Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 65–70.

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

Suffix Trees and Simple Sources

Julien Fayolle
Algorithms Project, INRIA (France)

November 17, 2003

Summary by Pierre Nicodème

Abstract

Using an intricate method, Jacquet and Szpankowski [2] compared the depth of insertion into suffix-trees and tries in the non-uniform Bernoulli model, as well as the average size of suffix-trees and tries under the same model. They proved that the depth of insertion has asymptotically the same probabilistic behaviour in both cases, and that the average sizes of a trie and a suffix-tree built with n keys are asymptotically equivalent. Julien Fayolle uses a simpler combinatorial approach to compare both tree structures. When considering a two-letters alphabet with letters probability p and $q = 1 - p$, he improves the asymptotic estimation for the expectations of external path length and size of the suffix-tree (more specifically, he obtains an asymptotic bound $O(n^{0.85})$ for the difference of the two expectations when $p \in [0.46, 0.54]$). The Lempel–Ziv compression algorithm and its variants use suffix-trees as underlying data-structure.

1. Introduction

We consider a memoryless source over an alphabet $\Sigma = \{\mathtt{0}, \mathtt{1}\}$, with $\mathbf{P}(\mathtt{0}) = p, \mathbf{P}(\mathtt{1}) = 1 - p = q$ and $p > q$.

Trie. Let X be a finite set of infinite words over Σ. A trie with input keys X is defined by

$$\operatorname{trie}(X) = \begin{cases} \emptyset & \text{if } |X| = 0, \\ \bullet & \text{if } |X| = 1, \\ \langle \bullet, \operatorname{trie}(X\backslash\mathtt{0}), \operatorname{trie}(X\backslash\mathtt{1}) \rangle & \text{elsewhere,} \end{cases}$$

where the symbol $\bullet$ represents a node and $X\backslash a = \{\, u : (w \in X \text{ and } w = au) \,\}$, for $a \in \Sigma$.

Suffix tree. The suffix-tree of n keys over an infinite random string Y is the trie built over the set X of the n first suffixes of Y.

Definitions. For any string $\omega \in \Sigma^\star$, let N_ω be the *number of keys of* X (or first n suffixes of Y) whose prefix is ω.

Given a trie $\mathcal{T}$, we only consider internal nodes; therefore,

- for any internal node ν of $\mathcal{T}$, if ω_ν is the word spelled by reading the labels of the edges of $\mathcal{T}$ from the root to ν, we have $N_{\omega_\nu} \geq 2$. We write $\omega_\nu \in \mathcal{T}$ in this case;
- reciprocally, if $N_\omega = 0$, there is no node in $\mathcal{T}$ accessed by reading ω and if $N_\omega = 1$, ω leads to a leaf; ($\omega \notin \mathcal{T}$ in both cases).

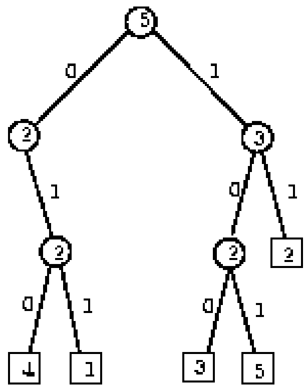

FIGURE 1. Trie for the keys `011...`, `110...`, `101...`, `010...` and `100...`; this is also the suffix tree built with the five first suffixes of the sequence `0110100....` Each internal node is numbered with the cardinality of leaves in its subtree; each leaf is numbered as starting position of the corresponding suffix in `0110100....`

Probabilistic model. In what follows, any random variable is understood as conditionned by a trie $\mathcal{T}$ over n keys, this tree being built either over a random set of infinite keys X, or over a random infinite sequence Y (suffix-tree). The generating function $\mathcal{L}(z)$ of a language $\mathcal{L}$ will always be considered in the weighted case, as $\mathcal{L}(z) = \sum_{\omega\in\mathcal{L}} \mathbf{P}(\omega) z^{|\omega|}$. We note in the following $\pi_\omega = \mathbf{P}(\omega)$ for all words ω.

External path length. Let L denote the external path length of a trie $\mathcal{T}$ build over a set of keys X. As clearly seen on the example of Figure 1, we have

$$L = \sum_{\omega\in\mathcal{T}} N_\omega \mathbf{1}_{\{N_\omega\geq 2\}} = \sum_{\omega\in\Sigma^\star} N_\omega \mathbf{1}_{\{N_\omega\geq 2\}},$$

which follows from the fact that $\mathbf{1}_{\{N_\omega\geq 2\}} = 0$ for $\omega \notin \mathcal{T}$.

Remarking that $\mathbf{E}(N_\omega \mathbf{1}_{\{N_\omega=0\}}) = 0$, this gives
(1)

$$\mathbf{E}(L) = \mathbf{E}\left(\sum_{\omega\in\Sigma^\star} N_\omega \mathbf{1}_{\{N_\omega\geq 2\}}\right) = \sum_{\omega\in\Sigma^\star} \Big(\mathbf{E}(N_\omega) - \mathbf{E}\left(N_\omega \mathbf{1}_{\{N_\omega=1\}}\right)\Big) = \sum_{\omega\in\Sigma^\star} \Big(\mathbf{E}(N_\omega) - \mathbf{P}(N_\omega = 1)\Big).$$

When $|X| = n$, we note respectively $\mathbf{E}^{(S_n)}(L)$ and $\mathbf{E}^{(T_n)}(L)$ the expectations of the external path length of a suffix-tree and of a trie. We will consider in the following $\mathbf{E}^{(S)}_L(z) = \sum_n \mathbf{E}^{(S_n)}(L) z^n$. We write in the same spirit $\mathbf{P}^{(S_n)}(N_\omega = 1), \mathbf{P}^{(T_n)}(N_\omega = 1)$, and define $\mathbf{E}^{(S)}_{N_\omega}(z) = \sum_n \mathbf{E}^{(S_n)}(N_\omega) z^n$ and $\mathbf{P}^{(S)}_{(N_\omega=1)}(z) = \sum_n \mathbf{P}^{(S_n)}(N_\omega = 1) z^n$.

Size. We obtain similarly for the size S of a trie $\mathcal{T}$

$$S = \sum_{\omega\in\mathcal{T}} \mathbf{1}_{\{N_\omega\geq 2\}} = \sum_{\omega\in\Sigma^\star} \mathbf{1}_{\{N_\omega\geq 2\}} \Rightarrow \mathbf{E}(S) = \sum_{\omega\in\Sigma^\star} \mathbf{P}(N_\omega \geq 2) = \sum_{\omega\in\Sigma^\star} \big(1 - \mathbf{P}(N_\omega = 0) - \mathbf{P}(N_\omega = 1)\big).$$

2. **Trie: External Path Length**

The external path length and size of a trie over n keys have been extensively studied. We have for a trie over n keys (see [1] for a proof; proofs for size and external path length are similar).

Theorem 1. *Asymptotically, for a memoryless source (p, q), the expectation of the external path length L is:*

- *if the source is periodic ($\log p / \log q$ rational)*
$$\mathbf{E}^{(T_n)}(L) = -\frac{n \log n}{p \log p + q \log q} + Kn + n\epsilon(n) + o(n),$$
where $\epsilon(n)$ is a periodic function of weak amplitude;
- *elsewhere (aperiodic source)*
$$\mathbf{E}^{(T_n)}(L) = -\frac{n \log n}{p \log p + q \log q} + Kn + o(n)$$

3. Suffix Tree: External Path Length

For any ω, the probability of occurrence of ω as one of the first n suffixes of the string Y is independent of the position; therefore $\mathbf{E}^{(S_n)}(N_\omega) = n \times \pi_\omega$.

We similarly have $\mathbf{E}^{(T_n)}(N_\omega) = n \times \pi_\omega = \mathbf{E}^{(S_n)}(N_\omega)$.

We want to compute the probability that ω occurs once and only once in the string. We use a variation of Guibas and Odlyzko's method. (See [3] p. 374).

We consider the autocorrelation set $\mathcal{A}_\omega$ of ω, defined as
$$\mathcal{A}_\omega = \{\, h : \omega.h = u.\omega \text{ and } |h| < |\omega| \,\}.$$
Let (a) $\mathcal{F}_\omega$, (b) $\mathcal{T}_\omega$ and (c) $\mathcal{W}_\omega$ be respectively the language of texts (a) whose first and lone occurrence of ω is at the end of the text (*First* occurrence in a text), (b) whose concatenation to ω do not create a new occurrence of ω (*Tail* following the last occurrence of ω) and (c) *Without* occurrence of ω. Let σ be any letter of the alphabet Σ.

We have two formal equations
$$\mathcal{W}_\omega.\sigma = \mathcal{W}_\omega + \mathcal{F}_\omega - \epsilon \quad \text{and} \quad \mathcal{W}_\omega.\omega = \mathcal{F}_\omega.\mathcal{A}_\omega,$$
where ϵ is the empty word (that also belongs to $\mathcal{A}_\omega$). Products and unions are unambiguous, and we obtain for the weighted generating functions (see Section 1)
$$\mathcal{W}_\omega(z) \times z = \mathcal{W}_\omega(z) + \mathcal{F}_\omega(z) - 1 \quad \text{and} \quad \mathcal{W}_\omega(z) \times \pi_\omega z^{|\omega|} = \mathcal{F}_\omega(z) \times \mathcal{A}_\omega(z).$$
Solving for $\mathcal{W}_\omega(z)$ and $\mathcal{F}_\omega(z)$, we obtain
$$\mathcal{F}_\omega(z) = \frac{\pi_\omega z^{|\omega|}}{\pi_\omega z^{|\omega|} + (1-z)\mathcal{A}_\omega(z)}.$$

Let $\leftarrow$ denote backwards reading of words of a language. For any $\mathcal{L}$, by reading backwards and forwards words, we have a bijection between $\mathcal{L}$ and $\overleftarrow{\mathcal{L}}$. This implies (with a memoryless source) that $\mathcal{L}(z) = \overleftarrow{\mathcal{L}}(z)$. But we have $\mathcal{W}_\omega(z) = \overleftarrow{\mathcal{W}_\omega}(z) = \overleftarrow{\mathcal{W}}_{\overleftarrow{\omega}}(z) = \mathcal{W}_{\overleftarrow{\omega}}(z)$ and therefore $\mathcal{F}_\omega(z) = \mathcal{F}_{\overleftarrow{\omega}}(z)$. This implies that
$$\mathcal{F}_\omega = \overleftarrow{\overleftarrow{\omega}.\mathcal{T}_{\overleftarrow{\omega}}} \qquad \Rightarrow \qquad \mathcal{T}_{\overleftarrow{\omega}}(z) = \frac{\overleftarrow{\mathcal{F}_\omega}(z)}{\overleftarrow{\omega}(z)} = \frac{\overleftarrow{\mathcal{F}_{\overleftarrow{\omega}}}(z)}{\omega(z)} = \mathcal{T}_\omega(z) = \frac{\mathcal{F}_\omega(z)}{\omega(z)} = \frac{\mathcal{F}_\omega(z)}{\pi_\omega z^{|\omega|}}.$$
We also have $\mathcal{O}_\omega(z) = \mathcal{F}_\omega(z)\mathcal{T}_\omega(z)$ for the generating function $\mathcal{O}_\omega(z)$ of texts with exactly one match with ω. Summing up over ω, we obtain the generating function $E_L^{(S)}(z)$ of expectations of

the external path length of a suffix-tree (see Jacquet and Szpankowski [2] for another proof)

$$(2) \quad \mathbf{E}_L^{(S)}(z) - \sum_{\omega\in\Sigma^\star} \mathbf{E}_{N_\omega}^{(S)}(z) = \sum_{\omega\in\Sigma^\star} \mathbf{P}_{(N_\omega=1)}^{(S)}(z) = \sum_{\omega\in\Sigma^\star} \mathcal{O}_\omega(z) = \sum_{\omega\in\Sigma^\star} \frac{\pi_\omega z^{|\omega|}}{\left(\pi_\omega z^{|\omega|} + (1-z)\mathcal{A}_\omega(z)\right)^2}.$$

4. External Path Length, Suffix Tree versus Trie

We consider (the index n corresponding to n keys) the differences

$$(3) \qquad \Delta_n = \mathbf{E}^{(T_n)}(L) - \mathbf{E}^{(S_n)}(L) = \sum_{\omega\in\Sigma^\star} \delta_\omega^{(n)} = \sum_{\omega\in\Sigma^\star} \mathbf{P}^{(T_n)}(N_\omega = 1) - \mathbf{P}^{(S_n)}(N_\omega = 1),$$

(since $\mathbf{E}^{(T_n)}(N_\omega) = \mathbf{E}^{(S_n)}(N_\omega)$ for all ω). We will prove that $\Delta_n = O(n^{0.85})$ for $0.5 < p < 0.54$.

The minimum μ_n of the fillup levels of both trees is $\alpha \log(n)$ for a given $\alpha > 0$ with probability one; all the following will be conditionned by the fact that $\mu_n = \alpha \log(n)$. With $|\omega| < \mu_n$, we have $\delta_\omega^{(n)} = 0$ and when $|\omega| \geq \mu_n$, asymptotically, $\pi_\omega = o(1)$.

4.1. Asymptotic contribution of the trie.

As a consequence of the preceding remark, we have

$$(4) \qquad \mathbf{P}^{(T_n)}(N_\omega = 1) = n\pi_\omega \times (1-\pi_\omega)^{n-1} \simeq n\pi_\omega \times e^{-n\pi_\omega}.$$

4.2. Asymptotic contribution of the suffix-tree.

Each function $\mathcal{O}_\omega(z)$ in Equation 2 is a rational function with dominant pole ρ_ω. We begin by isolating the dominant poles of these functions that give the asymptotic behaviour of the terms corresponding to the suffix-tree in Δ_n.

Lemma 1. *For $1 < R < 1/p$ and $|\omega| \geq \mu_n$, the set of poles of the functions $\mathcal{O}_\omega(z)$ contained in the disk centered at the origin and of radius R is exactly $\{\rho_\omega;\ \omega \in \Sigma^\star, |\omega| \geq \mu_n\}$.*

Proof. The proof is based on an application of the Rouché theorem; let $f(z) = \pi_\omega z^{|\omega|}$, and $g(z) = (1-z)\mathcal{A}_\omega(z)$. We are above the level μ_n and therefore, for $|z| < 1/p$, we have $|f(z)| = o(1)$. Let d be the smallest period of ω.

– If $d < |\omega|/2$, we have $\omega = u^r.v$, (with $|v| < |u| = d$ and v a prefix of u), and the second smallest period is at least $|\omega|/2$ (a consequence of the Wilf theorem). This gives (omitting the subscript ω)

$$\mathcal{A}(z) = 1 + S(z) + T(z), \qquad \text{with} \qquad S(z) = \pi_u z^{|u|} + \cdots + (\pi_u z^{|u|})^r,$$

where $T(z)$ is a polynomial of lowest degree $\geq \mu_n/2$; this implies $|T(z)| = o(1)$ for $|z| < 1/p$ and

$$|\mathcal{A}(z)| = \left| \frac{1}{1 - \pi_u z^{|u|}} \right| + o(1).$$

For all $d \geq 2$, this implies (up to negligible terms)

$$|\mathcal{A}(z)| \geq \frac{1}{1 + \pi_u |z|^{|u|}} \geq \frac{1}{1 + p|z|} \qquad \text{for} \qquad |z| < \frac{1}{p}.$$

We have the same lower bound for $d = 1$.

– If $d > |\omega|/2$, we have $S(z) = 0$ and $|\mathcal{A}(z)| = 1 + o(1)$ for $|z| < 1/p$.

Therefore, for any $R < 1/p$ (and we choose in the following $R > 1$), there exists a number N such that for $n > N$, on the circle $|z| = R$, we have $|\mathcal{A}_\omega(z)| > \kappa$ for a given $\kappa > 0$ and for all ω such that $|\omega| > \mu_n$; this implies $|f| < |g|$ over this circle. Moreover f and g are analytic everywhere, which implies that $f + g$ has as many zeros as g inside the disk $\mathcal{D}_R = \{z, |z| < R\}$, for any ω. The polynomial $\mathcal{A}(z)$ has no roots inside the disk $|z| < R$ and therefore $f(z) + g(z)$ has only one root inside the disk $\mathcal{D}_R$. □

For each ω, we compute

$$o_n^{(\omega)} = [z^n]\mathcal{O}_\omega(z) = \mathrm{Res}\left(\frac{\mathcal{O}_\omega(z)}{z^{n+1}}, 0\right) = I(\mathcal{C}_R) - \mathrm{Res}\left(\frac{\mathcal{O}_\omega(z)}{z^{n+1}}, \rho_\omega\right), \quad \text{where} \quad I(\mathcal{C}_R) = \int_{\mathcal{C}_R} \frac{\mathcal{O}_\omega(z)}{z^{n+1}} dz,$$

and $\mathcal{C}_R$ is the circle $|z| = R$. Considering

$$\mathcal{O}_\omega(z) = \frac{\pi_\omega z^{|\omega|}}{\left(\pi_\omega z^{|\omega|} + (1-z)\mathcal{A}_\omega(z)\right)^2},$$

for $|\omega| > \mu_n$ and $|z| = R$ ($R \in]1, 1/p[$) we have $\pi_\omega z^{|\omega|} = o(1)$ and $|(1-z)\mathcal{A}(z)| \geq (R-1)\kappa$. Therefore $I(\mathcal{C}_R) = O(R^{-n})$.

By bootstrapping, we have $\rho_\omega = 1 + \pi_\omega / A_\omega(1) + o(\pi_\omega)$. An expansion of the denominator of $\mathcal{O}_\omega$ in a neighborhood of ρ_ω gives

$$\text{(5)} \qquad \mathbf{P}^{(S_n)}(N_\omega = 1) = o_n^{(\omega)} = n\pi_\omega e^{-\frac{n\pi_\omega}{A_\omega(1)}} + O\left(n\pi_\omega^2 e^{-\frac{n\pi_\omega}{A_\omega(1)}}\right) + O(|\omega| n\pi_\omega^2) + C\left(\frac{1}{R}\right)^n.$$

Plugging Equations 4 and 5 into Equation 3 gives

$$\text{(6)} \qquad \Delta_n = \sum_{\omega\in\Sigma^\star, |\omega|>\mu_n} \delta_\omega^{(n)} \quad \text{with} \quad \delta_\omega^{(n)} \simeq \sum_{\omega\in\Sigma^\star, |\omega|>\mu_n} n\pi_\omega \left(e^{-n\pi_\omega/A_\omega(1)} - e^{-n\pi_\omega}\right).$$

4.3. **Bounding Δ_n by partitioning the motifs ω.** The aim of this section is to prove the validity of each entry of the following table (where $C_p = \log 2/\log(1/p)$ and $\Omega_s^{(n)}, \Omega_{i,p}^{(n)}, \Omega_{i,a}^{(n)}$, and $\Omega_l^{(n)}$ respectively are the set of short, intermediate periodic, intermediate aperiodic and long motifs when the number of input keys is n).

short patterns $\|\omega\| < \frac{5}{6}\log_{1/q} n$	intermediate patterns	long patterns $1.5\log_{1/p} n < \|\omega\|$
$\Delta_n^{(s)} = \sum_{\omega\in\Omega_s^{(n)}} \delta_\omega^{(n)} = o(1)$	periodic ($A_\omega(1) \geq 1 + 2^{-\|\omega\|/2}$) $\Delta_n^{(i,p)} = \sum_{\omega\in\Omega_{i,p}^{(n)}} \delta_\omega^{(n)} = O\big(n^{0.75C_p}\log n\big)$	$\Delta_n^{(l)} = \sum_{\omega\in\Omega_l^{(n)}} \delta_\omega^{(n)} = O\left(\sqrt{n}\right)$
($\mu_n < \|\omega\|$)	aperiodic ($A_\omega(1) < 1 + 2^{-\|\omega\|/2}$) $\Delta_n^{(i,a)} = \sum_{\omega\in\Omega_{i,a}^{(n)}} \delta_\omega^{(n)} = O\big(n^{0.75C_p}\big)$	

4.3.1. *Short patterns.* For these patterns, we have

$$n\pi_\omega > nq^{\frac{5}{6}\log_{1/q} n} = n^{1-\frac{5}{6}} = n^{1/6} \to \infty \qquad \text{and} \qquad \left|\Omega_s^{(n)}\right| = O(n^\alpha \log n) \quad \text{where } \alpha = \frac{5\log 2}{6\log(1/q)}.$$

Therefore $\Delta_n^{(s)}$ behaves as $n^\alpha \log n \times e^{-n^{1/6}}$ as n tends to infinity and is $o(1)$.

4.3.2. *Long patterns.* In this case, let $k_l(n) = 1.5\log_{1/p} n$; we have

$$n\pi_\omega \leq np^{k_l(n)} = np^{1.5\log_{1/p} n} = n^{-0.5} \to 0.$$

Expanding $\delta_\omega^{(n)}$ for small np_ω gives $\delta_\omega^{(n)} \simeq (n\pi_\omega)^2 (1 - 1/A_\omega(1))$. Therefore we have

$$\Delta_n^{(l)} \simeq \sum_{k\geq k_l(n)} \sum_{\omega\in\Sigma^k} n^2\pi_\omega^2 \left(1 - \frac{1}{A_\omega(1)}\right) \quad \text{and} \quad \sum_{\omega\in\Sigma^k} \pi_\omega^2 = \sum_{0\leq i\leq k} \binom{k}{i} p^{2i} q^{2(k-i)} = (p^2+q^2)^k < p^k.$$

This implies $\Delta_n^{(l)} = O(n^2 p^{k_l(n)}) = O(\sqrt{n})$.

4.3.3. *Intermediate patterns.* Julien Fayolle proves in [1] that $\sum_{\omega\in\Sigma^k} A_\omega(1) = 2^k + k - 1$. He defines the set of intermediate *periodic* patterns as

$$\Omega_{i,p}^{(n)} = \left\{\omega :\ k_s(n) < |\omega| \le k_l(n),\ A_\omega(1) \ge 1 + 2^{-k/2}\right\},$$

where $k_s(n) = 5/6 \log_{1/q} n$ (in the uniform case these patterns verify $d < |\omega|/2$). He also proves that $|\Omega_{i,p}^{(n)}| \le k 2^{k/2}$.

Summing up for the intermediate periodic patterns, we obtain

$$\Delta_n^{i,p} = \sum_{\omega\in\Omega_{i,p}^{(n)}} \delta_\omega^{(n)} < K \sum_{k=k_s(n)}^{k_l(n)} k 2^{k/2} = O\left(\log n \times e^{\frac{1.5\log 2\log n}{2\log(1/p)}}\right) = O(n^{0.85}) \quad \text{for} \quad p < 0.54.$$

The set $\Omega_{i,a}^{(n)}$ of intermediate *aperiodic* patterns is the complementary set of $\Omega_{i,p}^{(n)}$, inside the bounds $k_s(n) < |\omega| < k_l(n)$. We therefore have $1/A_\omega(1) \ge 1/(1+2^{-|\omega|/2}) \ge 1 - 2^{-|\omega|/2}$ and for $|\omega| = k$

$$\delta_\omega^{(n)} \le n\pi_\omega \left(e^{n\pi_\omega 2^{-k/2}} - 1\right).$$

We also have for $\omega \in \Omega_{i,a}^{(n)}$

$$n\pi_\omega 2^{-|\omega|/2} < np^{k_s(n)} 2^{-k_s(n)/2} \to 0 \quad \text{for} \quad \frac{5}{6}\frac{\log(p/\sqrt{2})}{\log(1/q)} + 1 < 0 \quad \text{or} \quad p < p_0 = 0.5469.$$

By expanding the exponential in the neighborhood of zero, we get $\delta_\omega^{(n)} \simeq (n\pi_\omega)^2 e^{-n\pi_\omega} 2^{-|\omega|/2}$, which gives (remarking that $x^2 e^{-x}$ is bounded on $\mathbb{R}^+$)

$$\Delta_n^{(i,a)} \le \sum_{k=k_s(n)}^{k_l(n)} K' 2^k 2^{-k/2} = O\left(e^{\frac{1.5\log 2\log n}{2\log(1/p)}}\right) = O\left(n^{0.85}\right) \quad \text{for} \quad p < p_0.$$

Remark that we also have $\Delta_n^{(i,p)} = O\left(n^{0.85}\right)$ for $p < p_0$.

4.4. **End result.** Summarizing the preceding results gives

Theorem 2. *For a suffix-tree with n keys, we have asymptotically for $p \le 0.54$*

$$\mathbf{E}(L_n^{(S)}) = \frac{n\log n}{p\log p + q\log q} + (K + \epsilon(n))n + O\left(n^{0.85}\right),$$

where $L_n^{(S)}$ is the external path length of the tree and $\epsilon(n)$ is a periodic function of small amplitude.

The same method applies for analysis of the asymptotic expectation of size.

Bibliography

[1] Fayolle (Julien). – Paramètres des arbres suffixes dans le cas de sources simples. – 2003. Master's thesis.
[2] Jacquet (P.) and Szpankowski (W.). – Autocorrelation on words and its applications. Analysis of Suffix Trees by String Ruler Approach. *Journal of Combinatorial Theory, Series A*, n° 66, 1994, pp. 237–269.
[3] Sedgewick (Robert) and Flajolet (Philippe). – *An Introduction to the Analysis of Algorithms.* – Addison-Wesley Publishing Company, 1996, 512p.