On the Limiting Distribution of Lempel Ziv'78 Redundancy for Memoryless Sources

Philippe Jacquet and Wojciech Szpankowski, *Fellow, IEEE,*

***Abstract*—We study the Lempel-Ziv'78 algorithm and show that its (normalized) redundancy rate tends to a Gaussian distribution for memoryless sources. We accomplish it by extending findings from our 1995 paper [4], in particular, by presenting a new simplified proof of the Central Limit Theorem (CLT) for the number of phrases in the LZ'78 algorithm. As in [4], we first analyze the asymptotic behavior of the total path length in the associated digital search tree (a DST) built from independent sequences. Then a renewal theory type argument yields CLT for LZ'78 scheme. Here, we extend our analysis of LZ'78 algorithm to present new results on the convergence of moments, moderate and large deviations, and CLT for the (normalized) redundancy. In particular we confirm that the average redundancy rate decays as $\frac{1}{\log n}$, and we find that the variance is of order $\frac{1}{n}$ where n is the length of the text.**

I. Introduction

The Lempel-Ziv compression algorithm [15] is a universal compression scheme. It partitions the text to be compressed into consecutive phrases such that the next phrase is the unique shortest prefix of the uncompressed text not seen before in the compressed portion of the text. The compression code for a word w over the alphabet $\mathcal{A}$ we denote as $C(w)$. It is known that for a large class of sources the average compression rate $|C(w)|/|w|$ tends to the source entropy rate h when $|w| \to \infty$. Our goal is to prove that the (normalized) redundancy rate

$$r(w) = \frac{|C(w)|}{|w|} - h$$

tends in probability and in moments to a normal distribution when w is generated by a memoryless source. In particular, we prove that

$$\mathbf{E}(r(w)) = O\left(\frac{1}{\log |w|}\right), \qquad \mathrm{Var}(r(w)) = O\left(\frac{1}{|w|}\right)$$

when $|w| \to \infty$.

It is convenient to organize the phrases (dictionary) of the Lempel-Ziv scheme in a *digital search tree* (DST) [8], [14] which represents a parsing tree. The root then contains an empty phrase. The first phrase is the first symbol, say "$a \in \mathcal{A}$" which is stored in a node appended to the root. The next phrase is either "$aa \in \mathcal{A}^2$" stored in another node that branches out from the node containing the first phrase "a" or a new symbol that is stored in a node attached to the root. This process repeats recursively until the text is parsed into full phrases (last incomplete phrase will be ignored here). A detailed description can be found in [3], [4], [14]; see also Figure 1.

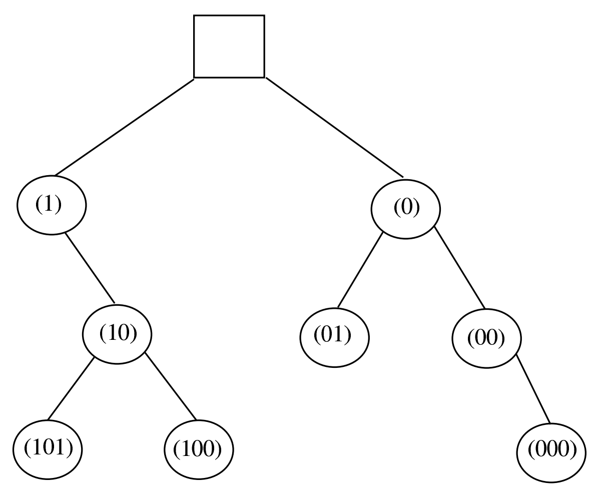

Fig. 1. A digital tree representation of the Lempel-Ziv parsing for the string 11001010001000100 ... into phrases (1)(10)(0)(101) ...

Let a text w be generated over an alphabet $\mathcal{A}$, and let $\mathcal{T}(w)$ be the associated digital search tree constructed by the algorithm. Each node in $\mathcal{T}(w)$ corresponds to a phrase in the parsing algorithm. Let $L(w)$ be the (total) path length in $\mathcal{T}(w)$, that is, the sum of all paths from the root to all nodes (i.e., the sum of phrases which is also the text length). We have $L(w) = |w|$ (if all phrases

Parts of this paper were presented at the 2011 IEEE International Symposium on Information Theory, St. Petersburg, Russia.

Philippe Jacquet, Bell Labs, Alcatel-Lucent, 91620 Nozay, France (e-mail: Philippe.Jacquet@inria.fr).

W. Szpankowski is with the Department of Computer Science, Purdue University, IN 47907, USA (e-mail: spa@cs.purdue.edu). His work was partially supported by the NSF Science and Technology Center for Science of Information Grant CCF-0939370, NSF Grants DMS-0800568 and CCF-0830140, AFOSR Grant FA8655-11-1-3076, NSA Grant H98230-11-1-0141.

are full). If we know the order of nodes creation in the tree $\mathcal{T}(w)$, then we can reconstruct the original text w; otherwise we construct a string of the same universal type, as discussed in [13].

The compression code $C(w)$ is a description of $\mathcal{T}(w)$, node by node in the order of creation; each node being identified by a pointer to its parent node in the tree and the symbol that labels the edge linking it to the parent node. The pointer to the kth node requires at most $\lceil \log_2 k \rceil$ bits, and the next symbol costs $\lceil \log_2 |\mathcal{A}| \rceil$ bits. We just assume that the total pointer cost is $\lceil \log_2(k) \rceil + \lceil \log_2 |\mathcal{A}|) \rceil$ bits. The compressed code length is

$$|C(w)| = \sum_{k=1}^{M(w)} \lceil \log_2(k) \rceil + \lceil \log_2(|\mathcal{A}|) \rceil \qquad (1)$$

where $M(w)$ is the number of full phrases needed to parse w. Clearly, $M(w)$ is also the number of nodes in the associated tree $\mathcal{T}(w)$. Notice that the code is self-consistent and does not need *a priori* knowledge of the text length, since the length is a simple function of the node sequence. We conclude from (1) that

$$|C(w)| = \xi_2(M(w)) \ .$$

where we define

$$\xi_Q(x) = x\lceil \log_Q(|\mathcal{A}|) \rceil + \sum_{0<k\le x} \lceil \log_Q(k) \rceil \ . \qquad (2)$$

for any integer Q and real x. Notice that $\xi_Q(M(w))$ is the code length if it is written in a Q-ary alphabet. It turns out that when x increases $\xi_Q(x)$ is (asymptotically) equivalent to $x(\lceil \log_Q x \rceil + \lceil \log_Q(|\mathcal{A}|))$. In fact, our results hold for any function which is asymptotically equivalent to $x(\alpha \log x + \beta)$ for some nonnegative numbers α and β. Actually, different implementation may add $O(M(w))$ to the code length (see [7] for precise formula) without changing our asymptotic findings. To simplify, we shall assume throughout that

$$|C(w)| = M(w)\left(\log(M(w)) + \log(|\mathcal{A}|)\right) . \qquad (3)$$

Using natural logarithm simply means that we measure the quantities of information in "nat" units.

In this paper we study the limiting distribution, large deviations, and moments of the number of phrases $M(w)$ and the redundancy $r(w)$ when the text of length $|w| = n$ is generated by a *memoryless source*. We prove the Central Limit Theorem (CLT) for the number of phrases and establish precise rate of decay of the LZ'78 code redundancy. Furthermore, we prove that the (normalized) redundancy rate of the LZ'78 code obeys the Central Limit Law with mean $O(1/\log n)$ and variance $O(1/n)$. The former result was already proved in our 1995 paper [4] while the the average redundancy was presented in [7], [12], but not the CLT for the redundancy which is new. However, the proof of the CLT in our 1995 paper was quite complicated; it involves a generalized analytic depoissonization over convex cones in the complex plane. In this paper we simplified and generalized it to present new comprehensive large deviations results. It should be pointed out that since our 1995 paper [4] no simpler, in fact, no new proof of CLT was presented except the one by Neininger and Ruschendorf [10] but only for *unbiased* memoryless sources (as in [1]). The proof of [10] applies the so called *contraction method*.

The paper is organized as follows. In the next section we present our main results concerning the LZ'78 algorithm followed by CLT and large deviations results for the path length in digital search trees which are at the heart of our proof. Most proofs are delayed till Sections IV and V.

II. Main Results

Let n be a nonnegative integer. We denote by M_n the number of phrases $M(w)$ and C_n the code length $C(w)$ when the original text w is of fixed length n.We shall assume throughout that the text is generated by a memoryless source over alphabet $\mathcal{A}$ such that the entropy rate is $h = -\sum_{a\in\mathcal{A}} p_a \log p_a > 0$ where p_a is the probability of symbol $a \in \mathcal{A}$. We respectively define the *compression rate* ρ_n:

$$\rho_n = \frac{C_n}{n} \ ,$$

and the *redundancy* r_n:

$$r_n = \rho_n - h \ .$$

We also define $h_2 = \sum_{a\in\mathcal{A}} p_a (\log p_a)^2$ and

$$\eta = -\sum_{k\ge 2} \frac{\sum_{a\in\mathcal{A}} p_a^k \log p_a}{1 - \sum_{a\in\mathcal{A}} p_a^k} \ . \qquad (4)$$

Finally, we introduce three functions over integer m:

$$\begin{aligned}
\beta(m) &= \frac{h_2}{2h} + \gamma - 1 - \eta + \Delta_1(\log m), \\
&+ \frac{1}{m}\left(\log m + \frac{h_2}{2h} + \gamma - \eta - \sum_{a\in\mathcal{A}} \log p_a - \frac{1}{2}\right), \\
v(m) &= \frac{m}{h}\left(\frac{h_2 - h^2}{h^2} \log m + c_2 + \Delta_2(\log m)\right) \\
\ell(m) &= \frac{m}{h}\left(\log m + \beta(m)\right),
\end{aligned}$$

where $\gamma = 2.718\ldots$ is the Euler constant, c_2 a constant, and $\Delta_1(x)$ and $\Delta_2(x)$ are weighted sum of periodic functions when $\log p_a$ for $a \in \mathcal{A}$ are *rationally related*, that is, $\log p_a$ are integer multiples of a real number; otherwise $\Delta_1(x)$ and $\Delta_2(x)$ converge to zero as $x \to \infty$ (see [4], [14] for details).

We prove the following theorem regarding the number of phrases M_n which improves our previous result from [4] by adding the convergence of moments.

Theorem 1. *Consider the LZ'78 algorithm over a sequence of length n generated by a memoryless source. The number of phrases M_n has mean $\mathbf{E}[M_n]$ and variance $\mathrm{Var}(M_n)$ satisfying*

$$\begin{aligned} \mathbf{E}(M_n) &= \ell^{-1}(n) + o(n^{1/2}/\log n) \qquad (5) \\ &= \frac{nh}{\log \ell^{-1}(n) + \beta(\ell^{-1}(n))} + o(n^{1/2}/\log n) \\ &\sim \frac{nh}{\log n}, \\ \mathrm{Var}(M_n) &\sim \frac{v(\ell^{-1}(n))}{(\ell'(\ell^{-1}(n)))^2} \sim \frac{(h_2 - h^2)n}{\log^2 n}. \qquad (6) \end{aligned}$$

Furthermore, the normalized number of phrases converges in distribution and moments to the the standard normal distribution $N(0,1)$. More precisely, for any given real x:

$$\lim_{n\to\infty} P(M_n < \mathbf{E}(M_n) + x\sqrt{\mathrm{Var}(M_n)}) = \Phi(x), \qquad (7)$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

In addition, for all nonnegative k

$$\lim_{n\to\infty} \mathbf{E}\left(\left(\frac{M_n - \mathbf{E}(M_n)}{\sqrt{\mathrm{Var}(M_n)}}\right)^k\right) = \mu_k \qquad (8)$$

where

$$\mu_k = \begin{cases} 0 & k \text{ odd} \\ \frac{k!}{2^{k/2}(\frac{k}{2})!} & k \text{ even} \end{cases} \qquad (9)$$

are the moments of $N(0,1)$.

We also have large and moderate deviations results for M_n. To the best of our knowledge these results are new (see also [4], [9]).

Theorem 2. *Consider the LZ'78 algorithm over a sequence of length n generated by a memoryless source.*

(i) [Large Deviations]. *For all $\frac{1}{2} < \delta < 1$ there exist $\varepsilon > 0$, $B > 0$ and $\beta > 0$ such that for all $y > 0$*

$$P(|M_n - \mathbf{E}(M_n)| > yn^\delta) \le A \exp\left(-\beta n^\varepsilon \frac{y}{(1 + n^{-\varepsilon} y)^\delta}\right)$$

for some $A > 0$.

(ii) [Moderate Deviation]. *There exists $B > 0$ such that*

$$P(|M_n - \mathbf{E}(M_n)| \ge x\sqrt{\mathrm{Var}(M_n)}) \le Be^{-\frac{x^2}{2}}$$

for all non-negative real $x < An^\delta$ with $\delta < \frac{1}{6}$.

Using these large deviations results, we shall conclude that the average compression rate converges to the entropy rate. Furthermore, our large deviation results allow us also to estimate the average redundancy

$$\mathbf{E}(r_n) = \frac{\mathbf{E}(C_n)}{n} - h$$

and its limiting distribution when $n \to \infty$. More precisely, in Section IV we prove the following.

Theorem 3. *The average compression rate converges to the entropy rate, that is,*

$$\lim_{n\to\infty} \frac{\mathbf{E}(C_n)}{n} = h \ . \qquad (10)$$

More precisely, for all $\frac{1}{2} < \delta < 1$

$$\begin{aligned} \mathbf{E}(C_n) &= \ell^{-1}(n)(\log \ell^{-1}(n) + \log|\mathcal{A}|) + O(n^\delta \log n) \\ &= \mathbf{E}(M_n)(\log \mathbf{E}(M_n) + \log|\mathcal{A}|) + o(n^{1/2+\varepsilon}), \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(C_n) &\sim \mathrm{Var}(M_n)(\log \mathbf{E}(M_n) + \log|\mathcal{A}| + 1)^2 \\ &\sim (h_2 - h^2)n. \end{aligned}$$

Furthermore,

$$\frac{C_n - \mathbf{E}(C_n)}{\sqrt{\mathrm{Var}(C_n)}} \xrightarrow{d} N(0,1)$$

and in moments, where $N(0,1)$ represents the standard normal distribution.

In order to establish the limiting distribution for the redundancy we also need the corresponding large deviation results for the code length C_n that we formulate next and prove in Section IV.

Theorem 4. *Consider the LZ'78 algorithm over a memoryless source.*

(i) [Large Deviations] *For all $\frac{1}{2} < \delta < 1$ there exist $\varepsilon > 0$, $B > 0$ and $\beta > 0$ such that for all $y > 0$*

$$P(|C_n - \mathbf{E}(C_n)| > yn^\delta \log(n/\log n)) \le \qquad (11)$$
$$\le A \exp\left(-\beta n^\varepsilon \frac{y}{(1 + n^{-\varepsilon} y)^\delta}\right)$$

for some $A > 0$.

(ii) [Moderate deviation] *There exists $B > 0$ such that for n*

$$P(|C_n - \mathbf{E}(C_n)| \ge x\sqrt{\mathrm{Var}(C_n)}) \le Be^{-\frac{x^2}{2}} \qquad (12)$$

for all non-negative real $x < An^\delta$ *with* $\delta < \frac{1}{6}$.

The next finding is a direct consequence of Theorems 3 and 4.

Corollary 1. *The redundancy rate* r_n *satisfies for all* $\frac{1}{2} < \delta < 1$*:*

$$\begin{aligned} \mathbf{E}(r_n) &= \frac{\mathbf{E}(C_n)}{n} - h \\ &= h\frac{\log(|\mathcal{A}|) - \beta(\ell^{-1}(n))}{\log \ell^{-1}(n) + \beta(\ell^{-1}(n))} + O(n^{\delta-1}\log n) \\ &\sim h\frac{\log(|\mathcal{A}|) - \beta\left(h\frac{n}{\log n}\right)}{\log n}, \end{aligned} \tag{13}$$

and

$$\mathrm{Var}(r_n) \sim \frac{(h_2 - h^2)}{n} \ .$$

Furthermore,

$$\frac{r_n - \mathbf{E}(r_n)}{\sqrt{\mathrm{Var}(r_n)}} \stackrel{d}{\to} N(0,1)$$

and the convergence also holds in moments.

In passing, we observe that the above Corollary follows from a trivial derivation:

$$\mathbf{E}(r_n) + O(n^{\delta-1}) = \frac{\ell^{-1}(n)(\log(\ell^{-1}(n)) + \log|\mathcal{A}|) - nh}{n}$$

$$= \frac{\ell^{-1}(n)\log|\mathcal{A}| + \ell^{-1}(n)\log \ell^{-1}(n) - h\ell(\ell^{-1}(n))}{n}$$

$$= \frac{\ell^{-1}(n)}{n}\left(\log|\mathcal{A}| - \beta(\ell^{-1}(n))\right)$$

where we use the fact that $\ell(\ell^{-1}(n)) = n$ by definition of the inverse function and

$$h\ell(\ell^{-1}(n)) = \ell^{-1}\log \ell^{-1}(n) + \beta(\ell^{-1}(n))$$

by the definition of $\ell(n)$.

The average redundancy estimate was first proved in [7], [12] but we provide here a new simplified proof. The variance and the limiting distribution of the redundancy are new. Notice that the estimate for the mean redundancy is smaller and more precise than previously obtained average estimates of $O(\frac{\log\log n}{\log n})$ obtained via probabilistic methods [11].

In Figure 2 we show a histograms of r_n for different values of n. We see that the mean of r_n decreases when n increases, in theory like $1/\log n$. However, the variance decreases much faster, like $1/n$.

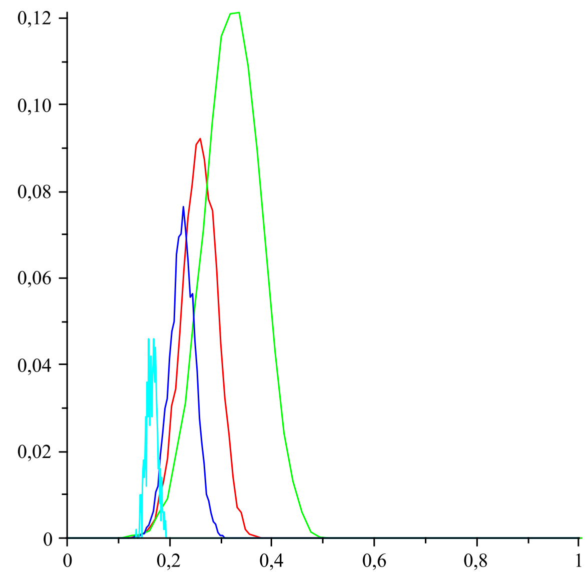

Fig. 2. Simulation of the redundancy distribution for $n = 400$ (green), $n = 1000$ (red), $n = 2000$ (blue) and $n = 10000$ (cyan) with $p_a = 0.9$

III. From Lempel-Ziv to Digital Search Tree

In this section we make a connection between the Lempel-Ziv algorithm and digital search trees using a renewal argument [2].

Our goal is to derive an estimate on the probability distribution of M_n. We assume that our original text is a prefix of an infinite sequence X generated by a memoryless source over the alphabet $\mathcal{A}$. We build a Digital Search Tree (DST) by parsing the infinite sequence X up to the mth phrase; see Figure 1 for an illustration. Thus the associated DST is constructed over m strings (phrases).

Let L_m be the total path length in the associated DST after inserting m (independent) strings. The quantity M_n is exactly the number of strings needed to be inserted to increase the path length of the associated DST to n. This observation leads to the following identity valid for all integers n and m:

$$P(M_n > m) = P(L_m < n) \ . \tag{14}$$

This so called *renewal equation* [2] allows us to study the number of phrases M_n through the path length L_m of the associated digital search tree built over m fixed independent strings.

We now use generating functions to find a functional equation for the distribution of L_m. Let $L_m(u) = \mathbf{E}(u^{L_m})$ be the moment generating function of L_m. In the following, $\mathbf{k}$ is a tuple $(k_1, \ldots k_{|\mathcal{A}|})$ where k_a for $a \in \mathcal{A}$ stands for the number of strings that starts with symbol a. Since inserted strings in DST are independent, we conclude that

$$L_{m+1}(u) = u^m \sum_{\mathbf{k}} \binom{m}{\mathbf{k}} \prod_{a\in\mathcal{A}} p_a^{k_a} L_{k_a}(u), \tag{15}$$

where

$$\binom{m}{\mathbf{k}} = \frac{m!}{\prod_{a\in\mathcal{A}} k_a!}.$$

Next, we introduce the exponential generating function

$$L(z,u) = \sum_m \frac{z^m}{m!} L_m(u)$$

leading to the following partial functional-differential equation

$$\frac{\partial}{\partial z} L(z,u) = \prod_{a\in\mathcal{A}} L(p_a uz, u)\ . \tag{16}$$

It is clear from the construction that $L(z,1) = e^z$, since $L_m(1) = 1$ for all integer m. Via the cumulant formula, we also know that for all integers m and for t complex (sufficiently small for which $\log(L_m(e^t)$ exists) we have

$$\log(L_m(e^t)) = t\mathbf{E}(L_m) + \frac{t^2}{2}\mathrm{Var}(L_m) + O(t^3)\ . \tag{17}$$

Notice that the term $O(t^3)$ is not uniform in m. In passing, we remark that $\mathbf{E}(L_m) = L'_m(1)$ and $\mathrm{Var}(L_m) = L''_m(1) + L'_m(1) - (L'_m(1))^2$.

In [4] we proved the following result (cf. Theorem 1 of [4]) that we adopt here.

Theorem 5. *Consider a digital search tree built over m independent strings. Then*

$$\begin{aligned} \mathbf{E}(L_m) &= \ell(m) + O(1), \\ \mathrm{Var}(L_m) &= v(m) + o(m) \end{aligned}$$

for large m.

We aim now at showing that the limiting distribution of the path length is normal for $m \to \infty$. In order to accomplish it, we need one important technical result proved in Section V.

Theorem 6. *For all $\delta > 0$ and for all $\delta' < \delta$ there exists $\varepsilon > 0$ such that $\log L_m(e^{tm^{-\delta}})$ exists for $|t| \le \varepsilon$, and*

$$\log L_m(e^{tm^{-\delta}}) = O(m), \tag{18}$$

$$\begin{aligned} \log L_m(e^{tm^{-\delta}}) &= \frac{t}{m^\delta}\mathbf{E}(L_m) \\ &+ \frac{t^2}{2m^{2\delta}}\mathrm{Var}(L_m) + t^3 O(m^{1-3\delta'}) \end{aligned} \tag{19}$$

for large m.

Provided Theorem 6 is granted, we are ready to prove our main results concerning the path length L_m.

Theorem 7. *Consider a digital search tree built over m sequences generated by a memoryless source. Then*

$$\frac{L_m - \mathbf{E}(L_m)}{\sqrt{\mathrm{Var} L_m}} \to N(0,1)$$

in probability and in moments. More precisely, for any given real number x:

$$\lim_{m\to\infty} P(L_m < \mathbf{E}(L_m) + x\sqrt{\mathrm{Var}(L_m)}) = \Phi(x), \tag{20}$$

and for all nonnegative integer k and $\varepsilon > 0$

$$\mathbf{E}\left(\left(\frac{L_m - \mathbf{E}(L_m)}{\sqrt{\mathrm{Var} L_m}}\right)^k\right) = \mu_k + O(m^{-\frac{1}{2}+\varepsilon}) \tag{21}$$

where μ_k are centralized moments of the normal distribution given by (9).

Proof: We apply Levy's continuity theorem or equivalently Goncharov's result [14] asserting that $\frac{L_m - \mathbf{E}(L_m)}{\sqrt{\mathrm{Var} L_m}}$ tends to the standard normal distribution if for complex τ

$$L_m\left(\exp\left(\frac{\tau}{\sqrt{\mathrm{Var}(L_m)}}\right)\right) e^{-\tau E(L_m)/\sqrt{\mathrm{Var}(L_m)}} \to e^{\tau^2/2}. \tag{22}$$

To prove it we apply several times our main technical result Theorem 6 with

$$t = \frac{\tau m^\delta}{\sqrt{\mathrm{Var} L_m}} = O(m^{-1/2-\varepsilon+\delta}) \to 0$$

where $\delta < 1/2$ and $\varepsilon > 0$. Thus by Theorem 6 we find

$$\begin{aligned} \log L_m\left(\exp\left(\frac{\tau}{\sqrt{\mathrm{Var}(L_m)}}\right)\right) &= \frac{\tau E[L_m]}{\sqrt{\mathrm{Var} L_m}} \\ &+ \frac{\tau^2}{2} + O(m^{-\frac{1}{2}+\varepsilon'}) \end{aligned} \tag{23}$$

for some $\varepsilon' > 0$. Thus by (22) the normality result follows.

To establish the convergence in moments, we use (23) in the Cauchy formula applied on a circle of radius R encircling the origin, that is,

$$\mathbf{E}\left(\left(\frac{L_m - \mathbf{E}(L_m)}{\sqrt{\mathrm{Var} L_m}}\right)^k\right) =$$

$$= \frac{1}{2i\pi}\oint \frac{d\tau}{\tau^{k+1}} L_m\left(\exp\left(\frac{\tau}{\sqrt{\mathrm{Var}(L_m)}}\right)\right) e^{-\tau\mathbf{E}(L_m)/\sqrt{\mathrm{Var}(L_m)}}$$

$$= \frac{1}{2i\pi}\oint \frac{d\tau}{\tau^{k+1}} \exp\left(\frac{\tau^2}{2}\right)(1 + O(m^{-\frac{1}{2}+\varepsilon'})$$

$$= \mu_k + O\left(R^{-k}\exp(R^2/2)m^{-\frac{1}{2}+\varepsilon'}\right).$$

This completes the proof. ■

We also have some large deviation results for the path length presented next.

Theorem 8. *Consider a digital search tree built over m sequences generated by a memoryless source.*

(i) [Large deviation]. *Let $\frac{1}{2} < \delta < 1$. Then there exist $\varepsilon > 0$, $B > 0$, and $\beta > 0$ such that for all $x \geq 0$:*

$$P(|L_m - \mathbf{E}(L_m)| > xm^\delta) \leq B\exp(-\beta m^\varepsilon x). \quad (24)$$

(ii) [Moderate deviation]. *There exists $B > 0$ such that*

$$P(|L_m - \mathbf{E}(L_m)| \geq x\sqrt{\mathrm{Var}(L_m)}) \leq Be^{-\frac{x^2}{2}} \quad (25)$$

for non-negative real $x < Am^\delta$ with $\delta < 1/6$ and $A > 0$.

Proof: We apply the Chernov bound. Let $t > 0$ be a non-negative real number. We have the identity

$$P\left(L_m > \mathbf{E}(L_m) + xm^\delta\right) = P\left(e^{tL_m} > e^{(\mathbf{E}(L_m)+xm^\delta)t}\right).$$

Using Markov's inequality we find

$$P(e^{tL_m} > e^{(\mathbf{E}(L_m)+xm^\delta)t}) \leq \frac{\mathbf{E}(e^{tL_m})}{e^{(\mathbf{E}(L_m)+xm^\delta)t}}$$
$$= L_m(e^t)\exp(-t\mathbf{E}(L_m) - xm^\delta t)\ .$$

Here we take

$$\delta' = \frac{\delta + 1/2}{2} > \frac{1}{2}, \qquad \varepsilon = \delta' - \frac{1}{2} > 0$$

since $\delta > 1/2$. We now apply Theorem 6 with t' and δ' and we set $t = t'm^{-\delta'}$ to obtain

$$\log L_m(e^t) = tE(L_m) + O(t^2\mathrm{Var}(L_m)).$$

By Theorem 5 we conclude

$$\log L_m(e^t) - t\mathbf{E}[L_m] = O(m^{-\varepsilon}). \quad (26)$$

We complete the lower bound by setting $tm^\delta = t'm^\varepsilon$ with $\beta = t'$.

To obtain an upper bound we follow the same route only considering $-t$ instead of t and using

$$P\left(L_m < \mathbf{E}(L_m) - xm^\delta\right)$$
$$= P\left(e^{-tL_m} > e^{-(\mathbf{E}(L_m)-xm^\delta)t}\right)$$
$$\leq L_m(e^{-t})\exp(t\mathbf{E}(L_m) - xm^\delta t).$$

To prove part (ii) of **moderate deviation**, we apply again Theorem 6 with

$$t = \frac{xm^{\delta'}}{\sqrt{\mathrm{Var}L_m}} \quad (27)$$

where $\delta < \delta' < \frac{1}{6}$. Then by Theorem 6 (with formally δ and δ' interchanged)

$$\log L_m\left(\exp(\frac{x}{\sqrt{\mathrm{Var}(L_m)}})\right) =$$
$$= \mathbf{E}(L_m)\frac{x}{\sqrt{\mathrm{Var}(L_m)}} + \frac{x^2}{2\mathrm{Var}(L_m)}\mathrm{Var}(L_m)$$
$$+ \frac{x^3m^{3\delta'}}{(\mathrm{Var}(L_m))^{\frac{3}{2}}}O(m^{1-3\delta}).$$

Observe that the error term for $x = O(m^\delta)$ is

$$O(m^{-\frac{1}{2}+3\delta'}(\log m)^{-3/2}) = o(1)$$

since $\delta' < 1/6$ leading to

$$\log L_m\left(\exp\left(\frac{x}{\sqrt{\mathrm{Var}(L_m)}}\right)\right) - \mathbf{E}(L_m)\frac{x}{\sqrt{\mathrm{Var}(L_m)}}$$
$$= \frac{x^2}{2} + o(1). \quad (28)$$

Therefore, by Markov inequality for all $t > 0$,

$$P(L_m > \mathbf{E}(L_m) + x\sqrt{\mathrm{Var}(L_m)}) \leq$$
$$\leq \exp(\log L_m(e^t) - t\mathbf{E}(L_m) - xt\sqrt{\mathrm{Var}(L_m)}).$$

Using (28) we find

$$P(L_m > \mathbf{E}(L_m) + x\sqrt{\mathrm{Var}(L_m)}) \leq$$
$$\exp(\log L_m(e^t) - t\mathbf{E}(L_m) - xt\sqrt{\mathrm{Var}(L_m)})$$
$$= \exp\left(\frac{x^2}{2} + o(1) - x^2\right) \sim \exp(-\frac{x^2}{2})$$

where we set $t = x/\sqrt{\mathrm{Var}(L_m)}$ in the last line. This completes the proof of the lower bound, while for the upper bound we follow the footsteps with t replaced by $-t$. ■

IV. Proofs of Main Theorems

In this section we prove our main results, namely Theorems 1 and 2 as well as Theorem 3 and 4. We start with the large deviation results.

A. Proof of Theorem 2

We start with Theorem 2(i). By (14) we have

$$P(M_n > \ell^{-1}(n) + yn^\delta) \;=\; P(M_n > \lfloor \ell^{-1}(n) + yn^\delta \rfloor)$$
$$= P(L_{\lfloor \ell^{-1}(n)+yn^\delta \rfloor} < n)\ .$$

Observe that $\mathbf{E}(L_m) = \ell(m) + O(1)$, hence

$$\mathbf{E}(L_{\lfloor \ell^{-1}(n)+yn^\delta \rfloor}) = \ell(\ell^{-1}(n) + yn^\delta) + O(1). \quad (29)$$

Since the function $\ell(\cdot)$ is convex and $\ell(0) = 0$, we have for all real numbers $a > 0$ and $b > 0$

$$\ell(a+b) \geq \ell(a) + \frac{\ell(a)}{a}b, \tag{30}$$

$$\ell(a-b) \leq \ell(a) - \frac{\ell(a)}{a}b. \tag{31}$$

Applying inequality (30) to $a = \ell^{-1}(n)$ and $b = yn^\delta$ we arrive at

$$n - \mathbf{E}(L_{\lfloor \ell^{-1}(n)+yn^\delta \rfloor}) \leq -y\frac{n}{\ell^{-1}(n)}n^\delta + O(1). \tag{32}$$

Thus

$$P(L_{\lfloor \ell^{-1}(n)+yn^\delta \rfloor} < n) \leq P(L_m - \mathbf{E}(L_m) < -xm^\delta + O(1))$$

by identifying

$$m = \lfloor \ell^{-1}(n) + yn^\delta \rfloor, \qquad x = \frac{n}{\ell^{-1}(n)}\frac{n^\delta}{m^\delta}y. \tag{33}$$

We now apply several times Theorem 8 from the previous section regarding the path length L_m. That is, for all $x > 0$ and for all m, there exist $\varepsilon > 0$ and A such that

$$P(L_m - \mathbf{E}(L_m) < xm^\delta) < Ae^{-\beta x m^\varepsilon} \ . \tag{34}$$

In other words,

$$P(L_m - \mathbf{E}(L_m) < xm^\delta + O(1)) \leq Ae^{-\beta x m^\varepsilon + O(m^{\varepsilon-\delta})} \leq A'e^{-\beta x m^\varepsilon}$$

for some $A' > A$ we find

$$P(M_n > \ell^{-1}(n) + yn^\delta) \leq A' \exp(-\beta x m^\varepsilon). \tag{35}$$

We know that $\ell^{-1}(n) = \Omega(\frac{n}{\log n})$. Thus with x defined as (33) we have

$$x = O((\log n)^{1+\delta})\frac{y}{(1+yn^{\delta-1}\log n)^\delta} \leq \beta' \frac{n^{\varepsilon_1}y}{(1+yn^{-\varepsilon_2})^\delta}$$

for some $\beta' > 0$. Setting $\varepsilon_1 < \varepsilon$ and $\varepsilon_2 < \varepsilon$ for some $\varepsilon > 0$ we establish the upper bound.

In a similar fashion, we have

$$P(M_n < \ell^{-1}(n) - yn^\delta) = P(L_{\lfloor \ell^{-1}(n)-yn^\delta \rfloor} > n) \tag{36}$$

and

$$\mathbf{E}(L_{\lfloor \ell^{-1}(n)-yn^\delta \rfloor}) = \ell(\ell^{-1}(n) - yn^\delta) + O(1). \tag{37}$$

Using inequality (31) we obtain

$$n - \mathbf{E}(L_{\lfloor \ell^{-1}(n)-yn^\delta \rfloor}) \geq y\frac{n}{\ell^{-1}(n)}n^\delta + O(1). \tag{38}$$

In conclusion,

$$P(L_{\lfloor \ell^{-1}(n)-yn^\delta \rfloor} > n) \leq P(L_m - \mathbf{E}(L_m) > xm^\delta + O(1))$$

by identifying

$$m = \lfloor \ell^{-1}(n) - yn^\delta \rfloor, \qquad x = \frac{n}{\ell^{-1}(n)}\frac{n^\delta}{m^\delta}y.$$

Observe that this case is easier since we have now $m < \ell^{-1}(n)$ and we don't need the correcting term $(1 + yn^\varepsilon)^{-\delta}$.

Now we can turn our attention to **moderate deviation** expressed in Theorem 2(ii) (with $\delta < 1/6$). It is essentially the same proof except that we consider

$$y\frac{s_n}{\ell'(\ell^{-1}(n))} \quad \text{with} \quad s_n = \sqrt{v(\ell^{-1}(n))}$$

instead of yn^δ, and we assume $y = O(n^{\delta'})$ for some $\delta' < \frac{1}{6}$. Thus

$$y\frac{s_n}{\ell'(\ell^{-1}(n))} = O(n^{\frac{1}{2}+\varepsilon}) = o(n)$$

for some $\varepsilon > 0$. By Theorem 1 we know that

$$\mathbf{E}(M_n) = \ell^{-1}(n) + o(s_n/\ell'(\ell^{-1}(n)),$$

thus for $y > 0$

$$P(M_n > \ell^{-1}(n) + y\frac{s_n}{\ell'(\ell^{-1}(n))}) = P(L_m < n) \tag{39}$$

with $m = \lfloor \ell^{-1}(n) + y\frac{s_n}{\ell'(\ell^{-1}(n))} \rfloor$. We use the estimate

$$\ell(a+b) = \ell(a) + \ell'(a)b + o(1)$$

when $b = o(a)$ and $a \to \infty$. Thus

$$\ell\left(\ell^{-1}(n) + y\frac{s_n}{\ell'(\ell^{-1}(n))}\right) = n + ys_n + o(1) \ . \tag{40}$$

Since $\sqrt{v(m)} = s_n + O(1)$ we have

$$n = \mathbf{E}(L_m) - y\sqrt{v(m)} + O(1).$$

By Theorem 8 we know that

$$P(L_m < \mathbf{E}(L_m) - y\sqrt{v(m)} + O(1)) \leq A\exp(-y^2/2),$$

where the term $O(1)$ inducing a term

$$\exp\left(O\left(\frac{y^2}{v(m)}\right)\right) = \exp(o(1))$$

which is absorbed in A since $\delta < 1/6$. The proof for $y < 0$ follows a similar path.

B. Proof of Theorem 1

We first show that for all $\frac{1}{2} < \delta < 1$

$$\mathbf{E}(M_n) = \ell^{-1}(n) + O(n^\delta).$$

Indeed, noticing that for any random variable X

$$|\mathbf{E}(X)| \le \mathbf{E}(|X|) = \int_0^\infty P(|X| > y)dy,$$

we set $X = M_n - \ell^{-1}(n)$ to find from Theorem 2(i)

$$|\mathbf{E}(M_n) - \ell^{-1}(n)| \le$$
$$\le n^\delta + n^\delta \int_1^\infty P(|M_n - \ell^{-1}(n)| > yn^\delta)dy = O(n^\delta).$$

By the renewal equation (14), for a given y we have

$$P(M_n > \ell^{-1}(n) + y\frac{s_n}{\ell'(\ell^{-1}(n))}) =$$
$$P(L_{\lfloor \ell^{-1}(n)+y\frac{s_n}{\ell'(\ell^{-1}(n))}\rfloor} < n).$$

Let $m = \lfloor \ell^{-1}(n) + y\frac{s_n}{\ell'(\ell^{-1}(n))}\rfloor$. We know that

$$n - \mathbf{E}(L_m) = -ys_n + O(1)$$

and

$$s_n = \sqrt{v(\ell^{-1}(n))} = \sqrt{\mathrm{Var}(L_m)}(1 + o(1)).$$

Therefore

$$P\left(M_n > \ell^{-1}(n) + y\frac{s_n}{\ell'(\ell^{-1}(n))}\right) =$$
$$P\left(L_m < \mathbf{E}(L_m) + y\sqrt{\mathrm{Var}(L_m)}(1 + +o(1))\right).$$

Hence

$$P\left(L_m < \mathbf{E}(L_m) + y(1 + o(1))\sqrt{\mathrm{Var}(L_m)}\right) \ge$$
$$P\left(M_n > \ell^{-1} + y\frac{s_n}{\ell'(\ell^{-1}(n))}\right)$$

since for all y' we have

$$\lim_{m\to\infty} P\left(L_m < \mathbf{E}(L_m) + y'\sqrt{\mathrm{Var}(L_m)}\right) = \Phi(y'),$$

and therefore by continuity of $\Phi(x)$

$$\lim_{m\to\infty} P\left(L_m < \mathbf{E}(L_m) + y(1 \pm o(1))\sqrt{\mathrm{Var}(L_m)}\right)$$
$$= \Phi(y)\ .$$

Thus

$$\lim_{m\to\infty} P\left(M_n > \ell^{-1}(n) + y\frac{s_n}{\ell'(\ell^{-1}(n))}\right) = 1 - \Phi(y)$$

and following the same footsteps we also establish the matching lower bound

$$\lim_{m\to\infty} P(M_n < \ell^{-1}(n) - y\frac{s_n}{\ell'(\ell^{-1}(n))}) = \Phi(y).$$

This proves two things: first that

$$(M_n - \ell^{-1}(n))\frac{\ell'(\ell(^{-1}(n))}{s_n}$$

tends to the normal distribution in probability. Second, since by the moderate deviation result the normalized random variable

$$(M_n - \ell^{-1}(n))\frac{\ell'(\ell(^{-1}(n))}{s_n}$$

has bounded moments, and then by the virtue of the dominated convergence and the convergence to the normal distribution

$$\lim_{n\to\infty} \mathbf{E}\left((M_n - \ell^{-1}(n))\frac{\ell'(\ell^{-1}((n))}{s_n}\right) = 0,$$
$$\lim_{n\to\infty} \mathbf{E}\left(\left((M_n - \ell^{-1}(n))\frac{\ell'(\ell(^{-1}(n))}{s_n}\right)^2\right) = 1.$$

In other words,

$$\mathbf{E}(M_n) = \ell^{-1}(n) + o(n^{1/2}\log n)$$
$$\mathrm{Var}(M_n) \sim \frac{v(\ell^{-1}(n))}{(\ell'(\ell^{-1}(n)))^2}.$$

which prove (5) and (6). This completes the proof of our main result Theorem 1.

C. Proof of Theorem 3

In this section we prove the central limit theorem and large deviation for the code length

$$C_n = M_n\left(\log M_n + \log(|\mathcal{A}|)\right)$$

as presented in (3).

Let now $\mu_n = \mathbf{E}(M_n)$ and $\sigma_n = \sqrt{\mathrm{Var}(M_n)}$. Define also

$$g(y) = y(\log|\mathcal{A}| + \log y), \quad g(M_n) = C_n.$$

Clearly, for any fixed $x > 0$

$$P(M_n \ge \mu_n + x\sigma_n) = P(g(M_n) \ge g(\mu_n + x\sigma_n)). \quad (41)$$

By Taylor's expansion, since $g''(y) = 1/y$, we have

$$g(\mu_n + x\sigma_n) = g(\mu_n) + x\sigma_n g'(\mu_n) + O\left(\frac{(x\sigma_n)^2}{\mu_n}\right) \quad (42)$$

where

$$g'(y) = 1 + \log|\mathcal{A}| + \log y.$$

By Theorem 1

$$\lim_{n\to\infty} P(M_n \ge \mu_n + x\sigma_n) = \Phi(x)\ ,$$

and then by (41)-(42) we find

$$P(C_n \ge g(\mu_n + x\sigma_n)) =$$

$$P\left(M_n \geq \mu_n + \sigma_n\left(x + O\left(\frac{x^2\sigma_n}{\mu_n g'(\mu_n)}\right)\right)\right)$$

which converges to $\Phi(x)$ since

$$\Phi\left(x + O\left(\frac{x^2\sigma_n}{\mu_n g'(\mu_n)}\right)\right) = \Phi(x) + O(n^{-1/2+\varepsilon}).$$

Via similar analysis we find

$$\lim_{n\to\infty} P\left(C_n \leq \mu_n(\log\mu_n + \log|\mathcal{A}|)-\right.$$

$$\left.-x\sigma_n(\log\mu_n + \log|\mathcal{A}| + 1)\right) = \Phi(x).$$

In other words the random variable

$$\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}$$

tends to the normal distribution in probability.

In order to conclude the convergence in moment, we use the moderate deviation result. Observe that by Theorem 4 proved next the normalized random variable satisfies

$$P\left(\left|\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}\right| \geq x\right) \leq Be^{-x^2/2}$$

for $x = O(n^\delta)$ when $\delta < \frac{1}{6}$. Thus

$$\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}$$

has bounded moments. Indeed, we have for $x = n^\varepsilon$ for $\varepsilon < 1/6$.

$$P\left(\left|\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}\right| \geq n^\varepsilon\right) \leq Be^{-n^{2\varepsilon}/2}\ .$$

Since $\sqrt{2n} \leq M_n \leq n$, we conclude that

$$g(\sqrt{n}) \leq g(n) \leq g(n) = O(n\log n).$$

Therefore, for all integer k:

$$\mathbf{E}\left(\left|\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}\right|^k\right) =$$

$$2k\int_0^\infty x^{k-1}P\left(\left|\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}\right| \geq x\right)dx$$

$$\leq 2k\int_0^{n^\varepsilon} x^{k-1}e^{-Bx^2/2}dx + O(n^k\log^k n)e^{-n^{2\varepsilon}/2} = O(1)\ .$$

In summary, the random variable

$$\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)} = \frac{C_n - \mu_n(\log\mu_n + \log|\mathcal{A}|))}{\sigma_n(\log\mu_n + \log|\mathcal{A}| + 1)}$$

has bounded moments. Therefore, by virtue of the dominated convergence and the convergence to the normal distribution

$$\lim_{n\to\infty}\mathbf{E}\left(\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}\right) = 0$$

$$\lim_{n\to\infty}\mathbf{E}\left(\frac{C_n - g(\mu_n)}{\sigma_n g'(\mu_n)}\right)^2 = 1.$$

In other words, for some $\varepsilon > 0$

$$\mathbf{E}(C_n) = g(\mu_n) + o(n^{-1/2+\varepsilon}) \quad (43)$$

$$\mathrm{Var}(C_n) \sim \mathrm{Var}(M_n)(\log\mu_n + \log|\mathcal{A}| + 1)^2 \quad (44)$$

which proves our variance estimate. ■

D. Proof of Theorem 4

We start with the moderate deviation results. We already know that for $x \leq An^{1/6}$ for some $A, B > 0$ we have

$$P(M_n \geq \mu_n + x\sigma_n) \leq Be^{-x^2/2}.$$

As in the previous section, we use $g(x) = x(\log|\mathcal{A}| + \log x)$ and note that $g(M_n) = C_n$. We also have

$$P(g(M_n) \geq g(\mu_n + x\sigma_n)) \leq Be^{-x^2/2}.$$

Since

$$g(\mu_n + x\sigma_n) = g(\mu_n) + g'(\mu_n)\sigma_n x\left(1 + O\left(\frac{x\sigma_n}{\mu_n g'(\mu_n)}\right)\right)$$

we arrive at

$$P\left(g(M_n) \geq g(\mu_n) + x\sigma_n g'(\mu_n)\left(1 + O(x\frac{\sigma_n}{g'(\mu_n)\mu_n}\right)\right)$$

$$\leq Be^{-x^2/2}.$$

But for $x = O(n^{1/6})$

$$x\frac{\sigma_n}{g'(\mu_n)\mu_n} \leq An^{1/6}\frac{\sigma_n}{g'(\mu_n)\mu_n} = O(n^{-1/3}) \to 0,$$

hence

$$P\left(g(M_n) \geq g(\mu_n) + x\sigma_n g'(\mu_n)\right) =$$

$$P(g(M_n) \geq g(\mu_n + \sigma_n x(1 + O(n^{-1/3})) \leq$$

$$Be^{-x^2/2(1+O(n^{-1/3}))} \leq B'e^{-x^2/2}$$

for some $B' > 0$. Therefore, from (43)-(44) we conclude

$$P(C_n > \mathbf{E}(C_n) + x\sqrt{\mathrm{Var}(C_n)}) \leq B''e^{-x^2/2}$$

and similarly

$$P(C_n < \mathbf{E}(C_n) - x\sqrt{\mathrm{Var}(C_n)}) \leq B''e^{-x^2/2}$$

for some $B'' > B'$, where B'' absorbs asymptotics of (43) and (44).

We now turn our attention to the large deviation result. We also have the fact that $M_n \leq n$. If $\mu_n \leq z \leq n$, we find

$$g(z) \leq g(\mu_n) + (z - \mu_n)g'(n) \ .$$

Thus for y such that $\mu_n + yn^\delta \leq n$:

$$\begin{aligned} g(M_n) \geq g(\mu_n) + yn^\delta g'(n) &\Rightarrow g(M_n) \geq g(\mu_n + yn^\delta) \\ &\Rightarrow M_n \geq \mu_n + yn^\delta. \end{aligned}$$

Thus by Theorem 2(i) we obtain

$$P(g(M_n) \geq g(\mu_n) + yn^\delta g'(n)) \leq A \exp\left(-\beta n^\varepsilon \frac{y}{(1 + n^{-\varepsilon} y)^\delta}\right).$$

The result also holds for $\mu_n + yn^\delta > n$ since $P(M_n > n) = 0$. On the other hand, for $\sqrt{2n} \leq v \leq \mu_n$ we have $g(v) \geq g(\mu_n) + (v - \mu_n)g'(\mu_n)$. Thus:

$$\begin{aligned} g(M_n) \leq g(\mu_n) - yn^\delta g'(\mu_n) &\Rightarrow g(M_n) \leq g(\mu_n - yn^\delta) \\ &\Rightarrow M_n \geq \mu_n - yn^\delta. \end{aligned}$$

By Theorem 2(i)

$$P(g(M_n) \geq g(\mu_n) - yn^\delta g'(\mu_n)) \leq A \exp\left(-\beta n^\varepsilon \frac{y}{(1 + n^{-\varepsilon} y)^\delta}\right).$$

Since $g'(\mu_n)$ and $g'(n)$ are both $O(\log n)$, the order n^δ is not changed in the large deviation result. ■

V. Technical Results

The main goal of this section is to prove Theorem 6. We accomplish it through several technical results using throughout analytic depoissonization [5].

A. Auxiliary Results

First, we will work with the Poisson model, that is, the exponential generating function $L(z, u)$ satisfying (16) (or its Poisson version $L(z, u)e^{-z}$) from which we extract information about the probability generating function $L_m(u)$ for large z and u in the vicinity of $u = 1$. Throughout we use analytic depoissonization of [5].

Define

$$\begin{aligned} X(z) &= \frac{\partial}{\partial u} L(z, 1), \qquad \widetilde{X}(z) = X(z)e^{-z}, \\ \widetilde{V}(z) &= e^{-z} \frac{\partial^2}{\partial u^2} L(z, 1) + \widetilde{X}(z) - (\widetilde{X}(z))^2, \end{aligned}$$

that is, the Poisson mean $\widetilde{X}(z)$ and the Poisson variance $\widetilde{V}(z)$. For a given fixed z, for any t, we then have

$$\log L(z, e^t) = z + \widetilde{X}(z)t + \widetilde{V}(z)\frac{t^2}{2} + O(t^3). \tag{45}$$

We first obtain some estimates on the Poisson mean $\widetilde{X}(z)$ and Poisson variance $\widetilde{V}(z)$ (cf. Lemma 1) by by applying Theorem 10 below that we prove in the last subsection. Then we derive some estimates on the derivative of $\log L(z, e^t)$ (cf. Lemma 2). Finally, we use depoissonization tool reviewed below in Theorem 9 to prove Theorem 6.

The main tool of this section is analytic depoissonization that we review next. To this end we will use the *diagonal exponential depoissonization* established in [5]. Let θ be a non-negative number smaller than $\frac{\pi}{2}$, and $\mathcal{C}(\theta)$ be complex cone around the positive real axis defined as $\mathcal{C}(\theta) = \{z : \ \arg(z)| \leq \theta\}$. We will use of the following theorem from [5] (cf. Theorem 8) known as the diagonal exponential depoissonization tool.

Theorem 9 (Jacquet and Szpankowski, 1998)**.** *Let u_k be a sequence of complex number, and $\theta \in]0, \frac{\pi}{2}[$. For all $\varepsilon > 0$ there exist $c > 1$, $\alpha < 1$, $A > 0$ and $B > 0$ such that:*

$$\begin{aligned} z \in \mathcal{C}(\theta) \ \& \ |z| \in [\frac{m}{c}, cm] &\Rightarrow |\log(L(z, u_m)| \leq B|z|, \\ z \notin \mathcal{C}(\theta) \ , \quad |z| = m &\Rightarrow |L(z, u_m)| \leq Ae^{\alpha m}. \end{aligned}$$

Then

$$L_m(u_m) = L(m, u_m)(1 + o(m^{-\frac{1}{2}+\varepsilon}))\cdot$$

$$\cdot \exp\left(-m - \frac{m}{2}\left(\frac{\partial}{\partial z}\log(L(m, u_m)) - 1\right)^2\right) \tag{46}$$

for $m \to \infty$.

In Theorems 12 and 13 of Section V-C we prove the following main technical result needed to establish Theorem 6.

Theorem 10. *Let $\delta \in]0, 1[$. There exist numbers $\theta \in]0, \frac{\pi}{2}[$, $\alpha < 1$, $A > 0$, $B > 0$ and $\varepsilon > 0$ such that for all complex t such $|t| \leq \varepsilon$:*

$$z \in \mathcal{C}(\theta) \Rightarrow |\log(L(z, e^{t|z|^{-\delta}}))| \leq B|z| \tag{47}$$

$$z \notin \mathcal{C}(\theta) \Rightarrow |L(z, e^{t|z|^{-\delta}})| \leq Ae^{\alpha|z|} \tag{48}$$

Granted Theorem 10, we now proceed to estimate the Poisson mean and variance that are further used to prove Theorem 11 – main technical result of this section – in which we obtain an estimate on $\log L_m(e^t)$.

We start with some bounds on the Poisson mean, variance, and $\log L(z, e^t)$.

Lemma 1. *Let δ be an arbitrary non negative number. There exists $\varepsilon > 0$ such that for $|t| \leq \varepsilon$ and $z \in \mathcal{C}(\theta)$*

the following estimates hold

$$\begin{aligned}\widetilde{X}(z) &= O(|z|^{1+\delta}),\\ \widetilde{V}(z) &= O(|z|^{1+2\delta}),\\ \log L(z,e^{t|z|^{-\delta}}) &= z+\widetilde{X}(z)\frac{t}{|z|^{\delta}}+\widetilde{V}(z)\frac{t^2}{2|z|^{2\delta}}\\ &\quad +O(t^3|z|^{1+3\delta})\ .\end{aligned}$$

Proof: We first notice that $\log L(z,1)=z$. We recall that $\widetilde{X}(z)$ and $\widetilde{V}(z)$ are respectively the first and second derivative of $L(z,e^t)$ with respect to t at $t=0$. By Cauchy's formula [5], [14]:

$$\widetilde{X}(z) = \frac{1}{2i\pi}\oint \log L(z,e^t)\frac{dt}{t^2}, \tag{49}$$

$$\widetilde{V}(z) = \frac{2}{2i\pi}\oint \log L(z,e^t)\frac{dt}{t^3}, \tag{50}$$

where the integrals are along the circle of center 0 and radius $\varepsilon|z|^{-\delta}$. On this integral loop the estimate $|\log L(z,e^t)|\leq B|z|$ holds, and therefore we have

$$|\widetilde{X}(z)| \leq \frac{B}{\varepsilon}|z|^{1+\delta}, \tag{51}$$

$$|\widetilde{V}(z)| \leq \frac{2B}{\varepsilon^2}|z|^{1+2\delta}, \tag{52}$$

which proves the first two assertions. For the third one, we need to assess the reminder

$$R(z,t)=\log L(z,e^t)-z-\widetilde{X}(z)t-\widetilde{V}(z)\frac{t^2}{2}.$$

We again use the Cauchy formula

$$R(z,t)=\frac{2t^3}{2i\pi}\oint \log L(z,e^{t'})\frac{dt'}{(t')^3(t'-t)}. \tag{53}$$

The above follows by noting that

$$\begin{aligned}R(z,t) &= \frac{1}{2i\pi}\oint \log L(z,e^{t'})\frac{dt'}{(t'-t)}\\ R(z,0) &= \frac{1}{2i\pi}\oint \log L(z,e^{t})\frac{dt}{t}\\ R'(z,0) &= \frac{1}{2i\pi}\oint \log L(z,e^{t})\frac{dt}{t^2}\\ R''(z,0) &= \frac{1}{2i\pi}\oint \log L(z,e^{t})\frac{dt}{t^3}.\end{aligned}$$

We integrate $R(z,t)$ around the circle of center 0 and radius $\varepsilon|z|^{-\delta}$. If we restrict $|t|\leq\frac{\varepsilon}{2}|z|^{-\delta}$, then $|t-t'|\geq\frac{\varepsilon}{2}|z|^{-\delta}$, and

$$|R(z,t)|\leq\frac{8B}{\varepsilon^3}|t|^3|z|^{1+3\delta}$$

which completes the proof. ∎

Let now

$$D(z,t):=\frac{\partial}{\partial z}\log L(z,e^t)$$

which is needed in (46) to apply the diagonal depoissonization. Our second technical lemma provides estimates on $D(z,t)$.

Lemma 2. *Let $\delta>0$. There exist $\varepsilon>0$ and $B'>0$ such that for all t such $|t|<\varepsilon$*

$$\begin{aligned}|D(m,tm^{-\delta)})| &\leq B',\\ D(m,tm^{-\delta}) &= 1+\widetilde{X}'(m)\frac{t}{m^{\delta}}+O(t^2m^{2\delta}),\\ \widetilde{X}'(m) &= O(m^{\delta})\end{aligned}$$

for $m\to\infty$.

Proof: The key point here is to show that $D(m,tm^{-\delta)})=O(1)$. In order to establish it, we again use the Cauchy formula:

$$D(m,tm^{-\delta})=\frac{1}{2i\pi}\oint \log L(z,e^{tm^{-\delta}})\frac{dz}{(z-m)^2}, \tag{54}$$

where the integration loop encircles m within a radius $O(m)$ and is included in the cone $\mathcal{C}(\theta)$. Let $|t|\leq\varepsilon$ so that (48) holds, namely $|\log L(z,e^{tm^{-\delta}})|\leq B|z|$ since $m=O(|z|)$. To this end the loop is chosen to be a circle of center m and radius $m\sin(\theta)$. Noticing that $|z|<m(1+\sin(\theta)$ we finally arrive at

$$|D(m,tm^{-\delta})|\leq B\frac{1+\sin(\theta)}{\sin(\theta)}\ . \tag{55}$$

From here the proof takes a similar path as in the previous lemma. By noticing that $D(m,0)=1$ we have

$$\begin{aligned}\widetilde{X}'(m)=\frac{\partial}{\partial t}D(m,0) &= \frac{1}{2i\pi}\oint D(m,t')\frac{dt'}{(t')^2}\\ D(m,tm^{-\delta}) &= 1+\widetilde{X}'(m)tm^{-\delta}+\\ &+\frac{tm^{-\delta}}{2i\pi}\oint D(m,t')\frac{dt'}{(t')^2(t'-tm^{-\delta})},\end{aligned}$$

where the integral loop is now the circle of center $tm^{-\delta}$ and radius $\varepsilon m^{-\delta}$. ∎

These two lemmas allow us to establish the following intermediate result.

Theorem 11. *There exists a number $A>0$ such that for all arbitrarily small $\delta'<\delta$, and for all complex t such that $|t|\leq A$, we have*

$$\begin{aligned}\log L_m(e^{tm^{-\delta}}) &= \widetilde{X}(m)\frac{t}{m^{\delta}}\\ &+ \left(\widetilde{V}(m)-m(\widetilde{X}'(m))^2\right)\frac{t^2}{2m^{2\delta}}\\ &+ O(t^3m^{1-3\delta+6\delta'})\end{aligned}$$

for $m \to \infty$

Proof: We apply (46) of Theorem 9. Let $\delta' < \delta$ be arbitrary small. We want to apply Lemma 1 with t' and δ' such that $t' = tm^{\delta'-\delta}$, so that the condition $|t'| \le \varepsilon$ is easily checked. From

$$\log L(m, e^{t'm^{-\delta'}}) = m + \widetilde{X}(m)\frac{t'}{m^{\delta'}} + \widetilde{V}(m)\frac{t'^2}{2m^{2\delta'}} + O(t'^3 m^{1+3\delta'})$$

we find

$$\log L(m, e^{tm^{-\delta}}) = m + \widetilde{X}(m)\frac{t}{m^{\delta}} + \widetilde{V}(m)\frac{t^2}{2m^{2\delta}} + O(t^3 m^{1-3\delta+6\delta'}) \ .$$

In order to apply (46) we need to estimate

$$\left(\frac{\partial}{\partial z}\log(L(m,u_m)) - 1\right)^2 = (D(m,u_m)-1)^2 ,$$

where $u_m = e^{tm^{-\delta}}$ Applying Lemma 2 with $t' = tm^{\delta'-\delta}$ we first find

$$D(m, e^{tm^{-\delta}}) = 1 + \widetilde{X}'(m)\frac{t}{m^{\delta}} + O(t^2 m^{-2\delta+4\delta'}) \ . \quad (56)$$

Then using $X'(m) = O(m^{\delta'})$ we arrive at

$$\left(D(m, e^{tm^{-\delta}}) - 1\right)^2 = (\widetilde{X}'(m))^2\frac{t^2}{m^{2\delta}} + O(t^3 m^{-3\delta+5\delta'}) \ .$$

Putting everything together and using (46) of Theorem 9, we finally achieve the expected estimate on $\log L_m(e^{tm^{-\delta}})$ [5], [14]. ■

The next result allows us to connect Poissonized mean and variance with the original mean and variance. It follows directly from Theorem 11.

Corollary 2. *For any* $\delta' > 0$

$$\begin{aligned} \mathbf{E}(L_m) &= \widetilde{X}(m) + o(1) = O(m^{1+\delta'}), \\ \mathrm{Var}(L_m) &= \widetilde{V}(m) - m(X'(m))^2 + o(1) = O(m^{1+2\delta'}) \end{aligned}$$

as $m \to \infty$.

B. Proof of Theorem 6

We are finally in the position to prove Theorem 6, granted Theorem 10 which we will establish in the next section. From Theorem 9 and Lemma 2, for any $\delta' > 0$, we obtain the estimate

$$\log L_m(e^{tm^{-\delta'}}) = O(m)$$

which proves (18) of Theorem 6. To prove (19) we estimate the reminder

$$R_m(t) = \log L_m(e^t) - \mathbf{E}(L_m)t - \mathrm{Var}(L_m)\frac{t^2}{2}.$$

By Cauchy formula, as in the proof of Lemma 1, we have

$$R_m(t) = \frac{2t^3}{2i\pi}\oint \log L_m(e^{t'})\frac{dt'}{(t')^3(t'-t)} \ , \quad (57)$$

where the integral is around a circle of center 0 and radius $\varepsilon|z|^{-\delta}$. If we restrict $|t| \le \frac{\varepsilon}{2}|z|^{-\delta}$, then $|t-t'| \ge \frac{\varepsilon}{2}|z|^{-\delta}$. As in the proof of Lemma 1, we find

$$R_m(t) = t^3 O\left(\frac{m^{1+3\delta'}}{\varepsilon^3}\right).$$

Therefore, for $\delta > \delta'$ we finally arrive at

$$R_m(tm^{-\delta}) = t^3 O\left(m^{1-3(\delta-\delta')}\right) = t^3 O\left(m^{1-3(\delta'')}\right) \quad (58)$$

for some $\delta'' > 0$, This completes the proof of Theorem 6.

C. Proof of Theorem 10

To complete our analysis we need to prove Theorem 10. We establish (47) and (48) in below Theorems 12 and 13, respectively.

We apply the so called *increasing domains* technique [4], [14]. This technique allows to establish a property over an area of complex plane (e.g., cone) by mathematical induction. Indeed, let R be a real number. We denote $\mathcal{C}_0(\theta)$ the subset of the linear cone $\mathcal{C}(\theta) = \{z \ : \ |\arg(z)| \le \theta\}$ consisting of complex numbers of modulus smaller than or equal to R. By extension, let k be an integer, and we denote by $\mathcal{C}_k(\theta)$ the subset of $\mathcal{C}(\theta)$ that consists of complex numbers of modulus smaller than or equal to $R\rho^k$ where

$$\rho = \min_{u\in\mathcal{U}(1), a\in\mathcal{A}}\left\{\frac{1}{p_a|u|}\right\} > 1 \quad (59)$$

for a neighborhood $\mathcal{U}(1)$ of $u = 1$. By construction if $z \in \mathcal{C}_k(\theta)$ for $k > 0$ and $u \in \mathcal{U}(1)$, then all $p_a uz$ for $a \in \mathcal{A}$ belong to $\mathcal{C}_{k-1}(\theta)$ so that a property that holds for $\mathcal{C}_{k-1}(\theta)$ can be extended to a larger subset of the cone, namely $\mathcal{C}_k(\theta)$, for quantities satisfying functional equations like (16). This is illustrated in Figure 3.

Our goal is to present a polynomial estimate $\log L(z,u)$ (i.e., $\log L(z,u) = O(z)$) for large z in a cone containing the real positive axis. The main problem is the existence of the logarithm of $L(z,u)$ in particular for complex values of z and u. Technically, we can only prove the existence and growth of $\log L(z,u)$ for complex u with small imaginary part as z increases. For this we fix an arbitrary non-negative real number $\delta < 1$, and we fix t and z complex so that

$$u(z,t) = e^{t|z|^{-\delta}} \ . \quad (60)$$

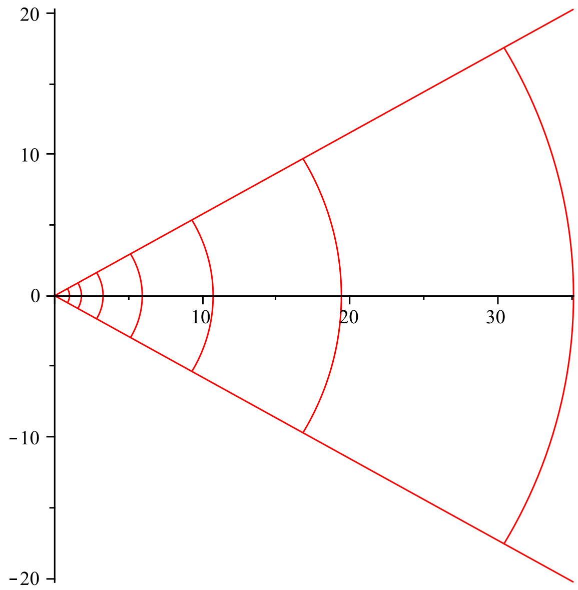

Fig. 3. Sets $\mathcal{C}_k(\theta)$ for $\theta = \frac{\pi}{6}$ for $k = 0, \ldots, 6$.

The key to our analysis is the following theorem that proves (47) of Theorem 10.

Theorem 12. *There exists a complex neighborhood $\mathcal{U}(0)$ of $t = 0$ and $B > 0$ such that for all $t \in \mathcal{U}(0)$ and for all $z \in \mathcal{C}(\theta)$ the function $\log L(z, u(z,t)$ exists and*

$$\log L(z, u(z,t) \leq B|z| \ . \qquad (61)$$

We prove it in several steps below. The road map for the proof is as follows: We first introduce the following function $f(z,u)$ that we call the *kernel* function defined as in [4]

$$f(z,u) = \frac{L(z,u)}{\frac{\partial}{\partial z}L(z,u)} = \frac{L(z,u)}{\prod_{a\in\mathcal{A}} L(p_a uz, u)}. \qquad (62)$$

Notice that formally

$$\frac{1}{f(z,u)} = \frac{\partial}{\partial z}\log L(z,u).$$

Indeed, if we show that the kernel function is well defined and is never zero in a convex set containing the real positive line, then we will prove that $\log L(z,u)$ exists since

$$\log L(z,u) = \int_0^z \frac{dx}{f(x,u)}. \qquad (63)$$

Furthermore, if we prove that the estimate $f(x,u) = \Omega(1)$, then

$$\log L(z,u) = \int_0^z \frac{dx}{f(x,u)} = O(z) \qquad (64)$$

as needed to establish (47) of Theorem 10.

Understanding the kernel function is therefore the key to our analysis. In passing we observe that the kernel function satisfies the following differential equation:

$$\frac{\partial}{\partial z}f(z,u) = 1 - f(z,u)\sum_{a\in\mathcal{A}}\frac{p_a u}{f(p_a uz, u)} \ . \qquad (65)$$

We proceed now with the **proof of Theorem 12**. We start with a trivial lemma, whose proof is left to the reader.

Lemma 3. *Let for (x,ε) real positive tuple a function $h(x,\varepsilon)$ be defined on an open set containing all tuples $(x,0)$ with $x \geq 0$. Assume that the function $h(x,\varepsilon)$ is real positive and continuously differentiable. If*

$$\forall x \geq 0: \ \frac{\partial}{\partial x}h(x,0) < 1,$$

then for all compact set $\mathcal{K}_x$ there exists a compact neighborhood of $\mathcal{U}(0)$ of 0: $(x_0, t) \in \mathcal{K}_x \times \mathcal{U}(0)$ so that the sequence defined for k integer

$$x_{k+1} = h(x_k, \varepsilon) \qquad (66)$$

converges to a bounded fixed point when $k \to \infty$.

Let us define the function $a(z,u)$ as

$$\frac{1}{f(z,u)} = 1 + a(z,u).$$

In the next two lemmas we prove that $a(z,u) = O(1)$ for u as in (60) which by (64) proves (47).

Lemma 4. *Let δ' be a real number such that $\delta' < \delta < 1$. For all number $\hat{a} > 0$ there exists a real number $\varepsilon > 0$ such that for all real t and $|t| < \varepsilon$ we have*

$$|a(z, u(z,t))| \leq \hat{a}\frac{|t|}{|z|^{\delta'}} \ . \qquad (67)$$

for all $z \in \mathcal{C}(\theta)$.

Proof: We apply the increasing domain technique with

$$\rho = \min_{u\in\mathcal{U}(1), a\in\mathcal{A}}\left\{\frac{1}{p_a|u|}\right\} > 1$$

for a compact neighborhood $\mathcal{U}(1)$ of $u = 1$ which is assumed to be small enough so that ρ is greater than 1. To proceed, we next make $u(z,t)$ independent of z in the subset $\mathcal{C}_k(\theta)$ of the kth increasing domain by introducing $u_k(t) = e^{t\nu^k}$ for $\nu = \rho^{-\delta}$, and we fix

$$\mu = \rho^{-\delta'} > \nu$$

for $\delta' < \delta$. In the following we will denote $f_k(z) = f(z, u_k(t))$, and $u_k = u_k(t)$, omitting variable t. Recall that the kernel function satisfies the differential equation:

$$f_k'(z) = 1 - f_k(z)\sum_{a\in\mathcal{A}}\frac{p_a u_k}{f(p_a u_k z, u_k)} \ . \qquad (68)$$

Let $a_k(z,t)=a(z,u_k(t))$. Since $L(z,1)=e^z$ for all z, hence $f(z,1)=1$. Since $\frac{\partial}{\partial u}f(z,u)$ is well defined and continuous, we can restrict the neighborhood $\mathcal{U}(1)$ such that $f(z,u)$ is non zero and therefore $a(z,u)$ is well defined for $z\in\mathcal{C}_0(\theta)=\{z\in\mathcal{C}(\theta):\ |z|<R\}$ and $u\in\mathcal{U}(1)$. Let a_0 be a non negative number such that

$$\forall u\in\mathcal{U}(1)\ ,\quad \forall z\in\mathcal{C}_0(\theta)\ :\quad |a_0(z,t)|\leq a_0|t| \tag{69}$$

Now we fix ε such that $a_0\varepsilon<1$. We aim at proving that there exists a number $\epsilon>0$ such that there exists an increasing sequence of non negative numbers a_k such that for all $z\in\mathcal{C}_k(\theta)$: and for all t such that $|t|\leq\epsilon$:

$$|a_k(z,t)|\leq a_k|t|\mu^k \tag{70}$$

and $\limsup_{k\to\infty}a_k<\infty$.

We now apply the increasing domain approach, Let $z\in\mathcal{C}_k(\theta)$. We denote

$$g_k(z)=\sum_{a\in\mathcal{A}}\frac{p_a u_k}{f(p_a u_k z,u_k)}\ . \tag{71}$$

Thus (68) can be rewritten as $f_k'(z)=1-g_k(z)f_k(z)$ and the differential equation can be solved by

$$f_k(z)=1+\int_0^z(1-g_k(x))\exp(G_k(x)-G_k(z))dx\ , \tag{72}$$

where $G_k(z)$ is a primitive of function $g_k(z)$.

We now will give some bounds on $g_k(z)$ for $z\in\mathcal{C}_k(\theta)$ and $|t|<\varepsilon$. For all $a\in\mathcal{A}$ we assume $p_a u_k z\in\mathcal{C}_{k-1}(\theta)$. We have $u_k(t)=u_{k-1}(\nu t)$ and we can use the recursion since $|\nu t|<\varepsilon$. In particular we have

$$\begin{aligned} g_k(z) &= \sum_{a\in\mathcal{A}}p_a u_k(1+a_{k-1}(p_a u_k z,\nu t)) && (73)\\ &= 1+b_k(z,t) && (74)\end{aligned}$$

with

$$b_k(z,t)=\sum_{a\in\mathcal{A}}p_a(u_k-1+u_k a_{k-1}(p_a u_k z,\nu t))\ . \tag{75}$$

Since both $|a_{k-1}(pu_k,\nu t)|$ and $|a_{k-1}(qu_k,\nu t)|$ are smaller than $a_{k-1}\nu\mu^{k-1}|t|$, and since $|u_k-1|\leq\beta\nu^k|t|$ for some β close to 1, we have $|b(z,t)|\leq b_k|t|$ with

$$b_k=(a_{k-1}\nu\mu^{k-1}+\beta\nu^k)(1+\beta\nu^k\varepsilon)\ . \tag{76}$$

Thus plugging in (72) we find

$$\begin{aligned} |f_k(z)-1| &\leq \int_0^z|b_k(x,t)|\exp(\Re(G_k(x)-G_k(z)))dx\\ &\leq \int_0^1 b_k|t||z|\exp(\Re(G_k(zy)-G_k(z)))dy\\ &\leq \frac{b_k|t|}{\cos(\theta)-b_k|t|}.\end{aligned}$$

Clearly,

$$\begin{aligned} \Re(G_k(yz)-G_k(z)) &= -\Re(z)(1-y)\\ &+ \int_y^1\Re(zb_k(zx,t))dx\\ &\leq -\cos(\theta)|z|+b_k|z|,\end{aligned}$$

hence

$$\left|\frac{1}{f_k(z)}-1\right|\leq\frac{\frac{b_k|t|}{\cos(\theta)-b_k|t|}}{1-\frac{b_k|t|}{\cos(\theta)-b_k|t|}}=\frac{b_k|t|}{\cos(\theta)-2b_k|t|}\ . \tag{77}$$

Therefore,

$$a_k\leq(a_{k-1}\frac{\nu}{\mu}+\beta\frac{\nu^k}{\mu^k})(1+\beta\nu^k\varepsilon)\frac{1}{\cos(\theta)-b_k\varepsilon}\ . \tag{78}$$

Let now $h(a_k,\varepsilon)$ be the right-hand side of (78). Notice that

$$\frac{\partial}{\partial a_k}h(a_k,0)=\frac{\nu}{\mu\cos(\theta)}<1$$

for small enough θ. Thus we are in the realm of Lemma 3. Moreover, $h(x,\varepsilon)$ is increasing. Since in Lemma 3 we can make ε small enough, hence $\limsup_{k\to\infty}a_k<\infty$ and (70) is proved, so the lemma. ■

We extend this lemma to a complex neighborhood of $t=0$ ($u=1$).

Lemma 5. *For all number $\alpha>0$ there exists $\varepsilon>0$, $\theta\in]0,\frac{\pi}{2}[$ such for complex t such that $|t|<\varepsilon$:*

$$|a(z,u(z,t))|\leq\alpha\frac{|t|}{|z|^{\delta'}} \tag{79}$$

for all $z\in\mathcal{C}$.

Proof: The proof is essentially the same as the previous lemma except that we have to extend the cone $\mathcal{C}(\theta)$ to a larger set $\mathcal{C}'(\theta)$ defined by $\{z:\ |\arg(z)|\leq\theta+\phi|z|^{\delta-1}\}$ so that if $z\in\mathcal{C}'(\theta)$, then for all $a\in\mathcal{A}$ also $p_a u(z,t)z$ belongs to $\mathcal{C}'(\theta)$ (with a small rotation of angle $\Im(\frac{t}{|z|^\delta})$ in the case two points outside $\mathcal{C}(\theta)$ may not satisfy the induction hypothesis). ■

We now establish (48) of Theorem 10.

Theorem 13. *Let $\theta\in]0,\frac{\pi}{2}[$. There exist numbers $A>0$, $\alpha<1$ and $\varepsilon>0$ such that for all complex t such $|t|\leq\varepsilon$:*

$$z\notin\mathcal{C}(\theta)\Rightarrow|L(z,u(z,t))|\leq Ae^{\alpha|z|}\ . \tag{80}$$

Proof: We proceed as with the previous proof: We first prove it for t real (near $t=0$) and then consider complex t. We take a neighborhood $\mathcal{U}(1)$ of $u=1$ (or $t=0$) and define ρ as in (59). We define $\bar{\mathcal{C}}(\theta)$ as the complementary of $\mathcal{C}(\theta)$ in the complex plan. We also introduce

$$\lambda = \min_{u\in\mathcal{U}(1), a\in\mathcal{A}} \{p_a|u|\}. \quad (81)$$

We set $R > 0$ and define $\bar{\mathcal{C}}_0(\theta)$ and $\bar{\mathcal{C}}_k(\theta)$ for $k > 0$ integer as subsets of $\bar{\mathcal{C}}(\theta)$:

$$\begin{aligned}\bar{\mathcal{C}}_0(\theta) &= \{z \in \bar{\mathcal{C}}(\theta), |z| \leq \lambda R\},\\ \bar{\mathcal{C}}_k(\theta) &= \{z \in \bar{\mathcal{C}}(\theta), \lambda R < |z| \leq \rho^k R\}.\end{aligned}$$

With these definitions, if $u \in \mathcal{U}(1)$ when z is in $\bar{\mathcal{C}}_k(\theta) - \bar{\mathcal{C}}_{k-1}(\theta)$, then both puz and quz are in $\bar{\mathcal{C}}_{k-1}(\theta)$. This determines the increasing domains in this case.

Since $L(z,1) = e^z$ and if $\alpha > \cos(\theta)$ then $|L(z,1)| \leq e^{\alpha|z|}$. There exist $A_0 > 0$ and ε such that for all t such $|t| \leq \varepsilon$ and for all $z \in \bar{\mathcal{C}}_0(\theta)$: $|L(z,e^t)| \leq A_0 e^{\alpha|z|}$. We also tune ε so that $\alpha\prod_k u_k(\varepsilon) < 1$.

We proceed doing the same analysis for $z \in \bar{\mathcal{C}}_1(\theta)$. But since $|L(z,1)|$ is strictly smaller than $e^{\alpha|z|}$ for all $z \in \bar{\mathcal{C}}_1(\theta)$, we can find $A_1 < 1$ and $\varepsilon > 0$ such that for all t such $|t| \leq \varepsilon$ and for all $z \in \bar{\mathcal{C}}_1(\theta)$ we have $|L(z,e^t)| \leq A_1 e^{\alpha|z|}$. In fact since

$$\min_{z\in\bar{\mathcal{C}}_1(\theta)} \{\frac{|e^z|}{e^{\alpha|z|}}\} \to 0$$

when $R \to \infty$ we can make A_1 as small as we want.

We now define $\alpha_k = \alpha\prod_{i=0}^{i=k} u_k(\varepsilon)$. We will prove by induction that there exists an increasing sequence $A_k < 1$ such that for t such $|t| \leq \varepsilon$:

$$z \in \bar{\mathcal{C}}_k(\theta) \Rightarrow |L(z, u_k(t))| \leq A_k e^{\alpha_k|z|} \ . \quad (82)$$

Our plan is to prove this property by induction. Assume it is true for all integers smaller equal to $k-1$, we then prove it is true for k. Assume $z \in \bar{\mathcal{C}}_k(\theta) - \bar{\mathcal{C}}_{k-1}(\theta)$. We use of the differential equation:

$$L(z,u_k) = L(z/\rho, u_k) + \int_{z/\rho}^{z} \prod_{a\in\mathcal{A}} L(p_a u_k x, u_k)dx \ .$$

Clearly,

$$|L(z,u_k)| \leq |L(\frac{z}{\rho}, u_k)| + |z| \int_{1/\rho}^{1} \prod_{a\in\mathcal{A}} |L(p_a u_k z y, u_k)|dy \ .$$

Using induction hypothesis

$$|L(\frac{z}{\rho}, u_k)| \leq A_{k-1} e^{\alpha_{k-1}|z|/\rho},$$

and for all $a \in \mathcal{A}$:

$$|L(p_a u_k y z, u_k)| \leq A_{k-1} e^{\alpha_{k-1} p_a |u_k||z|y} \leq A_{k-1} e^{\alpha_k p|z|y}$$

(we have $\alpha_{k-1}|u| \leq \alpha_{k-1}e^{\varepsilon} = \alpha_k$). Thus

$$|L(z,u_k) \leq A_{k-1} e^{\alpha_{k-1}|z|/\rho} + \frac{A_{k-1}^2}{\alpha_k}\left(e^{\alpha_k|z|} - e^{\alpha_k|z|/\rho}\right) \ .$$

This gives an estimate

$$A_k \leq \frac{A_{k-1}^2}{\alpha_k} + A_{k-1} e^{-\rho^{k-2}(\rho-1)\alpha_k R} \ .$$

Clearly the term in $e^{-\rho^{k-2}(\rho-1)\alpha_k R}$ can be made as small as we want by increasing R. If we choose A_1 such that

$$\frac{A_1}{\alpha_1} + e^{-\rho^{k-2}(\rho-1)\alpha_k R} < 1$$

for all k then we get $A_k \leq A_{k-1}$ and the theorem is proven for t real.

Second, we need to expand our proof to the case where t is complex and $|t| \leq \varepsilon$. To this end we use a similar trick as in the proof of Lemma 5. We expand $\bar{\mathcal{C}}(\theta)$ to

$$\bar{\mathcal{C}}'(\theta) = \{z : \ \arg(z)| \geq \theta + \phi R^{\delta-1} - \phi|z|^{\delta-1}\}$$

for $|z| > R\rho$ in order to assure that $p_a u_k z$ stays in $\bar{\mathcal{C}}'(\theta)$ for all $a \in \mathcal{A}$ when

$$z \in \bar{\mathcal{C}}'_k(\theta) - \bar{\mathcal{C}}'_{k-1}(\theta)$$

(absorbing a tiny rotation, if needed, that the factor u_k implies when t is complex). Of course, one must choose ϕ such that $\theta + \phi R^{\delta-1} < \frac{\pi}{2}$ and tune ε. ■

References

[1] D. Aldous, and P. Shields, A Diffusion Limit for a Class of Random-Growing Binary Trees, *Probab. Th. Rel. Fields*, 79, 509–542, 1988.

[2] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York 1968.

[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Second edition. John Wiley & Sons, New York, 2006.

[4] P. Jacquet and W. Szpankowski, Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees, *Theoretical Computer Science*, 144, 161–197, 1995.

[5] P. Jacquet and W. Szpankowski, Analytical depoissonization and its applications, *Theoretical Computer Science*, 201, 1–62, 1998.

[6] P. Jacquet, W. Szpankowski, and J. Tang, Average Profile of the Lempel-Ziv Parsing Scheme for a Markovian Source, *Algorithmica*, 31, 318–360, 2001.

[7] G Louchard, W Szpankowski, On the average redundancy rate of the Lempel-Ziv code. *IEEE Transactions on Information Theory*, 43, 2–8, 1997.

[8] D. Knuth, *The Art of Computer Programming. Vol. III Sorting and Searching,* (Second Edition), Addison-Wesley (1998).

[9] N. Merhav, Universal Coding with Minimum Probability of Codeword Length Overflow, *IEEE Trans. Information Theory*, 37, 556–563, 1991.

[10] R. Neininger and L. Ruschendorf, A General Limit Theorem for Recursive Algorithms and Combinatorial Structures, *The Annals of Applied Probability*, 14, No. 1, 378-418, 2004.

[11] E. Plotnik, M.J. Weinberger, and J. Ziv, Upper Bounds on the Probability of Sequences Emitted by Finite-State Sources and on the Redundancy of the Lempel-Ziv Algorithm, *IEEE Trans. Information Theory*, 38, 66-72 (1992).

[12] S. Savari, Redundancy of the Lempel-Ziv Incremental Parsing Rule, *IEEE Trans. Information Theory*, 43, 9–21, 1997.

[13] G. Seroussi, On Universal Types, *IEEE Trans. Information Theory*, 52, 171–189, 2006.
[14] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001.
[15] J. Ziv and A. Lempel, Compression of individual sequences via variable-rate coding, *IEEE Transactions on Information Theory*, 24, 530–536, 1978.