Algorithms Seminar 2000–2001,
F. Chyzak (ed.), INRIA, (2002), pp. 53–58.

Available online at the URL
`http://algo.inria.fr/seminars/`.

# Patricia Tries in the Context of Dynamical Systems

*Jérémie Bourdon*
GREYC, Université de Caen (France)

March 19, 2001

*Summary by Michel Nguyễn-Thế*

**Abstract**

Tries, a generalized form of digital trees, are a data structure widely used in numerous domains: algorithms for searching words, compression, dynamical hashing, ... Their interest and construction lie in the partitioning of a set of words. We present a compact form of tries, called Patricia tries, in which all unary nodes are suppressed (and thus do not intervene in the partitioning). We then study the means of the memory occupation and of the cost of inserting a word for that data structure when words are produced by a probabilistic source for which the dependencies between the emitted symbols can be very important.

## 1. Size and Path Length of Tries and Patricia Tries: Expressions for Expectations

We define the notions of tries and Patricia tries. We find general expressions for the expectations of the size and path length of tries and Patricia tries in the Bernoulli model, valid for any source.

1.1. **Operations on infinite words.** For a finite alphabet $\Sigma = \{a_1, a_2, \dots, a_r\}$, let $\Sigma^\infty$ be the set of infinite words on that alphabet, $\underline{\sigma} : \Sigma^\infty \to \Sigma^\infty$ the map that returns the first letter of a word, and $\underline{T} : \Sigma^\infty \to \Sigma^\infty$ the shift that returns the first suffix of a word. Let $\underline{T}_{[a]}$ denote the restriction of $\underline{T}$ to the set $\sigma^{-1}(\{a\})$ of words beginning with symbol $a$ and, for a finite prefix $w = a_1 \dots a_k$, let $\underline{T}_{[w]}$ denote the composition $\underline{T}_{[a_k]} \circ \underline{T}_{[a_{k-1}]} \circ \cdots \circ \underline{T}_{[a_1]}$. The notations $\sigma$ and $T$ are kept for operators acting on reals which will be used later.

1.2. **Tries.**

**Definition 1.** Let $X$ be a finite set of infinite words produced by the same source. A *trie* $\mathrm{Tr}(X)$ is a structure defined by the following rules:
($R_0$) If $X = \varnothing$ (the empty set), $\mathrm{Tr}(X)$ is the empty tree.
($R_1$) If $X = \{x\}$, $\mathrm{Tr}(X)$ consists of a single leaf node represented by $\square$ that contains $x$.
($R_2$) If $X$ is of cardinality greater than or equal to 2, $\mathrm{Tr}(X)$ is an *internal node* represented by $\bullet$ to which are attached $r$ subtrees:

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

The edge that attaches the subtrie $\mathrm{Tr}\big(\underline{T}_{[a_j]}X\big)$ is labelled by the symbol $a_j$. Notice a little abuse in ($R_2$): if there is no word in $X$ beginning with $a_j$, then $\underline{T}_{[a_j]}X$ is not defined, and we consider that is equal to the empty set. Hence $\mathrm{Tr}\big(\underline{T}_{[a_j]}X\big)$ is the empty tree, and it is as though there were no subtree corresponding to $a_j$ (see Figure 1).

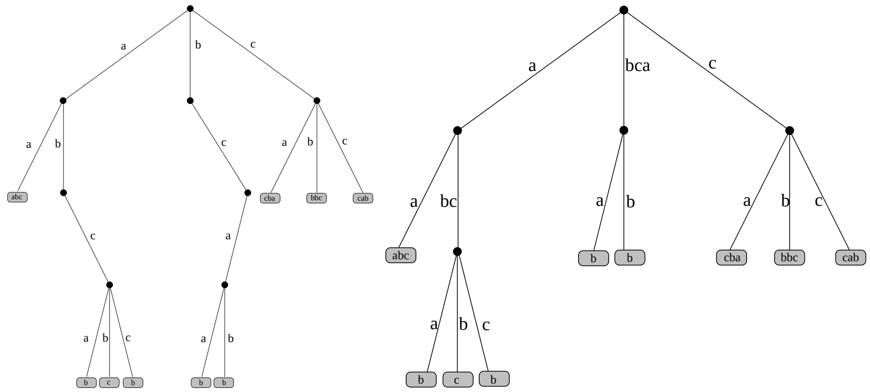


FIGURE 1. Standard trie and corresponding Patricia trie.

1.3. **Patricia Tries.** A Patricia trie is a trie from which all unary nodes are eliminated. Hence with any finite set $X$ of infinite words produced by the same source, we associate a Patricia trie $\mathrm{PaTr}(X)$. The first two rules are the same, but the last rule $(R_2')$ is more sophisticated:

$(R_2')$ If $X$ is of cardinality greater than or equal to 2, we have two cases:

$(R_{2,1}')$ if $\underline{\sigma}(X)$ consists of a single symbol, then $\mathrm{PaTr}(X)$ equals $\mathrm{PaTr}(\underline{T}X)$.

$(R_{2,2}')$ if $\underline{\sigma}(X)$ has at least two distinct symbols, $\mathrm{PaTr}(X)$ is an *internal node* generically represented by $\bullet$ to which are attached $r$ subtrees,

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

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

1.4. **Additive parameters.** The *depth* of a node in a tree is the number of edges of the path that connects it to the root. The *size* of a tree is the number of its internal nodes. The *path length* of a tree is the sum of the depths of all (nonempty) external nodes.

1.5. **Algebraic analysis of additive parameters.** In a standard trie built on the set $X = \{x_1, \ldots, x_n\}$, the structure of a node labelled by a prefix $w$ is a finite string called a *slice* given by

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

An additive parameter $\gamma$ on $X$ is defined by a toll parameter $\delta$ defined on finite strings and the recursive rule:

$$\gamma[X] = \begin{cases} 0, & \text{if } |X| \leq 1, \\ \delta[\underline{\sigma}(X)] + \sum_{m\in\Sigma} \gamma[\underline{T}_{[m]}X], & \text{if } |X| \geq 2, \end{cases}$$

Let $|s|$ and $\#(s)$ denote the number of symbols of the string $s$ and the number of distinct symbols of $s$, respectively. The parameters of interest are the size on tries and Patricia tries,

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

and the internal path length on tries and Patricia tries

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

| | |
|---|---|
| Size of Tr | $\widehat{S}(n) = \sum_{w\in\Sigma^*} \left(1-(1+(n-1)p_w)(1-p_w)^{n-1}\right)$ |
| Path Length of Tr | $\widehat{L}(n) = \sum_{w\in\Sigma^*} np_w\left(1-(1-p_w)^{n-1}\right)$ |
| Size of PaTr | $\widehat{S_P}(n) = \sum_{w\in\Sigma^*} \left(1-(1-p_w)^n - \sum_{i\in\Sigma}\left(\left(1-p_w(1-p_{[i\mid w]})\right)^n-(1-p_w)^n\right)\right)$ |
| Path Length of PaTr | $\widehat{L_P}(n) = \sum_{w\in\Sigma^*} np_w\left(1-(1-p_w)^{n-1} - \sum_{i\in\Sigma} p_{[i\mid w]}\left(1-p_w(1-p_{[i\mid w]})\right)^{n-1}\right)$ |

TABLE 1. Expectations of size and path length for tries (Tr) and Patricia tries (PaTr).

1.6. **Expectation of parameters.** Let $(\mathcal{P}_z, \mathcal{S})$ denote the Poisson model of rate $z$ relative to the source $\mathcal{S}$, and $p_w$ the probability that a given infinite word begins with the prefix $w$. If the cardinality of $X$ is a random Poisson variable of rate $z$, the length of the slice $\sigma\underline{T}_{[w]}X$ is also a random Poisson variable of rate $zp_w$. Hence the expectation of parameter $\gamma$ is a sum of expectations of parameter $\delta$, $\mathbf{E}[\gamma;\mathcal{P}_z,\mathcal{S}] = \sum_{w\in\Sigma^*}\mathbf{E}[\delta;\mathcal{P}_{zp_w},B_w]$.

The expectation of the parameter is given by $\mathbf{E}[\delta;\mathcal{P}_z,B] = e^{-z}\,\frac{\partial}{\partial u}F_\delta(z,u,p_1,\cdots,p_r)\big|_{u=1}$, where $F_\delta(z,u,x_1,\cdots,x_r) = \sum_{s\in\Sigma^*}\frac{z^{|s|}}{|s|!}u^{\delta(s)}x_1^{|s|_1}\cdots x_r^{|s|_r}$.

Using algebraic depoissonization [3], based on the equalities $\mathbf{E}[Y;\mathcal{P}_z] = e^{-z}\sum_{n\geq 0}\mathbf{E}[Y;\mathcal{B}_n]\frac{z^n}{n!}$ and thus $\mathbf{E}[Y;\mathcal{B}_n] = n![z^n]e^z\mathbf{E}[Y;\mathcal{P}_z]\frac{z^n}{n!}$, one can return to the Bernoulli model. Finally, the expectations of interest are given in Table 1.

## 2. Tools for the Asymptotics of the Expectations

2.1. **Mellin analysis and Dirichlet series.** To get asymptotics for the expressions found previously, we first note that they belong to the paradigm of harmonic sums. Their Mellin transforms are given in Table 2, where $\Lambda(s) = \sum_{w\in\mathcal{M}^*}p_w^s$ and

$$\Lambda_S(s) = -\sum_{w\in\Sigma^*}p_w^s - \sum_{w\in\Sigma^*}p_w^s\sum_{i\in\Sigma}\left[(1-p_{i|w})^s-1\right]$$

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

$$\Lambda_L(s) = \sum_{w\in\Sigma^*}p_w^s\sum_{i\in\Sigma}\left[(1-p_{[i|w]})^{s-1}-1\right]$$

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

with $\Lambda^{[k]}(s) = \sum_{w\in\Sigma^*}p_w^s\sum_{i\in\Sigma}p_{i|w}^k$, for $k\geq 1$,

2.2. **Dynamical sources.** We have to restrict ourselves to a class of dynamical sources $\mathcal{S}$ (see [4] for more details and [2] for its use in a study of standard tries),

($a$) a finite or denumerable alphabet $\Sigma$,
($b$) a topological partition of $\mathcal{I} := (0,1)$ with disjoint open intervals $\mathcal{I}_a$, for $a\in\Sigma$,
($c$) an encoding mapping $\sigma$ which is constant and equal to $a$ on each $\mathcal{I}_a$,

| | |
|---|---|
| Size of Tr | $S^*(s) = -\Lambda(-s)(s+1)\Gamma(s)$ |
| Path Length of Tr | $L^*(s) = -\Lambda(-s)\Gamma(s+1)$ |
| Size of PaTr | $S_P^*(s) = \Gamma(s)\Lambda_S(-s)$ |
| Path Length of PaTr | $L_P^*(s) = -\Gamma(s+1)\big(\Lambda(-s) + \Lambda_L(-s)\big)$ |

TABLE 2. Mellin transforms of expectations.

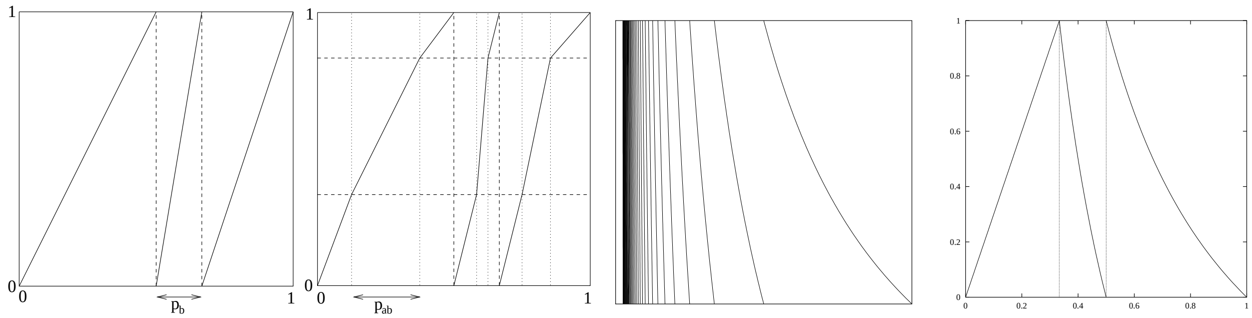


FIGURE 2. Memoryless source, Markov chain of order 1, continued fraction source, heteroclinal source.

($d$) a shift mapping $\mathcal{T}$ whose restriction to to $\mathcal{I}_a$ is a real analytic bijection from $\mathcal{I}_a$ to $\mathcal{I}$.

Besides, $\mathcal{T}$ has to satisfy more precise properties. If we let $h_a$ be the local inverse of $T$ restricted to $\mathcal{I}_a$ and $\mathcal{H}$ be the set $\mathcal{H} = \{\, h_a \mid a \in \Sigma \,\}$, then we add properties on bounds of the first derivatives, among which Rényi's condition which plays an important rôle in the study of conditional probabilities. This condition states that, if $h_a$ are the local inverse of $T$, supposed to be locally holomorphic, restricted to $\mathcal{I}_a$, then there exists a constant $K$ that bounds the ratio $|h_a''(x)/h_a'(x)|$ for all branch $h_a$ and all $x \in [0,1]$. With each $h_a$, that are only defined on $\mathcal{I}_a$, we associate its analytical extension $\tilde{h}_a$ to the whole set $\mathcal{I}$.

If $M$ maps $x \in [\,0,1\,]$ to $\big(\sigma(x), \sigma T(x), \sigma T^2(x), \dots\big) \in \Sigma^\infty$, $T$, and $\sigma$ are linked with the previously defined $\underline{T}$ and $\underline{\sigma}$ by $\underline{\sigma} M \equiv \sigma$ and $\underline{T} M \equiv MT$.

Figure 2 displays several types of dynamical sources:

*Memoryless sources.* We have affine branches of slope $1/p_a$ on intervals $\mathcal{I}_a := (q_a, q_{a+1})$, where $q_a = \sum_{i<a} p_i$.

*Markov chains.* Each $\mathcal{I}_a$ of a memoryless source is divided in $r$ intervals $\mathcal{I}_{a,b}$, $b \in \Sigma$, of length $p_{ab} = p_{[b|a]} \cdot p_a$ on which $T : \mathcal{I}_{a,b} \to \mathcal{I}_b$ has slope $\frac{p_a}{p_{ab}} = \frac{p_b}{p_{[b|a]}} \cdot \frac{1}{p_a}$. Notice that when the order $d$ of the Markov chain goes to infinity in a certain sense, one obtains at the limit a source with unbounded memory.

*Continued fractions.* With $\Sigma = \mathbb{N}$, $\mathcal{I}_a := \left(\frac{1}{a+1}, \frac{1}{a}\right)$, $T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$, and $\sigma(x) = \left\lfloor \frac{1}{x} \right\rfloor$, corresponding to a continued fraction source, we obtain a source with unbounded memory.

*Heteroclinal sources.* A source for which derivatives in different intervals can be of different signs is called *heteroclinal*. Otherwise the source is *homoclinal*, like the sources presented before.

2.3. **Ruelle operators, multi-secants and prefix probabilities.** In the context of dynamical systems, with transformations $T$ of local inverses $h_a$ are associated a transfer operator,

$$\mathcal{G}[f](x) := \sum_{a\in\Sigma} \left|h'_a(x)\right| f \circ h_a(x),$$

whose interest lies in the following property: if $X$ is a random variable with density function $f$, then the density of $T(X)$ is $\mathcal{G}[f]$. The Ruelle operator generalizes it by introducing a complex parameter $s$, interpreted in statistical physics as the temperature:

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

To deal with probabilities of prefixes of words $p_w$ and hence with fundamental intervals, we have to replace tangents with secants $H[h](x,y) := \left|\frac{h(x)-h(y)}{x-y}\right|$, leading to a first generalization $\mathbf{G}_s$ of the Ruelle operator, acting on functions $L$ of two complex variables:

$$\mathbf{G}_s[L](x) := \sum_{a\in\Sigma} \tilde{H_a}^s[h_a](x,y) L\big(h_a(x), h_a(y)\big).$$

To deal with conditional probabilities, we have to resort to a further generalization $\mathfrak{G}_s$ of the Ruelle operator involving multisecants instead of secants:

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

where the multisecants are defined by $\mathfrak{H}_s^{[m]}[h](x,y,z,t) = H[h]^{s-m}(x,y) H[h]^m(z,t)$, and $V$ by $V[h](x,y,z,t) = \big(h(x), h(y), h(z), h(t)\big)$.

Let $F$ be the distribution associated with the initial density $f$ of a source $(\mathcal{S}, f)$. The probability $p_w$ that a word begins with some prefix $w$ is $\left|F\big(h_w(0)\big) - F\big(h_w(1)\big)\right|$. For the special case $F = \mathrm{Id}$, it will be denoted $p_w^*$. Let $Q := H[F]$ be the secant of the initial distribution. Then the quasi-inverses of $\mathbf{G}_s$ and $\mathfrak{G}_s^{[k]}$ are related to Dirichlet series in the following way:

$$\Lambda(s) = \sum_{w\in\mathcal{M}^*} p_w^s = (\mathrm{Id} - \mathbf{G}_s)^{-1}[Q^s](0,1); \quad \Lambda^{[k]}(s) = \sum_{i\in\Sigma} \left(\mathrm{Id} - \mathfrak{G}_s^{[k]}\right)^{-1} \left[\mathfrak{H}_s^{[k]}[F]\right] \big(0, 1, h_i(0), h_i(1)\big).$$

Thanks to a theorem similar to the Perron–Frobenius theorem, we have the decomposition

$$(\mathrm{Id} - \mathbf{G}_s)^{-1} = \frac{\lambda(s)}{1-\lambda(s)} \mathbf{P}_s + (\mathrm{Id} - \mathbf{N}_s)^{-1}),$$

and a similar decomposition for the multi-secant operator. We deduce the asymptotics:

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

where $\Psi_1(x)$ is an eigenfunction associated with the dominant eigenvalue and chosen according to a proper normalization, and $\ell$ is the diagonal mapping of $L$. We get similar results for the $\Lambda^{[m]}$ that also have 1 as pole of order 1, and their respective residues $r_m$ are related to the dominant eigenfunctions $\Psi_1^{[m]}$ of the operators $\mathfrak{G}_1^{[m]}$, which allows us to find the singular expansion

$$\Lambda(s) = \Lambda^{[1]}(s) \asymp \frac{-1}{\lambda'(1)(s-1)} + C(\mathcal{S}),$$

where $C(\mathcal{S})$ is a constant depending on the source $\mathcal{S}$ and the initial density $f$. Using the equalities (1) and (2) we can then get asymptotics for $\Lambda_S(1)$ and $\Lambda_L(1)$.

| | |
|---|---|
| Size of Tr | $S(n) \approx \frac{1}{h(\mathcal{S})} n$ |
| Path Length of Tr | $L(n) \sim \frac{1}{h(\mathcal{S})} n \log n + \left( C(\mathcal{S}) - \frac{\gamma}{h(\mathcal{S})} \right) n$ |
| Size of PaTr | $S_P(n) \approx \frac{1}{h(\mathcal{S})} \big(1 - C_1(\mathcal{S})\big) n$ |
| Path Length of PaTr | $L(n) \sim \frac{1}{h(\mathcal{S})} n \log n + \left( C(\mathcal{S}) - \frac{\gamma + C_2(\mathcal{S})}{h(\mathcal{S})} \right) n$ |

TABLE 3. Asymptotics of expectations.

## 3. Results: Asymptotics

3.1. **General expressions.** Let $h(\mathcal{S}) = -\lambda'(1) = \lim_{\ell \to \infty} \sum_{w \in \mathcal{M}^\ell} p_w^* \big| \log p_w^* \big|$ be the entropy of fundamental intervals and, besides $C(\mathcal{S})$ encountered before, define the constants

$$
\begin{aligned}
C_1(\mathcal{S}) &= 1 - \sum_{k \geq 2} \frac{1}{k(k-1)} K^{[k]}(\mathcal{S}) = 1 - \lim_{\ell \to \infty} \sum_{w \in \mathcal{M}^\ell} p_w^* \sum_{w \in \mathcal{M}^\ell} \big(1 - p^*_{[i|w]}\big) \Big| \log\big(1 - p^*_{[i|w]}\big) \Big|, \\
C_2(\mathcal{S}) &= \sum_{k \geq 1} \frac{1}{k} K^{[k+1]}(\mathcal{S}) = \lim_{\ell \to \infty} \sum_{w \in \mathcal{M}^\ell} p_w^* \sum_{w \in \mathcal{M}^\ell} p^*_{[i|w]} \Big| \log\big(1 - p^*_{[i|w]}\big) \Big|.
\end{aligned}
$$

For random tries built from $n$ words emitted by a source $\mathcal{S}$, asymptotics of expectations are given in Table 3.

3.2. **Example.** For a memoryless source with probabilities $\{p_i\}$:

$$
\begin{aligned}
h(\mathcal{S}) &= \textstyle\sum_{i \in \mathcal{M}} p_i \big|\log p_i\big|, & C(\mathcal{S}) &= \frac{\sum_{i \in \mathcal{M}} p_i \log^2 p_i}{\left(\sum_{i \in \mathcal{M}} p_i \log p_i\right)^2}, \\
C_1(\mathcal{S}) &= 1 - \textstyle\sum_{i \in \mathcal{M}} (1 - p_i) \big|\log(1 - p_i)\big|, & C_2(\mathcal{S}) &= \textstyle\sum_{i \in \mathcal{M}} p_i \big|\log(1 - p_i)\big|.
\end{aligned}
$$

Similar formulae are available for Markov chains and continued fraction sources. Simulations are in agreement with theory.

## 4. Conclusion and Open Questions

For the average value of the size, a Patricia trie turns out to be better than a trie, and Rényi's condition is not necessary. For the average value of the path length, there is only a correcting term $C_2$ of order 2, and our proofs made use of Rényi's condition. An open question (see [1] for details) would be to know whether this correcting term remains valid for sources for which Rényi's condition does not hold, although all the natural sources we are aware of do satisfy that condition.

## Bibliography


[1] Bourdon (Jérémie). – Size and path length of Patricia tries: dynamical sources context. *Random Structures & Algorithms*, vol. 19, n° 3-4, 2001, pp. 289–315. – Special issue "Analysis of Algorithms" dedicated to Don Knuth.

[2] Clément (J.), Flajolet (P.), and Vallée (B.). – Dynamical sources in information theory: a general analysis of trie structures. *Algorithmica*, vol. 29, n° 1-2, 2001, pp. 307–369.

[3] Jacquet (Philippe) and Szpankowski (Wojciech). – Analytical de-Poissonization and its applications. *Theoretical Computer Science*, vol. 201, n° 1-2, 1998, pp. 1–62.

[4] Vallée (Brigitte). – Dynamical sources in information theory: fundamental intervals and word prefixes. *Algorithmica*, vol. 29, n° 1-2, 2001, pp. 262–306.