The height of the Lyndon tree

Lucas Mercier, Philippe Chassaing

THE HEIGHT OF THE LYNDON TREE

LUCAS MERCIER AND PHILIPPE CHASSAING

Abstract. We consider the set $\mathcal{L}_n$ of n-letters long Lyndon words on the alphabet $\mathcal{A} = \{0, 1\}$. For a random uniform element L_n of the set $\mathcal{L}_n$, the binary tree $\mathfrak{L}(L_n)$ obtained by successive standard factorizations of L_n and of the factors produced by these factorizations is the *Lyndon tree* of L_n. We prove that the height H_n of $\mathfrak{L}(L_n)$ satisfies

$$\lim_n \frac{H_n}{\ln n} = \Delta^\star,$$

in which the constant $\Delta^\star$ is solution of an equation involving large deviation rate functions related to the asymptotics of Eulerian numbers ($\Delta^\star \simeq 5.092\ldots$). The convergence is the convergence in probability of random variables.

1. Introduction

1.1. **Lyndon words and Lyndon trees.** We recall some notations of [Lot97] for readability. For an alphabet $\mathcal{A}$, $\mathcal{A}^n$ is the set of n-letters words, and the language, i.e. the set of finite words,

$$\{\emptyset\} \cup \mathcal{A} \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots,$$

is denoted by $\mathcal{A}^\star$. The length of a word $w \in \mathcal{A}^\star$ is denoted by $|w|$. A total order, $\prec$, on the alphabet $\mathcal{A}$, induces a corresponding lexicographic order, again denoted by $\prec$, on the language $\mathcal{A}^\star$: the word w_1 is smaller than the word w_2 (for the lexicographic order, $w_1 \prec w_2$) at one of the following conditions: either w_1 is a prefix of w_2, or there exist words p, v_1, v_2 in $\mathcal{A}^\star$ and letters $a_1 \prec a_2$ in $\mathcal{A}$, such that $w_1 = pa_1v_1$ and $w_2 = pa_2v_2$. For any factorization $w = uv$ of w, vu is called a rotation of w, and the set $\langle w\rangle$ of rotations of w is called the *necklace* of w. A word w is primitive if $|w| = \#\langle w\rangle$.

The notion of *Lyndon word* has many equivalent definitions, to be found, for instance, in [Lot97].

Definition 1 (Lyndon word)**.** A word w is Lyndon if w is primitive and is the smallest element of $\langle w\rangle$.

Example. The word $w = \mathtt{aabaab}$ is the smallest in its necklace

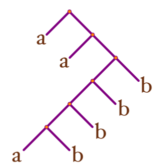
$$\langle w\rangle = \{\mathtt{aabaab}, \mathtt{abaaba}, \mathtt{baabaa}\}$$

but is not Lyndon; `baac` is not Lyndon, nor `acba` or `cbaa`, but `aacb` is Lyndon.

Here is a recursive characterization of Lyndon words:

2000 *Mathematics Subject Classification.* 60F05, 60C05, 68R15, 05A05.

Key words and phrases. Random word, Lyndon word, Lyndon tree, binary search tree, branching random walk, Yule process, Galton-Watson process.



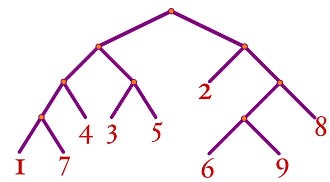


FIGURE 1. a. $\mathcal{A} = \{\mathtt{a}, \mathtt{b}\}$ and $\mathfrak{L}(\mathtt{a}^3\mathtt{b}^4)$, b. $\mathcal{A} = \{\mathtt{1}, \mathtt{2}, \dots, \mathtt{9}\}$ and $\mathfrak{L}(\mathtt{174352698})$.

Proposition 1. *One-letter words are Lyndon. A word w with length $n \geq 2$ is a Lyndon word if and only if there exists two Lyndon words u and v such that $w = uv$ and $u \prec v$.*

Among such decompositions of w, the decomposition with the longest second factor (or suffix) v is called the *standard* decomposition.

Example. $\mathtt{0011} = (\mathtt{001})(\mathtt{1}) = (\mathtt{0})(\mathtt{011})$ is a Lyndon word with two such decompositions. The latter is the standard decomposition.

The set of Lyndon words is denoted by $\mathcal{L}$, and we set $\mathcal{L}_n = \mathcal{L} \cap \mathcal{A}^n$. The Lyndon tree (cf. [HR03], also called standard bracketing tree by some authors, e.g. [Bar90]) of the Lyndon word w is a binary tree obtained by iteration of the standard decomposition:

Definition 2 (Lyndon tree)**.** For $w \in \mathcal{L}$, the Lyndon tree $\mathfrak{L}(w)$ of w is a labeled finite binary tree defined as follows:

- if $|w| = 1$, $\mathfrak{L}(w)$ has a unique node labeled w, and no edges;
- if (u, v) is the standard decomposition of w, then $\mathfrak{L}(w)$ is the binary tree with label w at its root, $\mathfrak{L}(u)$ as its left subtree and $\mathfrak{L}(v)$ as its right subtree.

Remark 1. The labels of the leaves of a Lyndon tree are letters. Also, the label of an internal node is the concatenation of the labels of its two children, and, if $|w| = n$, $\mathfrak{L}(w)$ is a rooted binary tree with n leaves, and $n-1$ internal nodes. In general, the height of a rooted tree $\mathfrak{T}$, denoted $h(\mathfrak{T})$, is the maximal distance between the root of $\mathfrak{T}$ and one of its leaves.

1.2. **Result.** The asymptotic behavior of the size of the right and left subtrees of $\mathfrak{L}(L_n)$, for n large, have been studied in [BCN05, CZA10], for L_n a random element of $\mathcal{L}_n$. The height $h(\mathfrak{L}(L_n))$ of $\mathfrak{L}(L_n)$ is of interest for analysis of algorithms and cryptanalysis , cf. [SSM92, SR03, BCN05], but it seems to have resisted analysis up to now.

For a 2-letter alphabet, say $\mathcal{A} = \{0, 1\}$, and for $n \geq 1$, let L_n denote a uniform random word in $\mathcal{L}_n$. Let $(A(n,k))_{n,k}$ denote the Eulerian numbers, i.e. $A(n,k)$ is the number of permutations σ of n symbols having exactly k descents (k places where $\sigma(i) \geq \sigma(i+1))$. Set

$$\begin{aligned}
\Xi(\theta) &= \lim_n \frac{1}{n} \ln(A(n, \lfloor \theta n \rfloor)/n!),\\
\Psi(\lambda,\mu,\nu) &= \ln\left(\tfrac{(1+\mu)^{1+\mu}}{\mu^\mu}\ \tfrac{(e\lambda\ln 2)^\nu \ln 2}{\nu^\nu 2^\lambda}\right) +\ \Xi(\lambda-\mu),\\
\Delta^\star &= \sup_{\lambda,\mu,\nu>0} \tfrac{(1+\nu+\mu)\ln 2+\Psi(\lambda,\mu,\nu)}{\lambda(\ln 2)^2}\\
&= 5.092\ldots
\end{aligned}$$

See Lemma 4 or [GK94, p. 299] for an expression of Ξ. We shall prove that:

Theorem 1.

$$\frac{h(\mathfrak{L}(L_n))}{\ln n} \xrightarrow{\mathbb{P}} \Delta^\star.$$

Conditionally given their lengths, the two factors of the standard decomposition of a uniform Lyndon word are not independent, and they are not uniform Lyndon words either, which seems to preclude a recursive approach to the proof of this Theorem. We shall rather use a coupling method: in Section 2 we sketch the main steps of the construction, on the same probability space, of a random Lyndon tree, and of two well studied trees, the binary search tree of a random uniform permutation, and a Yule tree, in such a way that the height of the Lyndon tree is closely related to some statistics of the two other trees. Then Theorem 1 follows from a large deviation result presented in Section 3.

2. Coupling results

2.1. **Reduction to a Bernoulli source.** If the word u is primitive but is not Lyndon, the Lyndon tree $\mathfrak{L}(u)$ of u is the Lyndon tree of the unique Lyndon word in the necklace $\langle u\rangle$ of u, in short $\mathfrak{L}(u)$ is the Lyndon tree of the Lyndon word of u. If u is periodic, we define the Lyndon word of u as the word $\mathtt{0}^{|u|-1}\mathtt{1}$, and $\mathfrak{L}(u)$ is defined accordingly. Then the following algorithm:

- let W_∞ be an infinite word of uniformly random characters, obtained through the binary expansion of a number U uniformly distributed on $[0,1]$;
- let W_n be the word W_∞ truncated after n letters, and let L_n be the Lyndon word of W_n.

produces a n-letters long random Lyndon word L_n. Conditionally, given that W_n is primitive, this random Lyndon word L_n is uniform on $\mathcal{L}_n$, but the unconditional distribution of L_n fails to be uniform due to the small probability that W_n is periodic. However, the total variation distance between the probability distribution of L_n and the uniform distribution on $\mathcal{L}_n$ is $\mathcal{O}\left(2^{-n/2}\right)$ (cf. e.g. [CZA10, Lemma 2.1]), thus any property that holds true asymptotically almost surely with respect to either distribution, holds true a.a.s. for both. From now on, we shall consider that L_n is produced by the previous algorithm.

2.2. **Poissonization.** In the first steps of the recursive construction of $\mathfrak{L}(L_n)$, the sizes of the factors of the successive standard decompositions are predicted by the positions of the longest runs of $\mathtt{0}$'s, and the structure of the top levels of $\mathfrak{L}(L_n)$ is given by the lexicographic comparisons between the suffixes of L_n beginning at these longest runs. But when n is large, the number of runs of $\mathtt{0}$'s is typically $n/4$, and several among these runs are tied for the title of the longest run. Actually the lengths of the runs behave pretty much as a sample of $n/4$ i.i.d. geometric random variables with parameter $1/2$, and, according to [BSW94], for any strictly increasing sequence n_k such that $\lim_k \log_2 n_k - \lfloor \log_2 n_k \rfloor = \alpha \in [0,1)$, the probability p_{m,n_k} that $m \geq 1$ among the n_k elements of such a sample are tied for the maximum is given, approximately, by

$$p_{m,n_k} \simeq \sum_{j\in\mathbb{Z}} e^{-2^{\alpha+j}} \frac{(2^{\alpha+j-1})^m}{m!}. \tag{1}$$

Thus the number of ties does not converge in distribution, but has a set of limit distributions indexed by $\alpha \in \mathbb{R}/\mathbb{Z}$.

Such a complex behavior does not bode well, so we shall rather analyze a transform of this problem, in the form of the Lyndon tree of a word with random length. Consider the finite word W^ℓ formed by a letter $\mathtt{1}$ followed by the truncation of W_∞ at the position τ_ℓ of the ℓth $\mathtt{0}$ in the first run of ℓ consecutive $\mathtt{0}$'s of W_∞. Then W^ℓ is primitive, and L^ℓ denotes the Lyndon word of W^ℓ, *i.e.*:

$$W^\ell = \mathtt{1}\,\underbrace{\mathtt{010110}\ldots\mathtt{1}\overbrace{\mathtt{000000}}^{\ell \mathtt{0}s}}_{\text{prefix of } W_\infty} \quad \text{and} \quad L^\ell = \overbrace{\mathtt{000000}}^{\ell \mathtt{0}s}\,\mathtt{1}\,\underbrace{\mathtt{010110}\ldots\mathtt{1}}_{\text{prefix of } W_\infty}.$$

If $\hat\tau_\ell$ is the position of the last $\mathtt{1}$ before τ_ℓ, L^ℓ is the concatenation of $\mathtt{0}^\ell\mathtt{1}$ and of the truncation of the word W_∞ at position $\hat\tau_\ell$.

Now, there exists a unique longest run of $\mathtt{0}$'s in W^ℓ as well as in L^ℓ, and this run is ℓ letters long, to be compared with the behavior revealed by (1). Moreover, if Z_k denotes the number of runs longer than $\ell-k-1$, then $Z_0 = 1$ and $(Z_j)_{0\leq j\leq \ell-1}$ is a Galton-Watson process with offspring distribution $2^{-k}\mathbb{1}_{k\geq 1}$, so that Z_j has a geometric distribution with parameter 2^{-j}, see for instance [Dev92]. The family tree of this Galton-Watson process gives a lot of information on $\mathfrak{L}(L^\ell)$, leading ultimately to the proof of Theorem 2 below. The replacement of $\mathfrak{L}(W_n)$ by $\mathfrak{L}(L^\ell)$, motivated by (1), has deeper consequences, initially unexpected to us: embedded in $\mathfrak{L}(L^\ell)$, appears a Yule family tree (see Section 2.5.2), and with this Yule tree come, besides several useful Galton-Watson trees, some Poisson point processes that leads to the Poisson-like formula (4). The asymptotic analysis of (4) finally leads to the computation of $\Delta^\star$. Thus the replacement of $\mathfrak{L}(W_n)$ by $\mathfrak{L}(L^\ell)$ can be seen as some kind of *poissonization*.

Note that τ_ℓ, the length of L^ℓ up to one unit, has a geometric distribution of order ℓ , cf. [BK02, p. 10], and, typically, grows exponentially fast with ℓ:

$$\mathbb{E}[|L^\ell|] = 2^{\ell+1} - 1.$$

Thus, expectedly, the typical height of the Lyndon tree of L^ℓ grows linearly with ℓ:

Theorem 2. $\frac{h(\mathfrak{L}(L^\ell))}{\ell} \xrightarrow[\ell\to\infty]{\mathbb{P}} \Delta^\star \ln 2.$

In Section 4 we deduce Theorem 1 from Theorem 2: choosing $\ell(n) = \log_2 n - \varepsilon_n$ in such a way that $\mathbb{P}\left(\tau_{\ell(n)} \geq n\right)$ is small, and that, with a large probability, $W^{\ell(n)}$ is a factor of W_n, we compare carefully $\mathfrak{L}(W^{\ell(n)})$ and $\mathfrak{L}(W_n)$. We set

$$\Delta^\bullet = \Delta^\star \times \ln 2.$$

Sections 2 and 3 are devoted to the proof of Theorem 2.

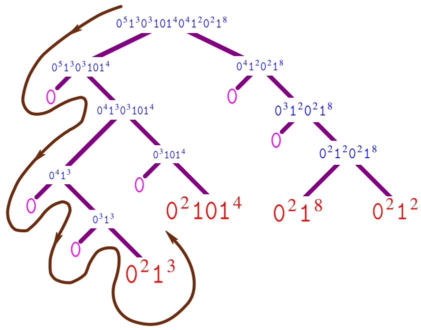


Figure 2. For $L^5 = 0^5 1^3 0^3 101^4 0^4 1^8 0^2 1^2$, the tree $\mathfrak{T}_5 = \mathfrak{L}^2(L^5)$ has 6 needles and 4 blades. In brown, its contour traversal.

2.3. **Reduction to a skeleton.** In the top levels of the tree $\mathfrak{L}(L^\ell)$, the successive standard decompositions of the Lyndon word L^ℓ, at the smallest suffixes of L^ℓ, split the word L^ℓ at the longest runs of 0's. For ℓ large enough, the longest runs are sparse enough to preserve some degree of independence between the factors. This is not true anymore at the lowest levels of the tree $\mathfrak{L}(L^\ell)$. For this reason, it is easier to split the study of the Lyndon tree in two parts: the first one focuses on the top of the tree, where the runs of 0's are still above a threshold a_ℓ, and the second part studies a forest of *shrubs* at the bottom of the tree, each of them labeled with a factor of L^ℓ that contains only runs of 0's shorter than the threshold. The top part is a tree itself (a subtree of $\mathfrak{L}(L^\ell)$), and each shrub of the forest at the bottom of $\mathfrak{L}(L^\ell)$ is rooted at (or grafted on) a leaf of the top tree. We follow here the same path as [BD08], our shrubs playing the same rôle as their *spaghetti-like subtrees*. Let us define by induction the tree above the threshold k, with $k \geq 1$:

Definition 3 (Top tree)**.** If w denotes a Lyndon word, $\mathfrak{L}^k(w)$ is a finite labeled binary tree built recursively according to the following set of instructions:

- if w has one factor 0^k or less (thus 0^{k+1} is not a factor of w), $\mathfrak{L}^k(w)$ is a single node, labeled w;
- otherwise, let (u, v) be the standard decomposition of w. Then the root of $\mathfrak{L}^k(w)$ has label w, the left subtree of $\mathfrak{L}^k(w)$ is $\mathfrak{L}^k(u)$ and the right subtree is $\mathfrak{L}^k(v)$.

$\mathfrak{L}^k(w)$ is called the top tree associated to w, with threshold k.

Set
$$\mathfrak{T}_\ell = \mathfrak{L}^{a_\ell}(L^\ell).$$
The threshold a_ℓ depends on ℓ. It has to be large enough for the top tree to retain the independence properties between the factors, but small enough that we can handle the shrubs, though they lack these nice independence properties. We shall assume that
$$a_\ell \uparrow \infty \quad \text{and} \quad a_\ell = o(\ell).$$

2.4. **A binary search tree.** Let $\mathfrak{s}_k(w)$ denote the suffix of the finite word w with length $|w| - k + 1$, and let $\sigma_w(k)$ be the rank of $\mathfrak{s}_k(w)$ once the sequence $(\mathfrak{s}_j(w))_{1 \leq j \leq |w|}$ is sorted in increasing lexicographic order. Then a word w is Lyndon if and only if $\sigma_w(1) = 1$, and in this case, according to [HR03], $\mathfrak{L}(w)$ is the binary search tree of the permutation σ_w. Note that according to [SF13, p. 362], it should rather be called the heap-ordered tree of σ_w, see Section 2.5.2 and Figure 6. The asymptotic behavior of the height of the binary search tree of a random uniform permutation (not different from that of the corresponding heap-ordered tree, cf. [SF13, p. 364, Th. 7.1]) is well studied [Dev86, Rob], but the distribution of σ_{L_n} or of σ_{L^ℓ} is all but uniform. In this Section, we produce a coupling between σ_{L^ℓ} and a random uniform permutation. In the next Sections we inspect the relations between the heap-ordered trees of these two random permutations.

Let us take a closer look at $\mathfrak{T}_\ell$: observe that if $w \in \mathcal{L}$, then $\mathtt{0}w \in \mathcal{L}$, and the two factors of the standard factorization of $\mathtt{0}w$ are $\mathtt{0}$ and w. Thus either a leaf v of $\mathfrak{T}_\ell$ has label $\mathtt{0}$, and v is called a *needle*, or the label of v is a factor of w beginning with $\mathtt{0}^{a_\ell}\mathtt{1}$, and v is called a *blade*. The number N_ℓ of blades of $\mathfrak{T}_\ell$ has a geometric distribution with parameter $2^{-\ell+a_\ell}$, and the set of blades has a natural order related to the contour traversal (see Figure 2), that allows to identify it to $[\![1, N_\ell]\!]$. Note that the number of needles has a simple expression in terms of a Galton-Watson process with geometric offspring distribution. In the analysis of the shape of $\mathfrak{T}_\ell$, the configuration of the needles is a special concern, and the following bound will help at some point. Let $g(v)$ denote the number of needles on the path between a blade v and the root $\emptyset$ of $\mathfrak{L}^k(w)$, and let $M(w)$ be the length of the longest run of $\mathtt{0}$'s in $\langle w\rangle$. Then

Lemma 1. *For any blade v of $\mathfrak{L}^k(w)$, $g(v) \le M(w) - k$.*

Proof. For any interior node ν of $\mathfrak{L}^k(w)$, let $g(\nu)$ denote the natural extension of g to the interior nodes of $\mathfrak{L}^k(w)$ and let $m(\nu)$ be the length of the longest run of $\mathtt{0}$'s in the label $f(\nu)$ of ν. Then, if ν is not a needle of $\mathfrak{L}^k(w)$, $g(\nu) + m(\nu)$ does not decrease on the edge towards the root, for

- either the father μ of ν has the label $\mathtt{0}f(\nu)$, in which case $g(\nu) = 1 + g(\mu)$ and $m(\nu) = -1 + m(\mu)$;
- or $f(\mu) \ne \mathtt{0}f(\nu)$, in which case $g(\nu) = g(\mu)$ and $m(\nu) \le m(\mu)$.

Thus $m(\emptyset) + g(\emptyset) = M(w) \ge m(v) + g(v) = g(v) + k$. □

We shall need some notations: in the contour traversal of $\mathfrak{T}_\ell$, there exists a sequence of $n_v - a_\ell \ge 0$ needles between a blade v and the previous blade (or between v and the root, if there exists no previous blade). The concatenation, starting at this sequence of needles, included, of the labels of the leaves in the order of the contour traversal of $\mathfrak{T}_\ell$, is a suffix $\mathfrak{s}(v)$ of L^ℓ that can be written $\mathtt{0}^{n_v}\mathtt{1}\,t_v$, the run $\mathtt{0}^{n_v}$ being maximal in the sense that $\mathtt{0}^{n_v+1}\mathtt{1}\,t_v$ is not a suffix of L^ℓ.

The words of the sequence $(t_v)_{1\le v\le N_\ell}$ have different lengths, being proper suffixes of each other, so they are all different, and we can give a reformulation of the algorithm that produces $\mathfrak{T}_\ell$, or more generally $\mathfrak{L}^k(w)$, in terms of the family $T_\ell = ((n_v, t_v))_{1\le v\le N_\ell}$ of the blades (with labels $\mathtt{0}^k\mathtt{1}\,t_v$), in which only the n_v's and the relative order of the t_v's matter. With this reformulation of the algorithm, a slight perturbation of the t_v's produces a new tree, $\mathfrak{S}_\ell$, that is easier to handle than $\mathfrak{T}_\ell$ due to its property of independence of labels, but that has essentially the same *profile* as $\mathfrak{T}_\ell$ (i.e. it has the same repartition of blades with respect to the height). Let $\epsilon^{(j)}$ denote the sequence of integers defined, for $j \in I$, by

$$\epsilon_i^{(j)} = \delta_{i,j}.$$

For $\mathcal{R}$ a totally ordered set, $\mathbb{N}_0 \times \mathcal{R}$ inherits a lexicographic order, $\prec$, from $\mathcal{R}$: $(n,t) \prec (m,u)$ if $n > m$ or if $n = m$ and $t < u$. Let $B = (l_i, r_i)_{1\le i\le N}$ (resp. $L = (l_i)_{1\le i\le N}$, $R = (r_i)_{1\le i\le N}$) be a finite sequence of elements of $\mathbb{N}_0 \times \mathcal{R}$ (resp. $\mathbb{N}_0$, $\mathcal{R}$), with no repetitions in the sequence R. Assume that $l_j \ge k$ for each j, and that (l_1, r_1) is the smallest element of B, for $\prec$.

Definition 4. The Lyndon tree $\mathfrak{L}^k(B)$ is defined by induction by:

1. If $N = 1$ and $l_1 = k$, $\mathfrak{L}^k(B)$ has no edge and its unique vertex, with label (k, r_1), is a blade.
2. Otherwise, consider the new sequence B' formed from $L - \epsilon^{(1)}$ and R and let i_0 denote the index of the smallest element in B', for $\prec$.
   - (a) If $i_0 = 1$, then $\mathfrak{L}^k(B)$ is the binary tree with a needle (labeled `0`) as its left child and $\mathfrak{L}^k(B')$ as its right child.
   - (b) If $i_0 \ge 2$, then the binary tree $\mathfrak{L}^k(B)$ has $\mathfrak{L}^k((l_i, r_i)_{1\le i\le i_0-1})$ for left subtree, and $\mathfrak{L}^k((l_{i+i_0}, r_{i+i_0})_{0\le i\le N-i_0})$ for right subtree.

Remark 2. Since $\sum(l_i+1)$ is strictly decreasing at each recursive call to instruction (2), and since the l_i's are not allowed to drop under level k, $\mathfrak{L}^k(B)$ is well-defined as long as the r_i's are distinct. The N blades of $\mathfrak{L}^k(B)$ are labeled $(k, r_i)_{1\le i\le N}$, and, during the contour traversal, they appear in this order.

Remark 3. For $\mathcal{R} = \{t_v \,|\, 1 \le v \le N_\ell\}$ and $T_\ell = ((n_v, t_v))_{1\le v\le N_\ell}$ defined in this section,

$$\mathfrak{L}^{a_\ell}(T_\ell) = \mathfrak{T}_\ell,$$

or more precisely, the shapes are the same, but the labels are different. When the label of some node of $\mathfrak{L}^{a_\ell}(T_\ell)$ is (k, t_v), the corresponding label of $\mathfrak{T}_\ell$ is the prefix of $\mathtt{0}^k\mathtt{1}t_v$ that stops with the last `1` before the next occurrence of $\mathtt{0}^k$.

For the analysis of $\mathfrak{T}_\ell$, the fact that the t_v's are suffixes of t_1, precluding any form of independence, is bothering. In order to fix the problem, in T_ℓ, we replace the sequence $(t_v)_{1\le v\le N_\ell}$ with a new sequence $(s_v)_{1\le v\le N_\ell}$ of infinite binary words, close to the t_v's but independent, defined as follows: let $(\zeta_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of uniform infinite words, independent of T_ℓ and let s_v be the concatenation of p_v, the prefix formed by the first a_ℓ letters of t_v, with ζ_v. When t_{N_ℓ} is shorter than a_ℓ letters, p_{N_ℓ} is completed with the appropriate number of `0`'s, before the concatenation with ζ_{N_ℓ}. This way, we obtain a new sequence $S_\ell = ((n_v, s_v))_{1\le v\le N_\ell}$, and we set

$$\mathfrak{S}_\ell \;=\; \mathfrak{L}^{a_\ell}\left(S_\ell\right).$$

Differences between $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$ occurs scarcely, only when at least a_ℓ letters are used to distinguish two suffixes, so that $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$ are close, in a sense stated precisely in Proposition 4. We have:

Proposition 2. *The probability distribution of S_ℓ is given by:*

1. *$n_1 = \ell$, and for $v \ge 2$, $n_v - a_\ell$ is a geometric random variable with parameter $\frac{1}{2}$, conditioned to be smaller than $\ell - a_\ell$;*
2. *s_v is a copy of W_∞;*
3. *For all v, n_v and s_v are independent;*
4. *N_ℓ is geometric with parameter $2^{a_\ell-\ell}$;*
5. *N_ℓ and the sequence $(n_v, s_v)_{v\in\mathbb{N}}$ are independent;*
6. *$(n_v, s_v)_{v\in\mathbb{N}}$ is a sequence of i.i.d. random variables.*

In terms of words, this can be rephrased as follows:

Proposition 3. *The sequence of words* $\tilde{S}_\ell = (\mathtt{0}^{n_v - a_\ell}\mathtt{1}s_v)_{2\leq v\leq N_\ell}$, *followed by the word* $\mathtt{0}^{n_1-a_\ell}\mathtt{1}s_1$, *is distributed as a sequence of copies of* W_∞, *observed until the first occurrence of the prefix* $\mathtt{0}^{\ell-a_\ell}$, *this first occurrence* $\mathtt{0}^{n_1-a_\ell}\mathtt{1}s_1$ *being eventually truncated of any* $\mathtt{0}$ *in excess of* $\mathtt{0}^{\ell-a_\ell}\mathtt{1}\dots$, *so that* $n_1 = \ell$.

Proof of Proposition 2. Consider the word $W'_\infty = \mathtt{0}^\ell\mathtt{1}W_\infty$ and let x_j, $j \geq 1$, be the jth letter in W'_∞, let $\tilde{n}_k$ denote the length of the kth maximal run of $\mathtt{0}$'s longer than $a_\ell - 1$ in W'_∞, let τ_k be the position of the letter $\mathtt{1}$ ending this kth run of $\mathtt{0}$'s, so that $x_{\tau_k-1}x_{\tau_k}$ ends the kth occurrence of the pattern $\mathtt{0}^{a_\ell}\mathtt{1}$ in W'_∞. Let $\tilde{N}_\ell$ be the number of runs of $\mathtt{0}$'s longer than $a_\ell - 1$ before the second run longer than $\ell - 1$ occurs, and let $\tilde{p}_k = x_{\tau_k+1}x_{\tau_k+2}\dots x_{\tau_k+a_\ell}$. Then $(x_{\ell+1+j})_{j\geq 1}$ is a Bernoulli process, the τ_j's are stopping times for the related filtration, and

$$\tilde{N}_\ell = N_\ell, \quad (\tilde{n}_v, \tilde{p}_v)_{1\leq v\leq N_\ell} = (n_v, p_v)_{1\leq v\leq N_\ell},$$

by definition. But since $\tau_k + a_\ell + 1 \leq \tau_{k+1}$, $(\tilde{n}_v, \tilde{p}_v)_{v\geq 1}$ is an i.i.d. sequence, with $\tilde{p}_v$ uniform on $\{\mathtt{0},\mathtt{1}\}^{a_\ell}$ and independent of $\tilde{n}_v$, and $\tilde{n}_v - a_\ell$ geometric. This entails the six points of Proposition 2. □

Due to Proposition 3, conditionally given N_ℓ, the ranks of the terms of the sequence $\tilde{S}_\ell = (\mathtt{0}^{n_v-a_\ell}\mathtt{1}s_v)_{2\leq v\leq N_\ell}$ with respect to the lexicographic order form a uniform permutation on $N_\ell - 1$ symbols. This permutation is quite close to the random non-uniform permutation induced by the family $(\mathfrak{s}(v))_{2\leq v\leq N_\ell}$ of suffixes of L^ℓ. As a consequence, the subtree $\mathfrak{U}_\ell$ of $\mathfrak{S}_\ell$ induced by the root and the blades, once the needles and the first blade erased, forms the *binary search tree* of a uniform permutation, a well studied random tree: for instance, a coupling between the Yule process and the binary search tree leads to a precise analysis of the depths of the leaves of the binary search tree, see [CKMR05]. The depths of blades in $\mathfrak{S}_\ell$, though they depend on their depths in $\mathfrak{U}_\ell$, are also affected by the positions of the needles, and we need to tweak the arguments of [CKMR05] in order to include the needles in their analysis.

In Section 2.5 we prove that the coupling between $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$ is tight enough that the depths of leaves of $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$ share the same asymptotic behavior, at some level of detail, see Proposition 4 below. The proof of Proposition 4 relies on a coupling between $\mathfrak{S}_\ell$ and a Yule process, described in Section 2.5.3. This coupling is also the key to the analysis of the depths of blades in $\mathfrak{S}_\ell$, which are expressed as functionals of the Yule process, see Section 2.6.

Set $\mathfrak{p}_{N_\ell} = t_{N_\ell}$, and if $v < N_\ell$, set $t_v = \mathfrak{p}_v\mathtt{0}^{n_{v+1}}\mathtt{1}t_{v+1}$, so that $(\mathtt{0}^{n_v}\mathtt{1}\mathfrak{p}_v)_{1\leq v\leq N_\ell}$ is a factorization of L^ℓ and $(\mathtt{0}^{a_\ell}\mathtt{1}\mathfrak{p}_v)_{1\leq v\leq N_\ell}$ is a sequence of Lyndon words. One obtains the Lyndon tree $\mathfrak{L}(L^\ell)$ when one grafts each shrub $\mathfrak{t}(v) = \mathfrak{L}(\mathtt{0}^{a_\ell}\mathtt{1}\mathfrak{p}_v)$ on $\mathfrak{T}_\ell$, $\mathfrak{t}(v)$ replacing the corresponding blade v of $\mathfrak{T}_\ell$.

Remark 4. Due to Proposition 4, the tree $\mathfrak{A}_\ell$ obtained by grafting the shrubs $\mathfrak{t}(v)$'s on the corresponding blades of $\mathfrak{S}_\ell$ (rather than $\mathfrak{T}_\ell$) is very close in height to $\mathfrak{L}(L^\ell)$. But $\mathfrak{t}(v)$ depends on $\mathfrak{S}_\ell$ only through the prefix p_v of $\mathfrak{p}_v$ and p_v is short compared to $\mathfrak{p}_v$ when a_ℓ grows (for $|p_v| = a_\ell$ while $\mathbb{E}[|\mathfrak{p}_v|] \simeq 2^{a_\ell}$). This has a crucial consequence for the study of the heights H_v's of the shrubs $\mathfrak{t}(v)$'s at the bottom of the tree: in Section 3.2 we shall see that $(H_v)_{1\leq v\leq N_\ell}$ behave like a sample of independent geometric random variables with parameter 1/2, essentially *independent* from $\mathfrak{S}_\ell$.

2.5. **The distance between $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$.** For a blade $v \in [\![1, N_\ell]\!]$, let h_v (resp. $\tilde{h}_v$) be its height in $\mathfrak{S}_\ell$ (resp. in $\mathfrak{T}_\ell$). Set

$$d_v = \left| h_v - \tilde{h}_v \right|, \qquad D_\ell = \max_{1 \le v \le N_\ell} d_v.$$

Then

Proposition 4.

$$\lim_\ell \mathbb{P}\left(D_\ell \ \ge \ \frac{\ell}{\sqrt{a_\ell}} \right) \ = \ 0.$$

The proof of Proposition 4 uses branching random walks arguments, as in [Big77], and a coupling of $\mathfrak{S}_\ell$ with a Yule process. For $1 \le i < j \le N_\ell$, the case when $t_i \prec t_j$ while $s_i \succ s_j$ is called an *inversion* between i and j. In order to bound D_ℓ, we need to track inversions. Let $\varpi_\ell = 2^{-a_\ell}$ and let U_i be the real number with dyadic expansion s_i (according to point (2) of Proposition 2, U_i is uniformly distributed on $[0, 1]$). We have:

Lemma 2. *If there is an inversion between i and j then s_i and s_j coincide on the first a_ℓ letters, and U_i and U_j are in the same dyadic interval with width ϖ_ℓ.*

Proof. Since t_i and t_j are at least a_ℓ-letters long by definition and since t_i (resp. t_j) coincides with s_i (resp. s_j) on the first a_ℓ letters, then the outcome of the comparisons $t_i \prec\succ t_j$ and $s_i \prec\succ s_j$ can be different only if it is decided after the first a_ℓ letters, which entails, by definition of the lexicographic order, that the 4 words have the same a_ℓ-letters long prefix. □

2.5.1. *A Galton-Watson process.* Lemma 2 allows to compute an upper bound on D_ℓ that depends only on $\mathfrak{S}_\ell$, and functionals of $\mathfrak{S}_\ell$ are more tractable than functionals of $\mathfrak{T}_\ell$ for several reasons. One of them is the description of $\mathfrak{S}_\ell$ in terms of a Galton-Watson process with geometric offspring distribution: due to points (1), (4) and (5) of Proposition 2, $(n_{v+1} - a_\ell + 1)_{1 \le v \le N_\ell - 1}$ is distributed as a sample of i.i.d. geometric random variables, observed until the last time when all the terms of the sequence are smaller than $\ell - a_\ell$. Then the first terms in T_ℓ or in S_ℓ, $(n_1, t_1) = (\ell, t_1)$ and (n_1, s_1), are seen as the ancestors, and the indices v such that $n_v \ge \ell - k$ form generation k. More precisely, if $n_v \ge \ell - k$, and if the next index that belongs to generation k is w, the offspring of $(\ell - k, t_v)$ at generation $k + 1$ is formed by $(\ell - k - 1, t_v)$ and by the nodes (n_j, t_j) or (n_j, s_j) such that $v < j < w$ and $n_j = \ell - k - 1$. Let us call the set

$$\begin{aligned}\mathcal{F}_{k,v} &= \{(\ell - k, t_v), (\ell - k - 1, t_v)\} \cup \{(n_j, t_j) \,|\, v < j < w \text{ and } n_j = \ell - k - 1\} \\ &= \{(\ell - k, s_v), (\ell - k - 1, s_v)\} \cup \{(n_j, s_j) \,|\, v < j < w \text{ and } n_j = \ell - k - 1\}\end{aligned}$$

the *family* of $(\ell - k, t_v)$ or of $(\ell - k, s_v)$. Due to the memoryless property of the geometric distribution, the probability p_n that $(\ell - k, s_v)$ has $O_{k,v} = n$ children satisfies

$$\mathbb{P}\left(O_{k,v} = n\right) = 2^{-n}\, \mathbb{1}_{n \ge 1},$$

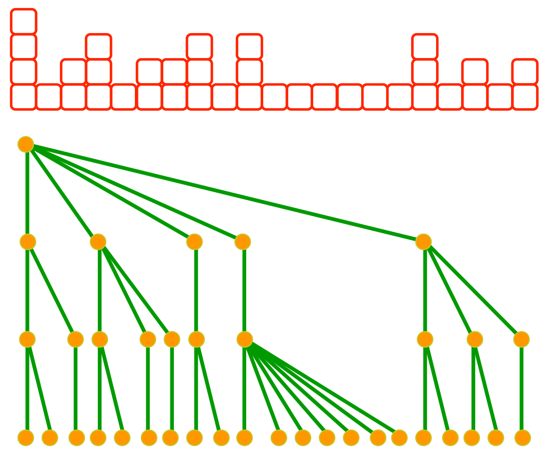


FIGURE 3. Here $\ell = 4$ and $a_\ell = 1$. Top figure: a sample $s = (n_v - a_\ell + 1)_{2\leq v\leq N_\ell} = 12312231311111131212$ of the geometric distribution, stopped before the first value larger than 3, value that is reduced to $n_1 = 4$, and is used as a prefix of s. Below, the corresponding Galton-Watson tree $\mathfrak{GW}_\ell$.

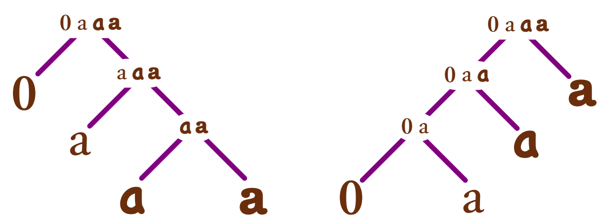


FIGURE 4. $\mathcal{S}_{k,w}$ and $\mathcal{T}_{k,w}$ in the worst case scenario, when the 3 scions in $\mathcal{O}_{k,w}$ are in the same dyadic interval, and $(t_j)_j$ is increasing while $(s_j)_j$ is decreasing, resulting in $(d(\omega)|\omega \in \mathcal{O}_{k,w}) = (1, 1, 2)$.

see for instance [Dev92, page 601] for an explanation. This process stops at generation $\ell - a_\ell$.

We call $\mathfrak{GW}_\ell$ the family tree of the Galton-Watson process described in this paragraph. The tree $\mathfrak{GW}_\ell$ depends only on the sequence $(n_j)_j$, and can be seen as the tree induced by some nodes of $\mathfrak{S}_\ell$, or of $\mathfrak{T}_\ell$, indifferently, including the blades. The differences between $\mathfrak{S}_\ell$ and $\mathfrak{T}_\ell$ can be analyzed at the level of the binary subtrees $\mathcal{S}_{k,w}$ (resp. $\mathcal{T}_{k,w}$) induced by the family $\mathcal{F}_{k,w}$ in $\mathfrak{S}_\ell$ (resp. in $\mathfrak{T}_\ell$), for the only comparisons that involve the t_j's or the s_j's are inside the families, the other comparisons being settled by inspections of the first terms n_j's of the labels, that are the same in S_ℓ and in T_ℓ.

The subtrees $\mathcal{S}_{k,w}$ and $\mathcal{T}_{k,w}$ have one needle, the other leaves beginning with one of the elements of the offspring $\mathcal{O}_{k,w} = \mathcal{F}_{k,w} \setminus \{(\ell - k, t_w)\}$. For such a leaf $\omega = (\ell - k - 1, t_j) \in \mathcal{O}_{k,w}$, let $d(\omega)$ denote the modulus of the difference between the height of ω in $\mathcal{T}_{k,w}$ and the height of $(\ell - k - 1, s_j)$ in $\mathcal{S}_{k,w}$. Then a bound for d_v is given by the sum of the $d(\omega)$ on the path between the blade v and the

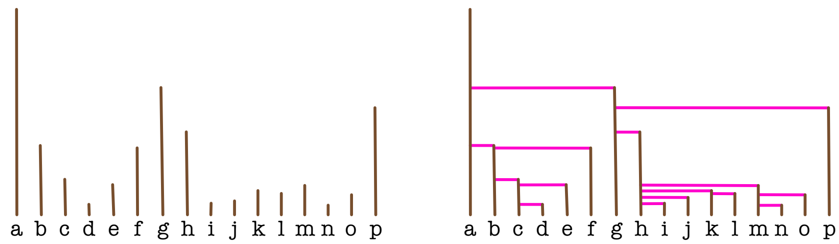


FIGURE 5. A sample of exponential random variables, until hitting t, and the related tree $\mathfrak{E}_t$ (in which the pink horizontal lines are to be seen as vertices rather than as edges).

root. Thus the Galton-Watson tree $\mathfrak{GW}_\ell$ underlies a branching random walk with positive steps $d(\omega)$'s, and the rightmost position of this branching random walk gives an upper bound for D_ℓ. Note that the independence of the steps $d(\omega)$ is questionable at this stage of the proof, see Section 2.5.3.

2.5.2. *A Yule process.* According to Lemma 2, there are differences between $\mathcal{T}_{k,w}$ and $\mathcal{S}_{k,w}$ only if some of the U_j's involved in the comparisons at step (2) of Algorithm 4 are in the same dyadic interval with width $\varpi_\ell = 2^{-a_\ell}$. An additional condition is that the results of these comparisons change the leader, i.e. the smallest element in B', at step (2) of Algorithm 4. But the leader does not change if it belongs to a different dyadic interval. It turns out that the family $\mathcal{F}_{k,v}$ can be seen as a Galton-Watson process itself, elements of the different dyadic intervals being the generations. Such nested Galton processes with geometric progeny typically appear in Yule processes. In this section, we define a Yule process in terms of the sequence T_ℓ, and in Section 2.5.3 we shall see how a bound for D_ℓ can be derived from this Yule process. In Section 2.6, we show how to represent the height of a blade of $\mathfrak{S}$ in term of this Yule process, a representation that is essential for the computation of the height of the Lyndon tree, through the large deviation results of Section 3.

A *Yule process* $\mathcal{Y}$ (cf. [AN72, page 109] or [Ald01]) models a population in which each individual lives forever, and gives birth to a new individual according to a Poisson process with rate ρ. We assume that the population starts at time 0 with a single individual, called the *ancestor*. One can keep track of the history of the population through the *Yule tree* $\mathfrak{Y}$ [CKMR05], a family tree of the population, in which a vertical life line is drawn downward, for each individual, starting at an ordinate given by *minus* the date of birth, on the left of the life line of its father, and is connected to the life line of its father by a dotted horizontal line. Let $\mathfrak{Y}_t$ denote the family tree $\mathfrak{Y}$ truncated at time t, i.e. at ordinate $-t$.

This representation comes handy for the description of a correspondence between the Yule tree and a sample $Y = (Y_n)_{n\geq 1}$ of i.i.d. exponential random variables with rate ρ. Consider the sequence $Z^{(t)}$ defined by:

$$T_t = \inf\{n \geq 1 \,|\, Y_n \geq t\}, \quad Z^{(t)} = (Z_k^{(t)})_{0\leq k\leq T_t-1} = (t, Y_1, Y_2, \ldots, Y_{T_t-1}).$$

Then picture each term $Z_k^{(t)}$ of the sequence $Z^{(t)}$ by a vertical line of the corresponding length $Z_k^{(t)}$, drawn at abscissa $T_t - k$, and connect, through an horizontal

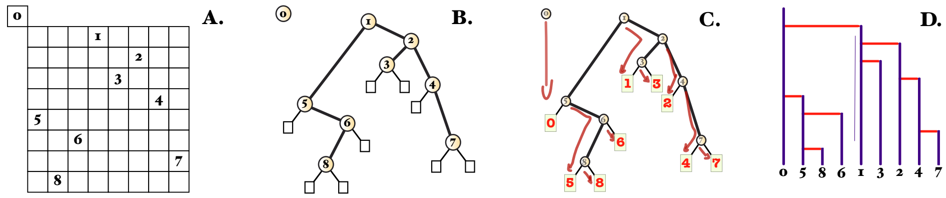


FIGURE 6. A. The permutation $\sigma = (5, 8, 6, 1, 3, 2, 4, 7)$. B. The heap-ordered tree $\mathfrak{H}_\sigma$. C. The Lyndon tree $\mathfrak{L}(0\sigma)$ built from $\mathfrak{H}_\sigma$. D. The Lyndon tree $\mathfrak{L}(0\sigma)$ built from $f(0\sigma)$, here $f(x) = 9 - x$.

line, the top of each line k, but the leftmost, to the next line on its left whose height tops the height $Z_k^{(t)}$, to obtain a family tree $\mathfrak{E}_t$.

Proposition 5. *$\mathfrak{E}_t$ and $\mathfrak{Y}_t$ have the same probability distribution.*

We do not know a reference for Proposition 5, that is part of the folklore on the topic. For the binary search tree, however, there exists a well studied analog: the shape of the *heap-ordered tree*[1] $\mathfrak{H}_\sigma$ of a random permutation σ has the same distribution as the shape of the related binary search tree. If we see σ as a word on the alphabet $[\![\mathtt{n}]\!]$, $\mathfrak{H}_\sigma$ is the Lyndon tree of the word 0σ and can also be obtained through the construction by vertical lines that leads to $\mathfrak{E}_t$, if one starts with the sample $(f(0), f(\sigma(1)), f(\sigma(2)), \ldots, f(\sigma(n)))$ (in which f is any positive strictly decreasing function) rather than with $Z^{(t)}$, cf. Fig. 6.

2.5.3. *Representation of $\mathfrak{S}_\ell$ in terms of a Yule process.* Let us denote by U_v the real number whose dyadic development is $0^{n_v - a_\ell} 1 s_v$. Due to Proposition 3,

$$U = (U_2, U_3, \ldots, U_{N_\ell}, U_1)$$

is distributed as a sequence of uniform random variables observed until a term belongs to $[0, 2^{-\ell + a_\ell}]$, this last term being eventually multiplied by a power of 2, so as to belong to $]2^{-\ell + a_\ell - 1}, 2^{-\ell + a_\ell}]$. As a consequence, the sequence X image of U by the mapping

$$X_i \;=\; -\log_2 U_i$$

is distributed as a sequence of exponential random variables observed until a term belongs to $[\ell - a_\ell, +\infty[$, this last term being eventually shifted by an integer, so as to belong to $[\ell - a_\ell, \ell - a_\ell + 1[$. Then the construction of Section 2.5.2, based on the sequence $(\ell - a_\ell, X_2, X_3, \ldots, X_{N_\ell})$, gives a Yule family tree $\mathfrak{Y}_{\ell - a_\ell}$ with lifetime $\ell - a_\ell$, and with intensity $\rho = \ln 2$. Then the tree induced by the points of $\mathfrak{Y}_{\ell - a_\ell}$ whose distance to the root is an integer between 0 and $\ell - a_\ell$ is $\mathfrak{GW}_\ell$. Also, removing the nodes of $\mathfrak{GW}_\ell$ splits $\mathfrak{Y}_{\ell - a_\ell}$ into connected components $\mathfrak{Y}^{(x)}$, one for each interior node $x = (\ell - k, s_v)$ of $\mathfrak{GW}_\ell$, connected components that are independent and distributed as $\mathfrak{Y}_1$ (still with intensity $\ln 2$). The set of leaves of $\mathfrak{Y}^{(x)}$ is $\mathcal{O}_{k,v}$.

[1]See for instance [SF13, p. 362].

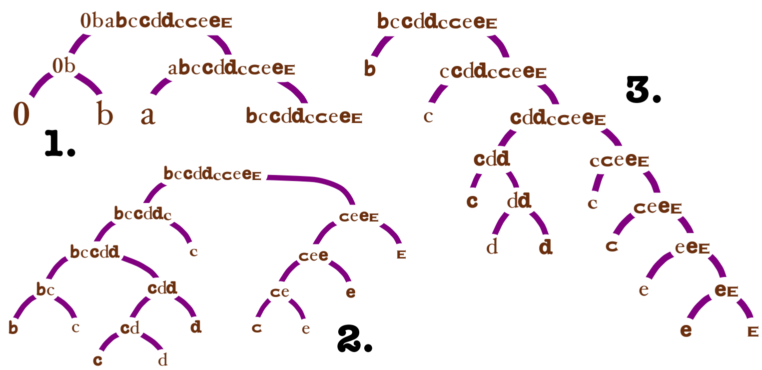


FIGURE 7. 1. $\mathcal{S}^{(2)}$ and $\mathcal{T}^{(2)}$ are equal in this example in which the blades of the family belong to 5 dyadic intervals `a`, `b`, `c`, `d`, `e` sorted in lexicographic order: the only `a` being interleaved between the 2 `b`'s, comparisons between the 2 `b`'s do not matter. Then we show two possible shapes (2. and 3.) for the subtree below `bccddcceee`, depending on the outcomes of comparisons $t_i \gtrless t_j$ and $s_i \gtrless s_j$. The fluctuation $d(\mathtt{E})$ is $7-2 = (k_c-1)_+ + (k_e-1)_+$, without $(k_d-1)_+$, for no `d` is involved in this part of the tree.

2.5.4. *Proof of Proposition 4.* For some element (n_j, s_j), $j > v$, of $\mathcal{O}_{k,v}$, $n_j = \ell - k - 1$. Also, s_j is the dyadic expansion of

$$\tilde{U}_j = 2^{\ell - a_\ell - k} U_j - 1 = 2^{\ell - a_\ell - k - X_j} - 1,$$

and $\tilde{U}_j$ belongs to the dyadic interval $[m2^{-a_\ell}, (m+1)2^{-a_\ell})$ if and only if

$$\tilde{X}_j = \{X_j\} = X_j - \ell + k + a_\ell + 1 \in I_m,$$

in which the intervals

$$I_m \;=\; \left[-\log_2\left(\tfrac{1}{2} + (m+1)2^{-a_\ell - 1}\right) \;;\; -\log_2\left(\tfrac{1}{2} + m2^{-a_\ell-1}\right)\right)$$

satisfy

$$\biguplus_{0 \le m < 2^{a_\ell}} I_m \;=\; [0,1).$$

In turn, the points of $\mathfrak{Y}^{(k,w)}$ with ordinates

$$y_m \;=\; -\log_2\left(\tfrac{1}{2} + m2^{-a_\ell-1}\right), \quad 0 \le m < 2^{a_\ell},$$

induce Galton-Watson subtrees $\mathfrak{GW}_{(k,w)}$ whose offspring distribution changes at each generation, and, at generation m, is a geometric distribution with parameter

$$p_m \;=\; \frac{\frac{1}{2} + m2^{-a_\ell-1}}{\frac{1}{2} + (m+1)2^{-a_\ell-1}} \;=\; 1 - \frac{1}{2^{a_\ell} + m + 1}.$$

Incidentally, $\mathfrak{Y}^{(k,w)}$ is distributed as $\mathfrak{Y}_1$ and it is built from the sample of exponential random variables $\tilde{X}_j$, for $(\ell - k - 1, s_j) \in \mathcal{O}_{k,w}$. Keeping in mind Lemma 2, note that differences between $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$ occur only at the level of the families of $\mathfrak{GW}_{(k,w)}$, because, if (n_i, s_i) and (n_j, s_j) are not in the same family but $\tilde{U}_i$ and $\tilde{U}_j$ are in the same dyadic interval, then

- either $n_j \neq n_i$, and the comparisons between (n_j, s_j) and (n_i, s_i), or between (n_j, t_j) and (n_i, t_i), have the same output, for only the n_j' s are involved,
- either $n_j = n_i$, but in the sequence S_ℓ, between (n_j, s_j) and (n_i, s_i), there exists some element (n_k, s_k) such that $n_k > n_i$, and the standard factorization will cut the label just before (n_k, s_k) in both trees, or before a similar element, independently of an eventual inversion between i and j,
- or (n_i, s_i) and (n_j, s_j) are in the same family of $\mathfrak{GW}_\ell$, i.e. $n_j = n_i = \ell-k-1$, and (n_i, s_i) and (n_j, s_j) are in the same subtree $\mathfrak{GW}^{(k,w)}$ but in the sequence S_ℓ, between (n_j, s_j) and (n_i, s_i), there exists some element (n_k, s_k) such that $\tilde{X}_k$ is in some interval I_{m_1} while $\tilde{X}_i$ and $\tilde{X}_j$ are in some I_{m_2} with $m_2 > m_1$. Then (n_i, s_i) and (n_j, s_j) land in different subtrees before an eventual inversion between them can produce a difference between $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$.

Inside such a family of $\mathfrak{GW}^{(k,w)}$, inversion may occur between any couple of leaves, and may produce differences between the corresponding subtrees $\mathfrak{T}^{(k,w)}$ and $\mathfrak{S}^{(k,w)}$ of $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$, so that the operation $\mathfrak{L}$ can produce any binary tree, as far as we know, only the number f of leaves of these 2 binary subtrees of $\mathfrak{T}^{(k,w)}$ and $\mathfrak{S}^{(k,w)}$ being given: thus the maximal depth of a leave is $f-1$ and the minimal depth is 1, if $f > 1$. In any case the difference between the depth of a leave in $\mathfrak{GW}^{(k,w)}$ and in the corresponding subtree of $\mathfrak{T}_\ell$ is bounded by $(f-2)_+$. Note that in a family at level m, the probability of difference between $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$ is thus bounded by

$$\mathcal{O}\left((1-p_m)^2\right) = \mathcal{O}\left(2^{-2a_\ell}\right),$$

which is pretty small. There is an exception, in which the bound is $(f-1)_+$ rather than $(f-2)_+$, when the only needle of $\mathfrak{S}^{(k,w)}$ happens, at some level m_0 that depends on the label of the ancestor of $\mathfrak{Y}^{(k,w)}$. This precludes the Markov property of branching random walks but stochastic monotonicity alleviates the problem.

Let us give a formal argument: consider

$$m_\ell(\theta) \;=\; \mathbb{E}\left[\sum_{1\leq v\leq N_\ell} e^{\theta d_v}\right],$$

and assign to each blade v the position $-d_v$. Let $Z_t^{(n)}$ denote the number of blades of generation n with a position to the left of t. Then, as in [Big77, page 634],

$$\mathbb{P}\left(D_\ell \geq t\right) = \mathbb{P}\left(Z_{-t}^{(\ell)} \geq 1\right) \;\leq \mathbb{E}[Z_{-t}^{(\ell)}] \;\leq e^{-\theta t} m_\ell(\theta). \tag{2}$$

Set

$$\begin{aligned}
p = 1-2^{-a_\ell} \;&=\; 1-\varepsilon \\
&\leq \min\left\{p_m \,|\, 0\leq m\leq 2^{a_\ell}-1\right\} \\
F(p,\theta) = \sum_{k\geq 1} k e^{\theta(k-2)_+} p(1-p)^{k-1} \;&= p(1-e^{-\theta}) + \frac{pe^{-\theta}}{\left(1-e^{\theta}\varepsilon\right)^2} \\
G(p,\theta) = \sum_{k\geq 1} k e^{\theta(k-1)_+} p(1-p)^{k-1} \;&= \frac{p}{\left(1-e^{\theta}\varepsilon\right)^2}
\end{aligned}$$

The key to the proof of Proposition 4, in the spirit of [Big77, Corollary (3.4)], is the next Lemma:

Lemma 3. *For $\theta \geq 0$,*

$$m_\ell(\theta) \ \leq \ \left(\tfrac{G(1-\varepsilon,\theta)}{F(1-\varepsilon,\theta)}\ F(1-\varepsilon,\theta)^{1/\varepsilon}\right)^{\ell-a_\ell}.$$

Given Lemma 3, setting $t = \frac{\ell}{\sqrt{a_\ell}}$ in (2), and

$$e^{2\theta_\ell} = 2^{a_\ell} = \tfrac{1}{\varepsilon},$$

and using

$$\begin{aligned} F(p,\theta_\ell) &= 1+\varepsilon+3\varepsilon\sqrt{\varepsilon}+2\varepsilon^2+2\varepsilon^2\sqrt{\varepsilon}+\dots \\ G(p,\theta_\ell) &= 1+2\sqrt{\varepsilon}+2\varepsilon+2\varepsilon\sqrt{\varepsilon}+\dots \end{aligned}$$

one obtains

$$\begin{aligned} \mathbb{P}\left(D_\ell \geq \tfrac{\ell}{\sqrt{a_\ell}}\right) &\leq e^{-\theta_\ell t} m_\ell(\theta_\ell) \\ &\leq \exp\left(-\ell\sqrt{a_\ell}\ln\sqrt{2} \ + \ \mathcal{O}(\ell)\right), \end{aligned}$$

which ends the proof of Proposition 4.

Proof of Lemma 3. Let us consider the words of $\{0,1\}^{a_\ell}$ as an increasing sequence $\left\{w_0 \prec w_1 \prec \dots \prec w_{-1+\frac{1}{\varepsilon}}\right\}$. Given $k \in [\![0, \ell-a_\ell]\!]$ and $m \in [\![0, \frac{1}{\varepsilon}[\![$, the factorizations of the sequences S_ℓ and T_ℓ in subsequences that start with some term lexicographically smaller than $(k+a_\ell, w_m)$ and end just before the next one are the same, due to Proposition 2. These factorizations transfer to the sequence $(X_j)_{1\leq j\leq N_\ell}$: here the factors are subsequences that begin with some X_j larger than $k+y_m$ and end just before the next one. The next level of factorization (fragmentation) is described by the subtrees of $\mathfrak{Y}_\ell$ with leaves at level $k+y_{m+1}$ (i.e. at time $\ell-k-y_{m+1}$) and ancestors at level $k+y_m$. Each of these subtrees $\mathfrak{h}$ is in bijection with binary subtrees of $\mathfrak{S}_\ell$ and of $\mathfrak{T}_\ell$ with the same number $\mathcal{O}_\mathfrak{h}$ of leaves (needles excluded), but eventually different shapes, leading to different depths of their leaves, relatively to the 2 subtrees. We argued that $\mathcal{O}_\mathfrak{h}$ has a geometric distribution with parameter $p_m \geq p$, independently of the other subtrees, and that for each leaf, the difference in depth is bounded by $(\mathcal{O}_\mathfrak{h}-2)_+$, in the absence of a needle, and by $(\mathcal{O}_\mathfrak{h}-1)_+$, in presence of a needle. The uniform bound $(\mathcal{O}_\mathfrak{h}-1)_+$ would be easier to handle, fitting perfectly in the context of branching random walks, but it is too crude for our purposes.

In each subtree $\mathfrak{Y}^{(k,w)}$, there are 2^{a_ℓ} levels of subtrees of type $\mathfrak{h}$, and there exists exactly one needle, in some subtree located on the right at a level $m(w)$ such that $\{X_w\} \in I_{m(w)}$: this level $m(w)$ depends on the levels of the tree $\mathfrak{Y}_\ell$ that are closer to the root of $\mathfrak{Y}_\ell$. Let us bound the differences in depth of each leaf of a subtree $\mathfrak{h}$ at level $m(w)$ of $\mathfrak{Y}^{(k,w)}$ by $d_\mathfrak{h} = (\mathcal{O}_\mathfrak{h}-1)_+$, and let us bound by $d_\mathfrak{h} = (\mathcal{O}_\mathfrak{h}-2)_+$ the differences in depth of each leaf of the subtrees $\mathfrak{h}$ of $\mathfrak{Y}^{(k,w)}$ at other levels than $m(w)$. Let us call $\mathfrak{Y}_k$ the part of $\mathfrak{Y}_\ell$ that is at a distance of root smaller or equal than k. Then, given $\mathfrak{Y}_k$, the conditional distribution of $\mathfrak{Y}^{(k,w)}$ endowed with elementary displacements $d_\mathfrak{h}$ from the roots to the leaves of each subtree $\mathfrak{h}$, is that of a non homogeneous branching random walk. Thus, if $\delta_j^{(k,w)}$ denotes the difference in depth for the leaf j between $\mathfrak{S}^{(k,w)}$ and $\mathfrak{T}^{(k,w)}$, then

$$\mathbb{E}\left[\sum_j e^{\theta\delta_j^{(k,w)}}\;\middle|\;\mathfrak{Y}_k\right] \leq \tfrac{G(p_{m(w)},\theta)}{F(p_{m(w)},\theta)} \prod_{m=0}^{2^{a_\ell}-1} F(p_m,\theta).$$

Note that F and G are decreasing in p, for the geometric distribution is stochastically decreasing in its parameter. Thus, by stochastic monotonicity,

$$\mathbb{E}\left[\sum_j e^{\theta\delta_j^{(k,w)}}\;\middle|\;\mathfrak{Y}_k\right] \leq \tfrac{G(1-\varepsilon,\theta)}{F(1-\varepsilon,\theta)}\ F(1-\varepsilon,\theta)^{1/\varepsilon}.$$

Now, let us denote by $d_i^{(k)}$ the difference in depth of a leaf $i=(k,w)$ between $\mathfrak{L}^{\ell-k}(T_\ell)$ and $\mathfrak{L}^{\ell-k}(S_\ell)$, and assume that j is a generic leaf of $\mathfrak{S}^{(k,w)}$ or $\mathfrak{T}^{(k,w)}$, indifferently. Then

$$d_j^{(k+1)} \leq d_i^{(k)} + \delta_j^i,$$

and

$$\begin{aligned}\mathbb{E}\left[\sum_j e^{\theta d_j^{(k+1)}}\right] &\leq \mathbb{E}\left[e^{\theta d_i^{(k)}}\sum_j e^{\theta\delta_j^i}\right]\\ &= \mathbb{E}\left[e^{\theta d_i^{(k)}}\,\mathbb{E}\left[\sum_j e^{\theta\delta_j^i}\;\middle|\;\mathfrak{Y}_k\right]\right]\\ &\leq \mathbb{E}\left[e^{\theta d_i^{(k)}}\right]\ \tfrac{G(1-\varepsilon,\theta)}{F(1-\varepsilon,\theta)}\ F(1-\varepsilon,\theta)^{1/\varepsilon}.\end{aligned}$$

Lemma 3 follows by induction. □

2.6. **Depths of leaves of the Lyndon tree in terms of the Yule process.** As observed for instance in [CKMR05, Lemma 2.1], $\mathfrak{Y}_{\ell-a_\ell}$ seen as a planar tree, with no edge length, or with edge length 1, is a random binary search tree. That is to say, $\mathfrak{U}_\ell$ is a random binary search tree, for the planar tree structures of $\mathfrak{Y}_{\ell-a_\ell}$ and $\mathfrak{U}_\ell$ depend only on the relative order of the words $\mathtt{0}^{n_v-a_\ell}\mathtt{1}s_v$ for $\mathfrak{U}_\ell$, and of the real numbers X_v for $\mathfrak{Y}_{\ell-a_\ell}$, through the same algorithm, and the mapping $\mathtt{0}^{n_v-a_\ell}\mathtt{1}s_v \longrightarrow X_v$ is monotone. Thus the depth of the blade v in the binary search tree $\mathfrak{U}_\ell$ induced by $\mathfrak{S}_\ell$ is the depth of the leaf v corresponding to X_v in $\mathfrak{Y}_{\ell-a_\ell}$.

Now, the difference between the depth of the blade v in $\mathfrak{U}_\ell$ and its depth in $\mathfrak{S}_\ell$ is the number of needles on the path to the root, whose expression is given by Proposition 6 below. For a given blade v of $\mathfrak{S}_\ell$, consider the marked point process Π_v formed by the vertices of $\mathfrak{Y}_{\ell-a_\ell}$ on the path from v to the root, the leaf v and the root excluded. This path is naturally identified with $[0,\ell-a_\ell]$, the leaf being at 0 and the root at $\ell-a_\ell$, for instance, and the mark being 0 or 1, respectively, according to the side of the branching, say left or right, leading to a decomposition

$$\Pi_v \;=\; \Pi_v^{(0)}\cup\Pi_v^{(1)}.$$

By convention, the mark for both points 0 and $\ell-a_\ell$ is one, and unless mentioned

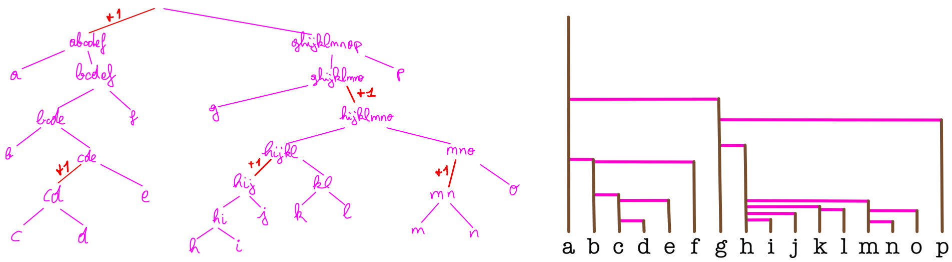


FIGURE 8. Trees $\mathfrak{S}_\ell$ and $\mathfrak{Y}_{\ell-a_\ell}$. In $\mathfrak{S}_\ell$, the sign $+1$ marks the presence of a needle, while in $\mathfrak{Y}_{\ell-a_\ell}$ the lengths of the sticks are rather $2^{Y_i} = 1/U_i$: a needle occurs above the branchings `ab`, `cd`, `hm`, `mn` because, in $\mathfrak{Y}_{\ell-a_\ell}$, the corresponding first stick is at least twice longer than the second, and above `hj` because the factor is larger than 4.

otherwise, they are not included in the point processes. For a point process $\pi = \{\xi_1 < \xi_2 < \cdots < \xi_{k-1} < \xi_k\}$ in some interval $[0, m]$, $G(\pi)$ denotes

$$G(\pi) \;=\; \sum_{r=1}^{k+1} \lfloor \xi_r - \xi_{r-1} \rfloor, \tag{3}$$

in which $\xi_{k+1} = m$ and $\xi_0 = 0$. We have

Proposition 6. *For* $1 \le v \le N_\ell$,

$$0 \le h_v - (\#\Pi_v + 1 + G(\Pi_v^{(1)})) \le 1.$$

Proof. Here $\#\Pi_v + 1$ accounts for the depth of v in $\mathfrak{U}_\ell$, $\lfloor \xi_1 \rfloor$ is $n_v - a_\ell$, and one can check that $\lfloor \xi_r - \xi_{r-1} \rfloor = a$ if the corresponding labels w_r and w_{r-1} satisfy $(n_{w_r} - a, s_{w_r}) \prec (n_{w_{r-1}}, s_{w_{r-1}})$ but $(n_{w_r} - a - 1, s_{w_r}) \succ (n_{w_{r-1}}, s_{w_{r-1}})$, in which case the corresponding edge in $\mathfrak{U}_\ell$ is obtained by erasing a needles of $\mathfrak{S}_\ell$. In order for the relation in Proposition 6 to be exact, the point ξ_{k+1} at the top should be X_1, while it is defined by $\xi_{k+1} = m = \ell - a_\ell$ here. The inequality $0 \le X_1 - \ell + a_\ell < 1$ entails that $0 \le h_v - (\#\Pi_v + 1 + G(\Pi_v^{(1)})) \le 1$. $\square$

The special rôle of $\Pi_v^{(1)}$ in Proposition 6 reflects the asymmetry of the Lyndon tree. Since $G(\Pi_v^{(1)})$ tends to be large when $\#\Pi_v^{(1)}$ is small, and small when $\#\Pi_v^{(1)}$ is large, the difference between the height and the saturation level should be smaller for the Lyndon tree than for the binary search tree.

From now on, we shall assume that the depth h_v of a leaf v of $\mathfrak{S}_\ell$ is given by

$$h_v = \#\Pi_v + 1 + G(\Pi_v^{(1)}),$$

since this does not affect the limit in Theorem 2.

3. PROOF OF THEOREM 2

In Section 3.1 we prove the key formula (4) about the Yule tree $\mathfrak{Y}_\ell$. In fine, through the couplings between $\mathfrak{T}_\ell$ and $\mathfrak{S}_\ell$ on one hand, between $\mathfrak{Y}_\ell$ and $\mathfrak{S}_\ell$ on the other hand (cf. Propositions 4 and 6), Proposition 7 allows to describe the profile of $\mathfrak{T}_\ell$, i.e. the repartition of the leaves of $\mathfrak{T}_\ell$ according to their depth. Section 3.2

presents a sketch of the proof of Theorem 2. The rest of Section 3 is devoted to a detailed proof of Theorem 2.

3.1. **A many-to-one formula.** A leaf v of $\mathfrak{Y}_\ell$ is said to be *of type* (m,n,A) if its left (resp. right) depth in $\mathfrak{Y}_\ell$ is m (resp. n) and if its point process $\Pi_v^{(1)}$ belongs to A. Let $\pi_{\ell,m,n,A}$ denote the average number of leaves of type (m,n,A) in $\mathfrak{Y}_\ell$ and let $\mathbb{U}_{n,\ell}$ denote the uniform probability distribution on the simplex $\{0<\xi_1<\xi_2<\dots<\xi_n<\ell\}$. Then

Proposition 7.

$$\pi_{\ell,m,n,A} \;=\; \frac{(\ell\ln 2)^{m+n}2^{-\ell}}{m!n!}\;\mathbb{U}_{n,\ell}(A). \tag{4}$$

Up to a factor 2^{m+n}, the right hand of (4) is the probability that two independent Poisson processes on $[0,\ell]$, with intensity $\rho/2=\ln\sqrt{2}$, $\Pi^{(0)}$ (resp. $\Pi^{(1)}$), have m (resp. n) points, and that $\Pi^{(1)}$ belongs to A. This could be seen as an elementary instance of the many-to-one formula for branching random walks, cf. [HH09].

Proof. Consider the probability p that a random direction of a random Yule tree $\mathfrak{Y}_\infty$ produces a leaf of type (m,n,A). On one hand, conditioning first on $\mathfrak{Y}_\infty$, the probability is the number of such leaves times the probability 2^{-m-n} of the path to one of these leaves, so that, going to expectations,

$$p=2^{-m-n}\pi_{\ell,m,n,A}.$$

On the other hand, consider the random walk, on $\mathfrak{Y}_\infty$, of a particle with life time ℓ: the particle chooses *left* or *right* at birth times that form Poisson process with intensity $\rho=\ln 2$ on $[0,\ell]$, and due to the colouring theorem (cf. Kingman, Ch. 5) the times when the particle goes left (resp. right) form two independent Poisson processes $\Pi^{(0)}$ and $\Pi^{(1)}$ with intensity $\rho/2$. Then p is the probability that $\#\Pi^{(0)}=m$, $\#\Pi^{(1)}=n$, and $\Pi^{(1)}\in A$. Thus:

$$p=\frac{(\ell\ln 2)^m 2^{-m-\ell/2}}{m!}\;\frac{(\ell\ln 2)^n 2^{-n-\ell/2}}{n!}\;\mathbb{U}_{n,\ell}\left(A\right).$$

□

When A_g is the set of point processes Π on $[0,\ell]$ such that $G(\Pi)=g$, we set

$$\pi_{\ell,m,n,A_g} \;=\; \pi_{\ell,m,n,g}.$$

3.2. **Sketch of proof.** To sketch the final argument, that follows [BD08], it will be convenient to think of $\Psi(\lambda,\mu,\nu)$ as the following limit:

$$\Psi(\lambda,\mu,\nu) \;=\; \lim_n n^{-1}\;\ln\left(\pi_{\ell,m,n,g}\right), \tag{5}$$

when the sequence $(\lambda_n,\nu_n,\mu_n)=(\ell/n,m/n,g/n)$ converges to (λ,ν,μ). Thus the number of blades of type (m,n,g) in $\mathfrak{S}_\ell$ would be, approximately:

$$e^{n\Psi(\lambda,\mu,\nu)}=e^{\ell\Psi(\lambda,\mu,\nu)/\lambda}.$$

Results in the direction of formula (5), sufficient for our purposes, are proven in Section 3.3. On each blade v of type (m,n,g) of $\mathfrak{S}_\ell$, with depth

$$m+n+g=\frac{1+\nu+\mu}{\lambda}\;\ell,$$

we graft the shrub $\mathfrak{t}(v)$. In Section 3.4, we prove that the maximum height of a set of k such shrubs behaves like the maximum of a sample of k independent geometric random variables with parameter 1/2, i.e. the maximum is essentially $\log_2 k$. As a consequence, the total height of the highest leaf in any shrub that is grafted on a blade of type (m, n, g) in $\mathfrak{S}_\ell$ is approximately

$$m + n + g + \log_2 \left(e^{n\Psi(\lambda,\mu,\nu)} \right) \ = \ \tfrac{(1+\nu+\mu)\ln 2 \ + \ \Psi(\lambda,\mu,\nu)}{\lambda \ln 2}\ \ell.$$

Set

$$\Delta(\lambda, \mu, \nu) = ((1 + \nu + \mu)\ln 2 \ + \ \Psi(\lambda, \mu, \nu))/\lambda \ln 2.$$

Then the highest leaf in the tree $\mathfrak{A}_\ell$ is approximately $\Delta^\bullet \ell$ high, in which

$$\Delta^\bullet \ = \ \sup_{\lambda,\mu,\nu>0} \Delta(\lambda, \mu, \nu). \tag{6}$$

The supremum is obtained for $\nu = 2\lambda \ln 2$, leading to

$$\Delta^\bullet = \sup_{\lambda,\mu>0} \Delta(\lambda, \mu, 2\lambda \ln 2).$$

Finally, due to Proposition 4, the same is true when the shrubs are grafted on the blades of $\mathfrak{T}_\ell$, producing $\mathfrak{L}(L^\ell)$ instead of $\mathfrak{A}_\ell$.

Remark 5. After scaling to recover the Lyndon tree based on a n-letters long Lyndon word, one finds that the maximal height occurs in a shrub grafted on a leaf of $\mathfrak{S}_{\log_2 n}$ whose left (resp. right) depth in $\mathfrak{U}_{\log_2 n}$ is approximately $2 \ln n$ (resp. $1.62.. \ln n$). This can be compared with the height $4.31.. \ln n$ of $\mathfrak{U}_{\log_2 n}$, induced at equal parts by left (resp. right) depths $2.15.. \ln n$. The contribution of needles is approximately $0.86.. \ln n$, leading to a height $4.48.. \ln n$ for the leaf of $\mathfrak{S}_{\log_2 n}$ on which the shrub with the highest top is grafted. Finally, this shrub is approximately $0.61.. \ln n$ high.

3.3. Asymptotic behavior of $n^{-1} \ln \pi_{\ell,m,n,g}$.

3.3.1. *Heuristic considerations.* For a finite or infinite sequence of non negative numbers $b = (b_j)_{j\in I\subset\mathbb{N}_0}$, let us denote

$$|b| \ = \ \sum_{j\in I} b_j, \quad \langle b\rangle \ = \ \sum_{j\in I} jb_j, \quad \text{and} \quad \mathcal{H}(b) \ = \ -\sum_{j\in I} b_j \ln(b_j),$$

and if the entries are integers, set:

$$b! \ = \ \prod b_i! \quad \text{and} \quad \binom{|b|}{b} \ = \ \frac{|b|\,!}{\prod b_i!},$$

whenever they are defined. If $|b| = 1$, $\mathcal{H}(b)$ is the Shannon entropy of b.

Under $\mathbb{U}_{n,\ell}$, rather than the ξ_j's, we shall consider the vector γ of gaps between the order statistics, defined by

$$\gamma_j = \xi_j - \xi_{j-1}, \ 1 \le j \le n+1, \ \text{and} \ \gamma = (\gamma_j)_j,$$

with the convention that $\xi_0 = 0$ and $\xi_{n+1} = \ell$. The random vector γ is uniformly distributed on the simplex

$$\mathcal{D}_{n,\ell} \ = \ \left\{ \sum_{j=1}^{n+1} x_j = \ell \quad \text{and} \quad \forall j, \ x_j \ge 0 \right\},$$

and its distribution is denoted $\mathbb{U}_{n,\ell}$ again. Also, set $\rho = (\rho_j)_{1\le j\le n+1}$, in which

$$\rho_j \;=\; \lfloor \gamma_j \rfloor.$$

Then $G = |\rho|$. The probability $\mathbb{U}_{n,\ell}(\rho = r)$ depends only on $|r|$ and is given by:

$$\mathbb{U}_{n,\ell}\left(\rho = r\right) = n!\,\ell^{-n}\mathbb{P}\left(\ell - |r| - 1 < \sum_{j=1}^{n} U_j < \ell - |r|\right),$$

in which the U_i's are a sequence of i.i.d. random variables uniform on $[0,1]$, $\mathbb{P}\left(\dots\right)$ is the Lebesgue measure of the domain $\{\rho = r\}$, and $\ell^n/n!$ is the Lebesgue measure of $\mathcal{D}_{n,\ell}$. In order to take advantage of the symmetric rôle of the ρ_j's, let $\beta = (\beta_j)_{j\ge 0}$ be defined by

$$\beta_j = \#\left\{1 \le i \le n+1 \,|\, \rho_i = j\right\},$$

so that

$$|\beta| = n+1, \text{ and } \quad \langle\beta\rangle = |\rho| = G(\xi).$$

Then

$$\mathbb{U}_{n,\ell}\left(\beta = b\right) \;=\; \frac{n!}{\ell^n}\binom{|b|}{b}\;\mathbb{P}\left(\ell - \langle b\rangle - 1 < \sum_{j=1}^{n} U_j < \ell - \langle b\rangle\right),$$

and (4) becomes

$$\pi_{\ell,m,n,g} = \frac{\ell^m(\ln 2)^{m+n}}{m!\,2^\ell}\mathbb{P}\left(\ell - g - 1 < \sum_{j=1}^{n} U_j < \ell - g\right)\sum_{\langle b\rangle = g}\binom{|b|}{b}. \tag{7}$$

Set

$$p_{n,m} = \mathbb{P}\left(m - 1 < \sum_{j=1}^{n} U_j < m\right) \quad \text{ and } \quad t_{n,g} = \sum_{\langle b\rangle = g, |b| = n+1}\binom{|b|}{b},$$

so that (7) can be written

$$\pi_{\ell,m,n,g} = \frac{\ell^m(\ln 2)^{m+n}}{m!\,2^\ell}\; p_{n,\ell-g}\; t_{n,g}.$$

From stronger results on the asymptotics of the Eulerian numbers (denoted here $(A(k,n))_{0\le k\le n}$), that can be found in [GK94][formulas (6.12) to (6.16), page 299][2], we know that

Lemma 4. *For $0 < \theta < 1$, the limit:*

$$\Xi(\theta) \;=\; \lim_n \tfrac{1}{n}\ln p_{n,\theta n}$$

exists and is given by

$$\Xi(\theta) \;=\; \ln\sinh\alpha - \alpha\coth\alpha + 1 - \ln\alpha,$$

in which α is given implicitly by $-\alpha^{-1} + 1 + \coth\alpha = 2\theta$.

[2]According to [Tan73], $A(k,n) = n!\; p_{n,k+1}$.

On the other hand, for distributions b such that

$$b_j \simeq c_j(n+1), \quad |c| = 1, \quad \text{and} \quad \langle c\rangle = \mu,$$

we expect that

$$\tfrac{1}{n}\ \ln\left(\binom{|b|}{b}\right) \;=\; \mathcal{H}\left(c\right) + o(1).$$

Following the lines of Lemma 8.3.1 in [Ash65, p. 238], one obtains that

$$\mathcal{H}\left(c\right) \le \mathcal{H}\left(d^{(\mu)}\right) = (1+\mu)\ln(1+\mu) - \mu\ln\mu,$$

in which $d^{(\mu)}$ is the geometric distribution with expectation μ. As usual in large deviation theory, only the leading term of $t_{n,\mu n}$, provided by $c = d^{(\mu)}$, contributes to the limit of $n^{-1}\ \ln t_{n,\mu n}$, so that we expect the following behavior:

$$n^{-1}\ \ln t_{n,\mu n} \quad\simeq\quad \mathcal{H}\left(d^{(\mu)}\right).$$

Together with (4), this would lead to the following expression for Ψ:

$$\Psi(\lambda,\mu,\nu) \;=\; \mathcal{H}\left(d^{(\mu)}\right) \;+\; \Phi(\lambda,\nu) \;+\; \Xi(\lambda-\mu),$$

in which

$$\Phi(\lambda,\nu) \quad=\quad \lim_n \tfrac{1}{n}\ \ln\frac{\ell^m(\ln 2)^{m+n}}{m!\,2^\ell} = \ln\left(\tfrac{(e\lambda\ln 2)^\nu \ln 2}{\nu^\nu 2^\lambda}\right).$$

The analysis of the contribution of the shrubs rests on more or less precise upper and lower bounds for $n^{-1}\ \ln\left(\pi_{\ell,m,n,g}\right)$, to which the rest of the Section is devoted.

3.3.2. *Lower bound for $n^{-1}\ln\pi_{\ell,m,n,g}$.* The lower bounds for $p_{n,g}$ and $t_{n,g}$ do not need to be very precise for our purposes: for $p_{n,g}$ the "lower bound" given by the existence of the limit in Lemma 4 is precise enough. With $\mu = g/n$, according to our heuristic considerations, the lower bound for $t_{n,g}$ provided by any term $\binom{|b|}{b}$ such that b is close to $d^{(\mu)}n$ should be good enough. The sequence $c^{(k)}$ defined by

- $c^{(k)}_j = d^{(\mu)}_j$ for $j \le k$.
- $c^{(k)}_{k+1} = \sum_{j>k} d^{(\mu)}_j = 1 - \sum_{i\le k} c^{(k)}_i$
- $c^{(k)}_i = 0$ for $i \ge k+2$.

satisfies

$$\begin{aligned}\mathcal{H}(c^{(k)}) - \mathcal{H}(d^{(\mu)}) \quad&=\quad \left(\tfrac{\mu}{1+\mu}\right)^{k+1}\left(k\ln\mu + (1+\mu)\ln\left(\tfrac{\mu}{1+\mu}\right)\right),\\ \langle c^{(k)}\rangle \quad&=\quad \mu\left(1-(\tfrac{\mu}{1+\mu})^{k+1}\right).\end{aligned}$$

Consider then the sequence of integers $b^{(n)}$ defined by

- $\forall i \le k, b^{(n)}_i = \lfloor c^{(k)}_i(1+n)\rfloor$,
- $b^{(n)}_{k+1} = n+1-\sum_{i\le k} b^{(n)}_i$,
- $b^{(n)}_i = 0$ pour $i \ge k+2$,

so that

$$\left|b^{(n)}_i - (n+1)nc^{(k)}_i\right| \le \begin{cases} 1 & \text{if} \quad i \le k,\\ k+1 & \text{if} \quad i = k+1,\\ 0 & \text{if} \quad i \ge k+2.\end{cases}$$

Thus,

$$\begin{aligned}\langle b^{(n)}\rangle &\leq (n+1)\langle c^{(k)}\rangle + (k+1)(b^{(n)}_{k+1} - (n+1)c^{(k)}_{k+1})\\ &\leq (n+1)\langle c^{(k)}\rangle + (k+1)^2\\ &\leq (n+1)\mu + (k+1)^2.\end{aligned}$$

On the other hand, since $b^{(n)}$ is obtained, from $(n+1)\,c^{(k)}$, by transport of mass from $i \leq k$ towards $k+1$, we have $\langle b^{(n)}\rangle \geq (n+1)\langle c^{(k)}\rangle$, and as a consequence

$$\langle b^{(n)}\rangle \quad \geq \quad (n+1)\mu\left(1 - (\tfrac{\mu}{1+\mu})^{k+1}\right).$$

Due to the variations of $\varphi(x) = x\ln x - x$, and due to the inequality:

$$|\ln(n!) - \varphi(n)| \leq \ln(4n), \tag{8}$$

that holds for $n \geq 1$, we obtain that:

$$\begin{aligned}|\ln\binom{|b^{(n)}|}{b^{(n)}} - (1+n)\mathcal{H}(d^{(\mu)})| &\leq (n+1)|\mathcal{H}(c^{(k)}) - \mathcal{H}(d^{(\mu)})| + 4k\ln(5n)\\ &= \mathcal{O}(\ln^2 n),\end{aligned}$$

if one chooses $k, n \geq 5$, and also $k = c\ln n$ for c large enough, namely $c\ln(1+\frac{1}{\mu}) > 1$, so that $|\mathcal{H}(c^{(k)}) - \mathcal{H}(d^{(\mu)})|$ is small enough. Now b^n satisfies $\langle b^n\rangle = \mu n + \delta$, instead of μn. This is corrected by transferring a mass δ from 1 to 0, say, i.e. considering $b^n + (\delta, -\delta, 0, 0, \dots)$, causing thus a variation $\delta\left(\ln(1+\frac{1}{\mu}) + \Theta_\mu(\frac{1}{n})\right)$ of $\ln\binom{|b|}{b}$. Here δ is $\mathcal{O}(\ln^2 n)$, thus at least one term of the sum $t_{n,g}$ is equal to $\exp(n\mathcal{H}(d^{(\mu)}) + \mathcal{O}(\ln^2 n))$, so that

$$t_{n,g} \geq \exp(n\mathcal{H}(d^{(\mu)}) + \mathcal{O}(\ln^2 n)). \tag{9}$$

Owing to (8), we also have

$$\left|n\Phi(\lambda,\nu) - \ln\frac{\ell^m(\ln 2)^{m+n}}{m!\,2^\ell}\right| \quad \leq \quad \ln(4m) = \ln(4\nu n). \tag{10}$$

Thus, owing to Lemma 4 and relations (9), (10), we obtain:

Proposition 8. *For $0 < \varepsilon < 0.05$, when T is large enough, the set*

$$L_{T,\varepsilon} = \{(m,n,g)\,|\,\pi_{T,m,n,g} > e^{\varepsilon n} \text{ and } m+n+g+\log_2\pi_{T,m,n,g} > (\Delta^\bullet - \varepsilon)T\}$$

is not empty.

Proof. First we prove that, for any $\varepsilon \in (0, 0.05)$, the set

$$D_\varepsilon = \left\{(\lambda,\mu,\nu) \in (0,+\infty)^3 \mid \Psi(\lambda,\mu,\nu) > 2\varepsilon \text{ and } \Delta(\lambda,\mu,\nu) > \Delta^\bullet - \varepsilon/2\right\}$$

contains a box. Consider the two functions

$$\tilde{\Delta}(\lambda,\mu) = \Delta(\lambda,\mu,2\lambda\ln 2) \text{ and } \tilde{\Psi}(\lambda,\mu) = \Psi(\lambda,\mu,2\lambda\ln 2).$$

When $\tilde{\Delta}$ reaches its maximum $\Delta^\bullet$, the corresponding value of $\tilde{\Psi}$, $\Psi^\bullet$, is positive. Actually, $\{\tilde{\Delta} \geq 5\ln 2\}$ entails $\{\tilde{\Psi} \geq 0.1\}$. The domain $\{(\lambda,\mu)|\tilde{\Delta} > 5\ln 2\}$ is bounded and bounded away from $\lambda = 0$. Both Ψ and Δ are Lipschitz continuous as functions of ν on any domain

$$\left\{(\lambda,\mu,\nu)|\tilde{\Delta} > 5\ln 2 \text{ and } |\nu - 2\lambda\ln 2| \leq \eta\right\}$$

with η not too large. This entails that D_ε contains a box. Thus it contains points of the form $(T,g,m)/n$ as long as T is large enough. If $(T,g,m)/n \in D_\varepsilon$, Lemma 4 and relations (9), (10) entail that if T is large enough then $(m,n,g) \in L_{T,\varepsilon}$. $\square$

3.3.3. *Uniform upper bound on each term of $t_{n,g}$.* For a sequence such that $|b| = n+1$ and $\langle b\rangle = \mu n$, the Markov inequality entails that

$$\sum_{i\le\tau_n} b_i \quad\ge\quad \tfrac{\tau_n-\mu}{\tau_n}\; n.$$

Set

$$\mathcal{F}_{\mu,n} \quad=\quad \left\{c \;\middle|\; c_i = 0 \text{ for } i\ge\tau_n+1 \text{ and } n\ge|c|\ge \tfrac{\tau_n-\mu}{\tau_n}\; n, \langle c\rangle\le\mu n\right\}.$$

Then

$$\inf_{\langle b\rangle=\mu n} b! \;\ge\; \inf_{c\in\mathcal{F}_{\mu,n}} c!$$

for the truncation, after the τ_nth term, of a sequence b such that $\langle b\rangle = \mu n$ produces an element c of $\mathcal{F}_{\mu,n}$ such that $b!\ge c!$.

Set $\tilde c_i = \frac{c_i}{n}$, so that $1\ge|\tilde c|\ge\frac{\tau_n-\mu}{\tau_n}$. For $c\in\mathcal{F}_{\mu,n}$, thanks to (8), we have

$$\ln(c!)\ge n(\ln(n)-1)(1-\tfrac{\mu}{\tau_n})\;-\;n\mathcal{H}\left(\tilde c\right)-(1+\tau_n)\ln(4n).$$

For $\tau_n=\sqrt{\mu n}$, this leads to:

$$\ln(c!)\ge n(\ln(n)-1)\;-\;n\mathcal{H}\left(\tilde c\right)-(1+2\sqrt{\mu n})\ln 4n.$$

Now, following the lines of Lemma 8.3.1 in [Ash65, p. 238], one obtains that

Proposition 9. *If $|a|\le 1$,*

$$\begin{aligned}\mathcal{H}\left(a\right) &\le (|a|+\langle a\rangle)\ln(|a|+\langle a\rangle)-2|a|\ln|a|-\langle a\rangle\ln\langle a\rangle,\\ &\le \mathcal{H}\left(d^{(\langle a\rangle)}\right)-2|a|\ln|a|.\end{aligned}$$

Proof. Set $d_i \;=\; |a|\;\frac{\beta^i}{(1+\beta)^{i+1}}\;1\!\!1_{i\ge0}$, so that $|d|=|a|$ and $\langle d\rangle=|a|\beta$. Then

$$\begin{aligned}\mathcal{H}\left(a\right) &= \sum_{i\ge0}a_i\ln\tfrac{d_i}{a_i}-\sum_{i\ge0}a_i(\ln\tfrac{|a|}{1+\beta}\;+\;i\ln\tfrac{\beta}{1+\beta})\\ &\le -\sum_{i\ge0}a_i(\ln\tfrac{|a|}{1+\beta}\;+\;i\ln\tfrac{\beta}{1+\beta})\\ &= -|a|\ln|a|+(|a|+\langle a\rangle)\ln(1+\beta)-\langle a\rangle\ln\beta.\end{aligned}$$

Choosing $\beta=\langle a\rangle/|a|$, i.e. $\langle d\rangle=\langle a\rangle$, leads to the first inequality. Finally, for $0\le y\le1$ and $x\ge0$, $(y+x)\ln(y+x)\le(1+x)\ln(1+x)$. $\square$

Using that $\mu\longrightarrow\mathcal{H}\left(d^{(\mu)}\right)$ is increasing, we obtain, for $c\in\mathcal{F}_{\mu,n}$,

$$\begin{aligned}\mathcal{H}\left(\tilde c\right) &\le \mathcal{H}\left(d^{(\langle\tilde c\rangle)}\right)-2|\tilde c|\ln|\tilde c|\\ &\le \mathcal{H}\left(d^{(\mu)}\right)+2(1-|\tilde c|)\le\mathcal{H}\left(d^{(\mu)}\right)+2\sqrt{\tfrac{\mu}{n}},\end{aligned}$$

so that

$$\ln(c!)\ge n\ln n-n\;-\;n\mathcal{H}\left(d^{(\mu)}\right)-(1+2\sqrt{\mu n})\ln 12n.$$

This leads to

Proposition 10. *For a sequence such that $|b| = n+1$ and $\langle b \rangle = \mu n = g$,*

$$\ln \binom{|b|}{b} \quad \leq \quad n\,\mathcal{H}\left(d^{(\mu)}\right) \; + \; (3 + 2\sqrt{g}) \ln(12n).$$

3.3.4. *Upper bound for $\#\left\{\langle b \rangle = \mu n\right\}$.* As in Section 3.3.1, let us see each sequence b such that $\langle b \rangle = \mu n$ as the distribution $\delta(r)$ of a sequence of integers $r = (r_j)_{1 \leq j \leq n+1}$ such that $|r| = \mu n$, in the sense that

$$b_j = \#\left\{1 \leq i \leq n+1 \,|\, r_i = j\right\}.$$

Now each sequence r can be partially sorted to form a new sequence $\sigma(r)$ as follows:

- the first terms of $\sigma(r)$ are the terms of r smaller than τ_n, sorted in increasing order;
- the other terms follow, and they retain their relative order in the sequence r.

Then $\delta \circ \sigma = \delta$ and the preimage of an element b such that $\langle b \rangle = \mu n$ has at least one element in $\sigma(\delta^{-1}(\{\langle b \rangle = \mu n\}))$. Thus

$$\#\left\{\langle b \rangle = \mu n\right\} \leq \#\sigma(\delta^{-1}(\{\langle b \rangle = \mu n\})) \leq (n+2)^{\tau_n + 1} \times (\mu n)^{\mu n / \tau_n},$$

the first term of the product on the right because an increasing sequence is well described by its distribution, here $(b_j)_{0 \leq j \leq \tau_n}$, the second term for the length of the second part of the sequence is shorter than $\mu n / \tau_n$ due to the Markov inequality, and each of its terms is smaller than μn and larger than τ_n. As long as $n \geq 2$, the choice $\tau_n = \sqrt{\mu n}$ leads to:

$$\#\left\{\langle b \rangle = g\right\} \;\leq\; \mu^{\sqrt{g}}(n+2)^{1+2\sqrt{g}} \leq (4ng)^{\sqrt{g}} \times 2n. \tag{11}$$

3.3.5. *Upper bound for $p_{n,\ell-g}$.* Trite computations lead to the next Proposition:

Proposition 11. *For $0.5 < \theta < 1$,*

$$\mathbb{P}\left(\theta n < \sum_{j=1}^{n} U_j\right) \;=\; \mathbb{P}\left((1-\theta)n > \sum_{j=1}^{n} U_j\right) \;\leq e^{n\Xi(\theta)}.$$

Proof. Set $S_n = \sum_{j=1}^{n} U_j$, $\varepsilon = 2\theta - 1 > 0$ and $t > 0$. From the Markov inequality, we obtain that:

$$\begin{aligned} \mathbb{P}\left(\theta n \leq S_n\right) &\leq \mathbb{E}\left[e^{t(2U-1)}\right]^n \; e^{-t\varepsilon n} \\ &\leq e^{nh(t)}, \end{aligned}$$

in which

$$h(t) = \ln\left(\frac{\sinh(t)\, e^{-t\varepsilon}}{t}\right).$$

Then

$$\frac{\partial h}{\partial t} = \coth(t) - \varepsilon - \frac{1}{t},$$

and since $t \to \coth(t) - \frac{1}{t}$ is continuous strictly increasing from -1 to 1, h is minimal when $t = \alpha$ in which α is defined by

$$\coth(\alpha) - \varepsilon - \alpha^{-1} = 0.$$

Finally

$$\begin{aligned} h(\alpha) &= \ln\sinh(\alpha) - \alpha\coth(\alpha) + 1 - \ln(\alpha) \\ &= \Xi(\theta). \end{aligned}$$

For $\theta < 0.5$ use that $1 - U_j$'s are uniform, and that $\Xi(\theta) = \Xi(1-\theta)$. □

Corollary 1. *For $n \geq 1$,*

$$\ln p_{n,\ell-g} \leq n(\Xi(\lambda-\mu) + \rho_n),$$

in which

$$\begin{aligned} \rho_n &= \Xi(\lambda-\mu-\delta_n) - \Xi(\lambda-\mu), \\ 0 \leq \delta_n &= \left|(-\infty,\, \lambda-\mu-\tfrac{1}{2}] \cap [0, \tfrac{1}{n}]\right| \leq \tfrac{1}{n}. \end{aligned}$$

3.3.6. *Conclusion.* Combining Proposition 10 and (11), we obtain that, as long as $n \geq 2$:

$$\ln t_{n,g} \quad \leq \quad n\,\mathcal{H}\left(d^{(\mu)}\right) \; + \; (3 + 2\sqrt{g})\ln(24ng),$$

With (10), this gives

Proposition 12. *For $(\lambda,\mu,\nu) = (\ell,g,m)/n$, as long as $n \geq 2$, one has:*

$$\begin{aligned} r_{\ell,m,n,g}\ln 2 \quad &= \quad \ln\pi_{\ell,m,n,g} \; - \; \ell\,\tfrac{\Psi(\lambda,\mu,\nu)}{\lambda} \\ &\leq \quad n\rho_n + \; (3+2\sqrt{g})\ln(24ng) + \ln(4m). \end{aligned}$$

3.4. **Contribution of the shrubs.** The proof of Theorem 2 has two final steps, contained in the next sections.

3.4.1. *Lower bound for $h(\mathfrak{L}(L^\ell))$.*

Proposition 13. *For any $\varepsilon > 0$,*

$$\lim_{\ell} \mathbb{P}\left(h(\mathfrak{L}(L^\ell)) \leq (\Delta^\bullet - \varepsilon)\ell\right) = 0.$$

Proof. Keeping in mind Section 2.5, we observe that

$$|h(\mathfrak{A}_\ell) - h(\mathfrak{L}(L^\ell))| \leq D_\ell.$$

Thus, according to Proposition 4, we only need to prove that

$$\lim_{\ell} \mathbb{P}\left(h(\mathfrak{A}_\ell) \leq (\Delta^\bullet - \varepsilon)\ell\right) = 0.$$

Working with $\mathfrak{A}_\ell$, we can use the representation of the depth of blades $\mathfrak{S}_\ell$ through a Yule process, as in Section 2.6. Let $\mathfrak{B}$ denote the infinite complete binary tree with the language $\mathcal{A}^\star = \{0,1\}^\star$ as set of vertices, and $\{(w,wa) \,|\, (w,a) \in \mathcal{A}^\star \times \mathcal{A}\}$ as edge set, $\emptyset$ being the root. We shall see the infinite family tree $\mathfrak{Y}$ of the Yule process as $\mathfrak{B}$ endowed with edge lengths that form an i.i.d. family of exponential random variables with parameter $\ln 2$. In this setting, we shall interpret the distance between $w \in \mathcal{A}^\star$ and the root $\emptyset$ as the date of the death of w, and the length of the last edge of the path from the root to w as the lifetime of w. An interior point of this last edge inherits the label w of the vertex sitting at the end of the edge.

For any $T > 0$, we shall consider a new tree $\mathfrak{D}_T$ defined as follows:

- the vertices of $\mathfrak{D}_T$ are the points of $\mathfrak{Y}$ at distance nT from the root (for any $n \geq 0$); they are almost surely interior points of some edges;
- a generic edge (x, y) of $\mathfrak{D}_T$ satisfies the following properties:
  - the distance between x and y in $\mathfrak{Y}$ is T;
  - the labels of x and y are of the form (w, ws) with $(w, s) \in \mathcal{A}^\star \times \mathcal{A}^\star$.

If points at distance nT from the root form the n-th generation of a population, then $\mathfrak{D}_T$ is the family tree of a Galton-Watson process with offspring distribution $2^{-T}(1-2^{-T})^{k-1}\, 1\!\!1_{k\geq 1}$ and average offspring size 2^T. Each individual $w \in \mathcal{A}^\star$ living at time nT possesses, attached to him, the Yule family tree $\mathfrak{Y}^{(w)}$ with life time T that describes his progeny between times nT and $(n+1)T$. Consider then, for each descendent $v = ws$ of w living at time $(n+1)T$, its marked point process $\Pi_{w,ws}$ on $[0, T]$ induced by $\mathfrak{Y}^{(w)}$: this allows to decide whether ws is a descendent of w of type (m, n, A) or not. When ws is a descendent of w of type (m, n, A) in $\mathfrak{Y}^{(w)}$, we say that the edge (w, ws) of $\mathfrak{D}_T$ is open. Erasing the closed edges of $\mathfrak{D}_T$, one obtains a forest $\mathfrak{F}$ whose connected components are independent Galton-Watson trees with average offspring size $\pi_{T,m,n,A}$. In the sequel, A is the set A_g of point processes Π on $[0, T]$ such that $G(\Pi) = g$, and we set

$$\pi_{T,m,n,A_g} \;=\; \pi_{T,m,n,g} \;=\; \pi_T.$$

With the help of Proposition 8, for T large enough, one can choose $(m_T, n_T, g_T) = (m, n, g) \in L_{T,\varepsilon}$, such that

$$\pi_T > e^{T\varepsilon/\lambda} > 1 \text{ and } \rho = m + n + g + \log_2 \pi_T > (\Delta^\bullet - \varepsilon)T. \tag{12}$$

As a consequence, the connected components of $\mathfrak{F}$ are supercritical. Since their offspring is bounded by a geometric distribution, according to the Kesten-Stigum Theorem, almost surely, one of the connected components, rooted at a random but finite distance $R \times T$ of the root of $\mathfrak{Y}$, is infinite and satisfies

$$\lim_k Z_k\; \pi_T^{-k} \;=\; C,$$

in which Z_k is the size of generation k of the said component, and C is random but almost surely positive. Finally, the number X_k of elements of type $(k-R)(m, n, g)$ at time kT in $\mathfrak{Y}$ is not smaller than Z_{k-R}, thus it satisfies a.s.

$$\liminf_k \tfrac{1}{k}\; \ln X_k \geq \ln \pi_T.$$

This entails that, for $\eta > 0$ and $c = -\eta + \ln \pi_T$,

$$\lim_k \mathbb{P}\left(X_k \leq e^{ck}\right) = 0.$$

Set $k = \lfloor(\ell - a_\ell)/T\rfloor$ and let S be the set of shrubs grafted on the blades of $\mathfrak{S}_\ell$ that belong to the progeny, at time $\ell - a_\ell$, of the X_k leaves of $\mathfrak{Y}_{kT}$. The set S contains $\tilde{Z}_\ell \geq X_k$ shrubs, and all the corresponding blades have a depth larger than

$$k\;(m + n + g),$$

according to Proposition 6. We shall prove that the highest among the $\tilde{Z}_\ell$ shrubs in S has a height M_ℓ given, approximately, by

$$M_\ell \simeq \log_2 \tilde{Z}_\ell \simeq \tfrac{ck}{\ln 2} \simeq \lfloor \ell/T \rfloor \log_2 \pi_T \simeq \tfrac{\Psi^\bullet \ell}{\ln 2},$$

then we shall use the fact that

$$h(\mathfrak{A}_\ell) \geq k\;(m + n + g) + M_\ell. \tag{13}$$

In general, a shrub contains at least one run of $\mathtt{1}$'s, but with a probability 2^{-a_ℓ}, thus the height H_v of a shrub $\mathfrak{t}(v)$ is larger than the length of, say, its last run of $\mathtt{1}$'s, i.e. stochastically larger than a geometric distribution with parameter $1/2$. Now, roughly speaking, the probability $p_{r,\alpha,a}$ that ar i.i.d. shrubs are all lower than $(1-\alpha)\log_2 r$ is smaller than the probability that ar i.i.d. geometric random variables are all smaller than $(1-\alpha)\log_2 r$, i.e.

$$
\begin{aligned}
p_{r,\alpha,a} &\leq (1 - 2^{-1-(1-\alpha)\log_2 r})^{ar} \\
&\leq \exp(-r^{-(1-\alpha)}/2)^{ar} = \exp(-ar^{\alpha}/2),
\end{aligned}
\tag{14}
$$

for $a > 0$, $0 < \alpha < 1$. However we are not interested here in a sample of $\tilde{Z}_\ell$ i.i.d. shrubs with the typical distribution, i.e. with the distribution associated with a Bernoulli word observed until the first occurence of $\mathtt{0}^{a_\ell}$: each shrub $\mathfrak{t}(v) \in S$ is selected according to the height $h_v \simeq m_v + n_v + g_v$ of the corresponding blade v, and h_v is determined by the operation of the algorithm $\mathfrak{L}^{a_\ell}$ on the sequence $S_\ell = ((n_w, s_w))_{1\leq w\leq N_\ell}$, see Definition 4, while $\mathfrak{t}(v) = \mathfrak{L}(\mathtt{0}^{a_\ell}\mathtt{1}\mathfrak{p}_v)$.

Now s_v and $\mathfrak{p}_v$ have the same a_ℓ-letters long prefix p_v, their respective suffixes, ζ_v and $\hat{\zeta}_v$, being independent. Furthermore $\hat{\zeta}_v$ is independent of S_ℓ, and as a consequence $\hat{\zeta}_v$ and h_v are independent. The probability that $\hat{\zeta}_v$ begins with the prefix $\mathtt{101}$ is thus $1/8$, and this insures that the last run of $\mathtt{1}$'s in $\mathfrak{p}_v$ is a factor of $\hat{\zeta}_v$ and does not depend on h_v. Let Z be the number of shrubs $\mathfrak{t}(v) \in S$ such that $\hat{\zeta}_v$ begins with $\mathtt{101}$: conditionally given $\tilde{Z}_\ell$, Z has a binomial distribution with parameters $\tilde{Z}_\ell$ and $1/8$. The maximal height $\mathfrak{H}_\ell$ in the sample of $\tilde{Z}_\ell$ shrubs is thus stochastically larger than the maximum of Z i.i.d. geometric random variables with parameter $1/2$, and satisfies, for any $\alpha > 0$,

$$
\begin{aligned}
\Pi_\ell &= \mathbb{P}\left(\mathfrak{H}_\ell \leq (1-\alpha)ck/\ln 2\right) \\
&\leq \mathbb{P}\left(Z \leq \tfrac{e^{ck}}{9}\right) + \mathbb{P}\left(Z \geq \tfrac{e^{ck}}{9} \text{ and } \mathfrak{H}_\ell \leq \tfrac{(1-\alpha)ck}{\ln 2}\right) \\
&\leq \mathbb{P}\left(Z \leq \tfrac{e^{ck}}{9}\right) + \exp(-e^{\alpha ck}/18),
\end{aligned}
$$

due to (14), and

$$
\mathbb{P}\left(Z \leq \tfrac{e^{ck}}{9}\right) \leq \mathbb{P}\left(X_k \leq e^{ck}\right) + \mathbb{P}\left(\mathrm{Bin}(e^{ck}, \tfrac{1}{8}) \leq \tfrac{e^{ck}}{9}\right).
$$

The probabilities on the right hand side vanish when ℓ grows, and so does Π_ℓ. Owing to (13),

$$
\mathbb{P}\left(h(\mathfrak{A}_\ell) \leq \gamma_\ell\right) \leq \Pi_\ell,
$$

in which

$$
\gamma_\ell = k\ \left(m + n + g\ + \tfrac{(1-\alpha)c}{\ln 2}\right).
$$

But due to $k = \lfloor(\ell - a_\ell)/T\rfloor$, and owing to our choice of (m, n, g), cf. (12),

$$
\begin{aligned}
\gamma_\ell &= \left(\rho - \tfrac{(1-\alpha)\eta + \alpha\ln\pi_T}{\ln 2}\right)k \\
&\geq (\Delta^\bullet - \varepsilon)(\ell - a_\ell) - \tfrac{(1-\alpha)\eta + \alpha\ln\pi_T}{\ln 2}\lfloor(\ell - a_\ell)/T\rfloor - (\Delta^\bullet - \varepsilon)T.
\end{aligned}
$$

Since α and η are arbitrary positive numbers, for a suitable choice,

$$
\gamma_\ell \geq (\Delta^\bullet - 2\varepsilon)\ell,
$$

for ℓ large enough, which concludes the proof. $\square$

3.4.2. *Upper bound for $h(\mathfrak{L}(L^\ell))$.*

Proposition 14. *For any $\varepsilon > 0$,*

$$\lim_\ell \mathbb{P}\left(h(\mathfrak{L}(L^\ell)) \geq (\Delta^\bullet + \varepsilon)\ell\right) = 0.$$

Proof. Again, according to Proposition 4, we only need to prove that

$$\lim_\ell \mathbb{P}\left(h(\mathfrak{A}_\ell) \geq (\Delta^\bullet + \varepsilon)\ell\right) = 0.$$

A shrub $\mathfrak{t}(v)$ is the Lyndon tree of a Bernoulli word observed until the first occurence of $\mathtt{0}^{a_\ell}$, and as such it comes with a sequence of i.i.d. geometric random variables (the lengths of the runs of $\mathtt{0}$'s) observed until the hitting time of $[a_\ell, +\infty)$: this sequence yields a Galton-Watson tree $\mathfrak{GW}_{a_\ell}(v)$ with geometric offspring distribution, that has a_ℓ generations, as described by Figure 3. Based on $\mathfrak{GW}_{a_\ell}(v)$, one can define a branching random walk $\mathfrak{BGW}_{a_\ell}(v)$ as follows: when the offspring of an individual has size N, each of the N scions jumps N steps in the same direction, say to the left. Then the height H_v of $\mathfrak{t}(v)$ is stochastically smaller than the leftmost position M_v of $\mathfrak{BGW}_{a_\ell}(v)$ at generation a_ℓ: the height of a binary tree with n leaves is smaller than $n-1$ and then one has to account for at most one needle in each family of $\mathfrak{GW}_{a_\ell}(v)$, so N leaves plus one needle entails that the depth of each of the N scions is at most the depth of its father plus N. Assume that the positions of the members of generation a_ℓ for $\mathfrak{BGW}_{a_\ell}(v)$ form a point process $(X_i)_{i\in I}$. Then, as usual, for $0 \leq \theta < \ln 2$, and $x \geq 3$,

$$\mathbb{P}\left(M_v \geq x\right) \leq\ e^{-\theta x}\mathbb{E}\left[e^{\theta M_v}\right],$$

and

$$\begin{aligned}\mathbb{E}\left[e^{\theta M_v}\right] &\leq \mathbb{E}\left[\sum_{i\in I} e^{\theta X_i}\right]\\ &= \mathbb{E}\left[N e^{\theta N}\right]^{a_\ell}\ =\ \left(\tfrac{2e^\theta}{(2-e^\theta)^2}\right)^{a_\ell}.\end{aligned}$$

It entails that

$$\begin{aligned}\mathbb{P}\left(M_v \geq x a_\ell\right) \leq&\ \left(\inf_{0\leq\theta<\ln 2} \tfrac{2e^{\theta(1-x)}}{(2-e^\theta)^2}\right)^{a_\ell}\\ &= \left(2^{-x-2}(x-1)^{1-x}(x+1)^{1+x}\right)^{a_\ell}\\ &\leq \left(2^{-x}(ex/2)^2\right)^{a_\ell},\end{aligned}$$

in which the equality of the second line holds for $x \geq 3$, while the last inequality holds for $x \geq 1$. Thus both the conditional probabilities given the first a_ℓ characters of the word, and, as a consequence, the conditional probability given the height of the root, satisfy:

$$\begin{aligned}\mathbb{P}\left(H_v \geq x a_\ell \mid .\ \right) &\leq 2^{a_\ell}\left(2^{-x}(ex/2)^2\right)^{a_\ell},\\ &\leq \left(2^{-x}e^2x^2\right)^{a_\ell}.\end{aligned}$$

As a consequence, the maximal height $\mathfrak{H}_S$ of any set S of shrubs selected according to the position of their roots in $\mathfrak{S}_\ell$ satisfies

$$\begin{aligned}\mathbb{P}\left(\tfrac{\mathfrak{H}_S}{\ell} \geq \xi/\ln 2\right) &\leq |S| \left(e^{-\xi\ell/a_\ell}(e\xi\ell/a_\ell \ln 2)^2\right)^{a_\ell} \\ &\leq e^{\ln|S| - \xi\ell + 2a_\ell \ln(\xi\ell)}, \end{aligned} \tag{15}$$

as long as $\xi\ell \geq 3a_\ell \ln 2$.

Let $Z_{\ell,m,n,g}$ denote the number of blades of type (ℓ, m, n, g) in $\mathfrak{S}_\ell$, and let $\mathfrak{H}_{\ell,m,n,g}$ (resp. $H_{\ell,m,n,g}$) denote the maximal height among the corresponding set of shrubs (resp. the height of the highest leaf of $\mathfrak{A}_\ell$ that has an ancestor among the $Z_{\ell,m,n,g}$ leaves of type (ℓ, m, n, g)). According to Propositions 4 and 6,

$$\begin{aligned} h(\mathfrak{L}(L^\ell)) \simeq h(\mathfrak{A}_\ell) &= \max\left\{H_{\ell,m,n,g} \,|\, m,n,g \in \mathbb{N}\right\} \\ &= \max\left\{m+n+g+\mathfrak{H}_{\ell,m,n,g} \,|\, m,n,g \in \mathbb{N}\right\}. \end{aligned}$$

As a consequence,

$$\mathbb{P}\left(h(\mathfrak{A}_\ell) \geq (\Delta^\bullet + 2\varepsilon)\ell\right) \leq \sum_{m,n,g\in\mathbb{N}} \mathbb{P}\left(H_{\ell,m,n,g} \geq (\Delta^\bullet + 2\varepsilon)\ell\right) \tag{16}$$

Set

$$\delta(\lambda,\mu,\nu) = \Delta^\bullet - \Delta(\lambda,\mu,\nu) \geq 0.$$

For an arbitrary choice of α, the Markov inequality yields that

$$\mathbb{P}\left(Z_{\ell,m,n,g} \geq \ 2^{\alpha\ell}\right) \leq e^{-\alpha\ell\ln 2 + \ln \pi_{\ell,m,n,g}},$$

Thus, if S stands for the set of blades of type (ℓ, m, n, g) in $\mathfrak{S}_\ell$, (15) entails that, for $\varepsilon > 0$,

$$\mathbb{P}\left(H_{\ell,m,n,g} \geq (\Delta^\bullet + 2\varepsilon)\ell\right) \leq e^{\alpha\ell\ln 2 - \xi\ell + 2a_\ell \ln(\xi\ell)} + \mathbb{P}\left(Z_{\ell,m,n,g} \geq \ 2^{\alpha\ell}\right)$$

in which

$$\xi = (\Delta^\bullet + 2\varepsilon - \tfrac{1+\mu+\nu}{\lambda})\ln 2 \ = (\delta(.) + 2\varepsilon + \tfrac{\Psi(.)}{\lambda \ln 2})\ln 2.$$

The choice

$$\alpha = \Delta^\bullet + \varepsilon - \tfrac{1+\mu+\nu}{\lambda}$$

leads to

$$\mathbb{P}\left(H_{\ell,m,n,g} \geq (\Delta^\bullet + 2\varepsilon)\ell\right) \leq 2^{-\varepsilon\ell + 2a_\ell \log_2(\xi\ell)} + 2^{-(\varepsilon+\delta(.))\ell + r_{\ell,m,n,g}}, \tag{17}$$

provided that $\xi\ell \geq 3a_\ell \ln 2$, since we need (15) to hold. In view of (16) and (17), we need

- to reduce the domain of summation in (16),
- and to check that $r_{\ell,m,n,g}$ is uniformly $o(\ell)$ on the reduced domain.

First point: for $\eta > 0$, set

$$T_{\ell,\eta} = \left\{(m,n,g) \,|\, n \geq \eta\ell,\, \tfrac{m+n}{\ell\ln 2} \in \left[\tfrac{1}{3}, 5\right],\ \text{and } \eta\ell + \ln \pi_{\ell,m,n,g} \geq 0\right\},$$

and note that $\#T_{\ell,\eta} \leq 25\ell^3$.

Lemma 5. *For η small enough, the probability that a leaf of $\mathfrak{A}_\ell$ higher than $(\Delta^\bullet + 2\varepsilon)\ell$ grows from a blade outside $T_{\ell,\eta}$ vanishes when ℓ grows, and*

$$\mathbb{P}\left(h(\mathfrak{A}_\ell) \geq (\Delta^\bullet + 2\varepsilon)\ell\right) \leq o(1) + \sum_{m,n,g\in T_{\ell,\eta}} \mathbb{P}\left(H_{\ell,m,n,g} \geq (\Delta^\bullet + 2\varepsilon)\ell\right).$$

Proof. First we prove that the average number $\pi_{\ell+}$ (resp. $\pi_{\ell-}$) of blades of type (m,n,g) with $m+n \geq 5\ell\ln 2$ (resp. with $m+n \leq (\ell\ln 2)/3$) vanishes when ℓ grows, and so does the probability that there exist such blades: due to (4),

$$\pi_{\ell+} \;=\; \sum_{k\geq 5\ell\ln 2} \frac{(\ell\ln 2)^k 2^{k-\ell}}{k!}, \qquad \pi_{\ell-} \;=\; \sum_{0\leq k\leq(\ell\ln 2)/3} \frac{(\ell\ln 2)^k 2^{k-\ell}}{k!}.$$

If X follows the Poisson distribution with parameter $\ell\ln 2$, we have:

$$\pi_{\ell+} \;=\; \mathbb{E}\left[2^X 1_{X\geq 5\ell\ln 2}\right], \qquad \pi_{\ell-} \;=\; \mathbb{E}\left[2^X 1_{X\leq(\ell\ln 2)/3}\right],$$

and, for any $t>0$,

$$\pi_{\ell+} \;\leq\; \mathbb{E}\left[2^X e^{t(X-5\ell\ln 2)}\right] \;=\; \exp[(e^{t+\ln 2}-1-5t)\ell\ln 2].$$

A suitable choice of t leads to $\lim_\ell \pi_{\ell+}=0$, and similar inequalities yields that $\lim_\ell \pi_{\ell-}=0$. In what follows, we consider only blades of type (m,n,g) with $a\ell \leq m+n \leq b\ell$, in which $a=(\ln 2)/3$ and $b=5\ln 2$.

Set

$$\pi_{\ell,m,n} = \sum_g \pi_{\ell,m,n,g} = \tfrac{(\ell\ln 2)^{m+n}2^{-\ell}}{m!n!}, \tag{18}$$

and let π_η be the expectation of the number of blades such that $\eta\ell + \ln\pi_{\ell,m,n,g} < 0$: it satisfies, for any $t\in(0,1)$,

$$\begin{aligned}
\pi_\eta \leq & \sum_{\substack{a\ell\,\leq\, m+n\,\leq\, b\ell,\\ (\ell-n)_+\,\leq\, g\,\leq\,\ell}} \pi_{\ell,m,n,g}\, e^{-t(\eta\ell+\ln\pi_{\ell,m,n,g})} \\
=&\; e^{-t\eta\ell} \sum_{\substack{a\ell\,\leq\, m+n\,\leq\, b\ell,\\ (\ell-n)_+\,\leq\, g\,\leq\,\ell}} \pi_{\ell,m,n,g}^{1-t} \\
\leq&\; \ell e^{-t\eta\ell} \sum_{a\ell\leq m+n\leq b\ell} \pi_{\ell,m,n}^{1-t} \\
\leq&\; b^{2t}\ell^{1+2t}e^{-t\eta\ell}\left(\sum_{a\ell\leq m+n\leq b\ell}\pi_{\ell,m,n}\right)^{1-t} \\
=&\; b^{2t}\ell^{1+2t}e^{-t\eta\ell}\mathbb{E}\left[2^X 1_{a\ell\leq X\leq b\ell}\right]^{1-t} \\
=&\; b^{2t}\ell^{1+2t}\exp(-\ell(t(\eta+b\ln 2)-b\ln 2)).
\end{aligned}$$

Choosing $1>t>(1+\frac{\eta}{b\ln 2})^{-1}$, we see that π_η vanishes as ℓ grows, and so does the probability of existence of blades such that $\eta\ell+\ln\pi_{\ell,m,n}<0$ and $a\ell\leq m+n\leq b\ell$.

For the bounds we have in mind, Ξ has to be Lipshitz continuous around $\arg\max\Delta$, i.e. its argument $\lambda-\mu$ has to be bounded away from 0 and 1. This should be easy because Ξ goes to $-\infty$ when its argument is close to 0 or 1, leaving then little chance to Δ to be at its maximum. There is chance, however, for Δ to be large even if Ξ is close to $-\infty$: if n is very small, and thus λ very large, Ξ/λ could be small, and Δ could eventually be close to its maximum even though Ξ is close to $-\infty$. The next Lemma fixes this problem: consider the set $\tilde{S}_\eta$ of blades such that $\eta\ell+\ln\pi_{\ell,m,n}\geq 0$ and $n\leq\eta\ell$, for $0<\eta<\ln 2$, and let $H_{\eta,\ell}$ denote the maximal height among the leaves of $\mathfrak{A}_\ell$ that have an ancestor in $\tilde{S}_\eta$. Then we have:

Lemma 6. *For any $\varepsilon > 0$, there exist $\eta > 0$ such that*

$$\lim_{\ell} \mathbb{P}\left(H_{\eta,\ell} \geq (1 + \ln 2 + \varepsilon)\ell\right) = 0.$$

Proof. Owing to (8) and (18),

$$0 \leq (\eta - \ln 2)\ell - \varphi(n) - \varphi(m) + (m+n)\ln(\ell \ln 2) + \ln 16mn.$$

Using that $c\varphi(\frac{x}{c}) = \varphi(x) - x \ln c$, for $c = \ell \ln 2$ we obtain:

$$\begin{aligned} 0 \leq\ & \eta\ell - c - c\varphi(\tfrac{m}{c}) - c\varphi(\tfrac{n}{c}) + 2\ln(2b\ell) \\ \leq\ & c\left(-1 - \varphi(\tfrac{m}{\ell \ln 2}) - \varphi(\tfrac{\eta}{\ln 2}) + \tfrac{\eta}{\ln 2} + \tfrac{2\ln(2b\ell)}{\ell \ln 2}\right) \end{aligned} \tag{19}$$

Since $x \longrightarrow -1-\varphi(x)$ reaches its maximum, 0, when $x = 1$, then η small and ℓ large entails that $|m - \ell \ln 2|$ has to be small, for the left hand size to be nonnegative.

More precisely, for any $\varepsilon > 0$, one can choose $\eta > 0$, independent of m, n, ℓ, in such a way that, for ℓ large enough, (19) holds only if $m \in \ell(\ln 2 \pm \varepsilon)$, by strict convexity and smoothness of φ: solve $1 + \varphi(u) = 2(\frac{\eta}{\ln 2} - \varphi(\frac{\eta}{\ln 2}))$. For η small enough, the 2 solutions r_1 and r_2 are close enough to 1, so that $r_1 \leq \frac{m}{\ell \ln 2} \leq r_2$ entails $m \in \ell(\ln 2 \pm 2\varepsilon)$. Then one can choose ℓ large enough, so that $\frac{2\ln(2b\ell)}{\ell \ln 2} \leq \frac{\eta}{\ln 2} - \varphi(\frac{\eta}{\ln 2})$. Then the depth $m+n+g$ of a blade $v \in \tilde{S}$ is at most $\ell(\ln 2+\varepsilon+\eta+1)$, since $g \leq \ell$.

On the other hand, the expected number $\tilde{\pi}_\eta$ of blades in $\tilde{S}$ is smaller than $\exp(\ell(\eta(\ln\ln 2) - \varphi(\eta)))$: again let X, Y be i.i.d. and Poisson distributed with parameter $\ell \ln 2/2$, so that we have:

$$\tilde{\pi}_\eta \ = \ \mathbb{E}\left[2^{X+Y} 1_{X \leq \eta\ell}\right] \ = \ 2^{\ell/2}\mathbb{E}\left[2^X 1_{X \leq \eta\ell}\right].$$

Again, for any $t > 0$, for instance for $t = \frac{\eta}{\ln 2}$,

$$\begin{aligned} \tilde{\pi}_\eta \leq\ & 2^{\ell/2}\mathbb{E}\left[2^X e^{t(\eta\ell - X)}\right] \ = \ \exp[\ell(t\eta + e^{-t}\ln 2)] \\ =\ & \exp[\ell(\eta(\ln\ln 2) - \varphi(\eta))]. \end{aligned} \tag{20}$$

Set

$$\tilde{\xi} = \eta(\ln\ln 2) - \varphi(\eta).$$

Then, due to (15), the probability that the taller shrub planted on a blade of $\tilde{S}$ is at least $\frac{3\tilde{\xi}\ell}{\ln 2}$ tall vanishes with ℓ: we have

$$\begin{aligned} \mathbb{P}\left(\mathfrak{H}_{\tilde{S}} \geq \tfrac{3\tilde{\xi}\ell}{\ln 2}\right) &\leq \mathbb{P}\left(|\tilde{S}| \geq e^{2\tilde{\xi}\ell}\right) + \mathbb{P}\left(\mathfrak{H}_{\tilde{S}} \geq \tfrac{3\tilde{\xi}\ell}{\ln 2} \text{ and } |\tilde{S}| \leq e^{2\tilde{\xi}\ell}\right) \\ &\leq e^{-\ell\tilde{\xi}} + e^{-\tilde{\xi}\ell + 2a_\ell \ln(3\tilde{\xi}\ell)}, \end{aligned}$$

as long as $\tilde{\xi}\ell \geq a_\ell \ln 2$. This leads to

$$\lim_{\ell} \mathbb{P}\left(H_{\eta,\ell} \geq (1 + \ln 2 + \varepsilon + \eta + \tfrac{3\tilde{\xi}}{\ln 2})\ell\right) = 0.$$

in which $\varepsilon + \eta + \frac{3\tilde{\xi}}{\ln 2}$ is arbitrarily close to 0 for η small. □

We proved that, with a large probability, there exists no blade with $\frac{m+n}{\ell \ln 2}$ outside $[a, b]$, and no blade with $\frac{m+n}{\ell \ln 2}$ inside $[a, b]$ but $\eta + \ln \pi_{\ell,m,n,g} < 0$. Then we proved that, with a large probability, a leaf of $\mathfrak{A}_\ell$ with depth larger than $(1 + \ln 2)\ell$ does not grow from a blade such that $\eta + \ln \pi_{\ell,m,n,g} \geq 0$ but $n \leq \eta\ell$. This concludes the proof since $\Delta^\bullet = 3.5... > 1 + \ln 2$. □

Inside $T_{\ell,\eta}$, we have $(5\ln 2)^{-1} \le \lambda \le \eta^{-1}$ and $0 \le \nu \le 5\lambda \ln 2$. We always have $0 \le \lambda - \mu \le 1$, but $\lambda - \mu$ has to be bounded away from 0 and 1, in order for $n\rho_n$ and $r_{\ell,m,n,g}$ to be small. This holds true due to

$$\begin{aligned}0 &\le \eta\ell + \ln \pi_{\ell,m,n,g}\\ &= \eta\ell + n\Psi(\lambda,\mu,\nu) + r_{\ell,m,n,g}\ln 2\\ &\le \eta\ell + n(\Psi(\lambda,\mu,\nu) + \rho_n)) + (3+2\sqrt{g})\ln(24ng) + \ln(4m).\end{aligned}$$

Dividing by ℓ, for ℓ large enough, this insures that on $T_{\ell,\eta}$,

$$\mathcal{H}\left(d^{(\mu)}\right) + \Phi(\lambda,\mu,\nu) + \Xi(\lambda - \mu - \delta_n)) \ge \tfrac{-2\eta}{5\ln 2},$$

in which all the terms on the left hand side are bounded on $T_{\ell,\eta}$, but eventually Ξ. Thus, for ℓ large enough, on $T_{\ell,\eta}$, Ξ is bounded away from $-\infty$, and $\lambda - \mu - \frac{1}{n}$ is bounded away from 0 and 1, which entails a Lipshitz condition on Ξ on the domain $T_{\ell,\eta}$. Thus, according to Proposition 12, on $T_{\ell,\eta}$, since $n \ge \eta\ell$, and $\lambda \le \eta^{-1}$,

$$\begin{aligned}\tfrac{1}{\ell}\, r_{\ell,m,n,g} \quad &\le \quad \tfrac{1}{\ell}(2n\rho_n + \ (6+4\sqrt{g})\ln(24ng) + 2\ln(4m))\\ &\le \quad \tfrac{10c\ln 2}{\eta\ell} + \ \tfrac{10(\ln(24)+2\ln\ell)}{\sqrt{\ell}} + \tfrac{2\ln(20\ell\ln 2)}{\ell}\end{aligned}$$

in which c is the Lipshitz coefficient for Ξ. Also

$$\begin{aligned}\xi &= (\Delta^\bullet + 2\varepsilon - \tfrac{1+\mu+\nu}{\lambda})\ln 2\\ &\le (\Delta^\bullet + 2\varepsilon - \tfrac{\ln 2}{3})\ln 2,\end{aligned}$$

so that $2a_\ell \log_2(\xi\ell)$ and $r_{\ell,m,n,g}$ are uniformly $o(\ell)$ on the domain $T_{\ell,\eta}$. Finally

$$\begin{aligned}\xi \quad &\ge \quad -\eta - \tfrac{1}{\ell}\ln \pi_{\ell,m,n,g} \ + \ (\delta(.) + 2\varepsilon)\ln 2 + \tfrac{\Psi(\lambda,\mu,\nu)}{\lambda}\\ &= \quad (\delta(.) + 2\varepsilon)\ln 2 - \eta - \tfrac{\ln 2}{\ell}\, r_{\ell,m,n,g}\end{aligned}$$

which ensures $\xi\ell \ge 3a_\ell \ln 2$, and as a consequence (15), on the domain $T_{\ell,\eta}$, for ℓ large enough. These inequalities yield that

$$\lim_{\ell} \sum_{(m,n,g)\in T_\ell} \mathbb{P}\left(H_{\ell,m,n,g} \ge (\Delta^\bullet + 2\varepsilon)\ell\right) = 0.$$

□

4. Depoissonization of the length

In this Section, we derive Theorem 1 from Theorem 2, i.e. we replace the random length $\tau_\ell + 1$ of W^ℓ by the deterministic length of W_n. A simple argument comes to mind: w.h.p. $\tau_{\log_2 n - \varepsilon_n} \le n - 1 \le \tau_{\log_2 n + \varepsilon_n}$ as long as

$$\varepsilon_n \to +\infty \text{ and } \varepsilon_n = o(\ln n)\,; \tag{21}$$

thus w.h.p. $h(\mathfrak{L}(W^{\log_2 n - \varepsilon_n})) \le h(\mathfrak{L}(W_n)) \le h(\mathfrak{L}(W^{\log_2 n + \varepsilon_n}))$. However, while $\ell \to h(\mathfrak{L}(W^\ell))$ is increasing, $n \to h(\mathfrak{L}(W_n))$ is not, see figure 10, and this line of proof fails. We do not even know whether $n \to h(\mathfrak{L}(W_n))$ is stochastically increasing.

In order to address these problems, let the integers α_n, β_n and γ_n be defined by

$$\alpha_n = \log_2 n - 3\varepsilon_n, \quad \beta_n = \lceil \log_2 n - 2\varepsilon_n \rceil, \quad \gamma_n = \lfloor \log_2 n \rfloor$$

in which ε_n meets conditions (21). Let us call *long run* of a word or of a necklace any maximal run that is at least α_n-letters long. Consider the infinite random

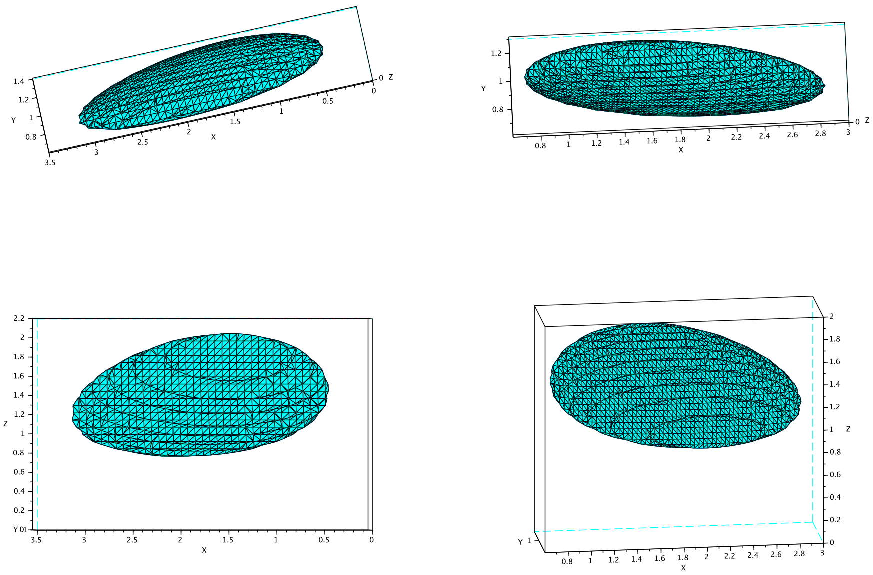


FIGURE 9. Domains $\Delta(\lambda,\mu,\nu) > 4.95$ and $\Delta(\lambda,\mu,\nu) > 5$, with $x = \coth(\lambda-\mu) - \frac{1}{\lambda-\mu}$, $y = \lambda$, and $z = \nu$.

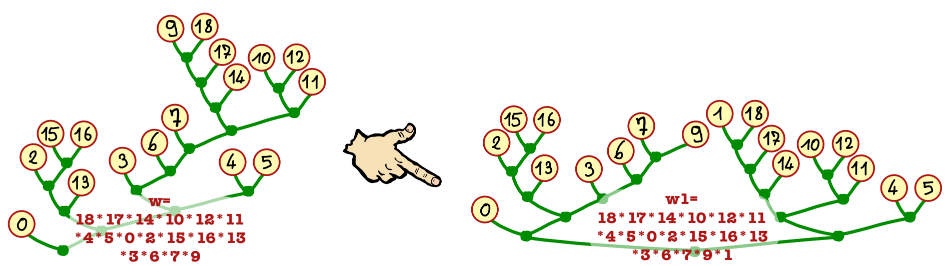


FIGURE 10. Here the alphabet is $\mathbb{N}$ and $6 = h(\mathfrak{L}(w\mathtt{1})) < h(\mathfrak{L}(w)) = 10$.

word W_∞ and its n-letters long prefix W_n. A sequence $Z = (Z_i)_{i\geq 1}$ of i.i.d. factors of W_∞ is defined as follows: for $i \geq 2$, Z_i begins just after the end of Z_{i-1} and, counting from the position of the first $\mathtt{1}$ of Z_i, Z_i stops with the α_n-th $\mathtt{0}$ of the first long run after that first $\mathtt{1}$; Z_1 stops according to the same rule, but starts with the first letter of W_∞. Each factor Z_i can be written $\mathtt{0}^{X_i} Y_i$, in which

- Y_i has the same distribution as W^{α_n}, cf. section 2.2,
- $\mathbb{P}(X_i = k) = 2^{-k-1}\mathbb{1}_{k\geq 0}$,
- Y_i and X_i are independent.

The Lyndon word of Y_i, denoted L_i in what follows, is obtained by a rotation in which the right factor is $\mathtt{0}^{\alpha_n}$; L_i has the same distribution as L^{α_n}, cf. section 2.2. The X_i's account for the overshoots of each run not shorter than α_n.

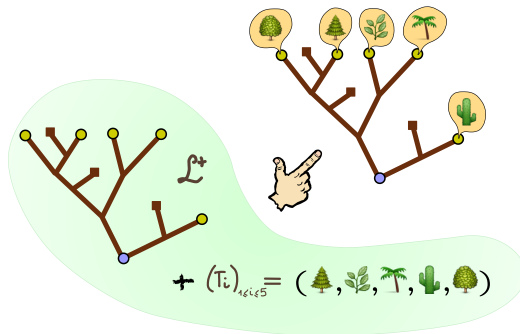


FIGURE 11. Permuted concatenation of $(T_i)_{1\leq i\leq 5}$. The underlying tree $\mathfrak{L}^+$ has 3 unmarked leaves and the permutation is circular.

Definition 5. Let $(T_i)_{1\leq i\leq k}$ be a finite sequence of planar trees, $\mathfrak{L}^+$ a planar tree with k marked leaves, naturally labeled by their rank (rank going clockwise, starting after the root), and σ a permutation of $\{1, 2 \ldots k\}$. Then the tree obtained when one grafts to $\mathfrak{L}^+$ the tree T_i, its root being substituted to the leaf of $\mathfrak{L}^+$ with label $\sigma(i)$, for all i, is called the *permuted concatenation* of $(T_1, \ldots, T_k)$ with underlying tree $\mathfrak{L}^+$.

For instance, for any integer-valued random variable N, $\mathfrak{L}(Z_1 Z_2 \ldots Z_N)$ is a permuted concatenation of the trees $\mathfrak{L}(L_i)$. In this case the permutation is a circular one, and the underlying tree is a binary tree with N marked leaves, and $\sum_{i=1}^N X_i$ unmarked leaves (here these unmarked leaves are needles); thus the height of the underlying tree is at most $N - 1 + \sum_{i=1}^N X_i$.

Consider the prefix $W_{[n]}$ of W_∞ obtained by concatenation of the factors Z_i that are contained or overlap W_n, say

$$W_{[n]} = Z_1 Z_2 \ldots Z_N.$$

4.1. **Lower bound.** According to [BK02, p. 111, (4.19)], the position of the α_n-th 0 of the rth long run has expectation $r2^{1+\alpha_n} - 2$ and a variance smaller than $r(2 + 2^{2+2\alpha_n})$. So by Chebyshev's inequality, w.h.p. there exist at least $2^{2\varepsilon_n}$ long runs in W_n, which entails successively that $N \geq 3$, that $\mathfrak{L}(L_2)$ is a subtree of $\mathfrak{L}(W_n)$ and that $h(\mathfrak{L}(W_n)) \geq h(\mathfrak{L}(L_2))$. For $\eta > 0$, set

$$p_{n-} = \mathbb{P}\left(\tfrac{h(\mathfrak{L}(W_n))}{\ln n} \leq \Delta^\star - \eta\right).$$

Thus

$$p_{n-} \leq \mathbb{P}\left(N \leq 2\right) + \mathbb{P}\left(\tfrac{h(\mathfrak{L}(L_2))}{\ln n} \leq \Delta^\star - \eta\right).$$

On the other hand, the height H_i of $\mathfrak{L}(L_i)$ has the same distribution as $h(\mathfrak{L}(L^{\alpha_n}))$, thus, according to Theorem 2, for any $\varepsilon > 0$, for instance if $0 < \varepsilon < \eta$,

$$\lim_n \mathbb{P}\left(\tfrac{h(\mathfrak{L}(L_2))}{\alpha_n} \leq (\Delta^\star - \varepsilon)\ln 2\right) = 0$$

which, together with the previous inequality, gives $\lim_n p_{n-} = 0$.

4.2. **Upper bound.** The upper bound for Theorem 1 follows at once from the next proposition:

Proposition 15. *For any $\eta > 0$, there exists a sequence (ε_n) that meets conditions* (21)*, and such that*

$$\lim_n \mathbb{P}\left(\frac{h(\mathfrak{L}(W_{[n]}))}{\ln n} \geq \Delta^\star + \eta\right) = 0, \tag{22}$$

and

$$\lim_n \mathbb{P}\left(h(\mathfrak{L}(W_n)) \geq h(\mathfrak{L}(W_{[n]}))\right) = 0. \tag{23}$$

Proof of (22). Let us prove that the sequence

$$p_{n+} = \mathbb{P}\left(\frac{h(\mathfrak{L}(W_{[n]}))}{\ln n} \geq \Delta^\star + 2\eta\right)$$

vanishes. Since $\mathfrak{L}(W_{[n]})$ is a permuted concatenation of $(\mathfrak{L}(L_i))_{1\leq i\leq N}$, in which the underlying tree, $\mathfrak{L}^{\alpha_n}(W_{[n]})$, has $\sum_{1\leq i\leq N} X_i$ needles unmarked, and N marked leaves (or blades), Lemma 1 entails that:

$$h(\mathfrak{L}_{\alpha_n}(W_{[n]})) \leq N + \max_{1\leq i\leq N} X_i,$$

which yields the following bound:

$$h(\mathfrak{L}(W_{[n]})) \leq N + \max_{1\leq i\leq N} X_i + \max_{1\leq i\leq N} H_i.$$

Thus, if $t_n = o(\ln n)$ and if n is large enough,

$$p_{n+} \leq \mathbb{P}\left(N \geq t_n\right) + t_n \mathbb{P}\left(\tfrac{H_1}{\ln n} \geq \Delta^\star + \eta\right) + \mathbb{P}\left(\max_{1\leq i\leq t_n} X_i > 2\log_2 t_n\right).$$

Since $N-1$ is smaller than the number X of occurrences of the pattern $\mathtt{0}^{\alpha_n}$ in W_n, that satisfies

$$\mathbb{E}\left[X+1\right] = (n-\alpha_n+1)2^{-\alpha_n} + 1 \leq 2^{3\varepsilon_n},$$

we have

$$\mathbb{P}\left(N \geq 3^{3\varepsilon_n}\right) \leq (\tfrac{2}{3})^{3\varepsilon_n}.$$

According to Theorem 2, for $\eta > 0$,

$$\lim_n \mathbb{P}\left(\frac{h(\mathfrak{L}(L^{\gamma_n}))}{\gamma_n} \geq (\Delta^\star + \eta)\ln 2\right) = 0.$$

We can thus choose a sequence ε_n that fulfills conditions (21) in such a way that $3^{3\varepsilon_n} = o(\ln n)$ and

$$\lim_n 3^{3\varepsilon_n} \mathbb{P}\left(\frac{h(\mathfrak{L}(L^{\gamma_n}))}{\gamma_n} \geq (\Delta^\star + \eta)\ln 2\right) = 0.$$

But since $h(\mathfrak{L}(L^k))$ is stochastically increasing in k, we have

$$\lim_n 3^{3\varepsilon_n} \mathbb{P}\left(\frac{h(\mathfrak{L}(L^{\alpha_n}))}{\log_2 n} \geq (\Delta^\star + \eta)\ln 2\right) = 0.$$

Finally

$$\mathbb{P}\left(\max_{1\leq i\leq t_n} X_i > 2\log_2 t_n\right) \leq 1 - (1 - \tfrac{2}{t_n^2})^{t_n} \leq -t_n \ln(1 - \tfrac{2}{t_n^2}),$$

and the choice $t_n = 3^{3\varepsilon_n}$ gives the desired result. $\square$

Proof of (23). Let us call *huge run* of a word any maximal run of $\mathtt{0}$'s of its necklace that is at least β_n-letters long. Let now N (resp. $\tilde{N}$) stand for the number of huge runs contained in the necklace of W_n (resp. of $W_{[n]}$); as in Section 4.1, w.h.p. N and $\tilde{N}$ grow at least like 2^{ε_n}. We shall prove that, w.h.p., $N = \tilde{N}$ and that, w.h.p., the Lyndon factors $\hat{L}_1, \hat{L}_2, \dots \hat{L}_N$ (resp. $\tilde{L}_1, \tilde{L}_2, \dots \tilde{L}_N$) induced by these huge runs have the same lexicographic order. The convention here is that $\hat{L}_1$ (or $\tilde{L}_1$ for that matter) begins with the first letter of the first huge run in W_n, and ends with the last letter before the second huge run in W_n.

This way, typically, $\hat{L}_N$ (resp. $\tilde{L}_N$) straddles the end of W_n (resp. $W_{[n]}$), for, in $\langle W_n\rangle$, or in $\langle W_{[n]}\rangle$ as well, w.h.p., the run of $\mathtt{0}$'s that straddles the end of the word is not huge:

- in W_n the length of this run is stochastically smaller than the sum of two X_i's, thus the probability that it is huge is smaller than $2^{-\beta_n/2}$;
- since $W_{[n]}$ ends with a run of exactly α_n $\mathtt{0}$'s, a run straddling the end of $W_{[n]}$ is huge only if the run of $\mathtt{0}$'s at the beginning of W_∞ is longer than ε_n, i.e. with probability $2^{-\varepsilon_n}$.

Thus $N = \tilde{N}$ w.h.p., and $\hat{L}_i = \tilde{L}_i$ for $1 \le i \le N-1$.

Also, w.h.p., there exists no huge run in the last $\beta_n + \sqrt{n}$ characters of W_n, the probability being bounded by

$$(1+\sqrt{n})2^{-\beta_n} = o(1),$$

and a $\sqrt{n}$-letters long pattern does not appear twice in W_n, the probability being bounded by

$$n^2 2^{-\sqrt{n}} = o(1).$$

It follows that, though different, $\hat{L}_N$ and $\tilde{L}_N$ begin with the same huge run, and have the same first $\sqrt{n}$ characters after their initial huge run. As a consequence, the outcome of the lexicographic comparison between $\hat{L}_N$ (resp. $\tilde{L}_N$) and $\hat{L}_i$, being known before reading the $\sqrt{n}$-th character of $\hat{L}_N$ (resp. $\tilde{L}_N$), is the same in both cases. So, $\mathfrak{L}(W_n)$ and $\mathfrak{L}(W_{[n]})$ are permuted concatenations of $(\mathfrak{L}(\hat{L}_i))_{1\le i\le N}$ (resp. $(\mathfrak{L}(\tilde{L}_i))_{1\le i\le N}$) with the same underlying tree and the same permutation, and the height of their leaves are the same, but perhaps if the label of the leaf is a letter from $\hat{L}_N$ or $\tilde{L}_N$.

Finally, we prove that w.h.p. at least one of the highest leaves, say $v^\star$, of $\mathfrak{L}(W_n)$ is not labeled with a letter from the factor $\hat{L}_N$. This is due to the following facts:

(1) if $1 \le x_n < y_n \le n$, then with a probability at least $\frac{y_n - x_n}{n}$, at least one of the highest leaves of $\mathfrak{L}(W_n)$ is between x_n and y_n;
(2) we can choose x_n, y_n so that, simultaneously, $1 - \frac{y_n - x_n}{n} = o(1)$, and, w.h.p., $\hat{L}_N$ does not overlap $[\![x_n, y_n]\!]$.

Since the previous highly likely events insure that all the leaves from $\hat{L}_i$, $i < N$ have the same height in $\mathfrak{L}(W_n)$ and in $\mathfrak{L}(W_{[n]})$, the height of $v^\star$ in $\mathfrak{L}(W_{[n]})$ is $h(\mathfrak{L}(W_n))$, which entails (23).

Let us prove points (1) and (2). Let w be a primitive word of $\{\mathtt{0},\mathtt{1}\}^n$ and let $\langle w\rangle$ denote its necklace. Conditionnally, given that $W_n \in \langle w\rangle$, W_n is uniform on $\langle w\rangle$. Let $\Pi = \{\Pi_i, 1 \le i \le k\} \subset [\![1,n]\!]$ be the set of positions of the labels of the highest leaves of $\mathfrak{L}(w)$. The ℓ-th element of $\langle w\rangle$ (obtained after the permutation of a ℓ-letters long suffix of w with the corresponding $n-\ell$-letters long prefix) fills the condition of point (1) iff $\ell \in [\![x_n, y_n]\!] - \Pi$. Now $[\![x_n, y_n]\!] - \Pi$ is an union of intervals of $\mathbb{Z}_n$ with the same width $y_n - x_n$ thus this reunion has exactly $y_n - x_n + 1$ elements if there exists only one highest leaf (this never happens for a binary tree), but else it has more elements. It follows that for any primitive word w,

$$\mathbb{P}\left(W_n \text{ meets condition (1)} | W_n \in \langle w \rangle\right) \geq \tfrac{y_n - x_n + 2}{n},$$

which takes care of point (1) for n large enough (recall e.g. [CZA10, Lemma 2.1]).

For point (2), consider the position L (resp. R) of the `0` with rank β_n in the first huge run of W_∞ (resp. the position of the `0` ranked β_n in the first huge run of W_∞, *going backward* from the last letter of W_n). For R to be fully defined you need to see W_∞ as a *doubly* infinite sequence of random characters. If $\hat{L}_N$ overlaps $[\![x_n, y_n]\!]$ then $L - \beta_n \geq x_n$ or $R \leq y_n$. Also L and $n+1-R$ follow the geometric distribution of order β_n (cf. [BK02, p. 17, (2.18)]) whose expectation (resp. variance) are

$$2^{1+\beta_n} - 2, \quad 2^{3+2\beta_n}(1 - (2\beta_n + 5)2^{-2-\beta_n} + 2^{-2-2\beta_n}),$$

with the consequence that, by Chebyshev's inequality, for n large enough,

$$\mathbb{P}\left(L \geq (1 + \rho_n\sqrt{2})2^{1+\beta_n}\right) \leq \rho_n^{-2}.$$

One can choose $\rho_n \simeq 2^{\varepsilon_n}$ so that $x_n = (1 + \rho_n\sqrt{2})2^{1+\beta_n} = n2^{-\varepsilon_n}$, and one can also take $n + 1 - y_n = x_n$, so that

$$\mathbb{P}\left(L \geq x_n\right) + \mathbb{P}\left(R \leq y_n\right) \leq 2\rho_n^{-2},$$

and $1 - \frac{y_n - x_n}{n} = 2^{1-\varepsilon_n} - \frac{1}{n}$, which proves point (2). □

Acknowledgements

The second author thanks Julien Clément for introducing him to the probabilistic analysis of Lyndon words and trees, 10 years ago[3], and Anne Briquet, Nicolas Broutin, Luc Devroye, Cyril Nicaud for renewing recently his interest after some years of unsuccessful attempts. Some discussions with Brigitte Chauvin were specially helpful.

References

[Ald01] David J. Aldous, *Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today*, Statist. Sci. **16** (2001), no. 1, 23–34. MR 1838600 (2003d:60171)

[AN72] Krishna B. Athreya and Peter E. Ney, *Branching processes*, Springer-Verlag, New York, 1972, Die Grundlehren der mathematischen Wissenschaften, Band 196.

[Ash65] Robert Ash, *Information theory*, Interscience Publishers John Wiley & Sons, 1965.

[Bar90] Hélène Barcelo, *On the action of the symmetric group on the free Lie algebra and the partition lattice*, J. Combin. Theory Ser. A **55** (1990), no. 1, 93–129.

[BCN05] Frédérique Bassino, Julien Clément, and Cyril Nicaud, *The standard factorization of Lyndon words: an average point of view*, Discrete Math. **290** (2005), no. 1, 1–25.

[BD08] N. Broutin and L. Devroye, *An analysis of the height of tries with random weights on the edges*, Combin. Probab. Comput. **17** (2008), no. 2, 161–202.

[Big77] J. D. Biggins, *Chernoff's theorem in the branching random walk*, J. Appl. Probability **14** (1977), no. 3, 630–636.

[BK02] N. Balakrishnan and Markos V. Koutras, *Runs and scans with applications*, Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], New York, 2002.

[3]According to Julien Clément, the question of the height of the Lyndon tree was raised by Jean Berstel.

[BSW94] J. J. A. M. Brands, F. W. Steutel, and R. J. G. Wilms, *On the number of maxima in a discrete sample*, Statist. Probab. Lett. **20** (1994), no. 3, 209–217.

[CKMR05] B. Chauvin, T. Klein, J.-F. Marckert, and A. Rouault, *Martingales and profile of binary search trees*, Electron. J. Probab. **10** (2005), no. 12, 420–435.

[CZA10] Philippe Chassaing and Elahe Zohoorian Azad, *Asymptotic behavior of some factorizations of random words*, 2010.

[Dev86] Luc Devroye, *A note on the height of binary search trees*, J. Assoc. Comput. Mach. **33** (1986), no. 3, 489–498.

[Dev92] Luc Devroye, *A limit theory for random skip lists*, Ann. Appl. Probab. **2** (1992), no. 3, 597–609.

[GK94] Eldar Giladi and Joseph B. Keller, *Eulerian number asymptotics*, Proc. Roy. Soc. London Ser. A **445** (1994), no. 1924, 291–303.

[HH09] Robert Hardy and Simon C. Harris, *A spine approach to branching diffusions with applications to L^p-convergence of martingales*, Séminaire de Probabilités XLII, Springer, 2009, pp. 281–330. MR 2599214 (2011f:60164)

[HR03] Christophe Hohlweg and Christophe Reutenauer, *Lyndon words, permutations and trees*, Theoret. Comput. Sci. **307** (2003), no. 1, 173–178, Words.

[Lot97] M. Lothaire, *Combinatorics on words*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1997.

[Rob] Matthew I. Roberts, *Almost sure asymptotics for the random binary search tree*, 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10), Discrete Math. Theor. Comput. Sci. Proc., AM, pp. 565–576.

[SF13] Robert Sedgewick and Philippe Flajolet, *An introduction to the analysis of algorithms*, Addison-Wesley-Longman, 2013.

[SR03] J. Sawada and F. Ruskey, *Generating Lyndon brackets. An addendum to: "Fast algorithms to generate necklaces, unlabeled necklaces and irreducible polynomials over* GF(2)*"*, J. Algorithms **46** (2003), no. 1, 21–26.

[SSM92] K. G. Subramanian, Rani Siromoney, and Lisa Mathew, *Lyndon trees*, Theoret. Comput. Sci. **106** (1992), no. 2, 373–383. MR 1192777 (94f:68152)

[Tan73] S. Tanny, *A probabilistic interpretation of Eulerian numbers*, Duke Math. J. **40** (1973), 717–722.

Institut Elie Cartan de Lorraine, Université de Lorraine, Campus Scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex France

E-mail address: mercier@phare.normalesup.org

E-mail address: chassaingph@gmail.com