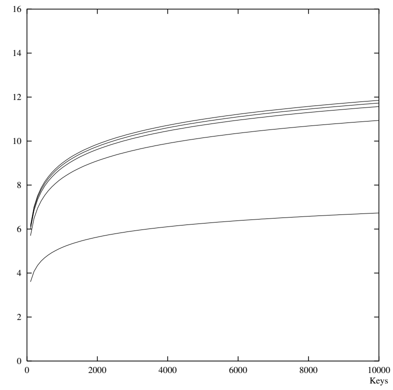


Figure 5: A plot of the formula for the average height and stack-size when the excluded symbol is respectively $1, 2, 3$ and 4.

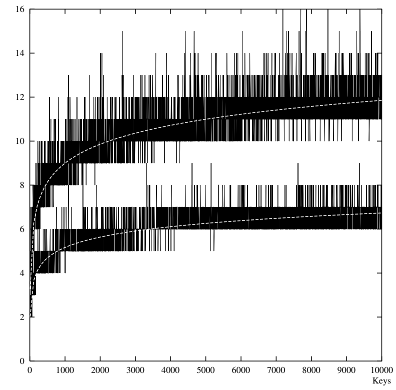


Figure 6: Plot of the formula for the height and for the 1-stack-size compared to the simulation results.

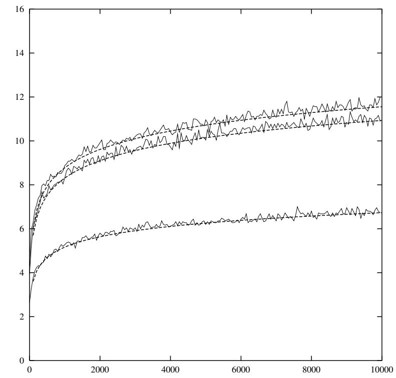


Figure 7: Plot of the formula for the average stack-size with different choices for the excluded symbol compared to the simulation results (averaged on 20 entries).

Figure 5 shows the fast convergence of the constants $c_m(\mathcal{S})$ to the constant $c(\mathcal{S})$ relative to the average height of the trie: the lowest curve corresponds to c_1, the second lowest corresponds to c_2,... and the highest curve corresponds to $c(\mathcal{S})$. This also confirms the conjecture that the optimal choice for the excluded symbol arises when it is the most probable. Figure 6 points out the double-exponential character of the distribution of height and stack-size. Finally, Figure 7 proves the good quality of our asymptotics for the average height (the constant $c(\mathcal{S})$ is known precisely) and for the 1-stack-size. The constants $c_m(\mathcal{S})$ are numericaly computed by using truncated Taylor series of both Ruelle operator and function f at the same point. This technique has been already applied in [2] for computing the "Vallée's" constant $c(\mathcal{S})$.

References

[1] Clément, J., Flajolet, P. and Vallée, B., Dynamical Sources in Information Theory: A General Analysis of Trie Structures, INRIA Rapport de recherche **3645**, 1999, to appear in *Algorithmica* (2000).

[2] Daudé, H., Flajolet, P. and Vallée, B., An average-case analysis of the Gaussian algorithm for lattice reduction. *Combinatorics, Probability and Computing*, 6 (1997), 397–433.

[3] De Bruijn, N.G., Knuth, D.E., and Rice, S.O., The average height of planted plane trees, *Graph Theory and Computing*, Academic Press, 15–22, 1972.

[4] Devroye, L. and Kruszewski P., *On the Horton-Strahler number for Random Tries*, RAIRO Theoretical Informatics and Applications **30**, 443-456, 1996.

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

[6] Flajolet, P., Gourdon, X. and Dumas, P., *Mellin transforms and asymptotics: Harmonic sums*, Theoretical Computer Science **144**, 3-58, 1995.

[7] Flajolet, P., Puech, C., Partial match retrieval of multidimensional data, *Journal de l'ACM* 33, 2 (1986) 371–407.

[8] Fredkin, E. H., Trie Memory, *Communications of the ACM* **3**, 490-500, 1990.

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

[10] Grothendieck, A. Produit tensoriels topologiques et espaces nucléaires, *Mem. Am. Math. Soc.*, 16 (1955).

[11] Grothendieck, A. La Théorie de Fredholm, *Bull. Soc. Math. France,* 84 pp 319–384.

[12] Jacquet, P. and W. Szpankowski, W., *Analytical Depoissonization and its Applications*, Theoretical Computer Science, in "Fundamental Study",**201**, No. 1-2, 1-62, 1998.

[13] Kirschenhofer, P. and Prodinger, H. *On the Recursion Depth of Special Tree Traversal Algorithms*, Information and Computation, **74**, 15–32, 1987.

[14] KEMP, R., *The average height of r-tuply rooted planted plane trees*, Computing **25**, 209-232, 1980.
[15] KNUTH, D. E., *The Art of Computer Programming*, Volume 3: Sorting and Searching, Addison-Wesley, 1973.
[16] KRASNOSELSKII, M. *Positive solutions of operator equations,* P. Noordhoff, Groningen, 1964.
[17] MAHMOUD, H. M., *Evolution of Random Search Trees*, Wiley-Interscience Series, 1992.
[18] NEBEL, M. E., The Stack-Size of Tries, a Combinatorial Study, Theoretical Computer Science, to appear.
[19] NEBEL, M. E., The Stack-Size of Uniform Random Tries Revisited, submitted, 2000.
[20] NEBEL, M. E., *On the Horton-Strahler Number for Combinatorial Tries*, R.A.I.R.O. Theoretical Informatics and Applications, to appear.
[21] RÉGNIER, M., Trie hashing analysis. *Proc. 4th Int.Conf. Data Eng.*, Los Angeles, CA, 1988, 377–387.
[22] RÉGNIER, M., On the average height of trees in in digital search and dynamic hashing, *Information Processing Letters*, 13, (1982), 64–66.
[23] RIVEST, R. L., Partial match retrieval algorithms, *SIAM Journal on Computing* 5 (1976) 19–50.
[24] SEDGEWICK R., *Algorithms*, Addison-Wesley, 1988.
[25] SZPANKOWSKI, W., On the height of digital trees and related problem. *Algorithmica* (6) 1991 256–277.
[26] SZPANKOWSKI, W., Some results on V–ary asymmetric tries *J. Algorithms*, vol. 9, 224–244, 1988.
[27] TRABB PARDO, L., Set representation and set intersection, Technical report, Stanford University, 1998.
[28] VALLÉE, B., Dynamical Sources in Information Theory: Fundamental Intervals and Word Prefixes, *Les cahiers du GREYC,* 2 (1999), to appear in *Algorithmica*, 2000.
[29] VIENNOT, X. G., *Trees Everywhere*, Proc. CAAP'90, LNCS **431**, 18–41, 1990.
[30] YAO, A., A note on the analysis of extendible hashing, *Information Processing Letters*, 11 (1980) 84–86.