CERIAS Tech Report 2002-05

AVERAGE PROFILE OF THE LEMPEL-ZIV PARSING SCHEME FOR A MARKOVIAN SOURCE

Philippe Flajolet[1], Wojciech Szpankowski[2], Jing Tang[3]
Center for Education and Research in
Information Assurance and Security
&
[2]Department of Computer Sciences,
Purdue University, West Lafayette, IN 47907-1398
[1]INRIA - Roquencourt
[3]Microsoft Corporation

AVERAGE PROFILE OF THE LEMPEL-ZIV PARSING SCHEME FOR A MARKOVIAN SOURCE*

October 2, 2000

Philippe Jacquet
INRIA
Rocquencourt
78153 Le Chesnay Cedex
France
Philippe.Jacquet@inria.fr

Wojciech Szpankowski
Dept. of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Jing Tang
Microsoft Corporation
One Microsoft Way, 1/2061
Redmond, WA 98052
U.S.A.
geneta@microsoft.com

Abstract

For a Markovian source, we analyze the Lempel-Ziv parsing scheme that partitions sequences into phrases such that a new phrase is the shortest phrase not seen in the past. We consider three models: In the **Markov Independent** model, several sequences are generated *independently* by Markovian sources, and the ith phrase is the shortest prefix of the ith sequence that was not seen before as a phrase (i.e., a prefix of previous $(i-1)$ sequences). In the other two models, only a single sequence is generated by a Markovian source. In the second model, called the **Gilbert-Kadota** model, a *fixed number* of phrases is generated according to the Lempel-Ziv algorithm, thus producing a sequence of a variable (random) length. In the last model, known also as the **Lempel-Ziv** model, a string of *fixed length* is partitioned into a variable (random) number of phrases. These three models can be efficiently represented and analyzed by *digital search trees* that are of interest to other algorithms such as sorting, searching and pattern matching. In this paper, we concentrate on analyzing the average profile (i.e., the average number of phrases of a given length), the typical phrase length, and the length of the last phrase. We obtain asymptotic expansions for the mean and the variance of the phrase length, and we prove that appropriately normalized phrase length in all three models tends to the standard normal distribution, which leads to bounds on the average redundancy of the Lempel-Ziv code. For Markov Independent model, this finding is established by analytic methods (i.e., generating functions, Mellin transform and depoissonization), while for the other two models we use a combination of analytic and probabilistic analyses.

Index Terms: Lempel-Ziv scheme, Markov source, digital search trees, data compression, phrase length, depth in a tree, Poisson transform, Mellin transform, analytic depoissonization, stochastic comparisons.

*This work was partially supported by NSF Grants NCR-9415491 and NCR-9804760, and NATO Collaborative Grant CRG.950060, and contract 1419991431A from sponsors of CERIAS at Purdue.

1 Introduction

The heart of many lossless data compression schemes is the incremental parsing algorithm due to Lempel and Ziv [29]. It partitions a sequence into variable phrases such that a new phrase is the shortest substring not seen in the past as a phrase. Fundamental information about the algorithm is contained in such parameters as the number of phrases, the phrase length, the number of phrases of a given size, and the longest phrase. In this paper, we study the length of a randomly selected phrase (which is equivalent to the so called *average profile* defined as the average number of phrases of a given size) and the length of the last phrase (cf. [13, 14, 24]) for Markov sources.

In the past, mostly *first order analysis* of these parameters were studied for memoryless sources with the exception of [1, 10, 14, 15, 21]. The first order analysis provides the first order asymptotics (e.g., is the redundancy of a code $o(n)$?). The second order analysis attempts to establish the rate of convergence, or even a full asymptotic expansion, large deviations behavior, deviation from the mean (e.g., central limit theorems), and so forth. We present here a second order analysis of the (typical) phrase length for the Lempel-Ziv parsing scheme in a Markovian setting. J. Ziv in his *1997 Shannon Lecture* [28] presented compelling arguments for "backing off" to a certain degree from the first-order asymptotic analysis of information systems in order to predict the behavior of real systems, where we always face *finite*, and often small, lengths (of sequences, files, codes, etc.) One way of overcoming these difficulties is to increase the accuracy of asymptotic analysis by replacing first-order analysis by full asymptotic expansions and more accurate analysis so that the approximate value of a quantity of interest is closer to the true value even for moderate and small lengths.

In this paper, we analyze three models of the Lempel-Ziv scheme in the Markovian settings. In the first one, called **Markov Independent model** or shortly MI model, we assume that there are m independent Markov sources defined on the same underlying probability space. The parsing is done with respect to the previous sequences. Namely, the zeroth phrase is an *empty* phrase, while the first phrase is a one character prefix of the first sequence. The ith phrase ($i \leq m$) is defined as the shortest prefix of the ith sequence not seen as a phrase (prefix) of the previous $(i-1)$ sequences. For example, for $m = 4$ sequences: $X(1) = 000000\ldots$, $X(2) = 1010101\ldots$, $X(3) = 1001101\ldots$ and $X(4) = 001100111\ldots$ we can construct the following Lempel-Ziv sequence: $(\epsilon)(0)(1)(10)(00)$ where ϵ is an empty phrase, and all phrases are shown in parentheses. We shall study two parameters, namely the length, D_m, of a randomly selected phrase, and the length I_m of the last phrase. In addition, one may investigate the length L_m of the Lempel-Ziv sequence. In the example above we have

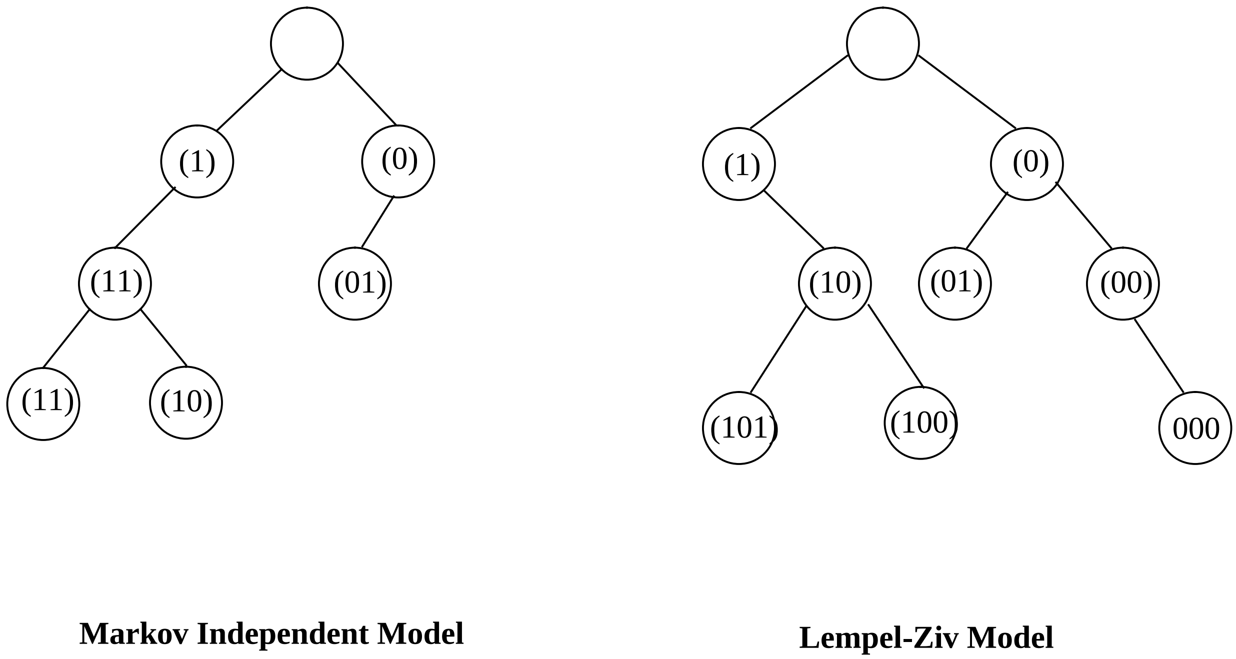

Figure 1: Digital tree representations for the MI model ($X(1) = 00000$, $X(2) = 01111$, $X_3 = 101010$, $X(4) = 111000$, $X(5) = 110111$, $X(6) = 111111$) and the LZ model ($X = 11001010001000100\ldots$) of the Lempel-Ziv algorithm.

$D_4 = 1\frac{1}{2}$, $I_4 = 2$ and $L_4 = 6$.

The next two models deal with a single sequence generated by a Markovian source. In the **fixed number of phrases model**, we partition the sequence according to the Lempel-Ziv algorithm until we obtain m full phrases (thus producing a variable and random length of the Lempel-Ziv sequence). For example, for $X = 11001010001000100\ldots$ we can construct $m = 5$ phrases as follows: $(\epsilon)(1)(10)(0)(101)(00)$. Such a model was also considered by Gilbert and Kadota [7], so we call it the **Gilbert-Kadota model** or shortly GK model. As before, we will be interested in the typical phrase length D_m and the last phrase length I_m. In the above example, we have $D_5 = 1\frac{4}{5}$, $I_5 = 2$, and in addition the length of the Lempel-Ziv sequence is $L_5 = 9$.

Finally, in the traditional **Lempel-Ziv model** or **fixed length model**, a sequence of fixed length, say n symbols, is partitioned according to the Lempel-Ziv algorithm. For example, the string $X = 110010100010$ of length $n = 12$ is parsed as $(\epsilon)(1)(10)(0)(101)(00)(01)(0)$. We shall study the length Δ_n of the randomly selected phrase (see Section 2 for a precise definition) and the length J_n of the last *full* phrase. The number of full phrases M_n is of significant interest for this model, but we will not investigate it here. In the example above,

$\Delta_{12} = 1\frac{5}{6}$, $J_{12} = 2$ and $M_{12} = 6$.

The above three models can be efficiently analyzed and uniformly represented by a *digital search tree*, a data structure that have been studied by its own right for more than thirty years (cf. [13, 17]). This tree is used to store strings in its nodes and can be described as follows: We consider m, possibly infinite, strings of symbols over a finite alphabet $\mathcal{A} = \{1, 2, \ldots, V\}$ (however, we often restrict our discussions to a binary alphabet $\mathcal{A} = \{0, 1\}$). The root contains the empty string ϵ. The first string occupies the right or the left child of the root depending whether its first symbol is "1" or "0". The remaining strings are stored in available nodes (that are directly attached to nodes already existing in the tree). The search for an available node follows the prefix structure of a string. The rule is simple: if the next symbol in a string is "1" we move to the right, otherwise move to the left. The resulting tree has m internal nodes. It corresponds to the MI model and the GK model, however, in the latter the strings are substrings (phrases) of one infinite string We can call such a digital search tree a suffix search tree (cf. Figure 1).

In the LZ model, we construct an analogous (suffix) digital tree except that the number of nodes varies and equals to the number of phrases M_n. More precisely, the empty phrase is stored in the root, and all other phrases are located in nodes. When a new phrase is created, the search starts at the root and proceeds down the tree as directed by the input symbols exactly in the same manner as in the digital search tree construction. For example, for the binary alphabet, "0" in the input string means move to the left and "1" means proceed to the right. The search is completed when a branch is taken from an existing tree node to a new node that has not been visited before. Then an edge and a new node are added to the tree. Phrases created in such a way are stored directly in nodes of the tree (cf. [14]). This is illustrated in Figure 1.

As mentioned before, in this paper we present second order analysis of the above three models of the Lempel-Ziv algorithm for a Markovian source. Among others, we compute precise asymptotic formulæ for the mean and the variance of the phrase length in the MI model. We also show that the appropriately normalized phrase length tends to a normal distribution with the rate of convergence of $O(1/\sqrt{\ln m})$. These results – which are at the heart of our findings – are established by analytic methods. The line of the attack can be briefly described as follows: We first derive a set of recurrence equations for the ordinary generating functions of the average profile (conditioned on the first symbol). These recurrence equations are too complicated to be solved directly, hence we derive a set of differential-functional equations on the so called Poisson transform of the average profile. In the Poisson model, the number of sequences m becomes a random variable N distributed as a Poisson with mean m. This process of replacing the deterministic input m by a Poisson variable is called *poissonization*. We shall use *analytic* poissonization since we replace m by a complex variable z. A typical set of differential-functional equations we have to deal with is of the following form

$$\frac{\partial \widetilde{B}^i(z,u)}{\partial z} + \widetilde{B}^i(z,u) = u\left(\widetilde{B}^1(p_{i,1}z,u) + \cdots + \widetilde{B}^V(p_{i,V}z,u)\right) + a(z,u), \qquad i = 1,2,\ldots,V,$$

where $\widetilde{B}^i(z,u)$ is the Poisson transform (cf. [10, 24]) of the average profile when all strings start with symbol $i \in \mathcal{A} = \{1,2,\ldots,V\}$, $a(z,u)$ is a given function, and $\mathsf{P} = \{p_{ij}\}_{i,j=1}^V$ is the underlying Markov chain. These differential-functional equations are reduced to a simple matrix functional equations of the Mellin transform $B_i^*(s)$ with respect to z of $\widetilde{B}^i(z,u)$ (cf. [6, 24]). A typical equation of the Mellin transform looks like

$$B_i^*(s) - (s-1)B_i^*(s-1) = B_1^*(s)p_{1,i}^{-s} + \cdots + B_V^*(s)p_{i,V}^{-s} + a^*(s), \qquad i = 1,2,\ldots,V.$$

We can solve exactly this matrix equation in a form of an infinite product of matrices. However, we develop a method to obtain relevant asymptotics without an explicit solution. It turns out that such asymptotics depend on singularity points of the matrix $\mathsf{Q}(s) = (\mathsf{I} - \mathsf{P}(s))^{-1}$ where $\mathsf{P}(s) = \{p_{ij}^{-s}\}_{i,j=1}^V$ for some complex s. Then through the inverse Mellin transform we obtain asymptotics of the Poisson transform $\widetilde{B}^i(z,u)$ for large z. We need to translate it into the asymptotics of the original generating function $B_m^i(u)$. This process is called *depoissonization*, and we shall use recent results of Jacquet and Szpankowski [11] on *analytic* depoissonization. Such analysis is an example of "analytic information theory" that applies complex-analytic tools (e.g., generating functions, Mellin transform, poissonization) to information theory problems (e.g., Lempel-Ziv schemes, minimax redundancy, computer networks).

To translate the results of the MI model to GK model and LZ model we shall use a combination of analytic, combinatorial and probabilistic methods. In particular, we construct two MI models that upper bound and lower bound stochastically the GK model. This will allow us to conclude the central limit theorem for the phrase length in the GK model, which will further lead to a similar result for the LZ model.

Finally, we should mention that our MI model is equivalent to the Markov model of *digital search trees* studied extensively in computer science. In fact, digital trees appear in a variety of computer and communications applications including searching, sorting, dynamic hashing, codes, conflict resolution protocols for multiaccess communications, and data compression (cf, [13, 17, 24]). Thus better understanding of their behavior is desirable and could lead to some algorithmic improvements. One parameter that is of interest to these applications is the *depth* of a randomly selected node (i.e., the length of the path from the root to the chosen node), and *depth of insertion*, which may represent the search time. Clearly, the depth and the depth of insertion are equivalent to the typical phrase length and the last phrase length in the MI model. The average profile of the MI model is the same as the average number of nodes at a given level in the associated digital tree.

Digital trees (which include tries, PATRICIA tries and digital search trees) have been studied extensively in the past for memoryless source (cf. [13, 10, 14, 16, 17, 20, 23]). Extensions to Markovian sources are scarce, and to the best of our knowledge only tries were analyzed (cf. [4, 9]). Lempel-Ziv model for memoryless sources was discussed in [10, 14, 15], while second order analyses for Markovian sources are very scarce. Savari [21] proposed the redundancy analysis of the LZ code for Markovian sources, and Wyner [27] derived the limiting distribution of the phrase length in the other Lempel-Ziv scheme (i.e., LZ'77), which is known to be considerable simpler to analyze than the Lempel-Ziv'78 scheme.

This paper is organized as follows. In the next section we present our main results for all three models, and discuss some of their consequences. In particular, we present tight bounds on the average redundancy of the Lempel-Ziv'78 code. The proof for the MI model can be found in Section 3, while Section 4 presents our analysis of the GK model. The proof of the LZ model is discussed after Theorem 3 in Section 2.

2 Main Results

We now present our main results for all three models, namely Markov Independent model, Gilbert-Kadota (fixed number of phrases) model, and Lempel-Ziv model. Most of the proofs are delayed till the next sections. Throughout, we assume that a sequence, say $X = (X_0, X_1, \ldots)$, is generated by a Markov source over a finite alphabet $\mathcal{A} = \{1, 2, \ldots, V\}$. More precisely:

(M) Markov Source

There is a Markovian dependency between consecutive symbols in a sequence, that is, the probability $p_{ij} = \Pr\{X_{k+1} = j | X_k = i\}0$ for all $k \geq 0$ describes the conditional probability of sampling symbol $j \in \mathcal{A}$ immediately after symbol $i \in \mathcal{A}$. We assume that the Markov chain is aperiodic, irreducible and that $p_{ii} > 0$ for $i \in \mathcal{A}$. We denote by $\mathsf{P} = \{p_{ij}\}_{i,j=1}^{V}$ the transition matrix, and by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_V)$ the stationary vector satisfying $\boldsymbol{\pi}\mathsf{P} = \boldsymbol{\pi}$. We say that the Markov chain is *stationary* if $\Pr\{X_k = i\} = \pi_i$ for all $k \geq 0$ and $i \in \mathcal{A}$. In general, X_{k+1} may dependent on last r symbols, and then we have rth order Markov chains, however, hereafter in this paper we only deal with $r = 1$.

2.1 Markov Independent Model – Stationary Source

Hereafter, we assume that m independent Markov sources generate m sequences, which are parsed with respect to previous ones according to the Lempel-Ziv algorithm, as described in the introduction. Equivalently, we build a digital search tree from these m sequences, as shown in Figure 1. Actually, it is more convenient to think in terms of this associated digital search tree (DST). In particular, the ith phrase length I_i is also the depth of the ith node in such a tree (where the depth of a node is understood as the number of nodes from the root to the ith node). When $i = m$ we shall refer to I_m as the *depth of insertion* or the *last phrase length*. The *typical depth* (typical phrase length) D_m is defined as the length of a randomly selected depth, that is

$$\Pr\{D_m = k\} = \frac{1}{m}\sum_{i=1}^{m}\Pr\{I_i = k\}.$$

Finally, we defined the *average* profile (in short: profile) B_m^k as the *average* number of nodes at level k of the DST or the average number of phrases of length k. Observe that $B_0^k = 0$ for all $k \geq 0$

There are simple relationships between just defined parameters. First of all, we notice that (cf. [13, 14, 23])

$$\Pr\{D_m = k\} = \frac{B_m^k}{m}. \tag{1}$$

This and the definition of the typical depth immediately imply

$$\Pr\{I_{m+1} = k\} = B_{m+1}^k - B_m^k, \tag{2}$$

with $\Pr\{I_0 = 0\} = 1$ and $\Pr\{I_0 = k\} = 0$ for all $k \geq 1$.

Throughout, we shall work with generating functions of the above quantities and the so called *Poisson transforms* that we define next. The ordinary generating functions are:

$$\begin{aligned}
D_m(u) &= \mathbf{E}[u^{D_m}] = \sum_{k\geq 0}\Pr\{D_m = k\}u^k, \qquad D_0(u) = 1,\\
I_m(u) &= \mathbf{E}[u^{I_m}] = \sum_{k\geq 0}\Pr\{I_m = k\}u^k, \qquad I_0(u) = 1,\\
B_m(u) &= \sum_{k\geq 0} B_m^k u^k \qquad\qquad B_0(u) = 0
\end{aligned}$$

for a complex u such that $|u| < 1$. The Poisson transforms are defined as follows:

$$\begin{aligned}
\widetilde{D}(z,u) &= \sum_{m\geq 0} D_m(u)\frac{z^m}{m!}e^{-z},\\
\widetilde{B}(z,u) &= \sum_{m\geq 0} B_m(u)\frac{z^m}{m!}e^{-z},\\
\widetilde{I}(z,u) &= \sum_{m\geq 0} I_m(u)\frac{z^m}{m!}e^{-z}.
\end{aligned}$$

The Poisson transform can be interpreted as the generating function in the so called *Poisson model* in which the *deterministic* number of sequences m is replaced by a random number of sequences distributed according to Poisson with mean $z = m$. We shall assume that z is a complex variable, and $\widetilde{B}(z,u)$ as well as $\widetilde{I}(z,u)$ are defined on the whole complex plane. We should also observe that by (2)

$$\frac{\partial \widetilde{I}(z,u)}{\partial z} + \widetilde{I}(z,u) = \frac{\partial \widetilde{B}(z,u)}{\partial z}. \tag{3}$$

Since also $D_m(u) = B_m(u)/m$, we can recover all results on the depth of insertion I_m as well as on the typical depth from the average profile B_m^k. Therefore, hereafter we concentrate on the analysis of the average profile.

To start the analysis, we derive a system of recurrence equations for the generating function of the average profile. Let $B_m^i(u)$ for $i \in \mathcal{A}$ be the ordinary generating function of the average profile when all sequences start with symbol i. Let also $\mathbf{p} = (p_1, \ldots, p_V)$ be the initial probability vector of the underlying Markov chain, that is, $\Pr\{X_0 = i\} = p_i$. (For the stationary Markov chain we have $\mathbf{p} = \boldsymbol{\pi}$.) Consider now the generating function $B_{m+1}(u)$ of the DST, in which the root contains an empty string and the other m independent Markov sequences are stored in V subtrees, which are digital search trees by themselves but of smaller size. Indeed, the probability that the first subtree contains j_1 sequences, the second subtree has j_2 sequences, and so on until the V subtree stores j_V sequences (out of m sequences) is equal to the multinomial distribution, that is,

$$\binom{m}{j_1, \ldots j_V} p_1^{j_1} \cdots p_V^{j_V}.$$

But, the ith subtree is again a digital search tree of size j_i containing only those sequences that start with symbol i. Hence, its average profile generating function must be $B_{j_1}^i(u)$. This leads to the following recurrence equation assuming $B_0(u) = 0$

$$B_{m+1}(u) = u \sum_{|\mathbf{j}|=m} \binom{m}{\mathbf{j}} p_1^{j_1} \cdots p_V^{j_V} \left(B_{j_1}^1(u) + \cdots + B_{j_V}^V(u) \right) + 1, \tag{4}$$

where $\mathbf{j} = (j_1, \ldots, j_V)$, $|\mathbf{j}| = j_1 + \cdots + j_V$ and for simplicity $\binom{m}{\mathbf{j}} = \binom{m}{j_1, \ldots j_V}$. Clearly, we can set up similar recurrences for the subtrees. That is,

$$B^i_{m+1}(u) = u \sum_{|\mathbf{j}|=m} \binom{m}{\mathbf{j}} p_{i1}^{j_1} \ldots p_{iV}^{j_V} \left(B^1_{j_1}(u) + \ldots + B^V_{j_V}(u) \right) + 1, \qquad \text{for all} \quad i \in \mathcal{A} \tag{5}$$

where $B^i_0(u) = 0$ for $i \in \mathcal{A}$.

If we can solve the above recurrences, then we can compute all moments and the distribution of the average profile, and consequently the characteristics of the typical depth and the depth of insertion. Indeed, after observing that $B_m(1) = m$, the average depth becomes $\mathbf{E}[D_m] = B'_m(1)$ and

$$\mathbf{Var}[D_m] = \frac{B''_m(1)}{m} + \frac{B'_m(1)}{m} - \left(\frac{B'_m(1)}{m} \right)^2,$$

where $B'_m(1)$ and $B''_m(1)$ are the first and the second derivatives of the generating function $B_m(u)$ calculated at $u = 1$. In passing, we should observe that $B'_m(1)$ and $B''_m(1)$ satisfy recurrences equations similar to the ones derived for $B_m(u)$, and we shall discuss them in details in the next section.

We should point out that the above recurrence equations are not easy to solve. Even, if in principle, one can write an explicit solution (cf. [14, 23] for memoryless sources), it is too complicated to gain any insights. Therefore, we must retreat to the asymptotic analysis. To accomplish this, we shall derive a functional-differential equation on the Poisson transforms $\widetilde{B}^i(z,u)$, which seem to have a simpler, or at least more compact, form. These functional-differential equations are next changed into a simple matrix recurrence in terms of the Mellin transform (cf. [6, 17, 24]). After solving this matrix equation (in fact, for the asymptotic analysis we do not even need to solve it explicitly), we apply the inverse Mellin transform to recover the Poisson transform $\widetilde{B}^i(z,u)$ for $z \to \infty$ in a cone around the real axis. This suffices, since by analytic depoissonization (cf. [10, 11]) we can extract asymptotic expression for the average profile B^i_m for $m \to \infty$, which further leads to our final results.

Before we present out findings, we must introduce some more notation. Let s be complex, and then

$$\mathsf{Q}(s) = \mathsf{I} - \mathsf{P}(s), \qquad \text{where} \quad \mathsf{P}(s) = \{p_{ij}^{-s}\}_{i,j=1}^{V},$$

where I is the identity matrix. Let now $\mathsf{Q}^*(s) = \mathbf{adj}[\mathsf{Q}(s)]$ be the adjoint matrix of $\mathsf{Q}(s)$, that is, $\mathsf{Q}^*(s) = (-1)^{i+j}\{Q^{j,i}(s)\}_{i,j\in\mathcal{A}}$ where $Q^{j,i}(s)$ is the (j,i) cofactor of $\mathsf{Q}(s)$ defined as $\mathsf{Q}^{-1}(s) = \mathsf{Q}^*(s)/\det \mathsf{Q}(s)$ (cf. [19]). Furthermore, we define the following constants

$$\beta \;:=\; [\det \mathsf{Q}''(s)]|_{s=-1},$$

$$\dot{\mathsf{Q}}^* := \dot{\mathsf{Q}}^*(s)|_{s=-1},$$
$$\vartheta := \boldsymbol{\pi} \sum_{i=1}^{\infty} \left(\mathsf{Q}^{-1}(-2) \cdots \mathsf{Q}^{-1}(-i)(\mathsf{Q}^{-1}(s))'|_{s=-i-1} \mathsf{Q}^{-1}(-i-2) \cdots \right) \mathsf{K},$$

where

$$\mathsf{K} := \left(\prod_{i=0}^{\infty} Q^{-1}(-2-i) \right)^{-1} \boldsymbol{\psi} \ , \tag{6}$$

and $\boldsymbol{\psi} = [1, 1, \cdots, 1]^T_{V\times 1}$ is the column vector consisting of all 1s. Finally

$$\omega := \det \begin{bmatrix} 1 & -p_{12} & ... & -p_{1V} \\ 1 & 1-p_{22} & ... & -p_{2V} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & -p_{V2} & ... & 1-p_{VV} \end{bmatrix}$$

In addition, we use the standard notation for the entropy of a Markov source. In particular,

$$h = -\sum_{i=1}^{V} \pi_i \sum_{j=1}^{V} p_{ij} \ln p_{ij},$$

and for a probability vector $\mathbf{p} = (p_1, \ldots, p_V)$

$$h_{\mathbf{p}} = -\sum_{i=1}^{V} p_i \ln p_i.$$

Also, we often use $\mathbf{p}(s) = [\pi_1^{-s}, \pi_2^{-s}, ..., \pi_V^{-s}]$, which becomes $\boldsymbol{\pi}$ when $s = -1$.

In Section 3.1 we prove the following main result for MI model with stationary Markov sources (i.e., $\mathbf{p} = \boldsymbol{\pi}$).

Theorem 1 *Consider a Markov stationary source with transition probabilities* $\mathsf{P} = \{p_{ij}\}_{i,j=1}^V$, *that is,* $\Pr\{X_t(\ell) = k\} = \pi_k$ *for all* $t = 0, 1, \ldots$ *and* $\ell = 1, 2, \ldots, m$.

(i) [Typical Depth/Phrase Length] *For large* m *the following holds*

$$\mathbf{E}[D_m] = \frac{1}{h}\left(\ln m + \gamma - 1 + h - h_{\boldsymbol{\pi}} - \frac{\beta}{2\omega h} - \vartheta + \delta_1(\ln m) \right) + O\left(\frac{\ln m}{m}\right) \tag{7}$$
$$\mathbf{Var}[D_m] = \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2 \right) \ln m + O(1) \ , \tag{8}$$

and

$$\frac{D_m - \mathbf{E}[D_m]}{\sqrt{\mathbf{Var} D_m}} \to N(0,1), \tag{9}$$

where $\gamma = 0.577...$ *is the Euler constant, and* $N(0,1)$ *represents the standard normal distribution. The function* $\delta_1(x)$ *is a fluctuating function with a small amplitude when*

$$\frac{\ln p_{ij} + \ln p_{1i} - \ln p_{1j}}{\ln p_{11}} \in \mathbb{Q}, \qquad i,j = 1, 2, \ldots, V, \tag{10}$$

where $\mathbb{Q}$ *is the set of rational numbers. If (10) does not hold, then* $\lim_{x\to\infty}\delta_1(x)=0$. *One can strengthen (9) as follows. If* $\mu_m=\mathbf{E}[D_m]$, *and* $\sigma_m=\sqrt{\mathbf{Var}D_m}$, *then for a complex* τ *the generating function* $D_m(u)=\mathbf{E}[u^{D_m}]$ *becomes*

$$e^{-\tau\mu_m/\sigma_m}D_m(e^{\tau/\sigma_m})=e^{\frac{\tau^2}{2}}\left(1+O\left(\frac{1}{\sqrt{\ln m}}\right)\right) \tag{11}$$

as $m\to\infty$, *thus the rate of convergence to the normal distribution is* $O(1/\sqrt{\ln m})$. *Also, there exist positive constants* A *and* $\alpha<1$ *such that*

$$\Pr\left\{\left|\frac{D_m-\mathbf{E}[D_m]}{\sqrt{\mathbf{Var}D_m}}\right|\geq k\right\}\leq A\alpha^k \tag{12}$$

uniformly in k.

(ii) [Depth of Insertion/Last Phrase Length] *The depth of insertion (or equivalently, the last phrase length)* I_m *behaves asymptotically as the typical phrase* D_m. *More precisely, for some* $A>0$ *and* $\alpha<1$

$$\mathbf{E}[I_m] = \frac{1}{h}\left(\ln m+\gamma+h-h_{\boldsymbol{\pi}}-\frac{\beta}{2\omega h}-\vartheta+\delta_2(\ln m)\right)+O\left(\frac{\ln m}{m}\right), \tag{13}$$

$$\mathbf{Var}[I_m] = \mathbf{Var}[D_m]+O(1), \tag{14}$$

$$e^{-\tau\mu_m/\sigma_m}I_m(e^{\tau/\sigma_m}) = e^{\frac{\tau^2}{2}}\left(1+O\left(\frac{1}{\sqrt{\ln m}}\right)\right) \tag{15}$$

where $\delta_2(x)$ *is a fluctuating function with the same property as* $\delta_1(x)$. *In addition, there exist positive constants* A *and* $\alpha<1$ *such that*

$$\Pr\left\{\left|\frac{I_m-\mathbf{E}[I_m]}{\sqrt{\mathbf{Var}I_m}}\right|\geq k\right\}\leq A\alpha^k \tag{16}$$

Remarks. (i) *Alternative Representation.* We can present main results of Theorem 1 in a different form, which is particularly useful for the proof of the limiting distribution and, more importantly, can lead to some further generalizations (cf. [4, 26]). This new derivation can be found in Appendix A. For matrix $\mathsf{P}(s)$, we define the principal left eigenvector $\boldsymbol{\pi}(s)$, the principal right eigenvector $\boldsymbol{\psi}(s)$ associated with the largest eigenvalue $\lambda(s)$ as

$$\boldsymbol{\pi}(s)\mathsf{P}(s) = \lambda(s)\boldsymbol{\pi}(s), \tag{17}$$

$$\mathsf{P}(s)\boldsymbol{\psi}(s) = \lambda(s)\boldsymbol{\psi}(s), \tag{18}$$

where $\boldsymbol{\pi}(s)\boldsymbol{\psi}(s)=1$. The transition matrix P of the underlying Markov source has positive diagonal transition probabilities, hence by the *Perron-Frobenius Theorem* the largest eigenvalue of $\mathsf{P}(s)$ is well defined and unique. Observe that $\boldsymbol{\pi}(-1)=\boldsymbol{\pi}=(\pi_1,\ldots,\pi_V)$,

$\boldsymbol{\psi}(-1) = \boldsymbol{\psi} = (1, \ldots, 1)$, and $\lambda(-1) = 1$. Also, for an vector $\mathbf{x}(s)$ we write $\dot{\mathbf{x}}(s) = \frac{d}{ds}\mathbf{x}(s)$ and $\ddot{\mathbf{x}}(s) = \frac{d^2}{ds^2}\mathbf{x}(s)$. In Appendix A we shall prove that

$$\begin{aligned}
\dot{\lambda}(-1) &= \boldsymbol{\pi}\dot{\mathsf{P}}(-1)\boldsymbol{\psi} = h, \\
\ddot{\lambda}(-1) &= \boldsymbol{\pi}\ddot{\mathsf{P}}(-1)\boldsymbol{\psi} + 2\dot{\boldsymbol{\pi}}(-1)\dot{\mathsf{P}}(-1)\boldsymbol{\psi} - 2\dot{\lambda}(-1)\dot{\boldsymbol{\pi}}(-1)\boldsymbol{\psi}.
\end{aligned}$$

Then (7)–(8) of Theorem 1 can be alternatively written as

$$\begin{aligned}
\mathbf{E}[D_m] &= \frac{1}{\dot{\lambda}(-1)}\left(\ln m + \gamma - 1 + \dot{\lambda}(-1) + \frac{\ddot{\lambda}(-1)}{2\dot{\lambda}^2(-1)} - \vartheta - \boldsymbol{\pi}\dot{\boldsymbol{\psi}}(-1) + \delta_1(\ln m)\right) \\
&+ O\left(\frac{\ln m}{m}\right), \qquad (19) \\
\mathbf{Var}[D_m] &= \frac{\ddot{\lambda}(-1) - \dot{\lambda}^2(-1)}{\dot{\lambda}^3(-1)} \ln m + O(1). \qquad (20)
\end{aligned}$$

In a similar fashion, we can write for I_m.

(ii) *Memoryless Source.* Let us compare the findings of Theorem 1 to those obtained for a memoryless source (cf. [14, 23]). The Markov source becomes a memoryless source if we assume $p_{ji} = \pi_i$ for $i, j = 1, 2, \ldots, V$. Observe that then $\omega = 1$, $\beta = -\sum_{i=1}^{V} \pi_i \ln^2 \pi_i$, $h_{\boldsymbol{\pi}} = h$, and

$$\begin{aligned}
\mathsf{Q}(s) &= \mathsf{I} - \boldsymbol{\psi} \otimes \mathbf{p}(s), \\
\mathsf{Q}^{-1}(s) &= \frac{1}{1 - \mathbf{p}(s)\boldsymbol{\psi}}[(1 - \mathbf{p}(s)\boldsymbol{\psi})\mathsf{I} + \boldsymbol{\psi} \otimes \mathbf{p}(s)], \\
\mathsf{Q}(-j)\boldsymbol{\psi} &= (1 - \mathbf{p}(-j)\boldsymbol{\psi})\boldsymbol{\psi},
\end{aligned}$$

where $\mathbf{p}(s) = (\pi_1^{-s}, \ldots, \pi_V^{-s})$, and $\otimes$ is the tensor product of vectors (e.g., the product $\boldsymbol{\psi} \otimes \mathbf{p}(s)$ is a matrix with the ith column equal to $(\pi_i^{-s}, \ldots, \pi_i^{-s})^T$). Thus

$$\dot{\mathsf{Q}}^* \boldsymbol{\psi} = (-\mathbf{p}'(s)\boldsymbol{\psi}\mathsf{I} + \boldsymbol{\psi} \times \mathbf{p}'(s))\boldsymbol{\psi} = 0.$$

We can also prove the following commutation laws

$$\mathsf{Q}^{-1}(-i)\mathsf{Q}^{-1}(-j) = \mathsf{Q}^{-1}(-j)\mathsf{Q}^{-1}(-i), \;\; \dot{\mathsf{Q}}^{-1}(-i)\mathsf{Q}^{-1}(-j) = \mathsf{Q}^{-1}(-j)\dot{\mathsf{Q}}^{-1}(-i)$$

for any $i, j \geq 2$. As a result, we find

$$\begin{aligned}
\sum_{i=2}^{\infty} \dot{\mathsf{Q}}^{-1}(-i)\mathsf{Q}(-i)\boldsymbol{\psi} &= \sum_{i=2}^{\infty} \dot{\mathsf{Q}}^{-1}(-i)(1 - \mathbf{p}(-i)\boldsymbol{\psi})\boldsymbol{\psi} \\
&= \sum_{i=2}^{\infty} \frac{\mathbf{p}'(-i)\boldsymbol{\psi}}{1 - \mathbf{p}(-i)\boldsymbol{\psi}}\boldsymbol{\psi},
\end{aligned}$$

and finally

$$\vartheta = -\sum_{k=1}^{\infty} \frac{\sum_{i=1}^{V} \pi_i^{k+1} \ln \pi_i}{1 - \sum_{i=1}^{V} \pi_i^{k+1}},$$

which coincides with the findings of [23]. In summary, our results for the Markovian source reduce to those of [23] when the source becomes memoryless.

(iii) *Fluctuating Function* $\delta(x)$. A few words of discussion about the fluctuating function $\delta(x)$ is in order. The amplitude of this function is very small, however, it increases with V. For example, for the unbiased memoryless source $|\delta_1(x)| \leq 10^{-6}$ for $V = 2$ (cf. [13, 17]). While this value may be safely ignored in the first order analysis, it is of prime interest to second order analysis. For example, the fluctuating function $\delta_1(x)$ determines the behavior of the Lempel-Ziv redundancy (cf. [15]). In view of this, one may ask for which Markov sources condition (10) holds. We know that for memoryless sources (10) becomes

$$\frac{\ln \pi_i}{\ln \pi_j} \in \mathbb{Q} \qquad i \in \mathcal{A}.$$

The questions whether we can find a non-degenerate Markov source (i.e., which is not a memoryless) that satisfies (10)? The answer is positive, and here is an example. Let $\mathsf{M}(b) = \{e^{-2\pi k_{ij}/b}\}_{i,j=1}^{V}$ for some integers k_{ij} and a positive b where $i, j \in \mathcal{A}$. The matrix $\mathsf{M}(b)$ is positive definite and its main eigenvalue $\lambda(b)$ is real positive with positive right eigenvector $\mathbf{r}(b) = (r_1(b), \ldots, r_V(b))$. Since $\lambda(b) \to 0$ as $b \to 0$ and $\lambda(b) \to V$ as $b \to \infty$, there exists b_0 such that $\lambda(b_0) = 1$. Define now

$$p_{ij} = \frac{r_j(b_0)}{r_i(b_0)} e^{-2\pi k_{ij}/b_0}$$

for $i, j \in \mathcal{A}$. Observe that

$$\sum_{j \in \mathcal{A}} p_{ij} = \frac{1}{r_i(b_0)} \sum_{j \in \mathcal{A}} r_j(b_0) e^{-2\pi k_{ij}/b_0} = \frac{r_i(b_0)}{r_i(b_0)} = 1,$$

since $\mathbf{r}(b_0)$ is the right eigenvector of M with $\lambda(b_0) = 1$. There $\mathsf{P} = \{p_{ij}\}_{i,j \in \mathcal{A}}$ generates a non-degenerated Markov source for which (10) holds. □

We now extend the above results into two directions, namely for a *non-stationary* Markov source and for the MI model with *binomial*(m, r) number of independent sources. Both extensions are crucial for our derivation of results for the GK model (i.e., with fixed number of phrases).

2.2 Markov Independent Model – Non-stationary Source

Let us start with a **non-stationary** Markov source. Observe that our basic set of recurrences (5) for the conditional generating functions $B_m^i(u)$ stays the same, and the only change in our global recurrence (4) for the cumulative generating function $B_m(u)$ reduces to replacing the stationary probability $\boldsymbol{\pi}$ by the initial distribution vector $\mathbf{p}$. As we shall see in Section 3, the asymptotics of the average profile largely depend on the asymptotics of the conditional average profile. This will translate in the same leading terms of the asymptotic expansions of the average depth (phrase length) $D_m(\mathbf{p})$, and the depth of insertion (last phrase length) $I_m(\mathbf{p})$. In fact, the difference is exhibited only in the $O(1)$ term.

We summarize our finding in the following corollary.

Corollary 1 [NON-STATIONARY MARKOV SOURCE] *Consider a Markov source with initial probability vector* $\mathbf{p} = (p_1, \ldots, p_V)$. *Then for large* m

$$\mathbf{E}[D_m(\mathbf{p})] = \frac{1}{h}\left(\ln m + \gamma - 1 + h - h_{\mathbf{p}} - \frac{\beta}{2\omega h} - \vartheta + \delta_3(\ln m)\right) + O\left(\frac{\ln m}{m}\right), \quad (21)$$

$$\mathbf{E}[I_m(\mathbf{p})] = \frac{1}{h}\left(\ln m + \gamma + h - h_{\mathbf{p}} - \frac{\beta}{2\omega h} - \vartheta + \delta_4(\ln m)\right) + O\left(\frac{\ln m}{m}\right), \quad (22)$$

$$\mathbf{Var}[I_m(\mathbf{p})] = \mathbf{Var}[D_m(\mathbf{p})] + O(1) = \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2\right)\ln m + O(1) \quad (23)$$

with the notation as in Theorem 1, where $\delta_3(x)$ *and* $\delta_4(x)$ *are fluctuating functions with small amplitudes. In addition,*

$$\frac{D_m(\mathbf{p}) - \mathbf{E}[D_m(\mathbf{p})]}{\sqrt{\mathbf{Var}D_m(\mathbf{p})}} \to N(0,1), \quad (24)$$

$$\frac{I_m(\mathbf{p}) - \mathbf{E}[I_m(\mathbf{p})]}{\sqrt{\mathbf{Var}I_m(\mathbf{p})}} \to N(0,1) \quad (25)$$

with the rate of convergence $O(1/\sqrt{\ln m})$. *Moreover, there exist positive constants* A *and* $\alpha < 1$ *such that*

$$\Pr\left\{\left|\frac{D_m(\mathbf{p}) - \mathbf{E}[D_m(\mathbf{p})]}{\sqrt{\mathbf{Var}D_m(\mathbf{p})}}\right| \geq k\right\} \leq A\alpha^k, \quad (26)$$

$$\Pr\left\{\left|\frac{I_m(\mathbf{p}) - \mathbf{E}[I_m(\mathbf{p})]}{\sqrt{\mathbf{Var}I_m(\mathbf{p})}}\right| \geq k\right\} \leq A\alpha^k \quad (27)$$

uniformly in k.

Finally, we consider the MI model, in which the number of sources M is a random variable distributed as $B(m, r) := binomial(m, r)$, that is,

$$\Pr\{M = k\} = \binom{m}{k} r^k (1-r)^{m-k}.$$

Let D_m^B and I_m^B (or $D_m^{B(r)}$ and $I_m^{B(r)}$) denote, respectively, the typical depth and the depth of insertion in such a model.

Corollary 2 [Random number of non-stationary Markov Sources] *Consider a Markov source with initial probability vector* $\mathbf{p} = (p_1, \ldots, p_V)$ *and random number, M, of sources distributed as the binomial*(m, r)*. Then for large* m

$$\mathbf{E}[D_m^B(\mathbf{p})] = \frac{1}{h}\left(\ln(mr) + \gamma - 1 + h - h_{\mathbf{p}} - \frac{\beta}{2\omega h} - \vartheta + \delta_5(\ln m)\right) + O\left(\frac{\ln m}{m}\right), \quad (28)$$

$$\mathbf{E}[I_m^B(\mathbf{p})] = \frac{1}{h}\left(\ln(mr) + \gamma + h - h_{\mathbf{p}} - \frac{\beta}{2\omega h} - \vartheta + \delta_6(\ln m)\right) + O\left(\frac{\ln m}{m}\right), \quad (29)$$

$$\mathbf{Var}[I_m^B(\mathbf{p})] = \mathbf{Var}[D_m^B(\mathbf{p})] + O(1) = \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2\right)\ln(mr) + O(1) \quad (30)$$

where $\delta_5(x)$ *and* $\delta_6(x)$ *are fluctuating functions with small amplitudes. In addition,*

$$\frac{D_m^B(\mathbf{p}) - \mathbf{E}[D_m^B(\mathbf{p})]}{\sqrt{\mathbf{Var} D_m^B(\mathbf{p})}} \to N(0,1), \quad (31)$$

$$\frac{I_m^B(\mathbf{p}) - \mathbf{E}[I_m^B(\mathbf{p})]}{\sqrt{\mathbf{Var} I_m^B(\mathbf{p})}} \to N(0,1) \quad (32)$$

with the rate of convergence $O(1/\sqrt{\ln m})$*. Finally, there exist positive constants* A *and* $\alpha < 1$ *such that*

$$\Pr\left\{\left|\frac{D_m^B(\mathbf{p}) - \mathbf{E}[D_m^B(\mathbf{p})]}{\sqrt{\mathbf{Var} D_m^B(\mathbf{p})}}\right| \geq k\right\} \leq A\alpha^k, \quad (33)$$

$$\Pr\left\{\left|\frac{I_m^B(\mathbf{p}) - \mathbf{E}[I_m^B(\mathbf{p})]}{\sqrt{\mathbf{Var} I_m^B(\mathbf{p})}}\right| \geq k\right\} \leq A\alpha^k \quad (34)$$

uniformly in k.

Proof. Let us only consider the typical depth D_m^{MIB}, where the superindex MIB indicates that we analyze the MI model with *binomial*(m, r) number of sources. The proof follows immediately from the fact that the generating function $D_m^{MIB}(u)$ satisfies

$$D_m^{MIB}(u) = \sum_{k=0}^{m} \binom{m}{k} r^k (1-r)^{m-k} D_k(u),$$

where $D_k(u)$ is the generating function of the typical depth in the MI model with k Markov sources. Observe now that the Poisson transform of D_m^B satisfies

$$\widetilde{D}^B(z, u) = \widetilde{D}(zr, u)e^{-zr}$$

where $\widetilde{D}(z,u)$ is the Poisson transform of the MI model with fixed number of sources (and already presented in Theorem 1 while the analysis can be found in Section 3). The moments can be also recovered from the following formula recently proved in [5, 12] (interestingly, analytic depoissonization was used to derive it, too)

$$\sum_{k=0}^{m} \binom{m}{k} r^k (1-r)^{m-k} \ln k = \ln(mr) - \frac{1-r}{2mr} + \sum_{i\geq 2} \frac{a_i}{m^i}$$

where the coefficients a_k are explicitly computable. ■

2.3 Fixed Number of Phrases Model — Gilbert-Kadota Model

In this subsection, we present our main findings for the Gilbert-Kadota model in which a single Markovian source generates a (possibly infinite) sequence that is partitioned according to the Lempel-Ziv algorithm until m full phrases are obtained. As before, we study the typical phrase length D_m and the last phrase length I_m. To avoid confusions, we often append an upper index MI or GK to D_m and I_m to denote the typical phrase length and last phrase length in the MI model and the GK model, respectively. Furthermore, as before, it is convenient to build a digital search tree out of these m phrases, as shown in Figure 1. We observe, however, that this time the DST is built from *suffixes* of a single Markovian sequence, thus we might call it a **suffix** digital search tree. Clearly, the typical phrase length D_m^{GK} becomes the typical depth, and the last phrase length I_m^{GK} corresponds to the depth of insertion in the associated DST.

The GK model introduces some tricky statistical dependency between phrases. The recurrence (4) and the differential-functional equation (5) do not hold any more, however, the relationship (3) between the typical depth and the depth of insertion is still true. To analyze GK model, we use *stochastic dominance*, that is, we (asymptotically) bound in a stochastic sense defined below the depth of insertion I_m^{GK} by the depth of insertion in the modified MI model. More precisely, in the GK model, we delete K phrases, thus making a "gap" of significant size so that the newly inserted phrase resembles the one in the MI model, hence results of MI model can be applied.

To present more succinctly our analysis, we introduce some new notation. We say that I'_m *stochastically dominates* I_m and write $I_m \leq_{\rm st} I'_m$ if for every k we have

$$\Pr\{I_m \geq k\} \leq \Pr\{I'_m \geq k\}.$$

The *asymptotic stochastic dominance* denoted as $I_m \preceq_{\rm st} I'_m$ is defined next.

Definition 1 (i) *Let X and Y be two integer random variables, and $\varepsilon > 0$. We say that X is at distance ε from Y and write it as $d(X, Y) \leq \varepsilon$ if for all integers k*

$$|\Pr\{X \geq k\} - \Pr\{Y \geq k\}| < \varepsilon. \tag{35}$$

(ii) *We say that the sequence of random variables Y_m asymptotically dominates X_m or shortly*

$$X_m \preceq_{\text{st}} Y_m$$

if

$$\limsup_{m\to\infty} \max_k \left(\Pr\{X_m \geq k\} - \Pr\{Y_m \geq k\}\right) = 0 \ . \tag{36}$$

The last definition is illustrated well by the following simple result.

Lemma 1 *If $X_m \leq_{\text{st}} Y'_m$ and $\lim_{m\to\infty} d(Y_m, Y'_m) = 0$, then $X_m \preceq_{\text{st}} Y_m$.*

Proof. By assumptions, for all integers k and m we have $\Pr\{X_m \geq k\} \leq \Pr\{Y'_m \geq k\}$ and $\lim_{m\to\infty} \max_k |\Pr\{Y'_m \geq k\} - \Pr\{Y_m \geq k\}| = 0$. Thus (36) follows. ■

In the next section, we establish certain inequalities between the MI model and the GK model, that we review briefly here. For some $K < m$ we denote by I_{m-K+1} the depth of insertion to a DST tree that is built from any subset of size $m - K$ of m original phrases. It is easy to see that in both models we have the following (deterministic) inequality

$$I_{m-K+1} \leq I_{m+1} \leq I_{m-K+1} + K, \tag{37}$$

provided *the same* phrase is inserted. The left-hand size is quite obvious, while the right-hand size is a consequence of the fact that a new phrase can be incremented at most by one symbol. In other words, the DST tree does not have unary nodes (i.e., nodes with degree one).

In view of this, we can work on I_{m-K} in which K phrases are (conveniently) deleted loosing up dependencies between phrases. We consider now the MI model such that *all* phrases start with a given, but otherwise arbitrary symbol, say $a \in \mathcal{A}$. In other words, we consider a non-stationary model with the initial vector $\mathbf{p}_a$ that contains all zeros except 1 at the position corresponding to symbol a, that is, $\mathbf{p}_a = (0, \ldots, 1, \ldots, 0)$. We denote $I_m^{MI}(\mathbf{p}_a)$ the depth of insertion in this model. We also consider the GK model conditioned on the fact that the mth phrase starts with symbol a. We denote $I_{m,K}^{GK}(\mathbf{p}_a)$ the depth of insertion of the mth phrase when K phrases are deleted before it. We shall prove in Section 4 that there exists $K = O(1)$ such that

$$I_{m-K}^{MIB(r)}(\mathbf{p}_a) \preceq_{\text{st}} I_{m,K}^{GK}(\mathbf{p}_a) \preceq_{\text{st}} I_{m-K}^{MI}(\mathbf{p}_a) + K, \tag{38}$$

where $I_{m-K}^{MIB(r)}(\mathbf{p}_a)$ is the depth of insertion in the MI model with the *binomial*$(r, m-K)$ number of phrases for some $0 < r < 1$. Thus, based on our results from the previous section, we shall be able to prove the following theorem.

Theorem 2 *Consider a Markov source with initial probability vector* $\mathbf{p}$. *Then for large* m

$$\mathbf{E}[D_m^{GK}(\mathbf{p})] = \mathbf{E}[I_m^{GK}(\mathbf{p})] + O(1) = \frac{1}{h}\ln m + O(1), \tag{39}$$

$$\mathbf{Var}[D_m^{GK}(\mathbf{p})] = \mathbf{Var}[I_m^{GK}(\mathbf{p})] + O(1) = \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2\right)\ln m + O(1) \tag{40}$$

with the notation as in Theorem 1, and

$$\frac{D_m^{GK}(\mathbf{p}) - \mathbf{E}[D_m^{GK}(\mathbf{p})]}{\sqrt{\mathbf{Var}D_m^{GK}(\mathbf{p})}} \to N(0,1), \tag{41}$$

$$\frac{I_m^{GK}(\mathbf{p}) - \mathbf{E}[I_m^{GK}(\mathbf{p})]}{\sqrt{\mathbf{Var}I_m^{GK}(\mathbf{p})}} \to N(0,1) \tag{42}$$

with the rate of convergence $O(1/\sqrt{\ln m})$. *In addition, the normalized* $D_m^{GK}(\mathbf{p})$ *and* $I_m^{GK}(\mathbf{p})$ *converge in moments to the corresponding moments of the standard normal distribution.*

2.4 Lempel-Ziv Model

Finally, we deal with the Lempel-Ziv model, in which a Markov sequence of fixed length n is partitioned in (a random number) M_n of (full) phrases. As before, I_i represents the ith phrase for $1 \le i \le M_n$. We write J_n for the last full phrase, which also becomes $J_n = I_{M_n}$. The typical phrase length Δ_n is defined as follows:

$$\Pr\{\Delta_n = k\} = \sum_{m=M_{\min}}^{M_{\max}} \frac{1}{m}\sum_{i=1}^{m} \Pr\{I_i = k \ \&\ M_n = m\}, \tag{43}$$

where $M_{\min} = O(\sqrt{n})$ is the minimum number of phrases and $M_{\max} = O(n/\log_2 n)$ is the maximum number of phrases (cf. [14]). In passing, we should observe that there is a relationship between the phrase length I_i and the number of phrases M_n. Indeed,

$$M_n = \max\{m : \sum_{i=1}^{m} I_i^{GK} \le n\},$$

where in the above we explicitly show that the phrase length I_i^{GK} is the one corresponding to the phrase length in the GK model.

Using Theorem 2, we shall prove below the following result. We shall write below $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Theorem 3 *Let a Markov source generates a single sequence of length n. Then, for large n*

$$\frac{\Delta_n - \mathbf{E}[\Delta_n]}{\sqrt{\mathbf{Var}\Delta_n}} \to N(0,1). \tag{44}$$

In addition, Δ_n converges in moments, and in particular

$$\mathbf{E}[\Delta_n] \sim \mathbf{E}[J_n] \sim \frac{1}{h}\ln(nh/\ln n), \tag{45}$$

$$\mathbf{Var}[\Delta_n] \sim \mathbf{Var}[J_n] \sim \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2\right)\ln(nh/\ln n). \tag{46}$$

provided the number of phrases M_n converges to its mean exponentially fast.

Proof. Let $\mu(n) = \frac{nh}{\ln(n)}$ and $\delta_n = \Pr\{M_n \notin ((1-\varepsilon)\mu(n), (1+\varepsilon)\mu(n))\}$. Observe that $\delta_n \to 0$ as $n \to \infty$ (cf. [29]).

We now prove that for any $\varepsilon > 0$ and for all set of integers B

$$\limsup_{n\to\infty} \max_B \left(\Pr\{\Delta_n \in B\} - \Pr\{D_{\lfloor(1+\varepsilon)\mu(n)\rfloor} \in B\}\right) = 0 \tag{47}$$

and

$$\limsup_{n\to\infty} \max_B \left(\Pr\{D_{\lfloor(1-\varepsilon)\mu(n)\rfloor} \in B\} - \Pr\{\Delta_n \in B\}\right) = 0. \tag{48}$$

We rewrite (43) as

$$\Pr\{\Delta_n \in B\} = \sum_{m=1}^{n} \frac{1}{m}\sum_{\ell=1}^{m} \Pr\{I_\ell^{GK} \in B \;\&\; M_n = m\}$$

for any set of integers B. Then

$$\Pr\{\Delta_n \in B\} \le \delta_n + \sum_{m=\lceil(1-\varepsilon)\mu(n)\rceil}^{\lfloor(1+\varepsilon)\mu(n))\rfloor} \frac{1}{m}\sum_{\ell=1}^{m} \Pr\{I_\ell^{GK} \in B \;\&\; M_n = m\},$$

where δ_n is defined above. We have the following chain of inequalities:

$$\begin{aligned}
\Pr\{\Delta_n \in B\} &\le \delta_n + \sum_{m=\lceil(1-\varepsilon)\mu(n)\rceil}^{\lfloor(1+\varepsilon)\mu(n))\rfloor} \frac{1}{m} \sum_{\ell=1}^{\lfloor(1+\varepsilon)\mu(n)\rfloor} \Pr\{I_\ell^{GK} \in B \;\&\; M_n = m\} \\
&\le \delta_n + \sum_{m=\lceil(1-\varepsilon)\mu(n)\rceil}^{\lfloor(1+\varepsilon)\mu(n))\rfloor} \frac{1}{(1-\varepsilon)\mu(n)} \sum_{\ell=1}^{\lfloor(1+\varepsilon)\mu(n)\rfloor} \Pr\{I_\ell^{GK} \in B \;\;\&\;\; M_n = m\} \\
&\le \delta_n + (\frac{1+\varepsilon}{1-\varepsilon}) \sum_{m=\lceil(1-\varepsilon)\mu(n)\rceil}^{\lfloor(1+\varepsilon)\mu(n))\rfloor} \Pr\{D^{GK}_{\lfloor(1+\varepsilon)\mu(n)\rfloor} \in B \;\&\; M_n = m\} \\
&\le \delta_n + \left(\frac{1+\varepsilon}{1-\varepsilon}\right)\delta_n + \left(\frac{1+\varepsilon}{1-\varepsilon}\right)\Pr\{D^{GK}_{\lfloor(1+\varepsilon)\mu(n)\rfloor} \in B\}
\end{aligned}$$

In a similar manner, we prove a lower bound

$$\Pr\{\Delta_n \in B\} \geq \left(\frac{1+\varepsilon}{1-\varepsilon}\right)\delta_n + \left(\frac{1+\varepsilon}{1-\varepsilon}\right)\Pr\{D^{GK}_{\lfloor(1-\varepsilon)\mu(n)\rfloor} \in B\}.$$

The above two inequalities prove (47) and (48). The convergence in moments follows from the above and the assumed exponential convergence of M_n to its mean. ■

Remark. We should point out that Merhav [18] proved that for Markov sources

$$\Pr\{M_n \geq \mu(n)(1+\varepsilon)\} \leq (1+o(1))e^{-\alpha n}$$

for a constant $\alpha > 0$ and $\varepsilon = O(1/\sqrt{\log n})$. □

As a consequence of Theorems 2 and 3, we can derive bounds on the average redundancy rate $\bar{R}_n$ of the Lempel-Ziv code for Markovian sources. To recall, consider a Markovian sequence of length n for which the Lempel-Ziv code is ℓ_n. Then the redundancy rate is defined as

$$R_n = \frac{\ell_n - nh}{n}.$$

We denote by $\bar{R}_n = \mathbf{E}[R_n]$ the average redundancy rate. Using the approach of [15], we obtain from Theorem 2 the following bounds (we assumed $K = 1$ in (38))

$$\begin{aligned} h\left(2 - \ln r - \gamma + h_{\mathbf{r}} + \frac{\beta}{2\omega h} + \vartheta - \delta_3(\ln n)\right) \quad &\leq \quad \bar{R}_n \ln n + o(1) \leq \\ &\leq \quad h\left(2 - \gamma - h + h_{\mathbf{p}_a} + \frac{\beta}{2\omega h} + \vartheta - \delta_2(\ln n)\right), \end{aligned}$$

where $r = \sum_{a\in\mathcal{A}} \min_i\{p_{ia}\}$ and $\mathbf{r}$ is a vector of size V whose jth component is equal to $\min_i\{p_{ij}\}/r$ (cf. Lemma 11). These bounds should be compared to Savari's upper bound for Markov sources (cf. [21]).

3 Analysis of Markov Independent Model

As mentioned before, the analysis of MI model is at the heart of our contribution to analytic information theory. In view of this, we present here a detailed proof. It is based on such analytic techniques as: analytic poissonization, Mellin transform, singularities of a complex matrix, and analytic depoissonization.

3.1 Poissonization and Mellin Transforms: Analysis of Moments

We first consider the *stationary* Markov source. The generating function $B_m(u)$ of the average profile satisfies (4) with the initial vector $\mathbf{p} = \boldsymbol{\pi}$. Observe that the conditional generating functions $B^i_m(u)$ fulfill the system of recurrence equations (5). We shall first deal with (5). There is no easy way to solve these recurrences, and therefore, we transform them to the Poisson model, in which m is replaced by a Poisson random variable with mean (complex) z that becomes m when z is restricted to positive integers. Let

$$\widetilde{B}^i(z,u) = \sum_{n=1}^{\infty} B^i_n(u)\frac{z^n}{n!}e^{-z}, \quad i \in \mathcal{A}$$

be the Poisson transform of $B^i_m(u)$. In addition, we shall write $\widetilde{B}^i_z(z,u) := \frac{\partial}{\partial z}\widetilde{B}^i(z,u)$ for the derivative of $\widetilde{B}^i(z,u)$ with respect to z. After some simple algebra, we have the following Poissonized differential–functional equations of recurrences (4) and (5)

$$\widetilde{B}_z(z,u) + \widetilde{B}(z,u) = u[\widetilde{B}^1(\pi_1 z,u) + \cdots + \widetilde{B}^V(\pi_V z,u)] + 1, \tag{49}$$

and

$$\widetilde{B}^i_z(z,u) + \widetilde{B}^i(z,u) = u[\widetilde{B}^1(p_{i1} z,u) + \cdots + \widetilde{B}^V(p_{iV} z,u)] + 1 \qquad \text{for all} \quad i \in \mathcal{A}. \tag{50}$$

Let us now concentrate on the evaluation of the first two moments of the depth, that is, we compute the first two derivatives of $\widetilde{B}(z,u)$ with respect to u at $u=1$. We obtain the following two systems of functional equations after taking into account that $\widetilde{B}^i(z,1) = z$, $\pi_1 + \cdots + \pi_V = 1$, and $\sum_{j=1}^V p_{ij} = 1$:

$$\widetilde{B}_{zu}(z,1) + \widetilde{B}_u(z,1) = z + [\widetilde{B}^1_u(\pi_1 z,1) + \cdots + \widetilde{B}^V_u(\pi_V z,1)], \tag{51}$$

$$\begin{aligned}
\widetilde{B}^1_{zu}(z,1) + \widetilde{B}^1_u(z,1) &= z + [\widetilde{B}^1_u(p_{11} z,1) + \cdots + \widetilde{B}^V_u(p_{1V} z,1)] \\
\cdots &= \cdots \\
\widetilde{B}^V_{zu}(z,1) + \widetilde{B}^V_u(z,1) &= z + [\widetilde{B}^1_u(p_{V1} z,1) + \cdots + \widetilde{B}^V_u(p_{VV} z,1)],
\end{aligned}$$

and

$$\widetilde{B}_{zuu}(z,1) + \widetilde{B}_{uu}(z,1) = 2[\widetilde{B}^1_u(\pi_1 z,1) + \cdots + \widetilde{B}^V_u(\pi_V z,1)] + [\widetilde{B}^1_{uu}(\pi_1 z,1) + \cdots + \widetilde{B}^V_{uu}(\pi_V z,1)], \tag{52}$$

$$\begin{aligned}
\widetilde{B}^1_{zuu}(z,1) + \widetilde{B}^1_{uu}(z,1) &= 2[\widetilde{B}^1_u(p_{11} z,1) + \cdots + \widetilde{B}^V_u(p_{1V} z,1)] + [\widetilde{B}^1_{uu}(p_{11} z,1) + \cdots + \widetilde{B}^V_{uu}(p_{1V} z,1)] \\
\cdots &= \cdots \\
\widetilde{B}^V_{zuu}(z,1) + \widetilde{B}^V_{uu}(z,1) &= 2[\widetilde{B}^1_u(p_{V1} z,1) + \cdots + \widetilde{B}^V_u(p_{VV} z,1)] + [\widetilde{B}^1_{uu}(p_{V1} z,1) + \cdots + \widetilde{B}^V_{uu}(p_{VV} z,1)].
\end{aligned}$$

Our goal is now to solve asymptotically (as $z \to \infty$ in a cone around $\Re(z) > 0$) the above two sets of functional equations. It is well known that equations like these are amiable to attack by the Mellin transform (cf. [6]). To recall, for a function $f(x)$ of real x, we define its Mellin transform $F^*(s)$ as

$$F^*(s) = \mathcal{M}[f(t);s] = \int_0^\infty f(t)t^{s-1}dt \ .$$

In some of our arguments we could use either Mellin transform of a complex variable function $f(z)$ or an analytical continuation argument. It is known (cf. [10]) that as long as $\arg(z)$ belongs to some cone around the real axis, the Mellin transform $F(s)$ of a function $f(x)$ of a real argument and its corresponding function of a complex argument is the same. Therefore, we work most of the time with the Mellin transform of a function of real variable as defined above.

In our case, a direct solution through Mellin transform does not work well, and therefore we factorize the Mellin transforms of the above functions as follows:

$$B_i^*(s) := \mathcal{M}[\widetilde{B}_u^i(z,1);s] = \Gamma(s)x_i(s), \quad i \in \mathcal{A} \tag{53}$$

$$B^*(s) := \mathcal{M}[\widetilde{B}_u(z,1);s] = \Gamma(s)x(s), \tag{54}$$

$$C_i^*(s) := \mathcal{M}[\widetilde{B}_{uu}^i(z,1);s] = \Gamma(s)v_i(s), \quad i \in \mathcal{A} \tag{55}$$

$$\tag{56}$$

$$C^*(s) := \mathcal{M}[\widetilde{B}_{uu}(z,1);s] = \Gamma(s)v(s), \tag{57}$$

where $\Gamma(s)$ is the Euler gamma function, and $x_i(s)$, $x(s)$, $v_i(s)$ and $v(s)$ are unknown. The lemma below establishes the existence of the above Mellin transforms.

Lemma 2 *The Mellin transforms $B_i^*(s)$, $B^*(s)$ and $C_i^*(s)$, $C^*(s)$ exist for $\Re(s) \in (-2,-1)$. In addition,*

$$\begin{aligned} x_i(-2) &= 1, \quad x(-2) = 1, \\ v_i(-2) &= 0, \quad v(-2) = 0. \end{aligned}$$

Proof. The proof is quite standard and replies on the Lemma 2 from [16]. We leave the details to the interested reader. ■

Now, we are ready to compute the Mellin transforms of $\widetilde{B}_u^i(z,1)$, $\widetilde{B}_{uu}^i(z,1)$ (cf. (51) and (52), respectively) with respect to z. We obtain

$$-(s-1)B^*(s-1) + B^*(s) = B_1^*(s)\pi_1^{-s} + \cdots + B_V^*(s)\pi_V^{-s}, \tag{58}$$

$$\begin{aligned} -(s-1)B_1^*(s-1) + B_1^*(s) &= B_1^*(s)p_{11}^{-s} + \cdots + B_V^*(s)p_{1V}^{-s}, \\ \cdots &= \cdots \\ -(s-1)B_V^*(s-1) + B_V^*(s) &= B_1^*(s)p_{V1}^{-s} + \cdots + B_V^*(s)p_{VV}^{-s}, \end{aligned}$$

and

$$-(s-1)C^*(s-1) + C^*(s) = 2[B_1^*(s)\pi_1^{-s} + \cdots + B_V^*(s)\pi_V^{-s}] + [C_1^*(s)\pi_1^{-s} + \cdots + C_V^*(s)\pi_V^{-s}], \tag{59}$$

$$
\begin{aligned}
-(s-1)C_1^*(s-1)+C_1^*(s) &= 2[B_1^*(s)p_{11}^{-s}+\cdots+B_V^*(s)p_{1V}^{-s}]+[C_1^*(s)p_{11}^{-s}+\cdots+C_V^*(s)p_{1V}^{-s}],\\
\cdots &= \cdots\\
-(s-1)C_V^*(s-1)+C_V^*(s) &= 2[B_1^*(s)p_{V1}^{-s}+\cdots+B_V^*(s)p_{VV}^{-s}]+[C_1^*(s)p_{V1}^{-s}+\cdots+C_V^*(s)p_{VV}^{-s}].
\end{aligned}
$$

In the above, we used the following two properties of the Mellin transform (cf. [6]):

$$
\begin{aligned}
\mathcal{M}[f(ax);s] &= a^{-s}F^*(s),\\
\mathcal{M}[f'(x);s] &= -(s-1)F^*(s-1).
\end{aligned}
$$

To solve these functional equations in a compact form, we define:

$$
\mathbf{x}(s)=\begin{bmatrix} x_1(s)\\ x_2(s)\\ \vdots\\ x_V(s)\end{bmatrix},\quad \mathbf{v}(s)=\begin{bmatrix} v_1(s)\\ v_2(s)\\ \vdots\\ v_V(s)\end{bmatrix} \tag{60}
$$

and

$$
\mathbf{b}(s)=\begin{bmatrix} B_1^*(s)\\ B_2^*(s)\\ \vdots\\ B_V^*(s)\end{bmatrix}=\Gamma(s)\mathbf{x}(s),\quad \mathbf{c}(s)=\begin{bmatrix} C_1^*(s)\\ C_2^*(s)\\ \vdots\\ C_V^*(s)\end{bmatrix}=\Gamma(s)\mathbf{v}(s). \tag{61}
$$

Using $\Gamma(s)=(s-1)\Gamma(s-1)$, the system of equations (58) and (59) become

$$
\begin{aligned}
\mathbf{x}(s)-\mathbf{x}(s-1) &= \mathsf{P}(s)\mathbf{x}(s),\\
\mathbf{v}(s)-\mathbf{v}(s-1) &= 2\mathsf{P}(s)\mathbf{x}(s)+\mathsf{P}(s)\mathbf{v}(s),
\end{aligned}
$$

where $\mathsf{P}=\{p_{ij}^{-s}\}_{i,j\in\mathcal{A}}$. Thus

$$
\mathbf{x}(s) = \mathsf{Q}^{-1}(s)\mathbf{x}(s-1)=\left(\prod_{i=0}^{\infty}\mathsf{Q}^{-1}(s-i)\right)\mathsf{K}, \tag{62}
$$

$$
\mathbf{v}(s) = 2\mathsf{Q}^{-1}(s)\mathsf{P}(s)\mathbf{x}(s)+\mathsf{Q}^{-1}(s)\mathbf{v}(s-1), \tag{63}
$$

where $\mathsf{Q}=\mathsf{I}-\mathsf{P}$ and I is the identity matrix, and K is defined in (6). The formula on K follows from Lemma 2 (i.e., $\mathbf{x}(-2)=(1,\ldots,1)^T$) and (62). In the next section we prove the convergence of the above infinite product (cf. Lemma 4), however, we shall not use this explicit infinite product solution anywhere in our further analysis.

Thus far we have obtained the Mellin transforms of the conditional generating functions $\widetilde{B}_i(z,1)$. In order to obtain the composite Mellin transform $B^*(s)$ and $C^*(s)$ of $\widetilde{B}_u(z,1)$ and

$\widetilde{B}_{uu}(z,1)$, respectively, we refer to (58) and (59). After some algebra, we finally obtain

$$B^*(s) = \mathbf{p}(s)\mathbf{b}(s) + \Gamma(s)x(s-1), \tag{64}$$

$$C^*(s) = 2\mathbf{p}(s)\mathbf{b}(s) + \mathbf{p}(s)\mathbf{c}(s) + \Gamma(s)v(s-1), \tag{65}$$

where $\mathbf{p}(s) = (\pi_1^{-s}, \ldots, \pi_V^{-s})$ in the stationary case, and $\mathbf{p}(s) = (p_1^{-s}, \ldots, p_V^{-s})$ in the nonstationary case. We shall see that the dominant asymptotics of $B^*(s)$ and $C^*(s)$ are determined by asymptotics of $\mathbf{b}(s)$ and $\mathbf{c}(s)$, which depend on singularities of $\mathsf{Q}(s)$ that we discuss next.

3.2 Singularities of the Matrix $\mathsf{Q}(s)$

We study here singularities of the matrix $\mathsf{Q}(s)$, which play central role in the asymptotic analysis of the depth. We prove the following lemma that characterizes the location of singularities of $\mathsf{Q}(s)$.

Lemma 3 *Let* $\mathsf{Q}(s) = \mathsf{I} - \mathsf{P}(s)$ *and* $\mathsf{P}(s) = \{p_{ij}^{-s}\}_{i,j\in\mathcal{A}}$. *Let* s_l *denote singularities of* $\mathsf{Q}(s)$, *where* $l \in \mathbb{Z}$ *is an integer. Then:*

(i) *Matrix* $\mathsf{Q}(s)$ *is nonsingular for* $\Re(s) < -1$, *and* $s_0 = -1$ *is a simple pole.*

(ii) *If and only if*

$$\frac{\ln p_{ij} + \ln p_{1i} - \ln p_{1j}}{\ln p_{11}} \in \mathbb{Q}, \qquad i,j \in \mathcal{A} \tag{66}$$

where $\mathbb{Q}$ *is the set of rational numbers, matrix* $\mathsf{Q}(s)$ *has simple poles on the line* $\Re(s) = -1$ *that can be written as*

$$s_l = -1 + l\theta\mathbf{i}$$

where $\mathbf{i} = \sqrt{-1}$ *and*

$$\theta = \frac{n_1}{n_2}\left|\frac{2\pi}{\ln p_{11}}\right|.$$

The integers n_1, n_2 *are such that* $\left\{\left|\frac{n_1}{n_2 \ln p_{11}}(\ln p_{ij} - \ln p_{1i} + \ln p_{1j})\right|\right\}_{ij=1}^{V}$ *is a set of relative primes.*

(iii) *Finally,*

$$\mathsf{Q}(-1 + l\theta\mathbf{i}) = \mathsf{E}^{-l}\mathsf{Q}(-1)\mathsf{E}^{l}$$

where $\mathsf{E} = \mathrm{diag}(1, e^{\theta_{12}\mathbf{i}}, \ldots, e^{\theta_{1V}\mathbf{i}})$ *is the diagonal matrix with* $\theta_{ik} = -\theta \ln p_{ik}$.

Proof. Observe that for $\Re(s) < -1$,

$$|1 - p_{ii}^{-s}| \geq 1 - |p_{ii}^{-s}| > 1 - p_{ii} = \sum_{j\neq i} p_{ij} \geq \sum_{j\neq i} |p_{ij}^{-s}|, \tag{67}$$

hence $\mathsf{Q}(s)$ is a strictly diagonal dominant matrix, and therefore nonsingular.

Now, we proceed with the proof of part (ii) of the lemma. For $b \neq 0$ such that $\mathsf{Q}(-1+b\mathbf{i})$ is singular, let $\mathbf{x} = [x_1, x_2, ..., x_V]^T \neq 0$ be a solution of $\mathsf{Q}(-1+b\mathbf{i})\mathbf{x} = 0$, where

$$\mathsf{Q}(-1+b\mathbf{i}) = \begin{bmatrix} 1-p_{11}e^{\xi_{11}\mathbf{i}} & -p_{12}e^{\xi_{12}\mathbf{i}} & \dots & -p_{1V}e^{\xi_{1V}\mathbf{i}} \\ -p_{21}e^{\xi_{21}\mathbf{i}} & 1-p_{22}e^{\xi_{22}\mathbf{i}} & \dots & -p_{2V}e^{\xi_{2V}\mathbf{i}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{i1}e^{\xi_{i1}\mathbf{i}} & -p_{i2}e^{\xi_{i2}\mathbf{i}} & \dots & -p_{iV}e^{\xi_{iV}\mathbf{i}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{V1}e^{\xi_{V1}\mathbf{i}} & -p_{V2}e^{\xi_{V2}\mathbf{i}} & \dots & 1-p_{VV}e^{\xi_{VV}\mathbf{i}} \end{bmatrix}$$

with $\xi_{ik} = -b\ln p_{ik}$. Without loss of generality, suppose $|x_1| = \max\{|x_1|, |x_2|, ..., |x_V|\} \neq 0$ (since $\mathsf{Q}(-1+b\mathbf{i})$ is singular). Then

$$(1-p_{11}e^{\xi_{11}\mathbf{i}})x_1 - p_{12}e^{\xi_{12}\mathbf{i}}x_2 - ... - p_{1V}e^{\xi_{1V}\mathbf{i}}x_V = 0,$$

implies

$$1-p_{11}e^{\xi_{11}\mathbf{i}} = p_{12}e^{\xi_{12}\mathbf{i}}x_2/x_1 + ... + p_{1V}e^{\xi_{1V}\mathbf{i}}x_V/x_1.$$

But as in (67)

$$|1-p_{11}e^{\xi_{11}\mathbf{i}}| \geq 1-p_{11},$$

and

$$|p_{12}e^{\xi_{12}\mathbf{i}}x_2/x_1 + ... + p_{1V}e^{\xi_{1V}\mathbf{i}}x_V/x_1| \leq p_{12} + ... + p_{1V} = 1-p_{11}.$$

Thus

$$1-p_{11}e^{\xi_{11}\mathbf{i}} = 1-p_{11},$$

$$p_{12}e^{\xi_{12}\mathbf{i}}x_2/x_1 + ... + p_{1V}e^{\xi_{1V}\mathbf{i}}x_V/x_1 = p_{12} + ... + p_{1V}.$$

This implies

$$e^{\xi_{11}\mathbf{i}} = e^{\xi_{1i}\mathbf{i}}x_i/x_1 = 1,$$

and $|x_i| = |x_j|$ for any $i, j = 1, 2, ..., V$, so that $e^{\xi_{ii}\mathbf{i}} = 1$ for all i. Define now ξ_i such that $x_i/x_1 = e^{-\xi_{1i}\mathbf{i}} = e^{\xi_i\mathbf{i}}$. Then

$$-p_{j1}e^{\xi_{j1}\mathbf{i}} - p_{j2}e^{\xi_{j2}\mathbf{i}}e^{\xi_2\mathbf{i}} - ... - p_{j(j-1)}e^{\xi_{j(j-1)}\mathbf{i}}e^{\xi_{j-1}\mathbf{i}} + (1-p_{jj})e^{\xi_j\mathbf{i}} - ..., -p_{jV}e^{\xi_{jV}\mathbf{i}}e^{\xi_V\mathbf{i}} = 0$$

for any $1 \leq j \leq V$. Note that since

$$-p_{j1} - p_{j2} - ... - p_{j(j-1)} + 1 - p_{jj} - \cdots - p_{jV} = 0,$$

we must have $e^{\xi_{ji}\mathbf{i}}e^{\xi_i\mathbf{i}}e^{-\xi_j\mathbf{i}}=1$, and thus

$$e^{\xi_{ji}\mathbf{i}}=e^{(\xi_j-\xi_i)\mathbf{i}}.$$

Hence $-b(\ln p_{ji}+\ln p_{1j}-\ln p_{1i})=2\pi n_{ji}$ for some integer n_{ji}, and as a consequence $(\ln p_{ij}+\ln p_{1i}-\ln p_{1j})/\ln p_{11}$ is rational for any $i,j=1,2,...,V$.

To prove the inverse part of (ii), suppose b is such that $|\frac{b}{2\pi}(\ln p_{ji}+\ln p_{1j}-\ln p_{1i})|$ are integers for any $i,j=1,2,...,V$. Then

$$\begin{aligned}
\mathsf{Q}(-1+b\mathbf{i}) &= \begin{bmatrix} 1-p_{11}e^{\xi_{11}\mathbf{i}} & -p_{12}e^{\xi_{12}\mathbf{i}} & \dots & -p_{1V}e^{\xi_{1V}\mathbf{i}} \\ -p_{21}e^{\xi_{21}\mathbf{i}} & 1-p_{22}e^{\xi_{22}\mathbf{i}} & \dots & -p_{2V}e^{\xi_{2V}\mathbf{i}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{i1}e^{\xi_{i1}\mathbf{i}} & -p_{i2}e^{\xi_{i2}\mathbf{i}} & \dots & -p_{iV}e^{\xi_{iV}\mathbf{i}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{V1}e^{\xi_{V1}\mathbf{i}} & -p_{V2}e^{\xi_{V2}\mathbf{i}} & \dots & 1-p_{VV}e^{\xi_{VV}\mathbf{i}} \end{bmatrix} \\
&= \begin{bmatrix} 1-p_{11} & -p_{12}e^{(\xi_1-\xi_2)\mathbf{i}} & \dots & -p_{1V}e^{(\xi_1-\xi_V)\mathbf{i}} \\ -p_{21}e^{(\xi_2-\xi_1)\mathbf{i}} & 1-p_{22} & \dots & -p_{2V}e^{(\xi_2-\xi_V)\mathbf{i}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{i1}e^{(\xi_i-\xi_1)\mathbf{i}} & -p_{i2}e^{(\xi_i-\xi_2)\mathbf{i}} & \dots & -p_{iV}e^{(\xi_i-\xi_V)\mathbf{i}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{V1}e^{(\xi_V-\xi_1)\mathbf{i}} & -p_{V2}e^{(\xi_V-\xi_2)\mathbf{i}} & \dots & 1-p_{VV} \end{bmatrix} \\
&= [\mathrm{diag}(1,e^{-\xi_2},e^{-\xi_3},...,e^{-\xi_V})]^{-1}\mathsf{Q}(-1)\mathrm{diag}(1,e^{-\xi_2},e^{-\xi_3},...,e^{-\xi_V})
\end{aligned}$$

Since $\mathsf{Q}(-1)$ is singular, so $\mathsf{Q}(-1+b\mathbf{i})$ is. Hence $s=-1+b\mathbf{i}$ is a pole of $\mathsf{Q}(s)$ if and only if $|\frac{b}{2\pi}(\ln p_{ji}+\ln p_{1j}-\ln p_{1i})|$ are integers for any $i,j=1,2,...,V$. Since $\{|\frac{\theta}{2\pi}(\ln p_{ij}+\ln p_{1i}-\ln p_{1j})|\}_{ij=1}^{V}$ is a set of relative primes, hence $b=l\theta$ for some integer l. Part (ii) is proved. Part (iii) can be inferred from the above proof. ∎

Observe that for the memoryless case, that is, when $p_{ji}=\pi_i$, condition (66) becomes $\frac{\ln \pi_i}{\ln \pi_j}\in\mathbb{Q}$ for all i,j. This agrees with previous known results (cf. [10]).

Finally, as a simple consequence of the above, we prove the convergence of the infinite product that appears (62).

Lemma 4 *The product*

$$\prod_{i=0}^{\infty}\mathsf{Q}^{-1}(s-i)$$

converges for $\Re(s)<-1$, *and it can be differentiated with respect to* s *term by term.*

Proof. For $\Re(s) < -1$, every factor of the above infinite product is non-singular, and $\|\mathsf{P}(s)\| \le Vp^{-s}$, where $p = \max_{i,j}\{p_{ij}\} < 1$. For k large enough such that $Vp^k < \frac{1}{2}$, we have $\|\mathsf{Q}(s-k)\| \le 1 + 2Vp^{-s+k}$. Since $\sum_{i=k}^{\infty} p^{-s+i} < \infty$, hence $|\prod_{i=0}^{\infty} \mathsf{Q}^{-1}(s-i)| \le \prod_{i=0}^{\infty} \|\mathsf{Q}^{-1}(s-i)\| < \infty$. ■

3.3 Asymptotic Expansions for the Moments in the Poisson Model

As outlined above, we seek the asymptotics of $\widetilde{B}_u(z,1)$ and $\widetilde{B}_{uu}(z,1)$ for large z, which further will lead through depoissonization to asymptotics of the first two moments of the depth. We derive asymptotic expansions of the moments in the Poisson model by applying the inverse Mellin transform. In particular,

$$\begin{aligned}
\widetilde{B}_u(z,1) &= \frac{1}{2\pi\mathbf{i}} \int_{-\frac{3}{2}-\mathbf{i}\infty}^{-\frac{3}{2}+\mathbf{i}\infty} B^*(s) z^{-s} ds, \\
\widetilde{B}_{uu}(z,1) &= \frac{1}{2\pi\mathbf{i}} \int_{-\frac{3}{2}-\mathbf{i}\infty}^{-\frac{3}{2}+\mathbf{i}\infty} C^*(s) z^{-s} ds.
\end{aligned}$$

The evaluation of the above integrals is quite standard (e.g., see [13, 17]): We extend the line of integration to a big rectangle right to the integration line, and observe that bottom and top lines contribute negligible because the gamma function decreases exponentially with the increase in the magnitude of the imaginary part. The right side positioned at, say d, contributes $|z|^{-d}$ for $d \to \infty$. Thus, the integral is asymptotically equal to minus the sum of residues positioned right to the line of the integration, that is, $(-\frac{3}{2} - \mathbf{i}\infty, -\frac{3}{2} + \mathbf{i}\infty)$. But, the residues of the above depend on the singularities of just studied $\mathsf{Q}(s)$ and gamma function. To estimate them, we expand the function under the integral around these singularities.

Let us start with the dominant singularity at $s_0 = -1$, and derive the Laurent expansion of $\mathbf{x}(s)$ and $\mathbf{v}(s)$. By Lemma 3, we have

$$\mathsf{Q}^{-1}(s) = \frac{1}{s+1}\mathsf{Q}_1 + \mathsf{Q}_2 + O(s+1),$$

where $\mathsf{Q}_1, \mathsf{Q}_2$ are $V \times V$ matrices. Since

$$\begin{aligned}
\mathbf{x}(s-1) &= \boldsymbol{\psi} + \dot{\mathbf{x}}(-2)(s+1) + O((s+1)^2), \\
\Gamma(s) &= \frac{-1}{s+1} + \gamma - 1 + O(s+1),
\end{aligned}$$

we obtain from (53), (62) and (55), (63)

$$\begin{aligned}
\mathbf{b}(s) &= \Gamma(s)\mathsf{Q}^{-1}(s)\mathbf{x}(s-1) = \frac{1}{(s+1)^2}\mathbf{a}_1 + \frac{1}{s+1}\mathbf{a}_2 + O(1) \\
\mathbf{c}(s) &= 2\Gamma(s)\mathsf{Q}^{-2}(s)\mathsf{P}(s)\mathbf{x}(s-1) + \Gamma(s)\mathsf{Q}^{-1}(s)\mathbf{v}(s-1), \\
&= \frac{1}{(s+1)^3}\mathbf{f}_1 + \frac{1}{(s+1)^2}\mathbf{f}_2 + O\left(\frac{1}{s+1}\right)
\end{aligned}$$

where $\mathbf{a}_1, \mathbf{a}_2, \mathbf{f}_1$ and $\mathbf{f}_2$ are vectors of constants for which explicit formulæ are presented below the next lemma. In addition, by (64), (65) and $x(s-1) = 1 + O(s+1)$, $v(s-1) = O(s+1)$, we have

$$B^*(s) = \frac{1}{(s+1)^2}\boldsymbol{\pi}\mathbf{a}_1 + \frac{1}{s+1}\left(\boldsymbol{\pi}\mathbf{a}_2 + \dot{\mathbf{p}}(-1)\mathbf{a}_1 - 1\right) + O(1), \tag{68}$$

$$C^*(s) = \frac{1}{(s+1)^3}\boldsymbol{\pi}\mathbf{f}_1 + \frac{1}{(s+1)^2}\left(\dot{\mathbf{p}}(-1)\mathbf{f}_1 + \boldsymbol{\pi}\mathbf{f}_2 + 2\boldsymbol{\pi}\mathbf{a}_1\right) + O\left(\frac{1}{s+1}\right), \tag{69}$$

where $\dot{\mathbf{p}}(-1) = \frac{d}{ds}\mathbf{p}(s)|_{s=-1} = (-\pi_1 \ln \pi_1, \ldots, -\pi_V \ln \pi_V)$.

To derive explicit expressions for the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{f}_1$ and $\mathbf{f}_2$ we need the following lemma, which proof is standard and omitted (detailed proof can be found in [25]).

Lemma 5 *Let us define*

$$\Pi = \begin{bmatrix} \pi_1 & \pi_2 & \ldots & \pi_V \\ \pi_1 & \pi_2 & \ldots & \pi_V \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \ldots & \pi_V \end{bmatrix} = \boldsymbol{\psi} \otimes \boldsymbol{\pi}$$

and let $\mathsf{Q}^* = \{q^*_{ji}\}_{i,j=1}^V$ *be the adjoint matrix of* $\mathsf{Q}(s)|_{s=-1}$. *Then*

$$\boldsymbol{\pi}\Pi = \boldsymbol{\pi}, \quad \Pi^2 = \Pi, \quad \Pi\boldsymbol{\psi} = \boldsymbol{\psi}, \tag{70}$$

$$\frac{d}{ds}\det \mathsf{Q}(s)|_{s=-1} = \frac{d}{ds}\det \mathsf{Q}(s)|_{s=-1+jb} = -\omega h, \quad q^*_{ji} = \omega\pi_i, \quad \mathsf{Q}^* = \omega\Pi, \tag{71}$$

$$\mathsf{Q}_1 = -\frac{1}{h}\Pi, \quad \mathsf{Q}_2 = -\frac{\dot{\mathsf{Q}}^*}{\omega h} - \frac{\beta}{2\omega h^2}\Pi, \quad \mathsf{Q}_1|_{s=-1+jb} = -\frac{1}{h}\mathsf{E}^{-1}\Pi\mathsf{E}. \tag{72}$$

$$\mathsf{Q}_1^2\dot{\mathsf{P}}(-1)\boldsymbol{\psi} = \frac{1}{h}\boldsymbol{\psi}, \quad \mathsf{Q}_1^2\dot{\mathsf{P}}(-1) = \frac{1}{h}\Pi, \tag{73}$$

where $s = -1 + b\mathbf{i}$ *is a pole of* $\mathsf{Q}^{-1}(s)$.

Using the above, we finally obtain after some tedious algebra

$$\begin{aligned}
\mathbf{a}_1 &= -\mathsf{Q}_1\boldsymbol{\psi} = \frac{1}{h}\boldsymbol{\psi}, \\
\mathbf{a}_2 &= -\frac{1}{h}(\gamma-1)\boldsymbol{\psi} + \frac{1}{\omega h}\dot{\mathsf{Q}}^*\boldsymbol{\psi} + \frac{\beta}{2\omega h^2}\boldsymbol{\psi} + \frac{1}{h}\boldsymbol{\pi}\dot{\mathbf{x}}(-2)\boldsymbol{\psi}, \\
\mathbf{f}_1 &= -2\mathsf{Q}_1^2\boldsymbol{\psi} = \frac{-2}{h^2}\boldsymbol{\psi}, \\
\mathbf{f}_2 &= 2\left(\frac{\gamma-1}{h^2} - \frac{\beta}{\omega h^3} - \frac{1}{h} - \frac{1}{h^2}\boldsymbol{\pi}\dot{\mathbf{x}}(-2)\right)\boldsymbol{\psi} - \frac{2}{\omega h^2}(\Pi\dot{\mathsf{Q}}^*\boldsymbol{\psi} + \dot{\mathsf{Q}}^*\boldsymbol{\psi}).
\end{aligned}$$

In summary, using (68) we obtain the following expansions on $B^*(s)$ and $C^*(s)$ around the dominant pole at $s_0 = -1$

$$
\begin{aligned}
B^*(s) &= \frac{1}{(s+1)^2}\frac{1}{h} + \frac{1}{s+1}\left(-\frac{1}{h}(\gamma-1) + \frac{1}{\omega h}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} + \frac{\beta}{2\omega h^2} + \frac{1}{h}\boldsymbol{\pi}\dot{\mathbf{x}}(-2) + \frac{h_{\boldsymbol{\pi}}}{h} - 1\right) + O(1),\\
C^*(s) &= \frac{-2}{h^2(s+1)^3} + \frac{2}{(s+1)^2}\left(\frac{-h_{\boldsymbol{\pi}}}{h^2} + \frac{\gamma-1}{h^2} - \frac{\beta}{\omega h^3}\ \frac{1}{h^2}\boldsymbol{\pi}\dot{\mathbf{x}}(-2) - \frac{2}{\omega h^2}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi}\right) + O\left(\frac{1}{s+1}\right).
\end{aligned}
$$

In Section 2 we introduced ϑ that now we can also represent as $\vartheta := \boldsymbol{\pi}\dot{\mathbf{x}}(-2)$.

Now, we deal with the asymptotics related to the non-dominant poles $s_l = -1 + l\theta\mathbf{i}$ for $l \neq 0$. By Lemma 3 we have

$$
\mathsf{Q}(s) = \frac{-1}{h}\frac{1}{s+1-l\theta\mathbf{i}} \times \mathsf{E}^{-1}\Pi\mathsf{E} + O(1).
$$

Therefore,

$$
\begin{aligned}
\mathbf{b}(s) &= -\frac{1}{h}\rho_l\boldsymbol{\psi}(l)\frac{1}{s+1-l\theta\mathbf{i}} + O(1),\\
\mathbf{c}(s) &= \frac{2}{h^2}\rho_l\boldsymbol{\psi}(l)\frac{1}{(s+1-l\theta\mathbf{i})^2} + O\left(\frac{1}{s+1-l\theta\mathbf{i}}\right),
\end{aligned}
$$

where $\rho_l = \Gamma(-1+l\theta\mathbf{i})\left(\boldsymbol{\pi}\mathsf{E}^l\mathbf{x}(-2+l\theta\mathbf{i})\right)$ and $\boldsymbol{\psi}(l) = \mathsf{E}^{-l}\boldsymbol{\psi}$. In summary, by (64) and (65) at $s = -1 + l\theta\mathbf{i}$ we obtain

$$
\begin{aligned}
B^*(s) &= -\frac{1}{h}\rho_l\mathbf{p}(-1+l\theta\mathbf{i})\boldsymbol{\psi}(l)\frac{1}{s+1-l\theta\mathbf{i}} + O(1),\\
C^*(s) &= \frac{2}{h^2}\rho_l\mathbf{p}(-1+l\theta\mathbf{i})\boldsymbol{\psi}(l)\frac{1}{(s+1-l\theta\mathbf{i})^2} + O\left(\frac{1}{s+1-l\theta\mathbf{i}}\right).
\end{aligned}
$$

Finally, we handle singularities in the half plane $\Re(s) > -1$. We consider two cases: $-1 < \Re(s) \le 0$ and $\Re(s) > 0$. Let $\mathcal{Z}_*$ be the set of singularities s^* of $\mathsf{Q}(s)$ lying in the strip $-1 < \Re(s^*) \le 0$, while $\mathcal{Z}_+$ be the set of singularities in $\Re(s) > 0$. For the pole $s^* \in \mathcal{Z}_*$ we have

$$
B^*(s) = \frac{1}{s-s^*}\boldsymbol{\pi}(s^*)\Gamma(s^*)R(s^*)\mathbf{x}(s^*-1) = \frac{1}{s-s^*}r(s^*)
$$

where $R(s^*)$ is the residue matrix of $Q^{-1}(s)$ at s^*. Note that $s = 0$ is the double pole. An application of the inverse Mellin transform gives for $z \to \infty$,

$$
\widetilde{B}_u(z,1) = \frac{1}{h}z\ln z + \frac{1}{h}\left(\gamma - 1 - \frac{\beta}{2\omega h} - \frac{1}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - \boldsymbol{\pi}\dot{\mathbf{x}}(-2) + h - h_{\boldsymbol{\pi}}\right)z + \delta_1(z) + O(\ln z),
$$

where

$$
\delta_1(z) = -\frac{1}{h}\sum_{l=0}\rho_l\boldsymbol{\pi}\boldsymbol{\psi}(l)z^{1-l\theta\mathbf{i}} + \sum_{s^*\in\mathcal{Z}_*} r(s^*)z^{-s^*}. \tag{74}
$$

Observe also that $r(0) + \sum_{s^* \in \mathcal{Z}_+} r(s^*) z^{-s^*} = O(\ln z)$. In a similar manner, we obtain

$$\begin{aligned}\widetilde{B}_{uu}(z,1) &= \frac{1}{h^2} z \ln^2 z + \frac{2}{h^2}\left(\gamma - 1 - \frac{\beta}{\omega h} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h_{\boldsymbol{\pi}} - \boldsymbol{\pi}\dot{\mathbf{x}}(-2)\right) z \ln z \\ &+ \frac{2}{h^2}\sum_{l=0} \rho_l \boldsymbol{\pi}(1 - l\theta\mathbf{i})\boldsymbol{\psi}(l) z^{1-l\theta\mathbf{i}} \ln z + O(z) \end{aligned} \tag{75}$$

as $z \to \infty$ in a cone around the real axis.

3.4 Analytic Depoissonization

The above asymptotic formulæ concern the behavior of the Poisson mean and the second factorial moment as $z \to \infty$. More precisely, we had to restrict the growth of z to a linear cone $S_\theta = \{z : |\arg(z)| \le \theta\}$ for some $|\theta| < \pi/2$. But our original goal was to derive asymptotics of the mean $\mathbf{E}[D_m]$ and the variance $\mathbf{Var}[D_m]$ in the MI model. To infer such a behavior from its Poisson model asymptotics, we must apply the so called *depoissonization lemma.* This lemma basically says that $m\mathbf{E}[D_m] \sim \widetilde{B}_u(m,1)$ and $m\mathbf{E}[D_m(D_m - 1)] \sim \widetilde{B}_{uu}(m,1)$ under some weak conditions that will be easy to verify in our case. The reader is referred to [10, 11, 12] for more details about depoissonization lemma. For completeness, however, we review some depoissonization results that are useful for our problem.

Let us consider a general problem: For a random variable X_n define g_n as a functional of the distribution of X_n (e.g., $g_n = \mathbf{E}[X_n]$ or $g_n = \mathbf{E}[X_n^2]$), or, in general, assume g_n is a sequence of n. In some situations (e.g., for limiting distributions we need to consider the generating function $G_n(u) = \mathbf{E}[u^{X_n}]$ for a random variable X_n) for a complex u which can be viewed as such a g_n (with a parameter u belonging to a compact set). Define the Poisson transform of g_n as $\widetilde{G}(z) = \sum_{n=0}^\infty g_n \frac{z^n}{n!} e^{-z}$ (or more generally: $\widetilde{G}(z,u) = \sum_{n=0}^\infty G_n(u) \frac{z^n}{n!} e^{-z}$ for u in a compact set). Assume that we know the asymptotics of $\widetilde{G}(z)$ for z large and belonging to a cone $S_\theta = \{z : |\arg(z)| \le \theta\}$ for some $|\theta| < \pi/2$. How can we infer asymptotics of g_n from $\widetilde{G}(z)$? An answer is given in the depoissonization lemma below (cf. [10, 11, 12]):

Lemma 6 (Depoissonization Lemma)
(i) *Let $\widetilde{G}(z)$ be the Poisson transform of a sequence g_n that is assumed to be an entire function of z. We postulate that for $0 < |\theta| < \pi/2$ the following two conditions simultaneously hold for some numbers $A, B, \xi > 0$, β, and $\alpha < 1$:*
(I) *For $z \in S_\theta$*

$$|z| > \xi \quad \Rightarrow \quad |\widetilde{G}(z)| \le B|z|^\beta \phi(|z|) \ , \tag{76}$$

where $\phi(z)$ is a slowly varying function (e.g., $\phi(z) = \log^d z$ for some $d > 0$),

(O) *For* $z \notin S_\theta$

$$|z| > \xi \quad \Rightarrow \quad |\widetilde{G}(z)e^z| \le A\exp(\alpha|z|) \ . \tag{77}$$

Then for large n

$$g_n = \widetilde{G}(n) + O(n^{\beta-1}\phi(n)) \ , \tag{78}$$

or more precisely:

$$g_n = \widetilde{G}(n) - \frac{1}{2}\widetilde{G}''(n) + O(n^{\beta-2}\phi(n)) \ .$$

(ii) *If the above two conditions, namely (I) and (O), hold for* $\widetilde{G}(z,u)$ *for* u *belonging to a compact set* $\mathcal{U}$*, then*

$$G_n(u) = \widetilde{G}(n,u) + O(n^{\beta-1}\phi(n)) \tag{79}$$

for large n *and uniformly in* $u \in \mathcal{U}$*.*

(iii) *Let* $g(z)$ *be an analytic continuation of a sequence* g_n *whose Poisson transform is* $\widetilde{G}(z)$*, and such that* $g(z) = O(z^\beta)$ *in a linear cone. Then, for some* θ_0 *and for all linear cones* S_θ *(*$\theta < \theta_0$*), there exists* $\alpha < 1$ *and* $A > 0$ *such that*

$$z \notin S_\theta \qquad \Rightarrow \qquad |\widetilde{G}(z)e^z| \le Ae^{\alpha|z|}.$$

In summary, when $g(z)$ *has a polynomial growth, then conditions (I) and (O) above are automatically satisfied and (78) holds.*

Now, we are equipped with the tool to depoissonize $\widetilde{B}_u(z,1)$ and $\widetilde{B}_{uu}(z,1)$, and obtain asymptotics for the mean $\mathbf{E}[D_m]$ and the variance $\mathbf{Var}[D_m]$. Observe that $\mathbf{E}[D_m] = O(m\ln m)$ and $\mathbf{Var}[D_m] = O(m\log^2 m)$, hence by Lemma 6 we can depoissonize the Poisson estimates. We obtain

$$\begin{aligned}\mathbf{E}[D_m] &= \frac{1}{h}\ln m + \frac{1}{h}\left(\gamma - 1 + h - h\boldsymbol{\pi} - \frac{\beta}{2\omega h} - \frac{1}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - \boldsymbol{\pi}\dot{\mathbf{x}}(-2))\right) \\ &+ \delta_1(m) + O\left(\frac{\ln m}{m}\right).\end{aligned} \tag{80}$$

To derive the variance, we observe that $\sum_{s^*\in\mathcal{Z}_*} r(s^*)m^{-s^*} = O(m^{-\delta})$ for some $\delta > 0$, thus such terms will not appear explicitly in the following formula where only $\Omega(\ln m)$ terms are considered. Again, by Lemma 6 we arrive at

$$\mathbf{Var}[D_m] = \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2\right)\ln m + O(1).$$

In conclusion, (7) and (8) of Theorem 1 are proved.

3.5 Limiting Distribution

Finally, we shall derive the limiting distribution of the depth D_m, just finishing the proof of Theorem 1. We repeat here the system of functional equations (50), that is,

$$\begin{aligned}
\widetilde{B}^1_z(z,u) + \widetilde{B}^1(z,u) &= u[\widetilde{B}^1(p_{11}z,u) + \cdots + \widetilde{B}^V(p_{1V}z,u)] + 1 \\
\cdots &= \cdots \\
\widetilde{B}^V_z(z,u) + \widetilde{B}^V(z,u) &= u[\widetilde{B}^1(p_{V1}z,u) + \cdots + \widetilde{B}^V(p_{VV}z,u)] + 1
\end{aligned}$$

Observe that $\widetilde{B}^i(z,1) - z = 0$, $\widetilde{B}(z,1) - z = 0$, $\widetilde{B}^i(z,u) - z = (u-1)A_i(u,z)$, and $\widetilde{B}(z,u) - z = (u-1)A(u,z)$, where $A_i(u,z)$ is a power series of u and thus analytic function of z. Let

$$\begin{aligned}
Z^*_i(u,s) &= \mathcal{M}[\widetilde{B}^i(z,u) - z; s] = \Gamma(s)\xi_i(u,s) = (u-1)A^*_i(u,s), \qquad i \in \mathcal{A} \\
Z^*(u,s) &= \mathcal{M}[\widetilde{B}(z,u) - z; s] = \Gamma(s)\xi(u,s) = (u-1)A^*(u,s)
\end{aligned}$$

be the Mellin transforms, where $\xi_i(u,s)$ and $\xi(u,s)$ are unknown functions.

Lemma 7 *The Mellin transforms $Z^*_i(u,s)$, $Z^*(u,s)$, $A^*_i(u,s)$ and $A^*(u,s)$ exist for $\Re(s) \in (-2,-1)$. In addition, $Z^*_i(u,-2) = u-1$, $A^*_i(u,-2) = 1$, $Z^*(u,-2) = u-1$, $A^*(u,-2) = 1$.*

Proof. By the same argument as in Lemma 2 of [16]. ∎

We proceed along the same lines as before, leaving out detailed explanations. After applying the Mellin transform to the above system of functional equations, we find

$$Z^*(u,s) - (s-1)Z^*(u,s-1) = u[Z^*_1(u,s)\pi_1^{-s} + \ldots + Z^*_V(u,s)\pi_V^{-s}],$$

$$\begin{aligned}
Z^*_1(u,s) - (s-1)Z^*_1(u,s-1) &= u[Z^*_1(u,s)p_{11}^{-s} + \cdots + Z^*_V(u,s)p_{1V}^{-s}] \\
\cdots &= \cdots \\
Z^*_V(u,s) - (s-1)Z^*_V(u,s-1) &= u[Z^*_1(u,s)p_{V1}^{-s} + \cdots + Z^*_V(u,s)p_{VV}^{-s}].
\end{aligned}$$

Let

$$\Gamma(s)\boldsymbol{\xi}(u,s) = \begin{bmatrix} Z^*_1(u,s) \\ Z^*_2(u,s) \\ \vdots \\ Z^*_V(u,s) \end{bmatrix} = (u-1)\begin{bmatrix} A^*_1(u,s) \\ A^*_2(u,s) \\ \vdots \\ A^*_V(u,s) \end{bmatrix} = (u-1)\mathbf{a}(u,s).$$

Then

$$\boldsymbol{\xi}(u,s) - \boldsymbol{\xi}(u,s-1) = u\mathsf{P}(s)\boldsymbol{\xi}(u,s),$$

which yields

$$\boldsymbol{\xi}(u,s) = [\mathsf{I} - u\mathsf{P}(s)]^{-1}\boldsymbol{\xi}(u,s-1),$$

and finally we arrive at

$$Z^*(u,s) = u\mathbf{p}(s)\Gamma(s)[\mathsf{I} - u\mathsf{P}(s)]^{-1}\boldsymbol{\xi}(u,s-1) + \Gamma(s)\xi(u,s-1).$$

Let now set $u = e^t$ for complex $t \to 0$ so that u is in the vicinity of $u = 1$. We denote by $s_k(t), k = 0, \pm 1, \pm 2, \ldots$ singularities of $\mathsf{Q}^{-1}(t,s) = (\mathsf{I} - e^t\mathsf{P}(s))^{-1}$. Then at $s = s_k(t)$

$$Z^*(e^t, s_k(t)) = e^t\boldsymbol{\pi}(s_k(t))\Gamma(s_k(t))\mathsf{R}_k\boldsymbol{\xi}(e^t, s_k(t)-1)\frac{1}{s - s_k(t)} + O(1), \tag{81}$$

where R_k is the residue matrix of $\mathsf{Q}^{-1}(u,s) = [\mathsf{I} - u\mathsf{P}(s)]^{-1}$ at $s = s_k(t)$. In addition, one must consider two poles of the gamma function $\Gamma(s)$ at $s_{-1} = -1$ and $s_0 = 0$. The latter pole contribute $O(1)$ while the former $-z\xi(u,-1)$. But, by Lemma 7 we know that $\xi(u,-1) = 1$, thus the total contribution of these two poles is $-z + O(1)$. By the inverse Mellin transform, we have

$$\widetilde{B}(e^t, z) = e^t\sum_{k=-\infty}^{\infty}\boldsymbol{\pi}(s_k(t))\Gamma(s_k(t))\mathsf{R}_k\boldsymbol{\xi}(e^t, s_k(t)-1)z^{-s_k(t)} + O(1)$$

as $z \to \infty$ in a cone. As before, the leading contribution to the asymptotics comes from the pole $s_0(t)$.

To obtain an asymptotic expansion for the original generating function $B_m(e^t)$ we apply the depoissonization lemma Lemma 6(ii). Since $\widetilde{B}(z, e^t) = O(z\log z)$, we conclude that $B_m(e^t) = \widetilde{B}(m, e^t) + O(\log m)$, where

$$\begin{aligned}\widetilde{B}(m, e^t) &= e^t\mathbf{p}(s_0(t))\Gamma(s_0(t))\mathsf{R}_0\boldsymbol{\xi}(e^t, s_0(t)-1)m^{-s_0(t)} \\ &+ e^t\sum_{k\neq 0}\mathbf{p}(s_k(t))\Gamma(s_k(t))\mathsf{R}_k\boldsymbol{\xi}(u, s_k(t)-1)m^{-s_k(t)} + O(1).\end{aligned}$$

Let (see (8), (20), and the Appendix)

$$v := \frac{\ddot{\lambda}(-1) - \dot{\lambda}^2(-1)}{\dot{\lambda}^3(-1)} = \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2\right).$$

Then

$$\begin{aligned}s_0(t) &= -1 - \frac{t}{h} - \frac{v^2t^2}{2} + O(t^3), \\ \mathsf{R}_0 &= -\frac{1}{h}\Pi + O(t), \\ \Gamma(s_0(t)) &= -\frac{h}{t} + O(1), \\ \boldsymbol{\xi}(s_0(t)-1) &= t\boldsymbol{\psi} + O(t^2), \\ \mathbf{p}(s_0(t)) &= \boldsymbol{\pi} + O(t).\end{aligned}$$

Indeed, we just observe that the expansion of $s_0(t)$ is obtained via the Lagrange inversion of $1-e^t\lambda(s)$, or better, of function $t+\log\lambda(s)$, at $s=-1$, which results in $t+(s+1)\dot\lambda(-1)+(s+1)^2(\frac{\ddot\lambda(-1)-(\dot\lambda(-1))^2}{2})+O(s+1)^3$. We again identify $\dot\lambda(-1)=h$. The residue R_0 is computed by using the fact that $\mathsf{Q}^{-1}(e^t,s)=(1-e^t\lambda(s))^{-1}\boldsymbol{\psi}(s)\otimes\boldsymbol{\pi}(s)+O(1)$. Observe also that

$$\lim_{t\to 0}\mathbf{p}(s_0(t))\Gamma(s_0(t))\mathsf{R}_0\boldsymbol{\xi}(e^t,s_0(t)-1)=\boldsymbol{\pi}\Pi\boldsymbol{\psi}=1.$$

Some remaining details can be found in [9].

We now set $t=\frac{\tau}{\sigma_m}=O(1/\sqrt{\ln m})$ for some fixed τ and $\sigma_m=\sqrt{\mathbf{Var}D_m}$. Then $m^{-1-s_0(t)}=e^{\tau\mu_m/\sigma_m+\frac{\tau^2}{2}}(1+O(t))$ and $D_m(e^t)=B(e^t)/m$ leading to

$$\begin{aligned}
e^{-\tau\mu_m/\sigma_m}D_m(e^{\tau/\sigma_m},m) &= e^{-\tau\mu_m/\sigma_m}\Bigg(e^{\tau\mu_m/\sigma_m+\frac{\tau^2}{2}}(1+O(t))\\
+\quad & e^{-t}m^{-1-s_0(t)}\sum_{k\neq 0}(u-1)(s_k(t)-1)\mathbf{p}(s_k(t))\mathsf{R}_k\mathbf{a}(s_k(e^t)-1,u)m^{s_0(t)-s_k(t)}+O\left(\frac{\log m}{m}\right)\Bigg)\\
=\quad & e^{\frac{\tau^2}{2}}\left(1+tO\left(\sum_{k\neq 0}(s_k(t)-1)\mathbf{p}(s_k(t))\mathsf{R}_k\mathbf{a}(s_k(e^t)-1,u)\right)\right).
\end{aligned}$$

In the above, we use the fact that $\Re(s_0(t))\le\Re(s_k(t))$ proved in [9], which allows to bound $|m^{s_0(t)-s_k(t)}|\le 1$. To complete the proof, it suffices to show that the sum appearing above is $O(1)$. Let $s_k(t)=x_k(t)+y_k(t)\mathbf{i}$ for any $M>0$,

$$\begin{aligned}
\left|\sum_{k\neq 0}(s_k(t)-1)\mathbf{p}(s_k(t))\mathsf{R}_k\mathbf{a}(s_k(e^t)-1)\right| &\le \sum_{k\neq 0}|(s_k(t)-1)|\|\mathbf{p}(s_k(t))\|\|\mathsf{R}_k\|\|\mathbf{a}(s_k(e^t)-1)\|\\
&\le \sum_{k\neq 0}\frac{1}{|y_k|^M}=O(1).
\end{aligned}$$

Here, we use the fact that $A_i(u,z)$ is infinitely differentiable, thus its Mellin transform satisfies $\lim_{y\to\infty}|y|^M A_i^*(u,x+y\mathbf{i})=0$.

In summary, we have just shown that

$$e^{-\tau\mu_m/\sigma_m}D_m(e^{\tau/\sigma_m})=e^{\frac{\tau^2}{2}}\left(1+O\left(\frac{1}{\sqrt{\ln m}}\right)\right),$$

which completes the proof of Theorem 1.

3.6 Non-Stationary MI Model

We now show how to adapt the above derivations to the non-stationary model, in which the initial distribution is $\mathbf{p}$ instead of $\boldsymbol{\pi}$. First of all, observe that $\mathbf{p}$ appears in equation (4) while the conditional generating functions $B_m^i(u)$ still satisfy (5). Thus in (49) we must replace

π_i by p_i, but again (50) stays unchanged. This leads to the following Mellin transforms of $B_u(z,1)$ and $B_{uu}(z,1)$ in the non-stationary case

$$B^*(s) = \mathbf{p}(s)\mathbf{b}(s) + \Gamma(s)x(s-1), \tag{82}$$

$$C^*(s) = 2\mathbf{p}(s)\mathbf{b}(s) + \mathbf{p}(s)\mathbf{c}(s) + \Gamma(s)v(s-1), \tag{83}$$

where $\mathbf{p}(s) = (p_1^{-s}, \ldots, p_V^{-s})$. Observe, however, that $\mathbf{b}(s)$ and $\mathbf{c}(s)$ are exactly the same as in the stationary MI model. Since, as we discussed before, the asymptotics of the mean and the variance in the Poisson model depend mostly on the asymptotics of $\mathbf{b}(s)$ and $\mathbf{c}(s)$, we may expect similar asymptotics results for the non-stationary model. Indeed, we obtain the following expansions of $B^*(s)$ and $C^*(s)$ around the dominant pole $s_0 = -1$

$$B^*(s) = \frac{1}{(s+1)^2}\mathbf{p}\mathbf{a}_1 + \frac{1}{s+1}\left(\mathbf{p}\mathbf{a}_2 + \mathbf{p}'(-1)\mathbf{a}_1 - 1\right) + O(1), \tag{84}$$

$$C^*(s) = \frac{1}{(s+1)^3}\mathbf{p}\mathbf{f}_1 + \frac{1}{(s+1)^2}\left(\mathbf{p}'(-1)\mathbf{f}_1 + \mathbf{p}\mathbf{f}_2 + 2\boldsymbol{\pi}\mathbf{a}_1\right) + O\left(\frac{1}{s+1}\right). \tag{85}$$

In view of the above, we conclude that the only term effected by the non-stationarity assumption is related to $\mathbf{p}'(-1)$ (and also the fluctuating function), which is responsible for replacing $h_{\boldsymbol{\pi}}$ by $h_{\mathbf{p}}$ in the final results. Similar conclusions hold for the limiting distribution. This proves Corollary 1.

4 Analysis of Fixed Number of Phrases (GK) Model

In this section we prove Theorem 2 using a combination of probabilistic and analytic techniques. We start our discussion with by introducing the so-called *tree-path* that plays a crucial role in the analysis. We study its property in Section 4.1, and in Section 4.2 we make a connection between the tree-path and the depth (i.e., phrase). Finally, in Section 4.3 we obtain the limiting distribution for the phrase while in Section 4.4 we establish the existence of the moments, thus proving Theorem 2.

4.1 Tree-Path in Digital Search Trees

We consider a DST tree $\mathcal{T}_m$ built over m strings regardless of the model of strings generation (e.g., MI, GK, or hybrids). For $k \le m$ we denote by $I_k(\mathcal{T}_m)$ the depth of insertion of the kth phrase in the tree $\mathcal{T}_m$. (Observe that $I_k(\mathcal{T}_k) = I_k(\mathcal{T}_m)$). If the tree $\mathcal{T}_m$ is known from the context, we often simplify the notation and write I_k.

We introduce now the *tree-path*. Let $w = x_1 x_2 \cdots x_k$ be a finite string whose length we also denote as $|w| = k$. We write $(w)_i$ for the prefix of w of length i, that is, $(w)_i = x_1 x_2 \cdots x_i$.

Assume now that $\mathcal{T}_m$ is given. The tree-path $C_m(w)$ associated with w is a "trace" (path) in $\mathcal{T}_m$ when one follows symbols of w along a path in the tree $\mathcal{T}_m$. More precisely:

Definition 2 *The tree-path $C_m(w)$ associated with a given string w in $\mathcal{T}_m$ is the largest integer $\ell \le |w|$ such that there exist $k \le m$ that satisfies: (i) $(w)_\ell$ is the prefix of phrase k, and (ii) $I_k(\mathcal{T}_m) = \ell$.*

We now outline some properties of the distribution of the tree-path when DST is random. The next lemma shows that the tree-path distribution satisfies a simple recurrence.

Lemma 8 (i) *Consider any model of phrase generations. Then for all integers $m > 1$*

$$\begin{aligned} \Pr\{C_m(w) \ge k\} &= \Pr\{C_{m-1}(w) \ge k\} + \\ &+ \Pr\{C_{m-1}(w) = k-1 \ \&\ (w)_k \text{ is prefix of } m\text{th phrase}\} \end{aligned} \tag{86}$$

for all $k \ge 0$.

(ii) *If the strings are generated according to the MI model, then (86) becomes*

$$\Pr\{C_m^{MI}(w) \ge k\} = \Pr\{C_{m-1}^{MI}(w) \ge k\} + \Pr\{C_{m-1}^{MI}(w) = k-1\} p_{x_1} p_{x_1 x_2} \cdots p_{x_{k-1} x_k} \tag{87}$$

where $\mathbf{p} = (p_1, \ldots, p_V)$ is the initial probability of generating the first symbol of the string $w = x_1 \cdots x_{|w|}$.

Proof. To prove (86) we observe that the tree-path in $\mathcal{T}_m$ is greater than or equal to k if and only if either it is greater than or equal to k in $\mathcal{T}_{m-1}$ (i.e., the mth insertion does not follow $(w)_k$) or the m insertion traces the word w up to $k-1$ and the kth prefix of w is a prefix of the mth phrase. ■.

We need a simple technical lemma whose proof requires *pathwise* comparison of two stochastic processes (trees).

Lemma 9 *Let w be a finite string. Consider two random DST trees $\mathcal{T}^1_{m_1}$ and $\mathcal{T}^2_{m_2}$ of respective size m_1 and m_2 with tree-paths $C^1_{m_1}(w)$ and $C^2_{m_2}(w)$. We assume that for all $w \in \mathcal{A}^{|w|}$*

$$C^1_{m_1}(w) \le_{\rm st} C^2_{m_2}(w).$$

If we insert to both trees the same **independent** *phrase (string), then the corresponding tree paths $C^1_{m_1+1}(w)$ and $C^2_{m_2+1}(w)$ still satisfy*

$$C^1_{m_1+1}(w) \le_{\rm st} C^2_{m_2+1}(w)$$

for all w.

Proof. We remark that we cannot use Lemma 8 since there is no easy way of bounding $\Pr\{C_m(w) = k-1\}$. Thus, we shall rely on another approach, namely stochastic dominance, in which the independence assumption plays a central role.

Let us fix a given string w. By the *pathwise stochastic dominance* theorem [22], there exists a probabilistic space on which a pair of DST trees $(\widetilde{\mathcal{T}}^1_{m_1}, \widetilde{\mathcal{T}}^2_{m_2})$ satisfies

- For $i = 1,\ 2$ the tree-path distribution of $\widetilde{C}^i_{m_i}(w)$ on $\widetilde{\mathcal{T}}^i_{m_i}$, is the same as the tree-path distribution of $C^i_{m_i}(w)$ on the original trees $\mathcal{T}^i_{m_i}$;
- $\widetilde{C}^1_{m_1}(w) \le \widetilde{C}^2_{m_2}(w)$ for *every* random event.

Now, we insert into both trees $\widetilde{\mathcal{T}}^1_{m_1}$ and $\widetilde{\mathcal{T}}^2_{m_2}$ the same independent random phrase. The path distribution after insertion becomes $\widetilde{C}^1_{m_1+1}(w)$ and $\widetilde{C}^2_{m_2+1}(w)$, respectively. It is easy to check via Lemma 8 that the distribution of $\widetilde{C}^i_{m_i+1}(w)$ will be the same as the distribution of $C^i_{m_i+1}(w)$. We consider the following two cases: either $\widetilde{C}^1_{m_1}(w) \le \widetilde{C}^2_{m_2}(w) - 1$ or $\widetilde{C}^1_{m_1}(w) = \widetilde{C}^2_{m_2}(w)$ for every w. In the first case we must have $\widetilde{C}^1_{m_1+1}(w) \le \widetilde{C}^2_{m_2+1}(w)$ after the insertion since the insertion of the new phrase can only increment by one unit the tree-path. In the second case, we also have $\widetilde{C}^1_{m_1+1}(w) = \widetilde{C}^2_{m_2+1}(w) = k$ since the insertion of the new phrase may either increment by one unit the tree-paths of w on *both trees* or change nothing on both tree-paths, depending whether $(w)_k$ is the kth length prefix of the new phrase. ■

In a typical application of this lemma, we shall assume that for any word w and sizes m_1 and m_2 the following

$$C^{GK}_{m_1}(w) \le_{\mathrm{st}} C^{MI}_{m_2}$$

implies

$$C^{GK+MI}_{m_1+1}(w) \le_{\mathrm{st}} C^{MI}_{m_2+1}$$

where $C^{GK+MI}_{m_1+1}$ denotes the tree path in the GK model in which a new *independent* phrase is inserted.

Now, we are in a position to establish main results of this subsection, namely lower and upper bounds on the tree path. Let $C^{GK}_m(aw)$ and $C^{MI}_m(aw)$ denote the tree-paths in the GK and MI models, respectively, when the associated words aw starts with a given symbol, say a. The following lemma gives an upper bound on $C^{GK}_m(aw)$ with respect to $C^{MI}_m(aw)$.

Lemma 10 *The tree path $C^{GK}_m(aw)$ in the GK model is stochastically bounded from the above by the tree path $C^{MI}_m(aw)$ in the MI model, in which all m phrases start with symbol a (i.e. $\mathbf{p} = \mathbf{p}_a$); that is,*

$$C^{GK}_m(aw) \le_{\mathrm{st}} C^{MI}_m(aw) \tag{88}$$

for all $w \in \mathcal{A}^{|w|}$ *and* $a \in \mathcal{A}$*.*

Proof. The proof is by induction on m. The property is true for $m = 1$. We now suppose it is true for $m-1$. Let us consider the path $C_m^{GK}(aw)$ in the GK model. We obtain by Lemma 8

$$\Pr\{C_m^{GK}(aw) \geq k+1\} = \Pr\{C_{m-1}^{GK}(aw) \geq k+1\}$$
$$+\sum_{b=1}^{V} \Pr\{C_{m-1}^{GK}(aw) = k \ \& \ \ (m-1)\text{th phrase ends with } b\} p_{ba} p_{ax_1} p_{x_1x_2} \cdots p_{x_{k-1}x_k} \ .$$

Since $p_{ba} \leq 1$, and

$$\sum_{b=1}^{b=V} \Pr\{C_{m-1}(aw) = k-1 \ \& \ (m-1)\text{th phrase ends with } b\} = \Pr\{C_{m-1}(aw) = k-1\}$$

we obtain

$$\begin{aligned} \Pr\{C_m^{GK}(aw) \geq k+1\} &\leq \Pr\{C_{m-1}^{GK}(aw) \geq k+1\} + \\ &+ \Pr\{C_{m-1}^{GK}(aw) = k\} p_{ax_1} p_{x_1x_2} \cdots p_{x_{k-1}x_k} \\ &= \Pr\{C_m^{GK+MI}(aw) \geq k+1\} \end{aligned}$$

The last equality directly follows from Lemma 8 with $p_a = 1$. Therefore $C_m^{GK}(aw) \leq_{\text{st}} C_m^{GK+MI}(aw)$. To complete the proof, we use the fact that

$$C_m^{GK+MI}(aw) \leq_{\text{st}} C_m^{MI}(aw), \tag{89}$$

which is a consequence of the induction hypothesis, $C_{m-1}^{GK}(aw) \leq_{\text{st}} C_{m-1}^{MI}(aw)$ and Lemma 9. Indeed in both models, $GK + MI$ and MI, the last phrase is statistically independent of the $m-1$ first phrases and therefore meets the conditions of Lemma 9. ■

Finally, we derive a lower bound on the tree path in the GK model. Below, we shall write $r(a) = \min_i\{p_{ia}\}$ and $r = \sum_{a \in \mathcal{A}} r(a)$. We denote by $C_m^{MIB(r)}(w)$ the path length in the MI model with *binomially*(m,r) distributed number of phrases. We denote $\mathbf{r}$ the probability vector consisting of $\frac{r(a)}{r}$ for $a \in \mathcal{A}$.

Lemma 11 *The tree path* $C_m^{GK}(w)$ *in the GK model is stochastically bounded from the below by the tree path* $C_{m-1}^{MIB(r)}(aw)$ *in the MI model, in which the first symbol of all phrases is distributed according to* $\mathbf{r}$*, and the number of phrases (strings) are binomialll*(m,r) *distributed with parameters* m *and* $r < 1$*; that is,*

$$C_{m-1}^{MIB(r)}(w) \leq_{\text{st}} C_m^{GK}(w) \ . \tag{90}$$

Proof. The proof is by induction, and we shall imitate our proof of Lemma 10 with a few changes. The property is true for $m = 2$, *i.e.*, the second phrase starts with symbol a with a probability smaller than $r(a)$ regardless of the actual value of the first phrase. We now suppose the property is true for $m-1$ and let us take an arbitrary symbol $a \in \mathcal{A}$. We have

$$\begin{aligned}
\Pr\{C_m^{GK}(aw) &\geq k+1\} = \Pr\{C_{m-1}^{GK}(aw) \geq k+1\} + \\
&+ \sum_{b=1}^{V} \Pr\{C_{m-1}^{GK}(aw) = k-1 \ \&\ (m-1)\text{th phrase ends with } b\} \times \\
&\times\ p_{ba}p_{ax_1}p_{x_1x_2}\cdots p_{x_{k-1}x_k} \\
&\geq \Pr\{C_{m-1}^{GK}(aw) \geq k\} + \Pr\{C_{m-1}^{GK}(aw) = k-1\}r \times \frac{r(a)}{r} p_{x_1x_2}\cdots p_{x_{k-1}x_k} \\
&=^{(A)} \Pr\{C_m^{GK+MIB(r)}(aw) \geq k+1\} \\
&\geq^{(B)} \Pr\{C_{m-1}^{MIB(r)}(aw) \geq k+1\}.
\end{aligned}$$

Equation (A) follows from Lemma 8 after noticing that the line above could be interpreted as the MI model, in which the m phrase is inserted with probability r and the initial symbol of every phrase has distribution $r(a)/r$. The inequality (B) is a consequence of the induction assumption and Lemma 9. Observe that we omit the first phrase (so we have $(m-1)$ in the last line of the above) since it does not fall under our assumptions, i.e., its first symbol is not distributed according to $\mathbf{r}$. ■

4.2 Bounds on the Phrase Length and Depth of Insertion

In this subsection, we translate the bounds on the tree path $C_m(w)$ into bounds on the depth of insertion I_m in the GK model. We start with a simple observation that relates the depth of insertion with the tree-path. We have

$$\begin{aligned}
\Pr\{I_m = |w| \quad &\& \quad w \text{ is a prefix of the } m\text{th phrase}\} \\
&= \quad \Pr\{C_{m-1}(w) = |w|-1 \ \&\ w \text{ is a prefix of the } m\text{th phrase}\},
\end{aligned}$$

which further implies

$$\Pr\{I_m \geq k\} = \sum_{|w|=k} \Pr\{C_{m-1}(w) \geq k-1 \ \&\ w \text{ is a prefix of the } m\text{th phrase}\}. \tag{91}$$

This and Lemma 9 lead immediately to the following claim.

Lemma 12 *Consider two random DST trees $\mathcal{T}_{m_1}^1$ and $\mathcal{T}_{m_2}^2$, of respective size m_1 and m_2, with tree-paths $C_{m_1}^1(w)$ and $C_{m_2}^2(w)$, and depths of insertion $I_{m_1}^1$ and $I_{m_2}^2$, respectively. If for all w*

$$C_{m_1}^1(w) \leq_{\text{st}} C_{m_2}^2(w),$$

then an **independent** *phrase inserted into both trees leads to the following inequality*

$$I^1_{m_1+1} \le_{\rm st} I^2_{m_2+1}.$$

Before we proceed with a formal derivation of the bounds on I_m, we present here a "guided tour" through the proof. The first step in establishing a bound for I_m^{GK} in the GK model is to break a strong dependency between phrases so that the precise results of the MI model can be applied. We accomplish it by deleting the last K phrases before inserting a new phrase. We denote by $I_{m,K}^{GK}$ the depth of insertion in the GK model when K last phrases are deleted. In order to make this idea useful, we need an inequality relating the depth I_m^{GK} and the depth $I_{m,K}^{GK}$. But in (37) of Section 2 we proved that

$$I_{m+1,K}^{GK} \le I_{m+1}^{GK} \le I_{m+1,K}^{GK} + K. \tag{92}$$

Unfortunately, we could not establish an easy bound on $I_{m,K}^{GK}$. However, in the previous section we proved lower and upper bounds on the tree paths; hence by Lemma 12 we can bound I_{m-K}^{GK+MI}, where I_{m-K}^{GK+MI} denotes the depth of insertion in the GK model when one inserts an *independent* phrase. The last step is to show that distributions of $I_{m,K}^{GK}$ and I_{m-K}^{GK+MI} are within distance $\varepsilon_m \to 0$.

We start the analysis by showing that $I_{m,K}^{GK}$ is within distance $\varepsilon_m \to 0$ from I_{m-K}^{GK+MI}, which is crucial to our analysis.

Lemma 13 *The random variable $I_{m,K}^{GK}$ is within distance $\varepsilon_m = O(m^{K\log\rho})$ from I_{m-K}^{GK+MI}, where $\rho < 1$ is the mixing coefficient of the underlying Markov chain. (We shall use a short-hand notation $I_{m,K}^{GK} \stackrel{d}{=} I_{m-K}^{GK+MI} + O(\varepsilon_m)$ in such a situation.)*

Proof. We shall use the fact that a Markov chain over a finite space is a ϕ-mixing process with exponentially decreasing mixing coefficient (cf. [3]). More precisely, let for some d and ℓ two events, say A and B, be defined on the sigma-algebras $\mathcal{F}_{-\infty}^{d}$ and $\mathcal{F}_{d+\ell}^{\infty}$, respectively (i.e., there is a gap of ℓ symbols between the events). Then there exists $\rho < 1$ such that (cf. [2, 24]

$$|\Pr\{A\&B\} - \Pr\{A\}\Pr\{B\}| \le \rho^\ell \Pr\{A\}\Pr\{B\}$$

We now associate A with the first $m-K-1$ phrases and B with the mth phrase. Actually, we consider $I_{m,K}^{GK}$, which can be viewed as event $A\&B$ while I_{m-K}^{GK+MI} is composed of two *independent* events, A and B. That is, if $\mathcal{E}_\ell$ denotes the event that K last phrases are of length at least ℓ symbols, then for any set D of integers

$$|\Pr\{I_{m,K}^{GK} \in D \mid \mathcal{E}_\ell\} - \Pr\{I_{m-K}^{GK+MI} \in D \mid \mathcal{E}_\ell\}| \le \rho^\ell \Pr\{I_{m-K}^{GK+MI} \in D \mid \mathcal{E}_\ell\}$$

In Lemma 14 below we prove that there exist $\alpha > 0$ such that $\Pr\{\text{not } \mathcal{E}_\ell\} < K\exp(-Am^\alpha)$ if $\ell = K\beta\log m$ for some $\beta > 0$. Thus

$$|\Pr\{I_{m,K}^{GK} \in D\} - \Pr\{I_{m-K}^{GK+MI} \in D\}| \le \varepsilon_m$$

with $\varepsilon_m = \rho^{K\beta\log m} + K\exp(-Am^\alpha) = O(m^{\beta' K \log\rho})$ where $\beta' > 0$. ■

Lemma 14 *There exist positive constants $A, \alpha, \beta > 0$ such that $\Pr\{I_m^{GK} \le \beta\log m\} \le \exp(-Am^\alpha)$ for all $m > 0$.*

Proof. By (91) we have

$$\Pr\{I_m^{GK} \ge k\} \ge 1 - \sum_{|w|=k} \Pr\{C_{m-1}(w) \le k-1\}. \tag{93}$$

To estimate $\Pr\{C_{m-1}(w) \le k-1\}$, we observe that by Lemma 8

$$\begin{aligned} \Pr\{C_m(w) = k \mid C_{m-1}(w) = k-1\} &= \sum_{a\in\mathcal{A}} \Pr\{\text{last phrase ends with a}\} P(a(w)_k), \\ \Pr\{C_m(w) = k-1 \mid C_{m-1}(w) = k-1\} &= \sum_{a\in\mathcal{A}} \Pr\{\text{last phrase ends with a}\}(1 - P(a(w)_{k-1})), \end{aligned}$$

where $P(aw)$ denotes the probability of the string aw induced by the underlying probabilistic model. Let now $\mu = \min_{a,b\in\mathcal{A}}\{p_{ab}\} > 0$. Then

$$\begin{aligned} \Pr\{C_m(w) = k \mid C_{m-1}(w) = k-1\} &\le P(a(w)_k) \le 1 \\ \Pr\{C_m(w) = k-1 \mid C_{m-1}(w) = k-1\} &\le 1 - \mu^{k+1}. \end{aligned}$$

But $\Pr\{C_m(w) = k\} \le \binom{m}{k}(1-\mu^{k+1})^{m-k}$, and hence

$$\Pr\{C_m(w) \le k\} \le k\binom{m}{k}(1-\mu^{k+1})^{m-k} \le k\binom{m}{k}\exp(-\mu^k(m-k)).$$

Set now $k = \lceil -\frac{\log m}{2\log\mu}\rceil$. Since $\binom{m}{k} \le \frac{m^k}{k!}$, the above becomes

$$\Pr\{C_m(w) \le k\} \le k\binom{m}{k}\exp\left(-\mu^k(m-k)\right) = \exp(-\eta\sqrt{m}),$$

where $\eta > 0$ is a constant. Finally, returning to (93) with $k = \lceil -\frac{\log m}{2\log\mu}\rceil$ and noticing that in this case $\sum_{|w|=k} 1 \le m^B$ for some $B > 0$, we obtain

$$\Pr\{I_m^{GK} \ge k\} \ge 1 - m^B\exp(-\eta\sqrt{m}),$$

which completes the proof. ■

Finally, we are in a position to establish an upper bound (cf. Theorem 4) and a lower bound (cf. Theorem 5) for the depth of insertion I_m^{GK}.

Theorem 4 *Let $I^{GK}_{m-K}(a)$ be the depth of insertion in the GK model when the mth phrase starts with symbol a, and $I^{MI}_{m-K}(\mathbf{p}_a)$ be the depth of insertion in the MI model with the initial probability vector $\mathbf{p}_a = (0, \ldots, 1, \ldots, 0)$ where 1 is at position $a \in \mathcal{A}$ (i.e., all strings start with symbol a). Then for any $\beta > 0$, there exists K such that $I^{GK}_m(a)$ is stochastically dominated by a random variable that is within distance $O(n^{-\beta})$ from $I^{MI}_{m-K}(\mathbf{p}_a) + K$*

Proof. Let K be a fixed integer. We have from (92)

$$I^{GK}_m(a) \le I^{GK}_{m,K}(a) + K \ .$$

We also have

$$I^{GK}_{m,K}(a) \stackrel{d}{=} I^{GI+MI}_{m-K}(a) + O(\varepsilon_m)$$

as a consequence of Lemma 13. Lemma 10 implies

$$I^{GI+MI}_{m-K}(a) \le_{\rm st} I^{MI}_{m-K}(\mathbf{p}_a),$$

which completes the proof. ■

The proof of the lower bound on I^{GK}_m follows the same footsteps as above, so we only sketch it here. As before, we shall write $I^{MIB(r)}_m(\mathbf{r})$ for the depth of insertion in the MI model in which first symbol in each phrase distributes according to vector $\mathbf{r}$ and the number of phrases is distributed according the the *binomial*(m, r) for some $r < 1$. The probability r and the probability vector $\mathbf{r}$ are defined above Lemma 11.

Theorem 5 *For any $\beta > 0$, there exists K such that $I^{GK}_m(a)$ stochastically dominates a random variable that is within distance $O(n^{-\beta})$ from $I^{MIB(r)}_{m-K}(\mathbf{r})$ for some $r < 1$.*

Proof. We have the following chain of inequalities

$$I^{GK}_m(a) \ge I^{GK}_{m,K}(a) \stackrel{d}{=} I^{GK+MI}_{m-K}(a) + O(\varepsilon_m) \succeq_{\rm st} I^{MIB(r)}_{m-K}(\mathbf{r})$$

which completes the proof. ■

4.3 Establishing the Limiting Distribution

We prove now that appropriately normalized I^{GK}_m converges in distribution to the standard normal distribution. Similar conclusion about the typical depth D^{GK}_m will follow directly via the Cesaro limit.

To simplify notation, let $L_m = \frac{\log m}{h}$ and $V_m = \frac{1}{h^3}\left(-\frac{\beta}{\omega} - \frac{2}{\omega}\boldsymbol{\pi}\dot{\mathsf{Q}}^*\boldsymbol{\psi} - h^2\right)\ln m$. We will prove that for all $x = O(1)$

$$\lim_{m\to\infty} \Pr\{\frac{I_m^{GK} - L_m}{\sqrt{V_m}} \ge x\} = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt.$$

By Theorem 4, there exist $\beta > 0$ and K such that the following upper bound holds for all k and m:

$$\Pr\{I_m^{GK} \ge k \mid \text{last phrase starts with } a\} \le \Pr\{I_{m-K}^{MI}(\mathbf{p}_a) \ge k - K\} + O(n^{-\beta}). \qquad (94)$$

Thus

$$\begin{aligned}\Pr\{I_m^{GK} \ge k\} &= \sum_{a\in\mathcal{A}} \Pr\{I_m^{GK} \ge k \mid \text{last } GK \text{ phrase starts with } a\} \\ &\times \ \Pr\{\text{last } GK \text{ phrase starts with } a\} \\ &\le \sum_{a\in\mathcal{A}} \Pr\{I_{m-K}^{MI}(\mathbf{p}_a) \ge k - K\}\Pr\{\text{last } GK \text{ phrase starts with } a\} + O(n^{-\beta}).\end{aligned}$$

By Corollary 1 we know that

$$\lim_{m\to\infty} \Pr\{\frac{I_m^{MI}(\mathbf{p}_a) - L_m}{\sqrt{V_m}} \ge x\} = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt.$$

Since $L_{m-K} = L_m + O(1/m)$, $V_{m-K} = V_m + O(1/m)$, and $\sum_{a\in\mathcal{A}} \Pr\{\text{last } GK \text{ phrase starts with } a\} = 1$, we conclude that

$$\limsup_{m\to\infty} \Pr\{\frac{I_m^{GK} - L_m}{\sqrt{V_m}} \ge x\} \le \lim_{m\to\infty} \frac{1}{\sqrt{2\pi}}\int_{x-O(1/m)}^\infty e^{-t^2/2}dt = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt. \qquad (95)$$

A similar argument works for the lower bound, however, this time we shall use Theorem 5 and Corollary 2. Certainly,

$$\Pr\{I_m^{GK} \ge k\} \ge \Pr\{I_{m-K}^{MIB(r)}(\mathbf{r}) \ge k\} + O(n^{-\beta})\ .$$

By Corollary 2, $(I_m^{MIB(r)}(\mathbf{p}_a) - L_m)/V_m \xrightarrow{d} N(0,1)$, hence by a similar line of reasoning as above we conclude that

$$\liminf_{m\to\infty} \Pr\{\frac{I_m^{GK} - L_m}{\sqrt{V_m}} \ge x\} \ge \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt,$$

which completes the proof for the limiting distribution of I_m^{GK}.

4.4 Establishing the Convergence of Moments

Finally, we prove the existence and convergence of moments of $(I_m^{GK} - L_m)/\sqrt{V_m}$. We accomplish this by showing that there exist constants A_1 and $\alpha_1 < 1$ such that uniformly for all integers ℓ

$$\Pr\left\{\left|\frac{I_m^{GK} - L_m}{\sqrt{V_m}}\right| \geq \ell\right\} \leq A_1\alpha_1^{\sqrt{\ell}}. \tag{96}$$

Indeed, above will prove the existence of the moments and by the *dominated convergence theorem* the moments will tend to the moments of the normal distribution as $n \to \infty$. Notice that in any model I_m cannot be greater than m and therefore there is no need to check the inequality for values of ℓ beyond m.

We present details of the derivations only for the case $\Pr\{I_m^{GK} - L_m \geq \ell\sqrt{V_m}\}$ since the case $\Pr\{I_m^{GK} - L_m \leq -\ell\sqrt{V_m}\}$ can be handled in a similar manner. By (92) we know that $I_m^{GK} \leq I_{m,K}^{GK} + K$ for a fixed K. But, Lemma 13 asserts that $I_{m,K}^{GK}$ is within distance $\varepsilon_m = O(m^{K\log\rho})$, where $\rho < 1$, from I_{m-K}^{GK+MI}. More precisely, for any set of integers B

$$\Pr\{I_{m,K}^{GK} \in B\} \leq (1+\varepsilon_m)\Pr\{I_{m-K}^{GK+MI} \in B\} + O(e^{-\eta\sqrt{m}})$$

for $\eta > 0$. From Theorem 4 we know also that

$$I_{m-K}^{GK+MI}(a) \leq_{\rm st} I_{m-K}^{MI}(\mathbf{p}_a),$$

where above we indicated that phrases starts with symbol a. Finally, Corollary 1 implies that there are constants A and $\alpha < 1$ such that

$$\Pr\left\{\left|\frac{I_m^{MI}(\mathbf{p}_a) - L_m]}{\sqrt{V_m}}\right| \geq \ell\right\} \leq A\alpha^{\ell}.$$

Putting everything together, we obtain

$$\begin{aligned}\Pr\{I_m^{GK} \geq L_m + \ell\sqrt{V_m}\} &\leq (1+\varepsilon_m)\sum_{a\in\mathcal{A}}\Pr\{I_{m-K}^{MI}(\mathbf{p}_a) \geq k - K\} \\ &\times \Pr\{\text{last } GK \text{ phrase starts with } a\} + O(e^{-\eta\sqrt{m}}) \\ &\leq A(1+\varepsilon_m)\alpha^{\ell} + O(e^{-\eta\sqrt{m}}) \leq A_1\alpha_1^{\sqrt{\ell}},\end{aligned}$$

since ℓ cannot be greater than m and therefore $O(e^{-\eta\sqrt{m}})$ can be dominated by $A_1\alpha_1^{\sqrt{\ell}}$ term. This prove the existence and convergence of moments, which completes the proof of Theorem 2.

Appendix A: Alternative Representation of Theorem 1 Results

In this appendix, we show how to prove our alternative representations (19)–(20) for the mean $\mathbf{E}[D_m]$ and $\mathbf{Var}[D_m]$. Instead of presenting a detailed derivations, as in Section 3, we rather sketch here the proof.

We concentrate on evaluating the mean. The starting point is (62), that is,

$$\mathbf{x}(s) = \mathsf{Q}^{-1}\mathbf{x}(s-1) = \sum_{k=0}^{\infty} \mathsf{P}^k(s)\mathbf{x}(s-1).$$

Before we apply the *spectral representation* to $\mathsf{P}^k(s)$, we need some notation. Let us denote by $\lambda(s), \mu_2(s), \ldots, \mu_V(s)$ the eigenvalues of $\mathsf{P}(s)$ with $|\lambda(s)| > |\mu_1(s)| \geq \cdots \geq |\mu_V(s)|$. The corresponding left eigenvectors are $\boldsymbol{\pi}(s), \boldsymbol{\pi}_2(s), \ldots, \boldsymbol{\pi}_V(s)$ while the right eigenvectors are $\boldsymbol{\psi}(s), \boldsymbol{\psi}_2(s), \ldots, \boldsymbol{\psi}_V(s)$. As in [9], we adopt an optional notation for the scalar product of vectors, namely, we either write as before $\mathbf{xy}$ for product of vectors $\mathbf{x}$ and $\mathbf{y}$ or $\langle \mathbf{x}, \mathbf{y} \rangle$. The latter notation is convenient when scalar products are often used, as in this appendix.

By *spectral representation* (cf. [19]), matrix $\mathsf{P}(s)$ can be represented as

$$\mathsf{P}^k(s)\mathbf{x}(s-1) = \lambda^k(s)\langle \boldsymbol{\pi}(s), \mathbf{x}(s-1)\rangle \boldsymbol{\psi}(s) + \sum_{i=2}^{V} \mu_i^k(s)\langle \boldsymbol{\pi}_i(s), \mathbf{x}(s-1)\rangle \boldsymbol{\psi}_i(s).$$

Thus $\mathbf{b}(s) = \Gamma(s)\mathbf{x}(s)$ becomes

$$\mathbf{b}(s) = \frac{\Gamma(s)\langle \boldsymbol{\pi}(s), \mathbf{x}(s-1)\rangle \boldsymbol{\psi}(s)}{1-\lambda(s)} + \sum_{i=2}^{V} \frac{\Gamma(s)\langle \boldsymbol{\pi}_i(s), \mathbf{x}(s-1)\rangle \boldsymbol{\psi}_i(s)}{1-\mu_i(s)}. \tag{97}$$

In order to obtain leading asymptotics of $B^*(s) = \mathbf{p}(s)\mathbf{b}(s) + \Gamma(s)x(s-1)$ (cf. (64)), we need Laurent's expansion of the above around the roots of $\lambda(s) = -1$. Observe that the second term of (97) contributed $o(m)$ since $\lambda(s)$ is the largest eigenvalue (cf. [9]), hence we further ignore this negligible term in our derivations. To simplify the presentation, we only deal here with the root $s_0 = -1$. We use our previous expansions for $\mathbf{x}(s-1)$ and $\Gamma(s)$ together with

$$\begin{aligned}
\frac{1}{1-\lambda(s)} &= \frac{-1}{\dot\lambda(-1)}\frac{1}{s+1} + \frac{\ddot\lambda(-1)}{2\dot\lambda^2(-1)} + O(s+1),\\
\boldsymbol{\psi}(s) &= \boldsymbol{\psi} + \dot{\boldsymbol{\psi}}(-1)(s+1) + O((s+1)^2).
\end{aligned}$$

This finally leads to

$$\begin{aligned}
B^*(s) &= \frac{-1}{\dot\lambda(-1)}\frac{1}{(s+1)^2}\\
&+ \frac{1}{s+1}\left(\frac{\langle \boldsymbol{\pi}, \dot{\mathbf{x}}(-2)\rangle}{\dot\lambda(-1)} - \frac{\gamma-1}{\dot\lambda(-1)} + \frac{\langle \mathbf{p}(-1), \dot{\boldsymbol{\psi}}(-1)\rangle}{\dot\lambda(-1)} + \frac{\ddot\lambda(-1)}{2\dot\lambda^2(-1)} - 1\right) + O(1).
\end{aligned}$$

After finding the inverse Mellin transform of the above and depoissonizing, we prove the alternative representation (19).

Finally, we turn our attention to the second factorial moment and the variance. We need to study $\mathbf{c}(s) = \Gamma(s)\mathbf{v}(s)$ where $\mathbf{v}(s) = 2\mathsf{Q}^{-1}(s)\mathsf{P}(s)\mathbf{x}(s) + \mathsf{Q}^{-1}(s)\mathbf{v}(s-1)$. As before, we obtain

$$\mathbf{c}(s) = \frac{2\Gamma(s)\langle \boldsymbol{\pi}(s), \mathbf{x}(s-1)\rangle \langle \boldsymbol{\pi}(s), \mathsf{P}(s)\boldsymbol{\psi}(s)\rangle \boldsymbol{\psi}(s)}{(1-\lambda(s))^2} + O\left((1-\lambda(s))^{-1}\right).$$

Similar algebra as above leads to

$$\begin{aligned} \mathbf{c}(s) &= \frac{-2}{\dot{\lambda}^2(-1)} \frac{1}{(s+1)^3} \\ &+ \frac{1}{(s+1)^2}\left(\frac{\ddot{\lambda}(-1)}{2\dot{\lambda}^3(-1)} + 2\frac{\gamma - 1 - \langle \boldsymbol{\pi}, \dot{\mathbf{x}}(-2)\rangle - \langle \mathbf{p}(-1), \dot{\boldsymbol{\psi}}(-1)\rangle - \dot{\lambda}(-1)}{\dot{\lambda}^2(-1)}\right) \\ &+ O\left(\frac{1}{s+1}\right). \end{aligned}$$

This is sufficient to prove (20), after some tedious algebra that was helped by MAPLE.

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] R. Bradley, Basic Properties of Strong Mixing Conditions, in *Dependence in Probability and Statistics* (Eds. E. Eberlein and M. Taqqu), 165–192, 1986.

[4] Julien Clément, Philippe Flajolet, Brigitte Vallée, J. Clement, P. Flajolet, and B. Vallée, Dynamic Sources in Information Theory: A General Analysis of Trie Structures, *Algorithmica*, 2000.

[5] P. Flajolet, Singularity Analysis and Asymptotics of Bernoulli Sums, *Theoretical Computer Science*, 215, 371–381, 1999.

[6] P. Flajolet, X. Gourdon, P. Dumas, Mellin Transforms and Asymptotics: Harmonic Sums, *Theoretical Computer Science*, 144, 3-58, 1995.

[7] E. Gilbert and T. Kadota, The Lempel–Ziv Algorithm and Message Complexity, *IEEE Trans. Information Theory*, 38, 1839–1842, 1992.

[8] Y. Hershkovits and J. Ziv, On Sliding-Window Universal Data Compression with Limited memory, *IEEE Trans. Information Theory*, 44, 66–78, 1997.

[9] P. Jacquet and W. Szpankowski, Analysis of Digital Tries with Markovian Dependency, *IEEE Trans. Information Theory*, 37, 1470–1475, 1991.

[10] P. Jacquet and W. Szpankowski, Asymptotic Behavior of the Lempel-Ziv Parsing Scheme and Digital Search Trees, *Theoretical Computer Science*, 144, 161-197, 1995.

[11] P. Jacquet and W. Szpankowski, Analytical Depoissonization Lemma and Its Applications, *Theoretical Computer Science*, 201, 1–62, 1998.

[12] P. Jacquet and W. Szpankowski, Entropy Computations Via Analytic Depoissonization, *IEEE Trans. on Information Theory*, 45, 1072-1081, 1999.

[13] D. Knuth, *The Art of Computer Programming. Sorting and Searching.* Vol. 3., Addison-Wesley 1973.

[14] G. Louchard and W. Szpankowski, Average Profile and Limiting Distribution for a Phrase Size in the Lempel-Ziv Parsing Algorithm, *IEEE Trans. Information Theory*, 41, 478-488, 1995.

[15] G. Louchard and W. Szpankowski, On the Average Redundancy Rate of the Lempel-Ziv Code, *IEEE Trans. Information Theory*, 43, 2–8, 1997.

[16] G. Louchard, W. Szpankowski and J. Tang, Average Profile of Generalized Digital Search Trees and the Generalized Lempel-Ziv Algorithm, *SIAM J. Computing*, 28, 935-954, 1999.

[17] H. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York 1992.

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

[19] B. Noble and J. Daniel, *Applied Linear Algebra*, Prentice Hall, Englewood Cliffs, 1988.

[20] B. Pittel, Asymptotic Growth of a Class of random Trees, *Ann. Probab.*, 13, 414–427, 1985.

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

[22] D. Stoyan, *Comparison methods for Queues and Other Stochastic Models*, John Wiley & Sons, Chichester 1983.

[23] W. Szpankowski, A Characterization of Digital search Trees From the Successful Search Viewpoint, *Theoretical Computer Science*, 85, 117–134, 1991.

[24] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences* John Wiley & Sons, New York, 2001.

[25] J. Tang, *Probabilistic Analysis of Digital Search Trees*, Ph.D. Thesis, Purdue University 1996.

[26] B. Vallée, Dynamical Sources in Information Theory : Fundamental intervals and Word Prefixes, *Algorithmica*, 2000.

[27] A. J. Wyner, The Redundancy and Distribution of the Phrase Lengths of the Fixed-Database Lempel-Ziv Algorithm, *IEEE Trans. Information Theory*, 43, 1439–1465, 1997.

[28] J. Ziv, Back from Infinity: A Constrained Resources Approach to Information Theory, *IEEE Information Theory Society Newsletter*, 48, 30–33, 1998.

[29] J. Ziv and A. Lempel, Compression of Individual Sequences via Variable-Rate Coding, *IEEE Trans. Information Theory*, 24, 530-536, 1978.