Tunstall Code, Khodak Variations, and Random Walks[*]

April 15, 2010

Michael Drmota[*]
Inst. Discrete Math. and Geometry
TU Wien
A-1040 Wien
Austria
michael.drmota@tuwien.ac.at

Yuriy A. Reznik
Qualcomm Inc.
5775 Morehouse Dr.
San Diego, CA 92121
U.S.A.
yreznik@qualcomm.com

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

Abstract

A variable-to-fixed length encoder partitions the source string into variable-length phrases that belong to a given and fixed dictionary. Tunstall, and independently Khodak, designed variable-to-fixed length codes for memoryless sources that are optimal under certain constraints. In this paper, we study the Tunstall and Khodak codes using variety of techniques ranging from stopping times for sums of independent random variables to Tauberian theorems and Mellin transform. After proposing an algebraic characterization of the Tunstall and Khodak codes, we present new results on the variance and a central limit theorem for dictionary phrase lengths. This analysis also provides a new argument for obtaining asymptotic results about the mean dictionary phrase length and average redundancy rates.

Index Terms: Variable-to-fixed length codes, Tunstall code, renewal theory, stopping time, analytic information theory, Mellin transform, Tauberian theorems.

1 Introduction

A variable-to-fixed length encoder partitions the source string over an m-ary alphabet $\mathcal{A}$ into a concatenation of variable-length phrases. Each phrase except the last one is constrained to belong to a given dictionary $\mathcal{D}$ of source strings; the last phrase is a non-null prefix of a dictionary entry. It is convenient to represent a uniquely parsable dictionary by a complete parsing tree $\mathcal{T}$, i.e., a tree in which every internal node has all m children nodes. The dictionary entries $d \in \mathcal{D}$ correspond to the leaves of the associated parsing tree. The encoder represents each parsed string by the fixed length binary code word corresponding to its dictionary entry. If the dictionary $\mathcal{D}$ has M entries, then the code word for each phrase has $\lceil \log_2 M \rceil$ bits. The best known variable-to-fixed length code is now generally attributed to Tunstall [34]; however, it was independently discovered by Khodak [14], Verhoeff [35], and possibly others. Known applications of Tunstall codes include error-resilient video and image coders, encoders for bi-level images, data recording, retrieval systems and approximating uniform distribution [1] (cf. also [3]).

Tunstall's algorithm is easy to visualize through evolving parsing trees in which every edge corresponds to a letter from the source alphabet $\mathcal{A}$. Start with a tree having a root node and m leaves corresponding to symbols from $\mathcal{A}$. At each iteration select the current leaf corresponding to a string of the *highest probability* and grow m children out of it, one for each symbol in $\mathcal{A}$. After J iterations, the parsing tree has J non-root internal nodes and $M = (m-1)J + 1$ leaves, which each corresponds to a distinct dictionary entry. The dictionary entries are prefix-free and can be easily enumerated. Note that a string x of the highest probability, is not usually unique since there are in principle imany different strings of the same probability (of the same type). Tunstall's algorithm adds these strings to the dictionary (parsing tree) one by one in an arbitrary (random) order.

Tunstall's algorithm has been studied extensively (cf. the survey article [1]). Simple bounds for its redundancy were obtained independently by Khodak [14] and by Jelinek and Schneider [13]. Tjalkens and Willems [31] were the first to look at extensions of this code to sources with memory. Savari and Gallager [22] proposed a generalization of Tunstall's algorithm for Markov sources and used renewal theory for an asymptotic analysis of average code word length and redundancy for memoryless and Markov sources. Savari [23] later published a non-asymptotic analysis of the Tunstall code for binary, memoryless sources with small entropies. Universal variable-to-fixed length codes were analyzed in [15, 16, 18, 32, 33, 36]; however, we are unaware of analyses of the minimax redundancy for variable-to-fixed and variable-to-variable length codes, and these problems remain open. In this paper, we offer a new perspective and generalized asymptotic analysis of the Tunstall and Khodak codes for known distributions. Among others, we establish the limiting distribution of the phrase length and provide a precise asymptotic analysis of the average redundancy of the Tunstall and Khodak codes.

In our analysis, we focus on Khodak's [14] construction of the variable-to-fixed length codes (see also [15]). Khodak independently discovered the Tunstall code using a rather different approach. Let p_i be the probability of the i-th source symbol and let $p_{\min} = \min\{p_1, \ldots, p_m\}$. Throughout we assume that the probabilities p_i are known. Khodak suggested choosing a real

number $r \in (0, p_{\min})$ and then growing a complete parsing tree until all its leaves d satisfy

$$p_{\min} r \leq P(d) < r, \quad d \in \mathcal{D}. \tag{1}$$

It follows that if y is a proper prefix of one or more entries of $\mathcal{D} = \mathcal{D}_r$, i.e., y corresponds to an internal node of $\mathcal{T} = \mathcal{T}_r$, then

$$P(y) \geq r. \tag{2}$$

Therefore, it is easier to characterize the internal nodes of the parsing tree $\mathcal{T}_r$ rather than its leaves. This algebraic characterization is crucial for our analysis, and will be used throughout this paper.

It is known (see, e.g., [13, Lemma 6], and [22, Lemma 2]) that the resulting parsing tree for the Khodak algorithm is exactly the same as a tree constructed by Tunstall's algorithm. However, one should observe that in Khodak's construction all strings with the same (highest) probability are added at once while in Tunstall's algorithm one by one in an arbitrary order. Therefore, in Khodak's algorithm the number of dictionary entries, M_r, does not attain all positive integers (there are certain jumps depending on the probabilities p_i.) The asymptotic relationship between r and the resulting number of entries M_r was studied in [22] and will be established here in a different way in Theorem 3.

Our main result presented in Theorem 2 establishes the central limit theorem for the phrase length $\mathcal{L} = \mathcal{L}_r$ of the Khodak algorithm. We will present two different proof methods. The first one, presented in Section 3 is based on the observation that $\mathcal{L}_r$ can be interpreted as the stopping time of a random walk (i.e., a sum of independent random variables) that directly provides the central limit theorem (cf. also recent paper [11]). Applying analytic techniques we also obtain asymptotic expansions for M_r as $r \to 0$ (see Theorem 3). Then the central limit theorem can be rewritten in terms of the number of phrases $M = M_r$. Recall that M_r, given by Khodak's condition (2), does not attain all positive integers. However, by using an "interpolation argument" we also derive a central limit theorem for the phrase length $\tilde{\mathcal{L}}_M$ of the original Tunstall code in terms of M (see Theorem 1).

Our second approach to $\mathcal{L} = \mathcal{L}_r$, presented in Section 4, is entirely analytic and applies tools such as generating functions, Mellin transform, and Tauberian theorems [27]. This analysis provides a precise asymptotic characterization of the the moment generating function for the phrase length $\mathcal{L}_r$. We note that this work directly extends recent analyses of fixed-to-variable codes (cf. [5, 10, 26, 27]) through tools of analytic algorithmics and hence belongs to the domain of analytic information theory. We point out that a slight modification of the Tunstall code (e.g., bounding its phrase length) may lead to considerable analytic difficulties that can be overcome by analytic tools [6]. Furthermore, our analytic approach allows us to estimate all moments and in principle the large deviations.

In passing, we should mention that in recent years we have seen resurgence of interest in variable-to-fixed-length codes, resulting in several faster techniques for their constructions (cf. [2, 21, 28, 30, 20]), as well as novel applications such as the use of Tunstall algorithm for the approximation of uniform distributions for random number generation and related problems [3]. We believe that our results will be useful for better understanding of these new techniques and applications.

The paper is organized as follows. In the next section we present our main results and their consequences. Section 3 is devoted to proofs of these results by a combination of renewal theory and Tauberian theorems. Finally, in Section 4 we briefly present a uniform approach using analytic techniques such as generating functions, functional equations, and Mellin transform techniques.

2   Main Results and Consequences

We consider a memoryless source over an m-ary alphabet $\mathcal{A}$. Let $p_i > 0$ be the probability of the ith letter of alphabet $\mathcal{A}$, $i \in \{1, \ldots, m\}$. Given a complete prefix free dictionary $\mathcal{D}$ and its corresponding complete parsing tree $\mathcal{T}$, the encoder partitions the source output sequence into a sequence of variable-length phrases. Let $d \in \mathcal{D}$ denote a dictionary entry (and a leaf in $\mathcal{T}$), $P(d)$ be its probability, and $|d|$ be its length. Since we assume that parsing tree is complete we have

$$\sum_{d \in \mathcal{D}} P(d) = 1.$$

Thus $(\mathcal{D}, P)$ is a probability space and every parameter on $\mathcal{D}$ becomes a random variable. Our main focus is on the random variable $\mathcal{L} = |d|$, the phrase length of a dictionary string.

Throughout we use the following additional notation. Let

$$H = p_1 \log(1/p_1) + \cdots p_m \log(1/p_m)$$

denote the *entropy* in *natural* units and

$$H_2 = p_1 \log^2(1/p_1) + \cdots p_m \log^2(1/p_m)$$

be a parameter needed to express the variance of the phrase length.

2.1   Central Limit Theorems for the Dictionary Length

We first consider the Tunstall code and present the central limit theorem for its phrase length $\tilde{\mathcal{L}}_M$. We prove it and all our findings discussed here in the next sections.

Theorem 1. *Let $\tilde{\mathcal{L}}_M$ denote the phrase length of the Tunstall code when the dictionary size is $M \geq 1$. Then for a biased source (i.e., when the probabilities p_i are not equal)*

$$\frac{\tilde{\mathcal{L}}_M - \frac{1}{H} \log M}{\sqrt{\left(\frac{H_2}{H^3} - \frac{1}{H}\right) \log M}} \to N(0, 1),$$

where $N(0, 1)$ denotes the standard normal distribution, and

$$\begin{aligned}
\mathbf{E}[\tilde{\mathcal{L}}_M] &= \frac{\log M}{H} + O(1), \\
\mathbf{Var}[\tilde{\mathcal{L}}_M] &\sim \left(\frac{H_2}{H^3} - \frac{1}{H}\right) \log M
\end{aligned}$$

for $M \to \infty$.

Remark: Observe that for the unbiased case (i.e., $p_1 = p_2 = \cdots - p_m = 1/m$), we have $H = \log m$ and $H_2 = \log^2 m$ which suggests that $\mathbf{Var}[\tilde{\mathcal{L}}_M] = O(1)$. This is actually true since the algorithms always tends to generate a complete m-ary tree so that the phrase lengths are always concentrated at one level or at two consecutive levels. Obviously, there is no central limit theorem in this case.

Since Khodak's construction induces Tunstall codes for special values of $M = M_r$, we obtain a corresponding property for $\mathcal{L}_r$. In fact, we will show in Section 4 that Theorems 1 and 2 are equivalent.

Theorem 2. *Let $\mathcal{L}_r$ denote the phrase length in Khodak's construction of the Tunstall code with a dictionary of size M_r over a biased memoryless source. Then*

$$\frac{\mathcal{L}_r - \frac{1}{H} \log M_r}{\sqrt{\left(\frac{H_2}{H^3} - \frac{1}{H}\right) \log M_r}} \to N(0, 1), \tag{3}$$

and

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log M_r}{H} + O(1), \tag{4}$$

$$\mathbf{Var}[\mathcal{L}_r] \sim \left(\frac{H_2}{H^3} - \frac{1}{H}\right) \log M_r \tag{5}$$

for $r \to 0$.

Note that Theorem 2 is implicit since there is no explicit dependence on r. Of course, it it obvious that $M_r \to \infty$ as $r \to 0$. In the next section we present precise results on the dependence of M_r on r.

2.2   The Dependence on r in Khodak's Construction

We consider two cases: We say that $\log(1/p_1), \ldots, \log(1/p_m)$ are *rationally related* if there exists a positive real number L such that $\log(1/p_1), \ldots, \log(1/p_m)$ are integer multiples of L, that is,

$$\log(1/p_j) = n_j L, \quad n_j \in \mathbb{Z}, \quad (1 \le j \le m).$$

Without loss of generality we can assume that L is as large as possible which is equivalent to $\gcd(n_1, \ldots, n_m) = 1$. For example, in the binary case $m = 2$ this is equivalent to the statement that the ratio $\log(1/p_1)/\log(1/p_2)$ is rational. Similarly we say that $\log(1/p_1), \ldots, \log(1/p_m)$ are *irrationally related* if they are not rationally related.

Now we are in the position to state our second main result concerning the dependence of M_r on r that plays the crucial role in the analysis.

Theorem 3. *Let $M = M_r$ denote the dictionary size in Khodak's construction of the Tunstall code.*

(i) *If $\log(1/p_1), \ldots, \log(1/p_m)$ are irrationally related, then*

$$M_r = \frac{m - 1}{rH} + o(1/r), \tag{6}$$

and

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log(1/r)}{H} + \frac{H_2}{2H^2} + o(1). \tag{7}$$

(ii) *If* $\log(1/p_1), \ldots, \log(1/p_m)$ *are rationally related, let* $L > 0$ *be the largest real number for which* $\log(1/p_j)$, $1 \leq j \leq m$ *are integer multiples of* L. *Then*

$$M_r = (m-1)\frac{Q_1(\log(1/r))}{rH} + O(r^{-(1-\eta)}) \tag{8}$$

for some $\eta > 0$, *where*

$$Q_1(x) = \frac{L}{1 - e^{-L}} e^{-L\langle \frac{x}{L} \rangle}, \tag{9}$$

and $\langle y \rangle = y - \lfloor y \rfloor$ *is the fractional part of the real number* y. *Furthermore,*

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log(1/r)}{H} + \frac{H_2}{2H^2} + \frac{Q_2(\log(1/r))}{H} + O(v^{-\eta}) \tag{10}$$

for some $\eta > 0$, *where*

$$Q_2(x) = L\left(\frac{1}{2} - \left\langle \frac{x}{L} \right\rangle\right) \tag{11}$$

is an oscillating function.

By combining (6) and (7) resp. (8) and (10) we can be even more precise. In the irrational case we have

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log M_r}{H} + \frac{\log H}{H} - \frac{\log(m-1)}{H} + \frac{H_2}{2H^2} + o(1)$$

and in the rational case we find

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log M_r}{H} + \frac{\log H}{H} - \frac{\log(m-1)}{H} + \frac{H_2}{2H^2} + \frac{Q_2(\log v) - \log(Q_1(\log v))}{H} + O(M_r^{-\eta}).$$

Note that (9) and (11) yield

$$Q_2(\log v) - \log\left(Q_1(\log v)\right) = -\log L + \log(1 - e^{-L}) + \frac{L}{2} = \log\left(\frac{\sinh(L/2)}{L/2}\right)$$

so that there is actually no oscillation. We find

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log M_r}{H} + \frac{\log H}{H} - \frac{\log(m-1)}{H} + \frac{H_2}{2H^2} + \frac{1}{H}\log\left(\frac{\sinh(L/2)}{L/2}\right) + O(M_r^{-\eta}).$$

As a direct consequence, we can derive a precise asymptotic formula for the average redundancy of the Tunstall and Khodak codes that is defined in [22] by

$$\mathcal{R}_M = \frac{\log M}{\mathbf{E}[\mathcal{L}_M]} - H. \tag{12}$$

The next result follows from the above derivations.

Corollary 1. *Let $\mathcal{D}_r$ denote the dictionary in Khodak's construction of the Tunstall code of size M_r. If $\log(1/p_1), \ldots, \log(1/p_m)$ are irrationally related, then*

$$\mathcal{R}_{M_r} = \frac{H}{\log M_r}\left(-\frac{H_2}{2H} - \log H + \log(m-1)\right) + o\left(\frac{1}{\log M_r}\right).$$

In the rationally related case, we have

$$\mathcal{R}_{M_r} = \frac{H}{\log M_r}\left(-\frac{H_2}{2H} - \log H + \log(m-1) - \log\left(\frac{\sinh(L/2)}{L/2}\right)\right) + O\left(\frac{1}{(\log M_r)^2}\right),$$

where $L > 0$ is the largest real number for which $\log(1/p_1), \ldots, \log(1/p_m)$ are integer multiples of L.

In passing we observe that the Corollary 1 is a special case of Theorems 5 and 12 of [22] for the Tunstall code. Observe also that the Tunstall code redundancy has some oscillations for the rational case which disappear for the Khodak code. This is explained in the next section.

3 Proofs of the Theorems

In this section we prove Theorem 1–3 using a combination of renewal theory (cf. Section 3.2) and analytic techniques (cf. Section 3.1). In the next Section 4, we present a unified and general proof that falls under the analytic information theory paradigm.

3.1 Proof of Theorem 3(i)

We prove here (6) and (8) of Theorem 3(i). Set $v = 1/r$ and let $A(v)$ devote the number of source strings with probability at least v^{-1}; i.e.,

$$A(v) = \sum_{y:P(y)\geq 1/v} 1. \tag{13}$$

Observe that $A(v)$ represents the number of internal nodes in Khodak's construction with parameter $r = v^{-1}$ of a Tunstall tree. Equivalently, $A(v)$ counts the number of strings y with the self-information $-\log P(y) \leq \log v$. The function $A(v)$ satisfies the following recurrence.

Lemma 1.

$$A(v) = \begin{cases} 0 & v < 1, \\ 1 + A(vp_1) + \cdots + A(vp_m) & v \geq 1. \end{cases} \tag{14}$$

Proof. By definition we have $A(v) = 0$ for $v < 1$. Now suppose that $v \geq 1$. Since every m-ary string is either the empty string or a string starting with a source letter $j \in \mathcal{A}$, $1 \leq j \leq m$, we directly find the recurrence $A(v) = 1 + A(vp_1) + \cdots + A(vp_m)$. □

Since $A(v)$ represents the number of internal nodes in Khodak's construction with parameter v^{-1} it follows that the dictionary size is given by

$$M_r = |\mathcal{D}_r| = (m-1)A(v) + 1.$$

Therefore, it is sufficient to obtain asymptotic expansions for $A(v)$ for $v \to \infty$.

Setting $F(x) = A(e^x)$ and $c_j = \log(1/p_j)$ the above recurrence relations for $A(v)$ rewrite to $F(x) = 0$ for $x < 0$ and to

$$F(x) = 1 + F(x - c_1) + \cdots + F(x - c_m) \qquad (x \geq 0).$$

Let α be the smallest positive real root of the equation $z^{c_1} + \cdots + z^{c_m} = 1$, set $\varphi = 1/\alpha$ and

$$C = \sum_{j=1}^{m} c_j \varphi^{-c_j}.$$

Choi and Golin [4] used analytic techniques (i.e., Laplace transform) to prove that as $x \to \infty$

$$F(x) \sim \frac{\varphi^x}{C \log \varphi},$$

if $c_1, \ldots, c_m$ are irrationally related, and

$$F(x) = \frac{L}{C(1 - \varphi^{-L})} \varphi^x + O(\rho^x),$$

if $c_1, \ldots, c_m$ are rationally related and $0 \leq \rho < \varphi$.

In our case we have $\alpha = 1/e$, $\varphi = e$, and $C = H$. Hence,

$$A(v) = F(\log v) \sim \frac{v}{H}$$

if $\log(1/p_1), \ldots, \log(1/p_m)$ are irrationally related, and

$$A(v) = F(\log v) = \frac{vL}{H(1 - e^{-L})} + O(v^{1-\eta})$$

if $\log(1/p_1), \ldots, \log(1/p_m)$ are rationally related. The asymptotic relations, (6) and (8) follow.

3.2 Proof of Theorem 2 and Theorem 3(ii)

Let us consider Khodak's formulation of the Tunstall code and let $\mathcal{L}_r$ denote the phrase length for the dictionary $\mathcal{D}_r$. The essential observation is that the phrase length $\mathcal{L}_r$ can be interpreted as the stopping time of a sum of independent random variables (i.e., a random walk).

Lemma 2. *Let X_j, $j \geq 1$, be independent random variables with probability distribution*

$$\mathbb{P}[X_j = \log(1/p_j)] = p_j, \quad (1 \leq j \leq m),$$

and set

$$S_n = \sum_{j=1}^{n} X_j.$$

to be a random walk. Let also $N(c)$ be the stopping time

$$N(c) = \min\{n : S_n > c\},$$

that is, the first time S_n exceeds threshold c. Then the distributions of $\mathcal{L}_r$ and $N(\log(1/r))$ coincide.

Proof. Consider the infinite m-ary tree $\mathcal{T}$ and define a discrete random walk on $\mathcal{T}$ in the following way. It starts at the root and at each step it goes to the j-th successor of the present node with probability p_j, $1 \leq j \leq m$. By the definition after n steps the logarithm $\log(1/P(x))$ of the probability $P(x)$ of the endpoint x (i.e., leaf x) is equal to S_n in distribution.

Recall that the leaves d of the parsing tree of $\mathcal{T}_r$ are precisely those nodes for which $P(d) < r$ and $P(y) \geq r$ for all y on the path from the root to d (that are different from d). Equivalently we have $\log(1/P(d)) > \log(1/r)$. Thus, the parsing tree $\mathcal{T}_r$ corresponds to all random walks that are stopped at those nodes where $\log(1/P(d)) > \log(1/r)$. Hence, the distribution of $\mathcal{L}_r$ and the stopping time $N(\log(1/r))$ coincide. $\qquad\qquad\square$

We also shall use Theorem 2.5 of Gut [7] who proved that for $c \to \infty$

$$\frac{N(c) - c/\theta}{\sqrt{\sigma^2 c/\theta^3}} \to N(0,1),$$

provided that first and second moments, $\mathbf{E}[X_j] = \theta$ and $\mathbf{Var}[X_j] = \sigma^2$, are non-zero and finite.

In our particular case we have

$$c = \log(1/r)$$

$$\theta = \mathbf{E}[X_j] = \sum_{j=1}^{m} p_j \log(1/p_j) = H, \quad \text{and}$$

$$\sigma^2 = \mathbf{Var}[X_j] = \sum_{j=1}^{m} p_j \log^2(1/p_j) - H^2 = H_2 - H^2.$$

Observe that $\sigma^2 >$ provided all probabilities p_i are not equal. Hence, Theorem 2 follows immediately.

Next we observe that X_j has a lattice distribution (in the terminology of [7]), that is, X_j contains only integer multiples of a positive real number L, if and only if $\log(1/p_1)$, ..., $\log(1/p_m)$ are rationally related. Thus we can apply [7, Theorem 2.6] and [7, Theorem 2.7'] and obtain, as $c \to \infty$,

$$\mathbf{E}[N(c)] = \frac{c}{\theta} + \frac{\mathbf{E}[X_1^2]}{2(\mathbf{E}[X_1])^2} + o(1),$$

$$\mathbf{Var}[N(c)] = \frac{\sigma^2 c}{\theta^3} + o(c),$$

if $\log(1/p_1), \ldots, \log(1/p_m)$ are irrationally related, and

$$\mathbf{E}[N(c)] = \frac{c}{\theta} + \frac{\mathbf{E}[X_1^2]}{2(\mathbf{E}[X_1])^2} + \frac{L}{\theta}\left(\frac{1}{2} - \left\langle\frac{c}{L}\right\rangle\right)o(1),$$

$$\mathbf{Var}[N(c)] = \frac{\sigma^2 c}{\theta^3} + o(c),$$

if $\log(1/p_1), \ldots, \log(1/p_m)$ are rationally related. In summary, the asymptotic relations (7) and (10) follow, given (6) and (8) just proved.

9

3.3 Proof of Theorem 1

We finally show that Theorem 1 can be deduced from Theorem 2. (The converse is obviously true.) This follows, informally, from the fact that Tunstall's code and Khodak's code are "almost equivalent." They ultimately produce the same parsing trees, however, they react differently to the probability tie when expanding a leaf. More precisely, when there are several leaves with the same probability, the Tunstall algorithm selects *one* leaf and expands it, then selects another leaf of the same probability, and continues doing it until all leaves of the same probability are expanded. The Khodak algorithm expands *all* leaves with the same probability simultaneously, in parallel; thus there are "jumps" in M_r when the parsing tree grows. This situation occurs for the rational case and for the irrational case.

Let's be more precise. Suppose that r is chosen in a way that there exists a word x with $P(x) = r$. In particular the dictionary $\mathcal{D}_r$ contains all external nodes d that are adjacent to internals x with $P(x) = r$. Now let $\tilde{\mathcal{D}}_M$ be the dictionary (of size M) of any Tunstall code where only some of these internal nodes x with $P(x) = r$ have been expanded. Then $\mathcal{D}_r$ is the Tunstall code where all nodes x with $P(x) = r$ have been expanded. Hence, by this coupling of the dictionaries we certainly have for the dictionary lengths $|\tilde{\mathcal{L}}_M - \mathcal{L}_r| \leq 1$. This also implies that $\mathbf{E}[\tilde{\mathcal{L}}_M] = \mathbf{E}[\mathcal{L}_r] + O(1)$ and $\mathbf{Var}[\tilde{\mathcal{L}}_M - \mathcal{L}_r] = O(1)$.

We also observe that the central limit theorem is not affected by this variation. Since $\mathcal{D}_r$ satisfies a central limit theorem (see Theorem 2) we find

$$\frac{\tilde{\mathcal{L}}_M - \frac{1}{H} \log M}{\sqrt{\left(\frac{H_2}{H^3} - \frac{1}{H}\right) \log M}} \to N(0, 1).$$

For the expected value and variance we have $\mathbf{E}[\tilde{\mathcal{L}}_M] = \frac{\log M}{H} + O(1)$ and

$$\mathbf{Var}[\tilde{\mathcal{L}}_M] = \mathbf{Var}[\mathcal{L}_r] + O\left(\sqrt{\mathbf{Var}[D_r]}\right)$$
$$\sim \left(\frac{H_2}{H^3} - \frac{1}{H}\right) \log M.$$

Indeed, more generally, let $Y_n = X_n + Z_n$ and we know that X_n satisfies a central limit theorem of the form

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

such that $\mathbf{Var}[X_n] \to \infty$ as well as $\mathbf{Var}[Z_n]/\mathbf{Var}[X_n] \to 0$ as $n \to \infty$. Then also Y_n satisfies a central limit theorem, i.e.

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

and we have

$$\mathbf{Var}[Y_n] = \mathbf{Var}[X_n] + \mathbf{Var}[Z_n] + O(\sqrt{\mathbf{Var}[X_n]\mathbf{Var}[Z_n]}) = \mathbf{Var}[X_n] \cdot (1 + o(1))$$

which follows from Cauchy-Schwarz's inequality

$$\mathbf{E}[(X_n - \mathbf{E}[X_n])(Z_n - \mathbf{E}[Z_n])] \leq (\mathbf{Var}[X_n])^{1/2}(\mathbf{Var}[Z_n])^{1/2}.$$

This completes the proof of Theorem 1.

4   A Unified Analysis via Mellin Transform Techniques

In this section we prove again Theorem 2 and 3 in a unified way via generating function, Mellin transform, and Tauberian techniques. These techniques constitute the main tools of analytic information theory [27].

As we notice in the previous section, analytic tools were already used to derive the number of dictionary entries through the function $A(v)$ defined in (13). More generally, in some applications, including a modified Tunstall code [6], one often analyses a generalized $A(v)$ function defined as follows

$$A(v) = \sum_{y:P(y)\geq 1/v;\ y\in\mathcal{C}} f(y),$$

where $f(v)$ is a function and $\mathcal{C}$ is an additional constraint. For example, in the modified Tunstall code discussed in [6] the phrase length is bounded by a K; hence in this case $\mathcal{C} = \{y \leq K\}$. It is shown in [6] that such a simple modification of the summation index of $A(v)$ leads to considerable challenges that can be overcome only by analytic tools.[†] In summary, we believe the methodology discussed below offers us significant advantages and expands its applicability beyond Tunstall code.

We should point out that the rationally related case of Theorems 2 and 3 is elementary (i.e., complex analysis is not used), while the irrational case requires non-trivial tools like Wiener's Tauberian theorem (cf. [4]). In fact, we can uniformly use the Mellin transform for the rational and irrational cases. However, in the sequel we concentrate on the more challenging irrational case. To simplify our presentation in this section, we only consider the binary case with $m = 2$ and $p_1 \neq p_2$. Extension to m-ary alphabet is straightforward. It amounts to replace $1 - p_1^{-s} - p_2^{-s}$ by $1 - \sum_{i=1}^{m} p_i^{-s}$. (e.g., see [27] for further explanations).

4.1   Combinatorics

In order to obtain the results for $M_r = (m - 1)A(v) + 1$ and $\mathcal{L}_r$ we analyze

$$A(v) = \sum_{y:P(y)\geq 1/v} 1$$

and the probability generating function

$$D(v, z) = \sum_{d\in\mathcal{D}_r} P(d)z^{|d|},$$

where we use the convention $v = 1/r$.

By Lemma 1 we already know that $A(v)$ satisfies the recurrence $A(v) = 1 + A(p_1v) + A(p_2v)$. Interestingly, $D(v, z)$ can be characterized in a similar way.

Lemma 3. *Let* $S(v, z)$ *be defined by*

$$S(v, z) = \sum_{y:P(y)\geq 1/v} P(y)z^{|y|}.$$

[†]In this particular case, one must consider infinite number of saddle points that may coincide with poles; details can be found in [6].

Then $S(v, z)$ satisfies the recurrence

$$S(v, z) = \begin{cases} 0 & v < 1, \\ 1 + p_1 S(vp_1, z) + p_2 S(vp_2, z) & v \geq 1. \end{cases} \tag{15}$$

Furthermore,

$$D(v, z) = 1 + (z - 1)S(v, z) \tag{16}$$

for all complex z.

Proof. The recurrence (15) can be derived in the same way as in the proof of Lemma 1. The relation (16) follows from the following general fact on trees. Let $\tilde{\mathcal{D}}$ be a uniquely parsable dictionary (e.g., leaves in the corresponding parsing tree) and $\tilde{\mathcal{Y}}$ be the collection of strings which are proper prefixes of one or more dictionary entries (e.g., internal nodes). Then for all complex z

$$\sum_{d \in \tilde{\mathcal{D}}} P(d) \left(1 + z + \cdots z^{|d|-1} \right) = \sum_{y \in \tilde{\mathcal{Y}}} P(y) z^{|y|}, \tag{17}$$

This can be deduced directly by induction and implies (16).

Alternatively we can use a result of [19], where it is shown that for every real-valued function G defined on strings over $\mathcal{A}$

$$\sum_{d \in \tilde{\mathcal{D}}} P(d) G(d) = G(\emptyset) + \sum_{y \in \tilde{\mathcal{Y}}} P(y) \sum_{s \in \mathcal{A}} \frac{P(ys)}{P(y)} (G(ys) - G(y))$$

where $\emptyset$ denotes an empty string. By choosing $G(x) = z^{|x|}$ we directly find

$$\sum_{d \in \tilde{\mathcal{D}}} P(d) z^{|d|} = z^0 + \sum_{y \in \tilde{\mathcal{Y}}} P(y) \sum_{s \in \mathcal{A}} P(s) \left(z \, z^{|y|} - z^{|y|} \right)$$

$$= 1 + (z - 1) \sum_{y \in \tilde{\mathcal{Y}}} P(y)$$

which again proves (17). $\qquad\square$

4.2 Mellin Transforms

The Mellin transform $F^*(s)$ of a function $F(v)$ is defined as (cf. [27])

$$F^*(s) = \int_0^\infty F(v) v^{s-1} dv,$$

if it exists. Using the fact that the Mellin transform of $F(ax)$ is $a^{-s} F^*(s)$, a simple analysis of recurrence (14) reveals that the Mellin transform $A^*(s)$ of $A(v)$ is given by

$$A^*(s) = \frac{-1}{s(1 - p_1^{-s} - p_2^{-s})}, \quad \Re(s) < -1.$$

In order to find asymptotics of $A(v)$ as $v \to \infty$ one can directly use the Tauberian theorem (for the Mellin transform) by Wiener-Ikehara[‡] [17, Theorem 4.1]. For this purpose we have to check that $s_0 = -1$ is the only (polar) singularity on the line $\Re(s) = -1$ and that $(s+1)A^*(s)$ can be analytically extended to a region that contains the line $\Re(s) = -1$. However, if $\log(p_1)/\log(p_2)$ is irrational this follows from a lemma of Schachinger [25] and Jacquet [9] (see Lemma 4 below). In particular, in the irrational case one finds

$$A(v) \sim \frac{v}{H}, \qquad (v \to \infty).$$

This proves the first part of Theorem 3.

In passing we point out that for the rational case, that is, $\log(1/p_1) = n_1 L$ and $\log(1/p_2) = n_2 L$ for coprime integers n_1, n_2 we just have to analyze the recurrence

$$G_n = 1 + G_{n-n_1} + G_{n-n_2},$$

where G_n abbreviates $A(e^{Ln})$. Equivalently we have $A(v) = G(\lfloor \log v \rfloor / L)$. Thus, from

$$G(n) = \frac{1}{(1 - e^{-L})(de^{-dL} + be^{-bL})} e^{Ln} + O(e^{Ln(1-\eta)})$$

for some $\eta > 0$ we directly obtain

$$A(v) = \frac{L e^{-L\langle \log v / L \rangle}}{(1 - e^{-L})} \frac{v}{H} + O(v^{(1-\eta)})$$

where, recall $\langle x \rangle$ is the fractional part of x.

Similarly, we can analyze the expected value of the path length, that is,

$$\mathbf{E}[\mathcal{L}_r] = \sum_{y: P(y) \geq 1/v} P(y).$$

Here the Mellin transform is given by

$$E^*(s) = \int_0^\infty \mathbf{E}[\mathcal{L}_{1/v}] v^{s-1} \, dv = \frac{-1}{s(1 - p_1^{1-s} - p_2^{1-s})} \qquad (\Re(s) < 0)$$

and it leads (in the irrational case) to the asymptotic equivalent

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log(1/r)}{H} + \frac{H_2}{2H^2} + o(1).$$

In the rational case, it is easy to see that

$$\mathbf{E}[\mathcal{L}_r] = \frac{\log(1/r)}{H} + \frac{H_2}{2H^2} + \frac{L}{H}\left(\frac{1}{2} - \left\langle \frac{\log(1/r)}{L} \right\rangle\right) + O(r^\eta)$$

[‡]One major assumption is that there are no singularities on the line $\Re(s) = -1$ despite $s_0 = -1$. In fact this Tauberian theorem is usually used to prove the prime number theorem. The function $-\zeta'(s)/\zeta(s)$ (where $\zeta(s) = \sum_{n \geq 1} n^{-s}$ denotes the Riemann zeta function) is (almost) the Mellin transform of the Chebyshev Ψ-function $\Psi(x) = \sum_{p^k \leq x} \log p$. Since $\zeta(s)$ has no zeroes on the line $\Re(s) = 1$, $s \neq 1$, it follows that $\Psi(x) \sim x$ $(x \to \infty)$ which is equivalent to the prime number theorem $\pi(x) = \sum_{p \leq x} 1 \sim x/\log x$.

for some $\eta > 0$. This improves the error term of the first proof of Theorem 3.

The analysis of $D(v, z)$ is more involved. Here we assume that z is a real number close to 1, say $|z - 1| \le \delta$. The Mellin transform with respect to v becomes

$$D^*(s, z) = \frac{1 - z}{s(1 - zp_1^{1-s} - zp_2^{1-s})} - \frac{1}{s}, \tag{18}$$

for $\Re(s) < s_0(z)$ where $s_0(z)$ denotes the real solution of $zp_1^{1-s} + zp_2^{1-s} = 1$. Note that $s = 0$ is no singularity of $D^*(s, z)$ if $z \ne 1$. As above it, follows from Wiener-Ikehara's Tauberian theorem that (for fixed $z \ne 1$)

$$D(v, z) = \frac{1 - z}{zs_0(z)H(s_0(z) - 1)} v^{-s_0(z)}(1 + o(1)), \qquad (v \to \infty), \tag{19}$$

where $H(s)$ abbreviates

$$H(s) = p_1^{-s} \log(1/p_1) + p_2^{-s} \log(1/p_2).$$

Assume for a moment that the error term in (19) is uniform in z (see next section for a detailed proof), then we can use the local expansion

$$s_0(z) = -\frac{z - 1}{H} + \left(\frac{1}{H} - \frac{H_2}{2H^3} \right)(z - 1)^2 + O(|z - 1|^3) \tag{20}$$

to obtain uniformly for $|z - 1| \le \delta$ as $v \to \infty$, and then

$$
\begin{aligned}
D(v, z) &= v^{-s_0(z)}(1 + O(s_0(z))) \\
&= v^{\frac{z-1}{H} + \left(\frac{1}{H} - \frac{H_2}{2H^3} \right)(z-1)^2 + O(|z-1|^3)} (1 + O(|z - 1|) + o(1)).
\end{aligned}
$$

Recall that $D(v, z) = \mathbf{E}[z^{\mathcal{L}_r}]$ (with $v = 1/r$) is the probability generating function of the dictionary length $\mathcal{L}_r$ and, therefore, it can be used to derive the limiting behavior. We can use the local expansion (20) with $z = e^{t/(\log v)^{1/2}}$ to obtain

$$
\begin{aligned}
v^{-s_0(z)} &= \exp \left(\log v \left(\frac{z - 1}{H} - \left(\frac{1}{H} - \frac{H_2}{2H^3} \right)(z - 1)^2 + O(|z - 1|^3) \right) \right) \\
&= \exp \left(\frac{1}{H} t \sqrt{\log v} + \frac{1}{H} \frac{t^2}{2} - \left(\frac{1}{H} - \frac{H_2}{2H^3} \right) t^2 + O(t^3/\sqrt{\log v}) \right) \\
&= \exp \left(\frac{1}{H} t \sqrt{\log v} + \left(\frac{H_2}{H^3} - \frac{1}{H} \right) \frac{t^2}{2} + O(t^3/\sqrt{\log v}) \right)
\end{aligned}
$$

Hence, we arrive at

$$\mathbf{E}\left[e^{t(\mathcal{L}_r - \frac{1}{H} \log v)/\sqrt{\log v}} \right] = e^{-(t/H)\sqrt{\log v}} \mathbf{E}\left[e^{\mathcal{L}_r t/\sqrt{\log v}} \right] = e^{\frac{t^2}{2}\left(\frac{H_2}{H^3} - \frac{1}{H} \right)} + o(1). \tag{21}$$

By Goncharov's theorem [27] this proves the normal limiting distribution as $v \to \infty$

Note that the above derivations also imply convergence of all (centralized) moments, rate of convergence in CLT, as well as exponential tail estimates. We choose not to present it here leaving details to the interested reader.

The main remaining problem is to show that the limit relation (19) holds uniformly for $|z - 1| \le \delta$. We present a proof in the following section.

4.3 Uniform Tauberian Theorems

In order to find the asymptotics of the Mellin transform $D(v, z)$ as $v \to \infty$ one uses the inverse transform of $D^*(s, z)$, that is (cf. [27])

$$D(v, z) = \frac{1}{2\pi i} \lim_{T \to \infty} \int_{\sigma - iT}^{\sigma + iT} D^*(s, z) v^{-s} \, ds, \qquad (22)$$

which is valid for $\sigma < s_0(z)$.

One problem with the integral (22) is that it is not absolutely convergent since the integrand is only of order $1/s$. To circumvent this problem, we resort to analyze another integral, namely

$$\begin{aligned} D_1(v, z) &= \int_0^v D(w, z) \, dw \\ &= \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \left(\frac{1 - z}{s(1 - zp_1^{1-s} - zp_2^{1-s})} - \frac{1}{s} \right) \frac{v^{-s+1}}{1 - s} \, ds \end{aligned}$$

Here the integrand is of order $O(1/s^2)$ assuring absolute convergence.

The usual procedure (cf. [27]) to prove asymptotics in this context is to shift the line of integration to the right and to collect residues of the polar singularities of $D^*(s, z)$ which are given by the set

$$\mathcal{Z}(z) = \{s \in \mathbb{C} : zp_1^{1-s} + zp_2^{1-s} = 1\} \qquad (23)$$

of all complex roots of $zp_1^{1-s} + zp_2^{1-s} = 1$.

The structure of $\mathcal{Z}(z)$ has been determined by Schachinger [25] and Jacquet [9] and is stated here (in a slightly extended form).

Lemma 4. *Suppose that $0 < p_1 < p_2 < 1$ with $p_1 + p_2 = 1$ and that z is a real number with $|z - 1| \leq \delta$ for some $0 < \delta < 1$. Let*

$$\mathcal{Z}(z) = \{s \in \mathbb{C} : p_1^{1-s} + p_2^{1-s} = 1/z\}.$$

Then
(i) All $s \in \mathcal{Z}(z)$ satisfy

$$s_0(z) \leq \Re(s) \leq \sigma_0(z),$$

where $s_0(z) < 1$ is the (unique) real solution of $p_1^{1-s} + p_2^{1-s} = 1/z$ and $\sigma_0(z) > 1$ is the (unique) real solution of $1/z + p_2^{1-s} = p_1^{1-s}$. Furthermore, for every integer k there uniquely exists $s_k(z) \in \mathcal{Z}(z)$ with

$$(2k - 1)\pi / \log(1/p_1) < \Im(s_k(z)) < (2k + 1)\pi / \log(1/p_1)$$

and consequently $\mathcal{Z}(z) = \{s_k(z) : k \in \mathbb{Z}\}$.

(ii) If $\log p_2 / \log p_1$ is irrational, then $\Re(s_k(z)) > \Re(s_0(z))$ for all $k \neq 0$.
(iii) If $\log p_2 / \log p_1 = r/d$ is rational, where $\gcd(r, d) = 1$ for integers $r, d > 0$, then we have $\Re(s_k(z)) = \Re(s_0(z))$ if and only if $k \equiv 0 \bmod d$. In particular $\Re(s_1(z)), \ldots, \Re(s_{d-1}(z)) > \Re(s_0(z))$ and

$$s_k(z) = s_{k \bmod d}(z) + \frac{2(k - k \bmod d)\pi i}{\log p_1},$$

that is, all $s \in \mathcal{Z}(z)$ are uniquely determined by $s_0(z)$ and by $s_1(z), s_2(z), \ldots, s_{d-1}(z)$, and their imaginary parts constitute an arithmetic progression.

15

One interesting consequence is that in the irrational case we have

$$\min_{|z-1|\le\delta} \left(\Re(s_k(z)) - \Re(s_0(z)) \right) > 0 \tag{24}$$

for all $k \neq 0$. This is due to the fact that $s_k(z)$ varies continuously in z.

By shifting the integral to the line $\Re(s) = \sigma' > \max\{1, \sigma_0(z)\}$ and collecting residues we obtain

$$D_1(v,z) = \frac{1-z}{zs_0(z)(1-s_0(z))H(s_0(z)-1)} v^{-s_0(z)+1}$$

$$+ \sum_{k \neq 0} \frac{1-z}{zs_k(z)(1-s_k(z))H(s_k(z)-1)} v^{-s_k(z)+1} + \frac{1-z}{1-2z} - 1$$

$$+ \frac{1}{2\pi i} \int_{\sigma'-i\infty}^{\sigma'+i\infty} \left(\frac{1-z}{s(1-zp_1^{1-s}-zp_2^{1-s})} - \frac{1}{s} \right) \frac{v^{-s+1}}{1-s} \, ds.$$

Due to the factors $s_k(z)(1-s_k(z))$ in the denominator the series is convergent. Thus, for every $\varepsilon > 0$ there exists K_0 such that

$$\sum_{|k| \ge K_0} \left| \frac{1-z}{zs_k(z)(1-s_k(z))H(s_k(z)-1)} v^{-s_k(z)+1} \right| \le \frac{\varepsilon}{2} v^{-s_0(z)+1}.$$

Furthermore, by (24) there exists v_0 such that

$$\sum_{0<|k|<K_0} \left| \frac{1-z}{zs_k(z)(1-s_k(z))H(s_k(z)-1)} v^{-s_k(z)+1} \right| \le \frac{\varepsilon}{2} v^{-s_0(z)+1}$$

for all $v \ge v_0$ and uniformly for all $|z| \le \delta$. Finally by shifting $\sigma' \to \infty$ it actually follow that

$$\frac{1}{2\pi i} \int_{\sigma'-i\infty}^{\sigma'+i\infty} \left(\frac{1-z}{s(1-zp_1^{1-s}-zp_2^{1-s})} - \frac{1}{s} \right) \frac{v^{-s+1}}{1-s} \, ds = 0$$

since the integral can be made arbitrarily small. Thus, as $v \to \infty$ and uniformly for $|z-1| \le \delta$

$$D_1(v,z) = \frac{1-z}{zs_0(z)(1-s_0(z))H(s_0(z)-1)} v^{-s_0(z)+1} \left(1 + |z-1|o(1) \right) + \frac{1-z}{1-2z} - 1.$$

Since $D(v,y)$ is monotone in v we can apply the elementary Tauberian Lemma 5 (proved below) to obtain

$$D(v,z) = \frac{1-z}{zs_0(z)H(s_0(z)-1)} v^{-s_0(z)} \left(1 + |z-1|^{1/2}o\left(1\right) \right),$$

where the convergence is again uniform for $|z-1| \le \delta$. Hence, we are actually in the situation of (19) and the central limit theorem follows.

To complete the proof, we need a Tauberian result presented below.

Lemma 5. *Suppose that $f(v,\lambda)$ is a non-negative increasing function in $v \ge 0$, where λ is a real parameter with $|\lambda| \le \delta$ for some $0 < \delta < 1$. Assume that $F(v,\lambda) = \int_0^v f(w,\lambda)\,dw$ has the asymptotic expansion*

$$F(v,\lambda) = g(\lambda) \frac{v^{\lambda+1}}{\lambda+1}(1 + \lambda \cdot o(1)) + C(\lambda)$$

16

as $v \to \infty$ and uniformly for $|\lambda| \leq \delta$, where $g(\lambda)$ and $C(\lambda)$ are continuous function depending just on λ with $g(\lambda) \neq 0$. Then

$$f(v, \lambda) = g(\lambda)v^\lambda(1 + |\lambda|^{\frac{1}{2}} \cdot o(1))$$

as $v \to \infty$ and again uniformly for $|\lambda| \leq \delta$.

Proof. Without loss of generality we may assume that $g(\lambda) = 1$. By the assumption

$$\left| F(v, \lambda) - C(\lambda) - \frac{v^{\lambda+1}}{\lambda+1} \right| \leq \varepsilon|\lambda|\frac{v^{\lambda+1}}{\lambda+1}$$

for $v \geq v_0$ and all $|\lambda| \leq \delta$. Set $v' = (\varepsilon|\lambda|)^{1/2}v$. By monotonicity we obtain (for $v \geq v_0$)

$$
\begin{aligned}
f(v, \lambda) &\leq \frac{F(v + v', \lambda) - F(v, \lambda)}{v'} \\
&\leq \frac{1}{v'}\left(\frac{(v+v')^{\lambda+1}}{\lambda+1} - \frac{v^{\lambda+1}}{\lambda+1} \right) + \frac{2}{v'}\varepsilon|\lambda|\frac{(v+v')^{\lambda+1}}{\lambda+1} \\
&= \frac{1}{v'(\lambda+1)}\left(v^{\lambda+1} + (\lambda+1)v^\lambda v' + O(v^{\lambda-1}(v')^2) - v^{\lambda+1} \right) + O\left(\frac{\varepsilon|\lambda|v^{\lambda+1}}{v'} \right) \\
&= v^\lambda + O\left(v^\lambda \varepsilon^{\frac{1}{2}}|\lambda|^{\frac{1}{2}} \right) + O\left(\frac{\varepsilon|\lambda|v^{\lambda+1}}{v'} \right) \\
&= v^\lambda + O\left(v^\lambda \varepsilon^{\frac{1}{2}}|\lambda|^{\frac{1}{2}} \right).
\end{aligned}
$$

In essentially the same way we obtain a corresponding lower bound (for $v \geq v_0 + v_0^{1/2}$). Hence, the result follows. $\square$

Acknowledgment

References

[1] J. Abrahams, "Code and parse trees for lossless source encoding," *Communications in Information and Systems* 1(2):113-146, April 2001.

[2] C. G. Boncelet, Block Arithmetic Coding for Source Compression, *IEEE Trans. Information Theory*, 39, 5, 1546-1554, 1993.

[3] F. Cicalese, L. Gargano, and U. Vaccaro, A note on Tunstall codes with applications to optimal approximation of uniform distributions, *IEEE Trans. Inform. Theory*, 52, 3772 - 3777, 2006.

[4] V. Choi and M. J. Golin, Lopsided trees. I. Analyses. *Algorithmica* 31 (2001), no. 3, 240–290.

[5] M. Drmota and W. Szpankowski, Precise Minimax Redundancy and Regrets, *IEEE Trans. Information Theory*, 50, 2686-2707, 2004.

[6] M. Drmota and W. Szpankowski, On the exit time of a random walk with the positive drift, 2007 Conference on Analysis of Algorithms, Juan-les-Pins, France, *Proc. Discrete Mathematics and Theoretical Computer Science*, 291-302, 2007.

[7] A. Gut, On the moments and limit distributions of some first passage times, *Annals Probab.*, **2**, 277–308, 1974.

[8] P. Henrici, *Applied and Computational Complex Analysis*, Vols. 1–3, John Wiley & Sons, New York, 1977.

[9] P. Jacquet, *Contribution de l'Analyse d'Algorithmes a l'Evaluation de Protocoles de Communication*, Thèse Université de Paris Sud-Orsay, Paris, 1989.

[10] P. Jacquet and W. Szpankowski, Markov Types and Minimax Redundancy for Markov Sources *IEEE Trans. Information Theory*, 50, 1393-1402, 2004.

[11] S. Janson, Renewal Theory in Analysis of Tries and Strings, preprint 2009.

[12] F. Jelinek, *Probabilistic Information Theory* (McGraw-Hill, New York, 1968).

[13] F. Jelinek and K. S. Schneider, On Variable-Length-to-Block Coding, *IEEE Trans. Inform. Theory*, **18** (6) (1972) 765–774.

[14] G. L. Khodak, Connection Between Redundancy and Average Delay of Fixed-Length Coding, *All-Union Conference on Problems of Theoretical Cybernetics* (Novosibirsk, USSR, 1969) 12 (in Russian)

[15] R. E. Krichevsky, Universal Data Compression and Retrieval. (Kluwer, Norwell, MA, 1993).

[16] R. E. Krichevsky and V. K. Trofimov, The Performance of Universal Encoding, *IEEE Trans. Information Theory*, **27** (1981) 199–207.

[17] J. Korevaar, A century of complex Tauberian theory, Bull. Amer. Math. Soc. (N.S.) 39, no. 4, 475–531, 2002.

[18] J. C. Lawrence, A New Universal Coding Scheme for the Binary Memoryless Source, *IEEE Trans. Inform. Theory*, 23 (4) (1977) 466–472.

[19] J. L. Massey, "The entropy of a rooted tree with probabilities," *Proc. of ISIT 1983*.

[20] M.D. Reavy, and C.G. Boncelet, An Algorithm for Compression of Bilevel Images, *IEEE Trans. Image Processing*, 10, 5, 669-676, 2001.

[21] Y.A. Reznik, A.V. Anisimov, On Direct Construction of Variable-Length-to-Block Codes, *Proc. 27-th Symp. Inform. Theory in the Benelux*, Noordwijk, Netherlands, 77-85, June 2006.

[22] S. A. Savari and Robert G. Gallager; Generalized Tunstall codes for sources with memory, IEEE Trans. Info. Theory, 43, 658 - 668, March 1997.

[23] S. A. Savari, Variable-to-Fixed Length Codes for Predictable Sources, Proc IEEE Data Compression Conference (DCC'98), Snowbird, UT, March 30 - April 1, 1998, 481–490.

[24] S. A. Savari, Renewal Theory and Source Coding, *Proceedings of the IEEE*, 88, no. 11, 1692-1702, November 2000.

[25] W. Schachinger, Limiting distributions for the costs of partial match retrievals in multi-dimensional tries. *Random Structures and Algorithms* 17 (2000), no. 3-4, 428–459.

[26] W. Szpankowski, Asymptotic Average Redundancy of Huffman (and other) Block Codes, *IEEE Trans. Information Theory*, 46, 2434-2443, 2000.

[27] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001.

[28] Tj. J. Tjalkens, Efficient and fast data compression codes for discrete sources with memory, Ph.D. Thesis, Eindhoven Univ. Techn., The Netherlands, 1987.

[29] Tj.J. Tjalkens, The Complexity of Minimum Redundancy Coding, *in Proc. 21-th Symp. Inform. Theory in the Benelux* (May 2000) 226-254.

[30] Tj. J. Tjalkens, A Comparative Study of Fixed-to-Variable Length and Variable-to-Fixed Length Source Codes, *in Proc. 25-th Symp. Inform. Theory in the Benelux* (Rolduc, Kerkrade, The Netherlands June 2-4, 2004) 81–88.

[31] T. J. Tjalkens and F. M. J. Willems, "Variable to fixed-length codes for Markov sources," *I.E.E.E. Trans. Inform. Theory* 33, 246-257, 1987.

[32] T. Tjalkens and F. Willems, A Universal Variable-to-Fixed Length Source Code Based on Lawrence's Algorithm, *IEEE Trans. Information Theory*, 38, 247-253, 1992.

[33] V. K. Trofimov, Universal Variable-Length to Block Codes for Bernoulli Sources, *Methods of Discrete Analysis*, 29 (Novosibirsk, 1976) 87–100 (in Russian)

[34] B. P. Tunstall, Synthesis of Noiseless Compression Codes, Ph.D. dissertation, (Georgia Inst. Tech., Atlanta, GA, 1967)

[35] J. Verhoeff, A new data compression technique, *Annals of Systems Research*, **6** (1977) 139-148.

[36] K. Visweswariah, S. R. Kulkarni, and S. Verdu, "Universal variable-to-fixed length source codes," IEEE Trans. Info. Theory, 47, 1461 - 1472, May 2001.