Minimax Redundancy for Large Alphabets

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

Marcelo J. Weinberger
Hewlett-Packard Laboratories
1501 Page Mill Road
Palo Alto, CA 94304, U.S.A.
Email: marcelo.weinberger@hp.com

***Abstract*—We study the minimax redundancy of universal coding for large alphabets over memoryless sources and present two main results: We first complete studies initiated in Orlitsky and Santhanam [11] deriving precise asymptotics of the minimax redundancy for *all ranges* of the alphabet sizes. Second, we consider the minimax redundancy of a source model in which some symbol probabilities are *fixed*. The latter model leads to an interesting binomial sum asymptotics with super-exponential growth functions. Our findings could be used to approximate numerically the minimax redundancy for various ranges of the sequence length and the alphabet size. These results are obtained by analytic techniques such as tree-like generating functions and the saddle point method.**

I. Introduction

Many applications of universal compression concern sources such as speech and image whose alphabets are large, often comparable to the length of the source sequences. Yet most analyses of universal schemes deal with finite, possibly binary, alphabets with exception of [1], [8], [10], [12], [11], [13]. In this work, we study the worst-case minimax redundancy (regret) for unbounded alphabets and present precise asymptotic results as the size of the alphabet and the length of the source sequence grow to infinity. To recall, the redundancy of universal codes for a class of sources determines by how much the actual code length exceeds the optimal code length. In the minimax scenario one designs the best code for the source with the worst redundancy. Such minimax redundancy comes in two flavors: average minimax or worst case minimax. We investigate here the latter.

A fixed-to-variable code $C_n : \mathcal{A}^n \to \{0,1\}^*$ is an injective mapping from the set $\mathcal{A}^n$ of all sequences of length n over the finite alphabet $\mathcal{A}$ of size $m = |\mathcal{A}|$ to the set $\{0,1\}^*$ of all binary sequences. A source P generates a sequence of length n, denoted as $x_1^n \in \mathcal{A}^n$, and we write $L(C_n, x_1^n)$ for the code length for x_1^n. The source entropy $H_n(P) = -\sum_{x_1^n} P(x_1^n)\log P(x_1^n)$ is the absolute lower bound on the expected code length, where $\log := \log_2$ throughout the paper will denote the binary logarithm. The *pointwise redundancy* is

$$R_n(C_n, P; x_1^n) = L(C_n, x_1^n) + \log P(x_1^n).$$

In practice, one can only hope to have some knowledge about a family of sources $\mathcal{S}$ that generates real data (e.g., memoryless sources $\mathcal{M}_0$). Following Davisson [3] and Shtarkov [14] we define the worst case (maximal) minimax redundancy $R_n^*(\mathcal{S})$ for family $\mathcal{S}$ as follows

$$R_n^*(\mathcal{S}) = \min_{C_n} \sup_{P\in\mathcal{S}} \max_{x_1^n} \left[L(C_n, x_1^n) + \log P(x_1^n)\right]. \qquad (1)$$

Our goal is to derive precise results for the worst case minimax redundancy $R_n^*(\mathcal{M}_0)$ for memoryless sources $\mathcal{M}_0$ when the alphabet size m varies with n. We also study this minimax redundancy when some of the parameters are *fixed*. Such *constrained* families of sources arise in applications in which we do have partial knowledge of the data generating mechanism and, consequently, we want to pay a redundancy corresponding to the smallest possible number of parameters (see, e.g., [17] for another example of a constrained family).

The worst case minimax redundancy, $R_n^*(\mathcal{S})$, for a family of sources $\mathcal{S}$ was studied by Shtarkov [14] who found that, ignoring the integer length constraint (cf. [4]),

$$R_n^*(\mathcal{S}) = \log\left(\sum_{x_1^n} \sup_{P\in\mathcal{S}} P(x_1^n)\right).$$

For $\mathcal{S} = \mathcal{M}_0$, we write $d_{n,m} := \log D_{n,m}(\mathcal{M}_0)$ for $R_n^*(\mathcal{M}_0)$, that is,

$$d_{n,m} = \log D_{n,m}(\mathcal{M}_0) = \log\left(\sum_{x_1^n} \sup_{P\in\mathcal{M}_0} P(x_1^n)\right).$$

In this paper, we first consider the minimax redundancy over the alphabet $\mathcal{A}\cup\mathcal{B}$ with $|\mathcal{A}| = m$ and $|\mathcal{B}| = M$, where the symbol probabilities of $\mathcal{B}$ are fixed while m may be large. We shall denote such a family of constrained (memoryless) sources by $\widetilde{\mathcal{M}}_0$.[1] We also write $d_{n,m,M} = \log D_{n,m,M}$, for such a minimax redundancy, and prove

$$D_{n,m,M} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} D_{k,m} \qquad (2)$$

where $p = 1 - P(\mathcal{B})$ and $D_{k,m} = 2^{d_{k,m}}$. In order to estimate it asymptotically we need a quite precise understanding of the asymptotic behavior of $D_{n,m}$ for large n and m.

The minimax redundancy $d_{n,m}$ for large alphabet size m was studied by Orlitsky and Santhanam [11] who established leading term asymptotics for $m = o(n)$ and $n = o(m)$, as

[1]Note that the families of sources $\mathcal{M}_0$ and $\widetilde{\mathcal{M}}_0$ are defined over different alphabets. In addition, the family $\widetilde{\mathcal{M}}_0$ is constrained in that the probabilities of symbols in $\mathcal{B}$ take fixed values.

well as bounds for $m = \Theta(n)$. In this paper, using techniques of analytic information theory, in Theorem 1 we first complete the study of [11] and provide precise asymptotics for all ranges of m. Then, in Theorem 2, we use this precise asymptotics to deal with the binomial sum (2) and extract asymptotics of $d_{n,m,M}$ for large n and unbounded m.

The study of the minimax redundancy over $\mathcal{A}\cup\mathcal{B}$ expressed in (2) leads to an interesting problem for the so called *binomial sums* defined in general as

$$S_f(n) = \sum_k \binom{n}{k} p^k (1-p)^{n-k} f(k),$$

where $0<p<1$ is a fixed probability and f is a given function. In general, asymptotics of f do not imply an asymptotic expansion for $S_f(n)$. In [5], [9], asymptotics of $S_f(n)$ were derived for the *polynomially* growing function $f(x) = O(x^b)$. In our case, when m grows with n, we encounter sub-exponential, exponential and super-exponential functions f; therefore, we need more precise information about f to extract precise asymptotics of $S_f(n)$. Our second main result, Theorem 2, presents asymptotics of (2). Our findings are obtained by analytic methods of analysis of algorithms [7], [16].

II. Main Results

Let us consider the minimax redundancy $D_{n,m,M} = D_{n,m,M}(\widetilde{\mathcal{M}}_0)$ over the alphabet $\mathcal{A}\cup\mathcal{B}$, where $|\mathcal{A}| = m$ and $|\mathcal{B}| = M$, for a class of constrained (some parameters are fixed) memoryless sources $\widetilde{\mathcal{M}}_0$. Specifically, the probabilities of symbols in $\mathcal{A}$, denoted by $p_1,\dots,p_m$, are allowed to vary (unknown), while the probabilities $q_1,\dots,q_M$ of the symbols in $\mathcal{B}$ are fixed (known). Furthermore, $q = q_1 + \cdots + q_M$ and $p = 1-q$. We assume that $0<q<1$ is fixed (independent of n). To simplify our notation, we also write $\mathbf{p} = (p_1,\dots,p_m)$ and $\mathbf{q} = (q_1,\dots,q_M)$.

Assume that a memoryless source generates a sequence of length n that, for simplicity, we denote as $x := x_1^n \in (\mathcal{A}\cup\mathcal{B})^n$. The minimax redundancy relative to $\widetilde{\mathcal{M}}_0$ takes the form

$$D_{n,m,M} = \sum_{x\in(\mathcal{A}\cup\mathcal{B})^n} \sup_{\mathbf{p}} P(x) = \sum_{x\in(\mathcal{A}\cup\mathcal{B})^n} \hat{P}_n(x), \qquad (3)$$

where $\hat{P}_n(x) = \sup_{\mathbf{p}} P(x)$ is the maximum-likelihood (ML) estimator of $P(x)$ over $\widetilde{\mathcal{M}}_0$. Our goal is to derive asymptotics of $D_{n,m,M}$ for large n and unbounded m.

Let us simplify (3). Consider $x \in (\mathcal{A}\cup\mathcal{B})^n$ and assume that i symbols are from $\mathcal{B}$ and the remaining $n-i$ from $\mathcal{A}$. We denote by $z\in\mathcal{B}^i$ the subsequence of x consisting of i symbols from $\mathcal{B}$. Similarly, $y\in\mathcal{A}^{n-i}$ is a subsequence of x over $\mathcal{A}$. For any such pair (y,z), there are $\binom{n}{i}$ ways of interleaving them, all leading to the same empirical probability. The ML probability $\hat{P}_n(x)$ of x can be proved to be

$$\hat{P}_n(x) = (1-q)^{n-i}\hat{P}_{n-i}(y)P_i(z),$$

that is, it is the product of the probability of the subsequence y under ML parameters times the (given) probability $P_i(z)$ of the subsequence s in alphabet $\mathcal{B}$. In summary, using (3),

$$\begin{aligned} D_{n,m,M} &= \sum_{x\in(\mathcal{A}\cup\mathcal{B})^n} \hat{P}_n(x) \\ &= \sum_{i=0}^n \binom{n}{i} \sum_{y\in\mathcal{A}^{n-i}} \sum_{z\in\mathcal{B}^i} (1-q)^{n-i}\hat{P}_{n-i}(y)P_i(z) \\ &= \sum_{i=0}^n \binom{n}{i} (1-q)^{n-i} q^i \sum_{y\in\mathcal{A}^{n-i}} \hat{P}_{n-i}(y) \end{aligned}$$

since $\sum_{z\in\mathcal{B}^i} P_i(z) = q^i$. But $D_{n-i,m} = \sum_{y\in\mathcal{A}^{n-i}} \hat{P}_{n-i}(y)$, which finally leads to Equation (2). This expression is our starting point to estimate $D_{n,m,M}$. For this we need a robust asymptotic expression for $D_{n,m}$, that is, the minimax redundancy relative to $\mathcal{M}_0$ for large m and a wide range of n, discussed next.

In view of the above, we focus now on finding asymptotics of $D_{n,m}(\mathcal{M}_0)$ for large m and $n\to\infty$. Recall that this minimax redundancy is also given by [15], [4]

$$D_{n,m} = \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1,\dots,k_m} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}, \qquad (4)$$

where k_i is the number of times symbol $i\in\mathcal{A}$ occurs in a string of length n.

It is argued in [15] that such a sum can be analyzed through the so–called *tree generating function*. Let us define

$$B(z) = \sum_{k=0}^\infty \frac{k^k}{k!} z^k = \frac{1}{1-T(z)}, \qquad (5)$$

where $T(z)$ satisfies $T(z) = ze^{T(z)}$ and also $T(z) = \sum_{k=1}^\infty \frac{k^{k-1}}{k!} z^k$ (cf. [16]). Defining a new tree-like generating function, namely

$$D_m(z) = \sum_{n=0}^\infty \frac{n^n}{n!} D_{n,m} z^n,$$

we notice that (4) and the convolution formula for generating functions (cf. [16]) immediately implies

$$D_m(z) = [B(z)]^m.$$

Let $[z^n]f(z)$ denote the coefficient of z^n in $f(z)$. Then, we finally arrive at

$$D_{n,m} = \frac{n!}{n^n}[z^n]\,[B(z)]^m. \qquad (6)$$

We can re-write it in a simpler form by defining $\beta(z) = B(z/e)$ and applying Stirling's formula, leading to

$$D_{n,m} = \sqrt{2\pi n}\left(1+O(n^{-1})\right)[z^n]\,[\beta(z)]^m \qquad (7)$$

since $[z^n]\beta(z) = e^{-n}[z^n]B(z)$.

We first recall the asymptotic expansion of $D_{n,m}(\mathcal{M}_0)$ for *fixed* m. To extract asymptotics for this case, we observe [2] that the singular expansion of $B(z)$ around its singularity $z = e^{-1}$ is

$$B(z) = \frac{1}{\sqrt{2(1-ez)}} + \frac{1}{3} - \frac{\sqrt{2}}{24}\sqrt{(1-ez)} + +O((1-ez)).$$

Then, an application of the Flajolet and Odlyzko *singularity analysis* [7], [16] yields [15] (cf. also [18], [19])

$$d_{n,m}(\mathcal{M}_0) := \log D_{n,m}(\mathcal{M}_0) = \frac{m-1}{2}\log\left(\frac{n}{2}\right) + \log\left(\frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\right) + \frac{\Gamma(\frac{m}{2})m}{3\Gamma(\frac{m}{2}-\frac{1}{2})}\cdot\frac{\sqrt{2}}{\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right) \quad (8)$$

for large n and fixed m, where Γ is the Euler gamma function.

Let us now focus on the asymptotic expansion of $D_{n,m}$ when m grows with n. In this case, the singularity analysis *does not* apply, and rather one must use the *saddle point method* [7], [16] for (7) since for m large the factor $\beta^m(z)$ grows.

We next summarize our first main findings delaying the proof until the next section.

Theorem 1: For memoryless sources $\mathcal{M}_0$ over an m-ary alphabet the minimax redundancy behaves asymptotically as follows:
(i) For $m = o(n)$

$$d_{n,m}(\mathcal{M}_0) = \frac{m-1}{2}\log\frac{n}{m} + \frac{m}{2}\log e + o(m). \quad (9)$$

(ii) For $m = \alpha n$, define z_0 as the smallest root of

$$z_0\frac{\beta'(z_0)}{\beta(z_0)} = \frac{1}{\alpha}. \quad (10)$$

Then,

$$d_{n,m}(\mathcal{M}_0) = n\log\left(\frac{\beta^\alpha(z_0)}{z_0}\right) - \frac{1}{2}\log(\alpha A(z_0)+z_0^{-2}) + O(1/n), \quad (11)$$

where

$$A(z) = \left[\frac{\beta'(z)}{\beta(z)}\right]' = \frac{\beta''(z)\beta(z)-[\beta'(z)]^2}{\beta^2(z)}.$$

(iii) For $n = o(m)$

$$d_{n,m}(\mathcal{M}_0) = n\log\frac{m}{n} - \log e + O(1/n). \quad (12)$$

Remark 1. The leading terms of the asymptotic expansions for $m = o(n)$ and $n = o(m)$ (i.e., (9) and (12)) were derived by Orlitsky and Santhanam in [11]. For the case $m = \alpha n$, the methodology of [11] allowed only to extract the growth rate, i.e., $d_{n,m} = \Theta(n)$ but not the constant in front of n. Numerical computations reveal that this constant, which is specified in (10) and (11), is well approximated by $1.2\alpha^{0.6}$ for $0.1 < \alpha < 10$ (e.g., 0.78 for $\alpha = 0.5$ and 1.19 for $\alpha = 1$).

Remark 2. For the case $m = o(n)$ if we additionally know that $m = o(\sqrt{n})$, then we can improve (9) to

$$d_{n,m}(\mathcal{M}_0) = \frac{m-1}{2}\log\frac{n}{m} + \frac{m}{2}\log e - \frac{1}{2} + O\left(\frac{m}{\sqrt{n}}\right).$$

Now, we are in a position to discuss the second main topic of this paper, namely, asymptotic expansion of the minimax redundancy $D_{n,m,M}$ relative to $\widetilde{\mathcal{M}}_0$, given by Equation (2). As mentioned, sums like (2) are known as the *binomial sum* [5], [9]. If $D_{k,m} = 2^{d_{k,m}}$ has a polynomial growth, (i.e., $D_{n,m} = 2^{d_{n,m}} = O(n^{(m-1)/2})$ when m is fixed), then we can use the asymptotic expansion derived in [5], [9] to conclude that $D_{n,m,M} \sim D_{np,m}$. However, when m varies with n as in our study, the problem is much more interesting. In particular, the polynomial growth of $D_{n,m,M}$ does not hold any more and we need to compute asymptotics anew. We summarize our second main result in the theorem below whose proof is sketched in the next section.

Theorem 2: Consider a family of memoryless sources $\widetilde{\mathcal{M}}_0$ over the $(m+M)$-ary alphabet $\mathcal{A}\cup\mathcal{B}$ with fixed probabilities $q_1,\ldots,q_M$ of the symbols in $\mathcal{B}$, such that $q = q_1 + \ldots q_M$ is bounded away from 0 and 1. Let $p = 1-q$. Then, the minimax redundancy $d_{n,m,M} = \log D_{n,m,M}$ takes the form
(i_0) If m is fixed, then

$$\begin{aligned} d_{n,m,M} &= d_{np,m} + O(1/n) \quad (14)\\ &= \frac{m-1}{2}\log\left(\frac{np}{2}\right) + \log\left(\frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\right) + O(1/n), \end{aligned}$$

where $d_{np,m}$ is given by (8) with n replaced by np.
(i) If $m_n = O(n^\delta)$, where we write m_n to explicitly show the dependence of m on n, then, for $0 < \delta < 1/2$,

$$d_{n,m,M} = \frac{m_{np}-1}{2}\log\left(\frac{np}{m_{np}}\right) + O\left(\frac{\log^2 n}{n^{1-2\delta}}\right), \quad (15)$$

while for $1/2 \le \delta < 1$,

$$d_{n,m,M} = \frac{m_{np}}{2}\log\left(\frac{np}{m_{np}}\right) + O\left(n^{2\delta-1}\log n\right). \quad (16)$$

(ii) If $m_n = \alpha n$, then

$$\begin{aligned} d_{n,m,M} &= n\log\left(\frac{\beta^\alpha(z_0)(1-p)}{z_0} + p\right) \quad (17)\\ &\quad -\frac{1}{2}\log(\alpha A(z_0) + z_0^{-2}) + O(1/n) \end{aligned}$$

where z_0 and $A(z)$ are defined in Theorem 1(ii).
(iii) Let $n = o(m)$ and let $\frac{m_k}{k} \le \frac{m_n}{n}$ for all $k \le n$. Then,

$$d_{n,m,M} = n\log\left(\frac{m_n}{n}\right) + n\log p + O(\max\{n^2/m, 1\}) \quad (18)$$

for large n.

In passing, let us explain intuitively the asymptotics behind Theorem 2. As discussed above, we deal here with the binomial sum (2) that in general can be written as

$$S_f(n) = \sum_k \binom{n}{k} p^k(1-p)^{n-k} f(k)$$

for a general function f. In our case, $f(k) = D_{k,m}$. Observe that when f grows polynomially, the maximum under the sum occurs around $k = np$, and to find asymptotics we need to sum only within the range $\pm\sqrt{n}$ around np. This basically explains case (i). When $m = \alpha n$, the growth of $f(k) = D_{k,m} = O(A^k)$ is exponential, and we need all the terms in order to recover the asymptotics. Finally, for case (iii) the function $f(k) = D_{k,m}$ grows super-exponentially, and the asymptotics of the binomial sum are determined by the last term, that is, $k = n$.

III. Analysis

In this section we prove Theorems 1 and 2 using analytic tools (Theorem 1) and elementary analysis (Theorem 2).

A. Proof of Theorem 1

We first prove Theorem 1. The starting point is Equation (6), which we re-write as follows:

$$\begin{aligned} D_{n,m} &= \frac{n!}{n^n}[z^n][B(z)]^m = e^n \frac{n!}{n^n}[z^n][B(z/e)]^m \\ &= \sqrt{2\pi n}(1+O(1/n))[z^n][\beta(z)]^m, \end{aligned}$$

where we used Stirling's formula an defined $\beta(z) := B(z/e)$. Thus, it suffices to extract the coefficient at z^n of $\beta^m(z)$.

In order to find asymptotics of $[z^n][\beta(z)]^m$ we apply the Cauchy coefficient formula [7], [16], that is,

$$\begin{aligned} [z^n][\beta(z)]^m &= \frac{1}{2\pi i}\oint \frac{\beta^m(z)}{z^{n+1}}dz \\ &= \frac{1}{2\pi i}\oint \exp[m\ln\beta(z)-(n+1)\ln z]dz \\ &= \frac{1}{2\pi i}\oint e^{h(z)}dz, \end{aligned}$$

where

$$h(z) = m\ln\beta(z)-(n+1)\ln z.$$

Since m is large, we apply the saddle point method [7], [16] to evaluate asymptotically the integral. Let z_0 be the unique root of $h'(z_0)=0$, that is,

$$z_0\frac{\beta'(z_0)}{\beta(z_0)} = \frac{n+1}{m}. \tag{19}$$

Observe that Taylor's expansion of $h(z)$ is

$$h(z) = h(z_0) + \frac{1}{2}(z-z_0)^2 h''(z_0) + O((z-z_0)^3)$$

where

$$h''(z_0) = mA(z_0) + n + 1/z_0^2,$$

and

$$A(z) = \left[\frac{\beta'(z)}{\beta(z)}\right]'.$$

To evaluate the integral we split it into two parts, I_1 and I_2, using the substitution $z = z_0 + e^{i\theta}$:

$$\begin{aligned} I(n) &= \frac{1}{2\pi i}\oint e^{h(z)}dz = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{h(z_0+e^{i\theta})}d\theta \\ &= \frac{1}{2\pi}\int_{-\theta_0}^{\theta_0} e^{h(z_0+e^{i\theta})}d\theta + \frac{1}{2\pi}\int_{-\theta\notin[-\theta_0,\theta_0]} e^{h(z_0+e^{i\theta})}d\theta \\ &= I_1(n) + I_2(n). \end{aligned}$$

We choose $\theta_0 = n^{-2/5}$. In order to assess the second integral $I_2(n)$ we observe, as in [6], that for $\theta \notin [-\theta_0, \theta_0]$

$$|\beta(e^{i\theta})| \le |\beta(e^{i\theta_0})|$$

leading to

$$I_2(n) = \frac{1}{2\pi}\int_{-\theta\notin[-\theta_0,\theta_0]} e^{h(z_0+e^{i\theta})}d\theta = O\left(I_1(n)e^{-cn^{1/5}}\right).$$

Thus, we are left with evaluating the integral $I_1(n)$. The standard Laplace's method and Gaussian integral lead to our final formula

$$\begin{aligned} I_1(n) &= \sqrt{2\pi n}\frac{1}{\sqrt{2\pi|h''(z_0)|}}\exp\left[m\ln\beta(z_0)-(n+1)\ln z_0\right] \\ &\quad \times \left(1+O\left(\frac{1}{\min(m,n)}\right)\right). \end{aligned} \tag{20}$$

To complete the proof of Theorem 1 we need to estimate z_0 for various ranges of m and n.

Case: $m = o(n)$.

In this case, it is easy to see that

$$z_0 = 1 - \frac{m}{2n+m}(1-O(\sqrt{m/n}))$$

and

$$\begin{aligned} h(z_0) &= \frac{m}{2}\log\frac{n}{m} + \frac{m}{2}\log e + o(m) \\ \sqrt{\frac{2\pi n}{2\pi h''(z_0)}} &= \frac{1}{\sqrt{2n/m(1+O(m/n))}}. \end{aligned}$$

This proves Theorem 1(i) after substituting in (20).

Case: $n = \alpha m$.

In this case, z_0 is an asymptotic solution of (10), and the polynomial factor of (20) becomes

$$\sqrt{\frac{2\pi n}{2\pi h''(z_0)}} = \frac{1}{\sqrt{\alpha A(z_0)+z_0^{-2}}} + O(1/n).$$

This completes the proof of Theorem 1(ii).

Case: $n = o(m)$.

In this case $z \approx 0$. More precisely,

$$z_0 = \frac{(n+1)e}{m}(1+O(n/m)),$$

and then

$$h(z_0) = n\log(m/n) + \log(m/n) - \log e + O(1/n).$$

Finally, after somewhat long calculations, we arrive at

$$\log\sqrt{n/h''(z_0)} = \log(n/m) + O(n/m).$$

Putting everything together, we prove Theorem 1(iii).

B. Sketch of the Proof of Theorem 2

In order to prove Theorem 2 we need to evaluate the binomial sum

$$S_f(n) = \sum_k \binom{n}{k} p^k(1-p)^{n-k}f(k) \tag{21}$$

for $f(k) = D_{k,m_k}$ that grows faster than any polynomial for $m \to \infty$. However, for completeness, we first present a simple derivation of asymptotics for polynomially (and sub-exponentially) growing f proving Theorem 2(i_0) and Theorem 2(i).

CASE: $m_n = o(n)$.

For this case, we only sketch the proof. A complete proof for sub-exponential growth (i.e., $D_{n,m} = O(e^{A\sqrt{n}})$) is available through depoissonization along the same line of arguments as in [9].

Let us first assume that $f(n) = D_{n,m}$ for fixed m, that is, $f(n) = Cn^{(m-1)/2}(1 + O(1/n))$, where in the sequel C is used to denote arbitrary constants that take appropriate values in each use. Expand $f(x)$ around $x = np$ to find

$$f(x) = f(np) + (x - np)f'(np) + \frac{(x - np)^2}{2} f''(np')$$

for some $0 < p' < p$. Observe now that $S_f(n)$ can be viewed as $S_f(n) = \mathbf{E}[f(X)]$ where X is a binomially distributed random variable. Thus

$$\begin{aligned} S_f(n) &= \mathbf{E}[f(X)] = f(np) + \frac{\mathbf{Var}\ (X)}{2} f''(np') \\ &= f(np) + O(nf''(n)) \end{aligned} \tag{22}$$

provided $nf''(n) = o(f(np))$. The last condition is obviously satisfied for m fixed, and hence Theorem 2(i$_0$) holds. A more detailed analysis can be found in [5], [9].

Let us now consider part (i) of Theorem 2, that is we assume that $m = O(n^\delta)$ for some $0 < \delta < 1$. Notice that $f(k) = D_{k,m_k} = Ck^{\frac{1-\delta}{2}(k^\delta - 1)}(1 + O(1/n))$. Then

$$f''(np) = O\left(f(np)\frac{\log^2 n}{n^{2(1-\delta)}}\right),$$

and Theorem 2(i) follows for $0 < \delta < 1/2$. Assuming now $1/2 < \delta < 1$, we need a different approach since the error term $O((\log^2 n)/n^{1-2\delta})$ dominates. Observe that we always have

$$\frac{1}{\sqrt{2\pi n}} f(np) \leq S_f(n) \leq n \max_k \left(\binom{n}{k} p^k (1-p)^{n-k} f(k)\right).$$

It is easy, however cumbersome, to compute the maximum of the right-hand side. Applying Stirling's formula, we find out that it is achieved at

$$k^* = np + O(n^\delta \log n)$$

and then the above inequalities become

$$\frac{1}{\sqrt{2\pi n}} f(np) \leq S_f(n) \leq C\sqrt{n} f(np) O(n^{2\delta-1} \log n)$$

which suffices to prove (16) of Theorem 2(i).

CASE: $m = \alpha n$.

This case is easy since

$$f(n) = D_{n,m_n} = \left(\frac{\beta^\alpha(z_0)}{z_0}\right)^n \sqrt{\alpha A(z_0) + z_0^{-1}}(1+O(1/n)).$$

This directly implies Theorem 2(ii).

CASE: $n = o(m)$.

Actually, in this case the proof is quite simple. Observe that

$$f(n) = D_{n,m} = C\left(\frac{m_n}{n}\right)^n (1 + O(1/n)).$$

Since we require $\frac{m_k}{k} \leq \frac{m_n}{n}$ for all $k \leq n$ we find

$$\begin{aligned} p^n D_{n,m_n} &\leq \sum_{k=1}^n \binom{n}{k} p^k q^{n-k} D_{k,m_k} \\ &\leq C \sum_{k=1}^n \binom{n}{k} \left(p\frac{m_n}{n}\right)^k q^{n-k} = C\left(p\frac{m_n}{n} + q\right)^n \end{aligned}$$

which completes the proof of Theorem 2(iii).

ACKNOWLEDGMENT

This work was partially done when W. Szpankowski was visiting Hewlett-Packard Laboratories, Palo Alto. This author was also supported by the NSF STC Grant CCF-0939370, NSF Grants DMS-0800568 and CCF-0830140, AFOSR Grant FA8655-08-1-3018, and Grant MNSW 1967/B/T02/2009/37 through Poznan University of Technology.

REFERENCES

[1] S. Boucheron, A. Garivier and E. Gassiat, "Coding on Countably Infinite Alphabets," *IEEE Trans. Information Theory*, 55, 358–373, 2009.

[2] R. Corless, G. Gonnet, D. Hare, D. Jeffrey and D. Knuth, "On the Lambert W Function," *Adv. Computational Mathematics*, 5, 329–359, 1996.

[3] L. D. Davisson, "Universal Noiseless Coding," *IEEE Trans. Information Theory*, 19, 783–795, 1973.

[4] M. Drmota and W. Szpankowski, "Precise Minimax Redundancy and Regret," *IEEE Trans. Information Theory*, 50, 2686–2707, 2004.

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

[6] P. Flajolet and W. Szpankowski, "Analytic Variations on Redundancy Rates of Renewal Processes," *IEEE Trans. Information Theory*, 48, 2911–2921, 2002.

[7] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2008.

[8] L. Györfi, I. Pali and E. der Meulen, "On Universal Noiseless Source Coding for Infinite Source Alphabets," *Europ. Trans. Telecommin. Related Technol.*, 4, 125–132, 1993.

[9] P. Jacquet and W. Szpankowski, "Entropy Computations via Analytic Depoissonization," *IEEE Trans. Information Theory*, 45, 1072–1081, 1999.

[10] J. Kieffer, "A Unified Approach to Weak Universal Source Coding," *IEEE Trans. Information Theory*, 24, 674–682, 1978.

[11] A. Orlitsky and N. Santhanam, "Speaking of Infinity," *IEEE Trans. Information Theory*, 50, 2215–2230, 2004.

[12] A. Orlitsky, N. Santhanam, and J. Zhang, "Universal Compression of Memoryless Sources over Unknown Alphabets," *IEEE Trans. Information Theory*, 50, 1469–1481, 2004.

[13] G. Shamir, "Universal Lossless Compression With Unknown Alphabets: The Average Case," *IEEE Trans. Information Theory*, 52, 4915–4944, 2006.

[14] Y. Shtarkov, "Universal Sequential Coding of Single Messages," *Problems of Information Transmission*, 23, 175–186, 1987.

[15] W. Szpankowski, "On Asymptotics of Certain Recurrences Arising in Universal Coding," *Problems of Information Transmission*, 34, 55–61, 1998.

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

[17] M.J. Weinberger and G. Seroussi, "Sequential Prediction and Ranking in Universal Context Modeling and Data Compression," *IEEE Trans. Inform. Theory*, 43, pp. 1697–1706, Sept. 1997.

[18] Q. Xie, A. Barron, "Minimax Redundancy for the Class of Memoryless Sources," *IEEE Trans. Information Theory*, 43, 647–657, 1997.

[19] Q. Xie, A. Barron, "Asymptotic Minimax Regret for Data Compression, Gambling, and Prediction," *IEEE Trans. Information Theory*, 46, 431–445, 2000.