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 [13] deriving precise asymptotics of the minimax redundancy for all ranges of the alphabet size. Second, we consider the minimax redundancy for a family of sources in which some symbol probabilities are fixed. The latter problem leads to an interesting binomial sum asymptotics with super-polynomial growth functions. Our findings can 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 images whose alphabets are large, often comparable to the length of the source sequences. Yet, the asymptotic analysis of universal schemes is typically concerned with a regime in which the alphabet size remains fixed as the sequence length grows. In contrast, [1], [8], [11], [14], [16] study universal compression for unknown (possibly infinite) alphabets. In this work, following the scenario in [13], we consider unbounded (but finite) alphabets. Specifically, we study the minimax worst-case redundancy (regret) in an asymptotic regime in which both the size of the alphabet and the length of the sequence grow to infinity, deriving precise asymptotic results for two families of memoryless sources. To recall, the redundancy of a code for a family of sources determines by how much the code length exceeds that of the optimal code designed for a specific source in the family. In the minimax scenario, one designs the best code for the source with the worst redundancy. Such redundancy comes in two flavors: average or worst-case. We investigate the latter, as defined next.

A fixed-to-variable code \( C_n : A^n \rightarrow \{0,1\}^* \) is an injective mapping from the set \( A^n \) of all sequences of length \( n \) over the finite alphabet \( A \) of size \( m = |A| \) to the set \( \{0,1\}^* \) of all binary sequences. A source \( P \) generates a sequence \( x_1, \ldots, x_n \) of length \( n \), denoted \( x^n \), and we write \( L(C_n, x^n) \) for the code length for \( x^n \). We assume that \( C_n \) satisfies the prefix condition. Therefore, the source entropy \( H_n(P) = -\sum x^n P(x^n) \log P(x^n) \) is the absolute lower bound on the expected code length, where \( \log := \log_2 \) will denote the binary logarithm throughout the paper. Given a sequence \( x^n \), the pointwise redundancy of \( C_n \) with respect to a specific source \( P \) is given by

\[
R_n(C_n, P; x^n) = L(C_n, x^n) + \log P(x^n).
\]

In practice, one can only hope to have some knowledge about a family of sources \( S \) that generates the data, such as the family of memoryless sources \( M_0 \). Following Davisson [3] and Shtarkov [17], we define the minimax worst-case (maximal) redundancy \( R_n^*(S) \) for a family \( S \) as

\[
R_n^*(S) = \min_{C_n} \sup_{P \in S} \left[ L(C_n, x^n) + \log P(x^n) \right].
\]

This quantity was studied by Shtarkov [17], who found that, ignoring the integer length constraint (cf. [4]),

\[
R_n^*(S) = \log \left( \sum_{x^n} \sup_{P \in S} P(x^n) \right),
\]

achievable with the normalized maximum-likelihood code, which assigns to each sequence a code length proportional to its maximum-likelihood probability over \( S \). In particular, for \( S = M_0 \), precise asymptotics of \( R_n^*(M_0) \) have been derived when the alphabet size \( m \) is fixed [18] (cf. also [21], [22]). This quantity, which we will denote \( d_{n,m} := \log D_{n,m} \), that is,

\[
d_{n,m} = \log D_{n,m} = \log \left( \sum_{x^n} \sup_{P \in M_0} P(x^n) \right),
\]

was also studied when both \( n \) and \( m \) are large, by Orlitsky and Santhanam [13]. Formulating this scenario as a sequence of problems in which \( m \) varies with \( n \), leading term asymptotics for \( m = o(n) \) and \( n = o(m) \), as well as bounds for \( m = \Theta(n) \), are established in [13].†

† We write \( f(n) = O(g(n)) \) if and only if \( |f(n)| \leq C|g(n)| \) for some positive constant \( C \) and sufficiently large \( n \). Also, \( f(n) = \Theta(g(n)) \) if and only if \( f(n) = O(g(n)) \) and \( g(n) = O(f(n)) \), \( f(n) = o(g(n)) \) if and only if \( \lim_{n \to \infty} f(n)/g(n) = 0 \), and \( f(n) = \Omega(g(n)) \) if and only if \( g(n) = O(f(n)) \).

* The material in this paper was presented in part at the 2010 IEEE International Symposium on Information Theory, Austin, Texas, USA, June 2010.
† This author’s contribution was partially made when visiting Hewlett-Packard Laboratories, Palo Alto, CA 94304, U.S.A. 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 DS 45-083/10 through Poznan University of Technology.
In this paper we first provide, in Theorem 1, precise asymptotics of $d_{n,m}$ for all ranges of $m$. Our findings are obtained by analytic methods of analysis of algorithms [7], [19]. Theorem 1 not only completes the study of [13] by covering all ranges of $m$ (including $m = \Theta(n)$), but also strengthens it by providing more precise asymptotics. Indeed, it will be shown that the error incurred by neglecting lower order terms may actually be quite significant, to the point that, for fixed $m = o(n)$, the first two terms of the asymptotic expansion for fixed $m$ given in [18] is a better approximation to $d_{n,m}$ than the leading term established in [13].

In addition, Theorem 1 also enables a precise analysis of the minimax worst-case redundancy in a more general scenario. Specifically, we consider the alphabet $\mathcal{A} \cup \mathcal{B}$, with $|\mathcal{A}| = m$ and $|\mathcal{B}| = M$, where the probabilities of symbols in $\mathcal{B}$ are fixed, while $m$ may be large. We shall denote this family of constrained (memoryless) sources by $\Tilde{\mathcal{M}}_0$. Such constrained families of sources arise in applications in which we do have partial knowledge of the data generating mechanism (see, e.g., [20] for another example of a constrained family). Since such knowledge reduces the number of free parameters in the probability model, one would expect to “pay” a smaller price for universality in terms of redundancy. In an average sense and for bounded $m$, Rissanen’s lower bound on the redundancy [15] is indeed proportional to the number $m - 1$ of free parameters. Moreover, it is easy to see that the leading term asymptotics of the redundancy for a (sequential) code that uses a fixed probability assignment for symbols in $\mathcal{A} \cup \mathcal{B}$ is the maximum-likelihood (ML) estimator of $\hat{f}^*_n$.

In the remainder of this paper, Section II presents our main results, which are proved in Section III.

II. Main Results

Assume that a source generates sequences of length $n$ over the alphabet $\mathcal{A} \cup \mathcal{B}$, with $|\mathcal{A}| = m$ and $|\mathcal{B}| = M$. Consider the minimax worst-case redundancy $R^*_n(\mathcal{M}_0) = d_{n,m,M}$ for a family $\mathcal{M}_0$ of constrained (i.e., some parameters are fixed) memoryless sources. Specifically, the probabilities of symbols in $\mathcal{A}$, denoted by $p_1, \ldots, p_m$, are allowed to vary (unknown), while the probabilities $q_1, \ldots, q_M$ 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 $p = (p_1, \ldots, p_m)$ and $q = (q_1, \ldots, q_M)$. The output sequence is denoted $x := x_1^n \in (\mathcal{A} \cup \mathcal{B})^n$.

Our goal is to derive asymptotics of $d_{n,m,M}$ for large $n$ and unbounded $m$. We start by stating, in Lemma 1 below, an expression relating $d_{n,m,M}$ to the minimax worst-case redundancy $d_{n,m}$ over $\mathcal{A}$. The lemma is stated in terms of $D_{n,m,M} = 2^{d_{n,m,M}}$ and $D_{n,m} = 2^{d_{n,m}}$.

Lemma 1:

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

where $p = 1 - P(\mathcal{B})$. As it turns out, in order to estimate this quantity asymptotically, we need a quite precise understanding of the asymptotic behavior of $D_{k,m}$ for large $k$ and $m$, as provided by Theorem 1.

The study of the minimax worst-case redundancy over $\mathcal{A} \cup \mathcal{B}$ expressed in (4) 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 $[5, [10]$, asymptotics of $S_f(n)$ were derived for the polynomially growing function $f(x) = O(x^k)$. This result applies to our case when $m$ is fixed, and leads to the conclusion that the asymptotics of $d_{n,m,M}$ are the same as those of $d_{n,p,m}$, an intuitively appealing result since the length of the sub-sequence over $\mathcal{A}$ is $np$ with high probability. But 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 $S_f(n)$. In our second main result, Theorem 2, we use the precise asymptotics derived in Theorem 1 to deal with the binomial sum (4) and extract asymptotics of $d_{n,m,M}$ for large $n$ and unbounded $m$.

In the remainder of this paper, Section II presents our main results, which are proved in Section III.
over $B$ with (given) probabilities $q_1/q, \ldots, q_M/q$. In summary, using (6), we obtain

$$D_{n,m,M} = \sum_{i=0}^{n} \binom{n}{i} p^{n-i} q^i \sum_{y \in A^{n-i}} \sum_{z \in B} \tilde{P}_{n-i}(y) P_i(z)$$

$$= \sum_{i=0}^{n} \binom{n}{i} p^{n-i} q^i \sum_{y \in A^{n-i}} \tilde{P}_{n-i}(y). \quad (7)$$

The proof is complete by noticing that the inner summation in (7) is precisely $D_{n-i,m}$.

As a starting point for estimating $D_{n,m,M}$, Lemma 1 requires a robust asymptotic expression for $D_{n,m}$, that is, the minimax worst-case redundancy relative to $\mathcal{M}_0$ for large $n$ and a wide range of $m$. We focus on $D_{n,m}$ next.

Observe that, by (3), $D_{n,m}$ takes the form

$$D_{n,m} = \sum_{k_1, \ldots, k_m = n} \left( \sum_{i=0}^{n} \binom{n}{i} \left( \frac{k_i}{n} \right)^{k_i} \cdots \left( \frac{k_m}{n} \right)^{k_m} \right) \sqrt{n}, \quad (8)$$

where $k_i$ is the number of times symbol $i \in A$ occurs in a string of length $n$. It is argued in [18] that the asymptotics of such a sum can be analyzed through its so-called tree-like generating function, defined as

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

Here, we will follow the same methodology, which by using (8) to define an appropriate recurrence on $D_{n,m}$ (involving both indexes, $n$ and $m$), and employing the convolution formula for generating functions (cf. [19]), relates $D_m(z)$ to the tree-like generating function of the constant 1, namely

$$B(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k.$$

This function, in turn, can be shown to satisfy (cf. [19])

$$B(z) = \frac{1}{1 - T(z)} \quad (9)$$

for $|z| < e^{-1}$, where $T(z)$ is the well-known tree function, which is a solution to the implicit equation

$$T(z) = z e^{T(z)} \quad (10)$$

with $|T(z)| < 1$. Specifically, the following relation is proved in [18].

**Lemma 2**: The tree-like generating function of $D_{n,m}$ satisfies, for $|z| < e^{-1}$,

$$D_m(z) = [B(z)]^m - 1$$

and, consequently,

$$D_{n,m} = \frac{n!}{n^n} [z^n] [B(z)]^m \quad (11)$$

where $[z^n]f(z)$ denotes the coefficient of $z^n$ in $f(z)$.

Defining $\beta(z) = B(z/e)$, $|z| < 1$, noticing that $[z^n]\beta(z) = e^{-n[z^n]}B(z)$, and applying Stirling’s formula, (11) yields

$$D_{n,m} = \sqrt{2\pi n} \left(1 + O(n^{-1})\right) [z^n] [\beta(z)]^m. \quad (12)$$

Thus, it suffices to extract asymptotics of the coefficient at $z^n$ of $[\beta(z)]^m$, for which a standard tool is Cauchy’s coefficient formula [7], [19], that is,

$$[z^n] [\beta(z)]^m = \frac{1}{2 \pi i} \oint \frac{\beta^m(z)}{z^{n+1}} \, dz \quad (13)$$

where the integration is around a closed path containing $z = 0$ inside which $\beta^m(z)$ is analytic.

Now, the fixed $m$ case is solved in [18] by use of the Flajolet and Odlyzko singularity analysis [7], [19], which applies because $[\beta(z)]^m$ has algebraic singularities. Indeed, using (9) and (10) it can be shown that the singular expansion of $\beta(z)$ around its singularity $z = 1$ is [2]

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

The singularity analysis then yields [18]

$$d_{n,m} := \log D_{n,m} = \frac{m-1}{2} \log \left(\frac{n}{2}\right) + \log \left(\frac{\sqrt{\pi}}{\Gamma\left(\frac{m}{2}\right)}\right)$$

$$+ \frac{\Gamma\left(\frac{m}{2}\right) \log e}{3 \Gamma\left(\frac{m}{2} - \frac{1}{2}\right)} \cdot \frac{\sqrt{2}}{\sqrt{n}} + O\left(\frac{1}{n}\right) \quad (14)$$

for large $n$ and fixed $m$, where $\Gamma$ is the Euler gamma function.

When $m$ grows with $n$, which is the case of interest in this paper, the singularity analysis does not apply. Instead, the growth of the factor $\beta^m(z)$ determines that the saddle point method [7], [19], which we briefly review next, can be applied to (13). We will restrict our attention to a special case of the method, where the goal is to obtain an asymptotic approximation of the coefficient $a_n := [z^n]g(z)$ for some analytic function $g(z)$, namely

$$a_n = \frac{1}{2 \pi i} \oint e^{h(z)} dz$$

where $h(z) := \log g(z) - (n + 1) \log z$, under the assumption that $h'(z)$ has a real root $z_0$.

The saddle point method is based on Taylor’s expansion of $h(z)$ around $z_0$ which, recalling that $h'(z_0) = 0$, yields

$$h(z) = h(z_0) + \frac{1}{2}(z-z_0)^2 h''(z_0) + O(h'''(z_0)(z-z_0)^3). \quad (15)$$

After choosing a path of integration that goes through $z_0$, and under certain assumptions on the function $h(z)$, it can be shown (cf., e.g., [19]) that the first term of (15) gives $e^{h(z_0)}$, the second term – after integrating a Gaussian integral – leads to $1/\sqrt{2\pi h''(z_0)}$, and finally the third term determines the error term in the expansion of $a_n$. The standard saddle point

---

3 As mentioned, Equation (2) ignores the integer length constraint of a code, and therefore $O(1)$ terms in (14) are arguably irrelevant. This issue is addressed in [4]; here, we focus on the probability assignment problem, which unlike coding does not entail an integer length constraint.
method described in [19, Table 8.4] then yields the following lemma.

Lemma 3: Assume the conditions required in [19, Table 8.4] hold and let \( z_0 \) denote a real root of \( h'(z) \). Then,

\[
d_n = \frac{e^{h(z_0)}}{\sqrt{2\pi |h''(z_0)|}} \left( 1 + O \left( \frac{h'''(z_0)}{(h''(z_0))^{\rho}} \right) \right)
\]

for any constant \( \rho < 3/2 \), provided the error term is \( o(1) \).

In order to control the error term, the conditions stated in [19, Table 8.4] include the requirement that, as \( n \) grows, \( h''(z_0) \rightarrow \infty \). It turns out, however, that more is known for our particular \( h(z) \): indeed, it will be further shown that the growth of \( h''(z_0) \) is at least linear. This additional property will allow us to extend Lemma 3 to the case \( \rho = 3/2 \). The modified lemma will be the main tool in the derivation of our first main findings, which we present and discuss next, deferring their proof to Section III.

Theorem 1: For memoryless sources \( \mathcal{M}_0 \) over an \( m \)-ary alphabet, where \( m \rightarrow \infty \) as \( n \) grows, the minimax worst-case redundancy \( d_{n,m} \) behaves asymptotically as follows:

(i) For \( m = o(n) \)

\[
d_{n,m} = \frac{m-1}{2} \log \frac{n}{m} + \frac{m}{2} \log e + \frac{m \log e}{3} \sqrt{\frac{m}{n}} \]

\[
\left( \frac{1}{m} - \frac{1}{4} \right) + O \left( \frac{m^2}{n} + \frac{1}{\sqrt{m}} \right).
\]

(ii) For \( m = \alpha n + \ell(n) \), where \( \alpha \) is a positive constant and \( \ell(n) = o(n) \),

\[
d_{n,m} = n \log B_\alpha + \ell(n) \log C_\alpha - \log \sqrt{A_\alpha}
\]

\[
- \frac{\ell(n)^2 \log e}{2n \alpha^2 A_\alpha} + O \left( \frac{\ell(n)^3}{n^2} + \frac{\ell(n)}{n} + \frac{1}{\sqrt{n}} \right),
\]

where

\[
C_\alpha := \frac{1}{2} + \frac{1}{2} \sqrt{1 + \frac{4}{\alpha}},
\]

\[
A_\alpha := C_\alpha + \frac{2}{\alpha},
\]

and

\[
B_\alpha = \alpha C_\alpha^{\alpha+2} e^{-\frac{1}{\alpha^2}}.
\]

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

\[
d_{n,m} = n \log \frac{m}{n} + \frac{3 n^2}{2 m} \log e - \frac{3 n}{2 m} \log e + O \left( \frac{1}{\sqrt{n}} + \frac{n^3}{m^2} \right).
\]

\[
4This expression for the error term in (16) is obtained with the choice \( \delta(n) = h''(z_0)^{-\rho/3} \) in [19, Table 8.4], provided certain conditions on \( h(z) \) are satisfied.

Discussion of Theorem 1

Related work. The leading terms of the asymptotic expansions for \( m = o(n) \) and \( n = o(m) \) (i.e., (17) and (22)) were derived by Orlitsky and Santhanam in [13]. The asymptotic expansion in (17) reveals that the error incurred by neglecting lower order terms may be significant. Consider the example in which \( n = 10^4 \) and \( m = 40 \) (or, approximately, \( m = n^{0.4} \)). Then, the leading term in (17) is only 5.5 times larger than the second term, and 131 times larger than the third term. The error from neglecting these two terms is thus 15.4% (assuming all other terms are negligible). Even for \( n = 10^8 \) and \( m = 1600 \), the error is still over 8%. It is interesting to notice that (17) is a “direct scaling” of (14): using Stirling’s approximation to replace \( \Gamma(x) \) in (14) by its asymptotic value \( \sqrt{2\pi/x} (x/e)^x \), and further approximating \( (1 + 1/x)^{x+1/2} \) with \( \sqrt{e} (1 + 1/(4x)) \), indeed yields exactly (17), up to the error terms. Thus, our results reveal that the first two terms of the asymptotic expansion for fixed \( m \) given by (14) are in fact a better approximation to \( d_{n,m} \) than the leading term of (17).

For the case \( m = \Theta(n) \), the methodology of [13] allowed only to extract the growth rate, i.e., \( d_{n,m} = \Theta(n) \), but not the constant in front of \( n \). The value of this constant, \( \log B_\alpha \), where \( B_\alpha \) is specified in (21) and (19), is plotted against \( \alpha \) in Figure 1. It is easy to see that, when \( \alpha \rightarrow 0 \), \( \log B_\alpha \approx (\alpha/2) \log(1/\alpha) \), in agreement with (17). Similarly, when \( \alpha \rightarrow \infty \), \( \log B_\alpha \approx \log \alpha \), in agreement with (22).

Finally, for the case \( n = o(m) \), our results confirm that the leading term is a good approximation to \( d_{n,m} \).

Convergence. Observe that the second order term in (17), which is \( \Theta(m) \), dominates \( -\log(n/m) \) whenever \( m = \Omega(n^a) \) for some \( a, 0 < a < 1 \). Hence, the leading term in the expansion is rather \( (m/2) \log(n/m) \) than \( (m-1)/2 \log(n/m) \). In the numerical example given for this case, the choice of a growth rate \( m = O(\sqrt{n}) \) is due to the fact that, otherwise, the error term \( O(m^2/n) \) may not even vanish, and it may dominate the constant, as well as the \( \sqrt{m/n} \) terms. For any given growth rate \( m = O(n^a) \), \( 0 < a < 1 \), an expansion in which the error term vanishes can be derived; however, no expansion has this
property for every possible value of $a$. The reason is that, as will become apparent in the proof of the theorem, any expansion will include an error term of the form $O(m(n/m)^{1/2})$ for some positive integer $j$. The same situation can be observed in (22), where one of the error terms becomes $O(n(n/m)^j)$ if a more accurate expansion is used.

A similar phenomenon is observed for the error term in (18), which is guaranteed to vanish only if $\ell(n) = o(n^{2/3})$, and it can otherwise dominate the constant term in the expansion. Again, for any given growth rate $\ell(n) = O(n^\alpha)$, an expansion in which the error term vanishes can be derived. Notice, however, that the case $\ell(n) \neq 0$ is analyzed only for completeness, since a typical application of (18) would in general involve approximating $d_{n,m_0}$ for a given pair of values $n_0, m_0$ which are deemed to fall in case (ii), by using (18) with $\alpha = n_0/m_0$ and $\ell(n) = 0$.

Now, we are in a position to revisit the second main topic of this paper, namely, the asymptotic expansion of the minimax worst-case redundancy $d_{n,m,M} = \log D_{n,m,M}$ relative to $\mathcal{M}_0$, using Lemma 1. As mentioned, the generic form of the sum in the lemma, given in Equation (5), is known as the binomial sum [5], [10]. If $D_{k,m}$ has a polynomial growth, (i.e., $D_{k,m} = 2^{k,m} = O(k^{(m-1)/2})$ when $m$ is fixed), then we can use the asymptotic expansion derived in [5], [10] to conclude that $D_{n,m,M} \sim D_{np,m}$. However, when $m$ varies with $n$ as in our study, the above expansion does not apply. In particular, the polynomial growth of $D_{n,m}$ does not hold any more and we need to compute asymptotics anew. We state and discuss our second main result in Theorem 2 below, whose proof is presented in Section III.

**Theorem 2:** Consider a family of memoryless sources $\tilde{\mathcal{M}}_0$ 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 worst-case redundancy $d_{n,m,M} = \log D_{n,m,M}$ takes the form:

(i) If $m$ is fixed, then

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

(23)

(ii) Let $m \to \infty$ as $n$ grows, with $m_n = o(n)$, where we write $m_n$ to explicitly show the dependence of $m$ on $n$. Assume:

(a) $m(x) := m_x$ is a continuous function, as well as its derivatives $m'(x)$ and $m''(x)$.

(b) $\Delta_n := m_{n+1} - m_n = O(m'(n))$, $m'(n) = O(m/n)$, and $m''(n) = O(m/n^2)$, where $m'(n)$ and $m''(n)$ are derivatives of $m(x)$ at $x = n$.

If $m_n = o(\sqrt{n}/\log n)$, then

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

(24)

Otherwise,

$$d_{n,m,M} = \frac{m_{np}}{2} \log \left( \frac{np}{m_{np}} \right) + \frac{m_{np} \log e}{2} + O \left( \log n + \frac{m_{np}^2 \log^2 n}{m_n} \right).$$

(25)

(ii) Let $m_n = \alpha n + \ell(n)$, where $\alpha$ is a positive constant and $\ell(n)$ is a monotonic function such that $\ell(n) = o(n)$. Then,

$$d_{n,m,M} = n \log (B_\alpha + 1 - \rho) - \log \sqrt{A_n} + O \left( \ell(n) + \frac{1}{\sqrt{n}} \right),$$

(26)

where $A_\alpha$ and $B_\alpha$ are defined in Theorem 1(ii).

(iii) Let $n = o(m_n)$ and assume $m_k/k$ is a nondecreasing sequence. Then,

$$d_{n,m,M} = n \log \left( \frac{pm_n}{n} \right) + O \left( \frac{n^2}{m_n} + \frac{1}{\sqrt{n}} \right).$$

(27)

**Discussion of Theorem 2**

**Assumptions.** For the assumption $\Delta_n = O(m'(n))$ in part (i) of the theorem to hold we need appropriate smoothness conditions (e.g., $\log m'(x)$ should be of bounded variation). In turn, for the assumption $m'(n) = O(m/n)$ to hold, it suffices to further assume that $m_n/n$ monotonically decreases for sufficiently large $n$, which is natural since $m_n/n = o(1)$ in this case. Finally, $m''(n) = O(m/n^2)$ requires natural convexity assumptions.\(^3\) In fact, when these assumptions cease to hold due to oscillations (which are not natural in our context, since our asymptotic analysis is best suited for some assumed, large size of the set $\mathcal{A}$, such as $m_n = n^a$ for some $a$), the claim of the theorem may not hold. In the “natural” $m_n = n^a$ case, with $a < 1/2$, we have $m_n = o(\sqrt{n}/\log n)$, $\Delta_n \approx m'(n) = an^{a-1} = O(m/n)$, and $m''(n) = -a(1-a)n^{a-2} = O(m/n^2)$. However, for $m_n = \sqrt{n}(\sin(n) + 2)$, we have $m'(n) = O(\sqrt{n}(\cos(n) + 2))$, the assumption $\Delta_n = O(m/n)$ breaks, and, as shown in Figure 2, Theorem 2(i) is invalid.

Similarly, the assumption of a monotonic increase of $m_k/k$ in case (iii) is also natural, since $n/m_n = o(1)$ in this case. We can replace this assumption by the weaker version $1 \leq \frac{m_k}{n} \leq \frac{C m_n}{n}$ for all $k \leq n$ and some $C > 0$, but then we can only show that

$$d_{n,m,M} = n \log \left( \frac{pm_n}{n} \right) + O(n).$$

As for case (ii), we have assumed that $\ell(n)$ is monotonic in order to prevent certain types of fluctuations. The result holds under a weaker assumption, though, namely that there exist constants $C$ and $a$ such that, for every pair of positive integers $i, j$, if $i < j$ we have

$$e^{\ell(i)+\ell(j)} i^a \leq C e^{\ell(j)+\ell(j)} j^a.$$

(28)

\(^3\)For example, if $m_n/n$ vanishes in a convex manner and $m_n$ is concave, then it is easy to see that $m''(n) = O(m/n^2)$.
Clearly, this condition is satisfied if \( \ell(n) \) is monotonic (and therefore so is \( |\ell(n)| \) for sufficiently large \( n \)). In any case, if \( g(n) \) is a monotonic function such that \( \ell(n) = O(g(n)) \), then the theorem holds with \( \ell(n) \) replaced with \( g(n) \) in the error term. If \( \ell(n) \) is a constant, denoted \( \ell \), then the constant term in (26) can be shown to be exactly \( \log(C_n/\sqrt{A_n}) \). If \( \ell'(n) := \ell(n) - (\log \sqrt{A_n})/(\log C_n) = \Omega(1) \), under the additional assumption that \( |\ell'(k)|/k \) is nonincreasing (which is again natural since \( \ell(n) = o(n) \)), the error term in (26) can be further shown to be \( \Theta(\ell'(n)) \).

Asymptotics. In passing, let us explain intuitively the asymptotics behind Theorem 2. As shown in Lemma 1, we deal here with the binomial sum, for which a general function \( f \), takes the form (5) (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 \( \pm \sqrt{m} \) around \( np \). This observation essentially explains case (i). When \( m = \Theta(n) \), the growth of \( f(k) = D_{k,m} = O(A^k) \) is exponential, and we need all the terms in the sum in order to extract the asymptotics. Finally, for case (iii), the function \( f(k) = D_{k,m} \) grows superexponentially, and the asymptotics of the binomial sum are determined by the last term, that is, \( k = n \). It is interesting to notice that, in case (ii), even the main asymptotic term differs from that of \( \log f(np) \), which is the value that intuition would indicate for a fixed \( m \).

III. PROOFS OF MAIN THEOREMS

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

A. Proof of Theorem 1

The starting point is Equation (12) which, as noted, follows from Lemma 2 and Stirling’s formula, and Cauchy’s coefficient formula (13), which takes the form

\[
[z^n][\beta(z)]^m = \frac{1}{2\pi i} \int e^{h(z)}dz,
\]

where

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

We will apply a modification of Lemma 3 in the evaluation of (29), for which we need to check that the necessary conditions are satisfied by the function \( h(z) \) of (30).

First we find an explicit real root, \( z_0 \), of the saddle point equation \( h' = 0 \), and show that it is unique in the interval [0, 1]. Differentiating (30), we have

\[
z \frac{\beta'(z_0)}{\beta(z_0)} = \frac{n + 1}{m}.
\]

Differentiating Equation (10), and using Equation (9), it is easy to see that

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

Thus, (31) takes the form

\[
\beta(z_0)^2 - \beta(z_0) = \frac{n + 1}{m}.
\]

By (9) and the definition of \( T(z) \), the range of \( \beta(z) \) for \( 0 < z < 1 \) is \([1, +\infty)\). Since the quadratic equation (33) has a unique real root in this range, we have

\[
\beta(z_0) = \frac{1}{2} \left[ 1 + \sqrt{1 + \frac{4(n + 1)}{m}} \right] := \frac{1}{\gamma_{n,m}}, \tag{34}
\]

and the uniqueness of a real root \( z_0 \) in \([0, 1)\) follows from the fact that \( \beta(z) \) is increasing in this interval. Moreover, by (9), (34) takes the form

\[
T\left(\frac{z_0}{e}\right) = 1 - \gamma_{n,m}.
\]

Therefore, by (10), we finally obtain the explicit expression

\[
z_0 = (1 - \gamma_{n,m})e^{\gamma_{n,m}} \tag{35}
\]

where, since

\[
\gamma_{n,m} = \frac{m}{2(n + 1)} \left( \sqrt{1 + \frac{4(n + 1)}{m}} - 1 \right), \tag{36}
\]

we have \( 0 < \gamma_{n,m} < 1 \) and also \( 0 < z_0 < 1 \). We then see that, by (30), (34), and (35), \( h(z_0) \) takes the form

\[
h(z_0) = -m \ln \gamma_{n,m} - (n + 1)[\ln(1 - \gamma_{n,m}) + \gamma_{n,m}] \tag{37}
\]

In addition, differentiating (30) twice, we obtain

\[
h''(z_0) = mA(z_0) + \frac{n + 1}{z_0^2},
\]

where

\[
A(z) = \frac{d}{dz} \left[ \frac{\beta'(z)}{\beta(z)} \right] = \frac{[\beta(z)^2 - \beta(z)] [2\beta(z)^2 - \beta(z) - 1]}{z^2}, \tag{38}
\]

the second equality being easily seen to follow from further differentiating (32). Thus, using (33),

\[
h''(z_0) = \frac{n + 1}{z_0^2} \left[ \frac{2(n + 1)}{m} + \beta(z_0) \right]
\]

which, again by (34) and (35), can be expressed in terms of \( \gamma_{n,m} \) as

\[
h''(z_0) = \frac{n + 1}{(1 - \gamma_{n,m})^2 e^{2\gamma_{n,m}}} \left[ \frac{2(n + 1)}{m} + \frac{1}{\gamma_{n,m}} \right]. \tag{39}
\]
Finally, taking another derivative in (38) and further using (32) and (33), after some additional computations, we obtain
\[
 h''(z_0) = \frac{n+1}{\gamma_{n,m}z_0} \left[ \frac{n+1}{m} \left( \frac{8}{\gamma_{n,m}} - 1 \right) - \frac{5}{\gamma_{n,m}} + 3 \right].
\] (40)

With these expressions on hand, we can now check the conditions required in Lemma 3 for the evaluation of (29).

The most intricate condition to be checked is that of “tail eliminations” (denoted (SP3) in [19, Table 8.4, (8.105)]). This condition is actually shown in [6, Lemma 5] to hold in more general cases than the function \( h(z) \) of (30). Also, proceeding along the lines of the proof of [19, Theorem 8.17]), it can be shown that Equation (16) of Lemma 3 holds with \( \rho = 3/2 \) if \( h''(z_0) \) grows at least linearly and if \( h'''(z_0) = o(h''(z_0))^{3/2} \).

Thus, (12) and the modified Lemma 3 yield
\[
d_{n,m} = h(z_0) \log e - \log \sqrt{\frac{h''(z_0)}{n}} + O \left( \frac{h''(z_0)}{(h''(z_0))^{3/2}} + \frac{1}{n} \right)
\] (41)
provided the error term is \( o(1) \) and \( h''(z_0) \) grows at least linearly. Consequently, to complete the proof of Theorem 1, we need to evaluate the right-hand side of (41). In view of (37) and (39), which give \( h(z_0) \) and \( h''(z_0) \) as functions of \( \gamma_{n,m} \), the solution depends on the possible growth rates of \( m \). We analyze next all possible cases.

**CASE: \( m = o(n) \).**

Letting \( m/n \to 0 \) in Equation (36), it is easy to see that
\[
\gamma_{n,m} = \sqrt{\frac{m}{n}} \left( 1 - \frac{1}{2} \sqrt{\frac{m}{n}} \right) + O \left( \frac{m^{3/2}}{n^{3/2}} \right).
\]
Substituting into (37) and (39), we obtain
\[
h(z_0) = \frac{m}{2} \ln \frac{n}{m} + \frac{m}{2} + \frac{m}{3} \sqrt{\frac{m}{n}} + O \left( \frac{m^2}{n} \right)
\]
and
\[
\ln \frac{h''(z_0)}{n} = \ln \frac{n}{m} + \ln 2 + \frac{1}{2} \sqrt{\frac{m}{n}} + O \left( \frac{m}{n} \right).
\] (42)
From (40), and noticing that, in this case, Equation (35) yields \( z_0 \to 1 \), we further obtain
\[
h''(z_0) = \Theta \left( \frac{n^3}{m^2} \right).
\] (43)

Theorem 1(i) follows from substituting these equations into (41), observing that (42) and (43) guarantee that the necessary conditions for the modified Lemma 3 to hold for \( h(z) \) are satisfied.\(^6\)

**CASE: \( m = \Theta(n) \).**

Since \( z_0 \) is given by (35) where, in this case, \( m = \alpha n + \ell(n) \) and \( \ell(n) = o(n) \), we can view \( z_0 \) as a function of \( m/(n+1) \), which we expand around \( \alpha \). The value of this function at \( \alpha \) is
\[
z_0 = (1 - C_\alpha^{-1}) e^{1/C_\alpha} = \alpha^{-1} C_\alpha^{-2} e^{1/C_\alpha}
\]
where \( C_\alpha \) is given by (19). It is then easy to see that
\[
z_0 = z_0 - z_0 \alpha^{-1} A_\alpha^{-1} \delta(n) + O(\delta(n)^2),
\]
where \( \delta(n) := (\ell(n) - \alpha)/(n+1) = o(1) \) and \( A_\alpha \) is given by (20). With this value of \( z_0 \) we can then compute, with a Taylor expansion around \( z_0 \),
\[
h(z_0) = n \ln(C_\alpha z_0^{-1}) + \ell(n) \ln C_\alpha - \ln z_0 - n\delta(n)^2 \frac{1}{2 C_\alpha A_\alpha} + O(n\delta(n)^3),
\]
\[
\ln \frac{h''(z_0)}{n} = \ln(A_\alpha z_0^{-2}) + O(\delta(n)),
\]
\[
h''(z_0) = O(n).
\]
Substitution into (41) completes the proof of Theorem 1(ii), after observing, again, that the necessary conditions for the modified Lemma 3 hold.

**CASE: \( n = o(m) \).**

Letting \( n/m \to 0 \) in Equation (36), it is easy to see that
\[
\gamma_{n,m} = 1 - \frac{n+1}{m} + \frac{2(n+1)^2}{m^2} + O \left( \frac{n^3}{m^2} \right).
\]
Substituting into (37) and (39), we obtain
\[
h(z_0) = (n+1) \ln \frac{m}{n+1} + \frac{3(n+1)^2}{2m} + O \left( \frac{n^3}{m^2} \right)
\]
and
\[
\ln \frac{h''(z_0)}{n} = 2 \ln \frac{m}{n+1} + 9 \frac{n+1}{m} + O \left( \frac{n^2}{m^2} \right).
\]
From (40), and noticing that, in this case, Equation (35) yields \( z_0 = \Theta(\gamma_{n,m}) = \Theta(n/m) \), we further obtain
\[
h''(z_0) = \Theta \left( \frac{n^3}{m^2} \right).
\]
Putting everything together, substituting into (41), and observing that the necessary conditions for the modified Lemma 3 hold, we prove Theorem 1(iii).\(^7\)

**B. Proof of Theorem 2**

By Lemma 1, in order to prove Theorem 2 we need to evaluate the binomial sum
\[
S_f(n) = \sum_k \binom{n}{k} p^k (1-p)^{n-k} f(k)
\] (44)
for \( f(k) = D_{k,m_k} \) that grows faster than any polynomial for \( m \to \infty \).

**CASE: \( m_n = o(n) \).**

We first observe that
\[
S_f(n) = E_X[f(X)],
\]
where \( E_X \) denotes expectation with respect to a binomially distributed random variable \( X \). Our basic evaluation technique

\(^6\)Taking more terms in the expansion of \( \gamma_{n,m} \), an \( O((m/n)^{1/2}) \) error term for \( h(z_0) \) can be obtained, where \( j \) is as large as desired. Thus, while no value of \( j \) guarantees a vanishing error for every \( m \), for each given \( m = O(n^\alpha) \), a choice of \( j \) exists that guarantees \( o(1) \) error.

\(^7\)We can take more terms in the expansion of \( \gamma_{n,m} \), also in this case, leading to an \( O((n/m)^{1/2}) \) error term for \( h(z_0) \).
will rely on the concentration of $X$ around its mean $np$. The following lemma is a straightforward consequence of this concentration.

**Lemma 4:** Let $g(k)$ be a function satisfying the following condition: There exist constants $C$ and $a$ such that, for every pair of positive integers $i, j$, with $i < j$, we have $|g(i)|v^i \leq C|g(j)|v^j$. Then, $S_{g}(n) = O(g(n))$ and $S_{1/|g|}(n) = \Omega(1/g(n))$.

**Proof:** By Hoeffding’s inequality [9], for any $\epsilon > 0$ we have

$$\Pr\{X < n(p - \epsilon)\} \leq e^{-\frac{1}{2}n\epsilon^2}.$$ 

Therefore,

$$S_{g}(n) \leq e^{-\frac{1}{2}n\epsilon^2} \max_{1 \leq k \leq n} |g(k)| + \max_{n(p - \epsilon) \leq k \leq n} |g(k)| \leq C|g(n)|[e^{-\frac{1}{2}n\epsilon^2} n^a + (p - \epsilon)^{-a}] \leq C'|g(n)|$$

for some constant $C'$, where the second inequality follows from the condition on $g$.

Similarly,

$$S_{1/|g|}(n) \geq \Pr\{X > n(p - \epsilon)\} \min_{n(p - \epsilon) \leq k \leq n} (1/|g(k)|) \geq \frac{(p - \epsilon)^a}{2C} \frac{1}{|g(n)|}.$$ 

Lemma 4 applies, e.g., to functions that vanish polynomially fast without excessive fluctuations. It holds trivially for non-decreasing functions.

One approach for taking advantage of the concentration of $X$ consists of applying Taylor’s theorem to $f(x)$ (the extension of $f(n)$ to the real line) around the mean $x = np$, and estimating $f''(n)$. However, notice that Theorem 1 does not provide enough information about $f''(n)$ to obtain such an estimate, since the behavior of $f''(n)$ could be dominated by the error term of $f(n)$. We circumvent this problem by appropriately defining functions $f_1$ and $f_2$ such that

$$f(n) = f_1(n)[1 + O(f_2(n))],$$

where $f_2(n)$ is a function that satisfies the condition of Lemma 4, and

$$\max_{0 \leq x \leq n} |f_1(x)| = O(f_1(n)).$$

It then follows from Lemma 4 and (45) that

$$S_{f}(n) = S_{f_1}(n) = O(f_1(n)f_2(n)).$$

Next, we estimate $S_{f_1}(n)$ by applying Taylor’s theorem to $f_1(x)$ around $x = np$, which yields

$$f_1(x) = f_1(np) + (x - np)f_1'(np) + \frac{(x - np)^2}{2}f_1''(x')$$

for some $x'$ that lies between $x$ and $np$. Letting

$$\xi(n) := \max_{0 \leq x \leq n} |f_1''(x)|,$$

we obtain

$$f_1(x) - f_1(np) - (x - np)f_1'(np) = \frac{(x - np)^2}{2}O(\xi(n)).$$

Taking expectations with respect to $X$ in (47), and noting that $E_X[X] = np$ and $\text{Var}[X] = npq$, yields

$$S_{f_1}(n) - f_1(np) = O(n\xi(n))$$

which, together with (46), implies

$$S_{f}(n) - f_1(np) = O(n\xi(n) + f_1(n)f_2(n)).$$

By (45) we then have

$$S_{f}(n) = f_1(np) \left[1 + O \left(\frac{n\xi(n)}{f_1(n)} + f_2(n)\right)\right].$$

As we will show, this bound leads to a precise asymptotic estimate of $S_{f}(n)$ provided that $\xi(n) = o(f_1(n))$. In this case, (48) implies

$$D_{n,m,M} = \log S_{f}(n) = \log f_1(np) + O \left(\frac{n\xi(n)}{f_1(n)} + f_2(n)\right).$$

In the fixed $m$ case we have, by (14),

$$f(n) = Kn^{(m-1)/2} \left[1 + O \left(1/\sqrt{n}\right)\right]$$

where $K$ is a constant. Thus, we can choose $f_1(n) = Kn^{(m-1)/2}$ and all the necessary conditions are obviously satisfied. Hence, Theorem 2(iii) holds. A more precise asymptotic expansion can be found using tools from [5], [10].

Let us now consider part (i) of Theorem 2, that is, we assume that $m \to \infty$ and $m = o(n)$. If we further assume, first, that $m = o(\sqrt{n})$, in view of (17), we can then choose

$$f_1(n) = \left(\frac{ne}{m}\right)^{\frac{m}{2}} \sqrt{\frac{m}{2n}},$$

which clearly satisfies (45), and

$$f_2(n) = O \left(\frac{m^2}{n} + \frac{1}{\sqrt{m}}\right)$$

which dominates the $O(\sqrt{m/n})$ term in (17) for this range of values of $m$, and vanishes polynomially fast. In order to check the applicability of (49), we need to estimate $\xi(n)/f_1(n)$, for which we will use two of the additional assumptions in this part of the lemma, namely that $O(m'(n)) = O(m/n)$ and $O(m''(n)) = O(m/n^2)$. Now, since for any function $g$ we have $g''/g = [(\log g')^2 + (\log g)''$, it is relatively simple to compute that

$$\frac{n f_1''(n)}{f_1(n)} = O \left(\frac{m^2 \log^2 n}{n}\right).$$

Moreover, due to the continuity of $m$, $m'$, and $m''$ (which implies the continuity of $f_1''(n)$, and to the fact that $[f_1(n)m^2 \log^2 n]/n^2$ is increasing for sufficiently large $n$, it is easy to see that (52) holds also when $\xi(n)$ replaces $f_1''(n)$ in the right-hand side. When $m = o(\sqrt{n}/\log n)$, we have $n\xi(n)/f_1(n) = o(1)$ and (24) follows from (49), (50), and (51).
We need a different approach for the remaining \( m = o(n) \) cases, since in those cases the error term \( O(m^2(\log^2 n)/n) \) does not vanish. Observe that we always have
\[
e^{-1/\{12np(1-p)\}} \sqrt{2\pi np(1-p)} f(np) \leq S_f(n)
\]
\[
\leq n \max_k \left( \binom{n}{k} p^k (1-p)^{n-k} f(k) \right) \tag{53}
\]
where we have used Stirling’s inequality to lower-bound the term corresponding to \( k = \lfloor np \rfloor \) in the sum (44). We need to find \( k = k^* \) that maximizes the right-hand side of (53). Let
\[
F(k) = \binom{n}{k} p^k (1-p)^{n-k} f(k).
\]
Then, \( k^* \) satisfies
\[
\frac{F(k^* + 1)}{F(k^*)} \approx 1. \tag{54}
\]
We first observe that for our \( f(n) = D_{n,m_n} \), using (17) and our assumption that \( \Delta_n = O(m_n/n) \), we obtain, after some computations,
\[
f(k + 1) \leq f(k) + \log k + \frac{\log m_n}{m_n}
\]
Thus, (54) takes the form
\[
\frac{n - k}{k + 1} = 1 - p - \log \frac{m_n}{m_n}
\]
which yields
\[
k^* = np + O(m_n \log(n/m_n)).
\]
Applying Stirling’s formula it can then be shown that
\[
\log F(k^*) = \log f(k^*) + O(\log n) + O\left( \frac{m_n^2}{n} \log^2 \frac{n}{m_n} \right) \tag{55}
\]
where the first error term is due to the \( 1/\sqrt{n} \) factor in the formula, and the second error term is due to the discrepancy between \( k^* \) and \( np \). In addition,
\[
\log f(k^*) = \frac{m_n p}{2} \log \frac{m_n p}{m_n} + \frac{m_n p}{2} \log e + O\left( \frac{m_n^2}{n} \log^2 \frac{n}{m_n} \right) \tag{56}
\]
where again the error term is due to the discrepancy between \( k^* \) and \( np \) and is easily seen to dominate other terms in (17). Equations (53), (55), and (56) imply (25) of Theorem 2(i), where the growth rate of \( m_n \) further determines the dominating error terms.

**CASE: \( m_n = \Theta(n) \).**

By (18), since \( \ell(n)/n = o(1) \),
\[
f(k) = D_{k,m_k} \leq A_\alpha^{-1/2} B_{\alpha q}^{k-\log f_1(k)}
\]
where \( f_1(k) = O(\ell(k) + 1/\sqrt{\ell}) \), and the inequality is needed because \( \ell(k) \) could be negative. Thus,
\[
S_f(n) \leq A_\alpha^{-1/2} (B_{\alpha p} + q)^n \cdot \sum_{k=0}^n \binom{n}{k} \left( \frac{B_{\alpha p}}{B_{\alpha p} + q} \right)^k \left( \frac{q}{B_{\alpha p} + q} \right)^{n-k} 2^{f_1(k)}.
\]
The above sum is upper-bounded by the binomial sum (with parameter \( B_{\alpha p}/(B_{\alpha p} + q) \) rather than \( p \) for the function \( 2^{C'(|\ell(k)| + 1/\sqrt{\ell})} \) for some constant \( C' \). Since \( \ell(n) \) is assumed monotonic, Condition (28) is satisfied (see discussion on Theorem 2), and therefore we can apply Lemma 4 to this new binomial sum, to obtain
\[
S_f(n) \leq A_\alpha^{-1/2} (B_{\alpha p} + q)^n O(2^{f_2(n)}) \tag{57}
\]
where \( f_2(n) = C'(|\ell(n)| + 1/\sqrt{n}) \). Since \( 2^{f_2(n)} \geq 1 \), we conclude that
\[
\log S_f(n) \leq n \log (B_{\alpha p} + q) - \log \sqrt{A_\alpha} + O(\ell(n) + 1/\sqrt{n}) \tag{58}
\]
where we notice that (58) is in fact an equality whenever \( \ell(n) \geq 0 \).

To obtain a matching lower bound, we have
\[
f(k) \geq A_\alpha^{-1/2} B_{\alpha q}^{k-\log f_1(k)}
\]
so that, proceeding as in the upper bound,
\[
S_f(n) \geq A_\alpha^{-1/2} (B_{\alpha p} + q)^n \cdot \sum_{k=0}^n \binom{n}{k} \left( \frac{B_{\alpha p}}{B_{\alpha p} + q} \right)^k \left( \frac{q}{B_{\alpha p} + q} \right)^{n-k} 2^{-f_1(k)}.
\]
We can now apply the second statement in Lemma 4, to obtain
\[
S_f(n) \geq A_\alpha^{-1/2} (B_{\alpha p} + q)^n \Omega(2^{-f_2(n)})
\]
which, after taking logarithms, yields the desired lower bound and, hence, Equation (26) of Theorem 2(iii). A more precise estimate is discussed in Remark 2.

When \( \ell(n) \) is a constant, denoted \( \ell \), the constant term in (18) includes an additional \( \ell \log C_\alpha \), which is added also in \( d_{n,m_n,M} \) and the error term becomes \( O(1/\sqrt{n}) \).

**CASE: \( n = o(m) \).**

By (22),
\[
f(k) = D_{k,m_k} = g_1(k)(1 + g_2(k))
\]
where \( g_2(k) = O(1/\sqrt{k} + k/m_k) \), and
\[
g_1(k) = \frac{m_k}{k} e^{3k^2/(2m_k)}
\]
\[
= \frac{m_k}{k} \left( 1 + \frac{3k}{2m_k} + O\left( \frac{k}{m_k^2} \right) \right)^k
\]
\[
= \frac{m_k}{k} \left( 1 + \frac{3}{2} + O\left( \frac{k}{m_k} \right) \right)^k.
\]
We first use our assumption that \( 1 \leq (m_k/k) \leq (m_n/n) \) for all \( k \leq n \) to obtain the upper bound
\[
S_{g_1}(n) = \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} g_1(k) \\
\leq \sum_{k=1}^{n} \binom{n}{k} \left( p \left( \frac{m_n}{n} + \frac{3}{2} + O \left( \frac{k}{m_k} \right) \right) \right)^k q^{n-k} \\
\leq \left( \frac{pm_n}{n} + K \right)^n
\]
for some constant \( K \), where we have upper-bounded the \( O(k/m_k) \) terms with a constant, since \( k/m_k = o(1) \). In addition, proceeding as in the derivation of (57),
\[
\sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} g_1(k)g_2(k) \\
\leq \left( \frac{pm_n}{n} + K \right)^n O \left( \frac{1}{\sqrt{n}} + \frac{n}{m_n} \right)
\]
where we have used again Lemma 4. Thus, \( S_f(n) \leq \left( \frac{pm_n}{n} + K \right)^n \left( 1 + O \left( \frac{1}{\sqrt{n}} + \frac{n}{m_n} \right) \right) \), or
\[
\log S_f(n) \leq n \log \left( \frac{pm_n}{n} + K \right) + O \left( \frac{1}{\sqrt{n}} + \frac{n}{m_n} \right) = n \log \frac{pm_n}{n} + O \left( \frac{1}{\sqrt{n}} + \frac{n^2}{m_n} \right). \quad (59)
\]
On the other hand, we can lower-bound the binomial sum (44) with the term corresponding to \( k = n \), namely \( p^n D_{m,n,m} \), to obtain
\[
\log S_f(n) \geq n \log p + d_{m,n,m}. \quad (60)
\]
Theorem 2(iii) then follows from (59), (60), and (22). If \( (m_k/k) \leq C(m_n/n) \), we obtain an additional term \( n \log C \), thus the error term is \( O(n) \).

Remark 1. Notice that one of the error terms generated by the “sandwich argument” of (53), used in the proof of (25), is \( O(\log n) \), independently of the value of \( m \). Therefore, this method is not suitable for the \( m = O(\log n) \) cases (addressed via a Taylor expansion in the proof of (24)) as this error term would dominate one of the other terms. Moreover, for fixed \( m \), the method cannot even provide the main asymptotic term, which is also \( O(\log n) \).

Remark 2. In part (ii), under the additional assumptions that \( \ell(n) := \ell(n) - (\log \sqrt{\alpha}/(\log C_\alpha) = \Omega(1) \) and \( |\ell(k)/k| \) is nonincreasing, we can further prove that the error term is \( \Theta(\ell(n)) \). Clearly, our assumptions imply that \( \ell(k) \) is constant sign. Assume \( \ell(k) > 0 \); a similar argument can be used for \( \ell(k) < 0 \). Then,
\[
f(k) = B_\alpha 2^{\ell(k)n} = \left[ B_\alpha 2^{\ell(k)/k} \right]^k = \left[ B_\alpha 2^{\ell'(n)/n} \right]^k.
\]

Therefore, using a bounding technique similar to part (iii), we obtain
\[
S_f(n) = \left[ B_\alpha p + q + \Omega(\ell'(n)/n) \right]^n
\]
and, after taking the logarithm,
\[
d_{m,n,M} = n \log (B_\alpha p + q) + \Omega(\ell'(n)).
\]
Together with (26), we conclude that the error term is \( \Theta(\ell'(n)) \).

REFERENCES