A Remark on Unified Error Exponents: Hypothesis Testing, Data Compression & Measure Concentration

Ioannis Kontoyiannis and Ali Devin Sezer

Abstract. Let A be finite set equipped with a probability distribution P, and let M be a "mass" function on A. A characterization is given for the most efficient way in which A^n can be covered using spheres of a fixed radius. A covering is a subset C_n of A^n with the property that most of the elements of A^n are within some fixed distance from at least one element of C_n, and "most of the elements" means a set whose probability is exponentially close to one (with respect to the product distribution P^n). An efficient covering is one with small mass $M^n(C_n)$. With different choices for the geometry on A, this characterization gives various corollaries as special cases, including Marton's error-exponents theorem in lossy data compression, Hoeffding's optimal hypothesis testing exponents, and a new sharp converse to some measure concentration inequalities on discrete spaces.

Let A be finite set equipped with a probability distribution P, and let M be a "mass" function on A. A characterization is given for the most efficient way in which A^n can be covered using spheres of a fixed radius. A covering is a subset C_n of A^n with the property that most of the elements of A^n are within some fixed distance from at least one element of C_n, and "most of the elements" means a set whose probability is exponentially close to one (with respect to the product distribution P^n). An efficient covering is one with small mass $M^n(C_n)$. With different choices for the geometry on A, this characterization gives various corollaries as special cases, including Marton's error-exponents theorem in lossy data compression, Hoeffding's optimal hypothesis testing exponents, and a new sharp converse to some measure concentration inequalities on discrete spaces.

1. Introduction

Let A be a finite set and P a probability distribution on A. Suppose that the distance (or "distortion") $\rho(x, y)$ between any two points $x, y \in A$ is measured by a given nonnegative function $\rho : A \times A \to [0, \infty)$, and for strings $x_1^n = (x_1, x_2, \dots, x_n)$ and $y_1^n = (y_1, y_2, \dots, y_n)$ in A^n let $\rho_n(x_1^n, y_1^n)$ be the corresponding coordinate-wise distance (or single-letter distortion measure) on $A^n \times A^n$:

$$\rho_n(x_1^n, y_1^n) = \frac{1}{n} \sum_{i=1}^{n} \rho(x_i, y_i).$$

Since A is a finite set, the function ρ is bounded above by

$$D_{\max} \overset{\triangle}{=} \max_{x,y \in A} \rho(x, y) = \max_{x_1^n, y_1^n \in A^n} \rho_n(x_1^n, y_1^n).$$

Without loss of generality we assume throughout that $P(a) > 0$ for all $a \in A$, and that for each $a \in A$ there exists a $b \in A$ with $\rho(a, b) = 0$ (otherwise we may consider $\rho'(x, y) = [\rho(x, y) - \min_{z \in A} \rho(x, z)]$ instead of $\rho(x, y)$).

Given a $D \geq 0$, we want to cover "most" of A^n using balls $B(y_1^n, D)$, where

$$B(y_1^n, D) = \{x_1^n \in A^n \ : \ \rho_n(x_1^n, y_1^n) \leq D\}$$

is the closed ball of radius D centered at $y_1^n \in A^n$. To be precise, given a set $C_n \subset A^n$, we write $[C_n]_D$ for the D-blowup of C_n,

$$[C_n]_D \overset{\triangle}{=} \bigcup_{y_1^n \in C_n} B(y_1^n, D).$$

A D-covering of A^n is a sequence of subsets C_n of A^n, $n \geq 1$, such that the P^n-probability of the part of A^n which is not covered by C_n within distance D has exponentially small probability,

$$(1) \qquad\qquad \Pr\{\text{"error"}\} \overset{\triangle}{=} 1 - P^n([C_n]_D) \ \approx \ 2^{-nE},$$

for some $E > 0$. We are interested in "efficient" coverings of A^n, that is, given a "mass function" $M : A \to (0, \infty)$, we want to find D-coverings $\{C_n\}$ that satisfy (1) and also have small mass

$$M^n(C_n) \overset{\triangle}{=} \sum_{y_1^n \in C_n} M^n(y_1^n) = \sum_{y_1^n \in C_n} \prod_{i=1}^{n} M(y_i).$$

Clearly there is a trade-off between finding coverings $\{C_n\}$ with small mass, and coverings with a good (i.e., large) error-exponent E as in (1). Typically, the better the error-exponent, the larger the C_n, and the bigger their mass would tend to be.

Motivated, in part, by the following example and by the applications illustrated in the examples of the following section, in our main result we give a precise characterization of this trade-off.

Example: Measure Concentration on the Binary Cube.

Consider the n-dimensional binary cube $A^n = \{0,1\}^n$. We measure distance on A^n by the proportion of mismatches between two binary strings x_1^n and y_1^n, i.e., we take $\rho_n(x_1^n, y_1^n)$ to be the Hamming distance,

$$(2) \qquad \rho_n(x_1^n, y_1^n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{x_i \neq y_i\}}, \quad x_1^n, y_1^n \in A^n,$$

which also coincides with the normalized graph distance when A^n is equipped with the nearest-neighbor graph structure. For simplicity, in this example we consider natural logarithms and exponentials.

A well-known measure concentration inequality [10, Prop. 2.1.1][9, Thm. 3.5] gives a precise lower bound on the sphere-covering error probability of an arbitrary C_n: For any $D \geq 0$, any product distribution P^n on A^n, and any $C_n \subset A^n$,

$$\Pr\{\text{``error''}\} = 1 - P^n([C_n]_D) \leq \frac{e^{-nD^2/2}}{P^n(C_n)}.$$

Therefore, if $\{C_n\}$ is *any* D-covering consisting of sets with $P^n(C_n) \approx e^{-nr}$ for some $r > 0$, then the union of the balls $B(y_1^n, D)$ centered at the points $y_1^n \in C_n$ covers all of A^n except for a set of probability no greater than

$$(3) \qquad \approx e^{-n\left(\frac{D^2}{2} - r\right)}.$$

It is then natural to ask, what is the *best* achievable error exponent among all D-coverings $\{C_n\}$ with probability no greater that $\approx e^{-nr}$? In other words, we are asking for small sets with the largest possible "boundary," sets C_n with "volume" $P^n(C_n)$ no greater than e^{-nr} but whose D-blowups $[C_n]_D$ cover as much of A^n as possible. As pointed in [6], this question can be thought of as the opposite of the usual isoperimetric problem.

Taking $M = P$ in the general setting described above, we obtain the answer to this question as a corollary to our general result in the following section; see Corollary 3.

2. Results

Given any $D \geq 0$ and any $R \in \mathbb{R}$, let $E(R, D)$ denote the best achievable error-exponent among all D-coverings with mass asymptotically bounded by 2^{nR}. Letting $\mathcal{C}(R)$ denote the collection of all sequences of subsets C_n of A^n with $\limsup_n \frac{1}{n} \log M^n(C_n) \leq R$, define,

$$E(R, D) \overset{\triangle}{=} \sup_{\{C_n\} \in \mathcal{C}(R)} \liminf_{n \to \infty} -\frac{1}{n} \log \left[1 - P^n([C_n]_D) \right],$$

where 'log' denotes the logarithm taken to base 2.

A weaker version of this problem was recently considered in [6], where it was shown that the probability of error can only decrease to zero if R is greater than $R(D; P, M)$,

$$(4) \quad R(D; P, M) \triangleq \inf_{X \sim P,\ E\rho(X,Y) \leq D} \left\{ H(P_{X,Y} \| P \times P_Y) + E[\log M(Y)] \right\},$$

where the infimum is taken over all jointly distributed random variables (X, Y) such that X has distribution P and $E\rho(X, Y) \leq D$, and $P_{X,Y}$ denotes the joint distribution of X, Y, P_Y denotes the marginal distribution of Y, and $H(\mu \| \nu)$ denotes the relative entropy between two probability measures μ and ν on the same finite set S,

$$H(\mu \| \nu) \triangleq \sum_{s \in S} \mu(s) \log \frac{\mu(s)}{\nu(s)}.$$

Therefore, the error-exponent $E(R, D)$ can only be nontrivial (i.e., nonzero) for $R > R(D; P, M)$. Also note that any $C_n \subset A^n$ has

$$\frac{1}{n} \log M^n(C_n) \leq \frac{1}{n} \log M^n(A^n) = \log M(A).$$

Hence, from now on we restrict attention to the range of interesting values for R between $R(D; P, M)$ and $R_{\max} \triangleq \log M(A)$.

Theorem. For all $D \in [0, D_{\max})$ and all $R(D; P, M) < R < R_{\max}$, the best achievable exponent of the error probability, among all D-coverings $\{C_n\}$ with mass asymptotically bounded by 2^{nR}, is

$$E(R, D) = E^*(R, D) \triangleq \inf_{Q\ :\ R(D; Q, M) > R} H(Q \| P),$$

where $R(D; P, M)$ is defined in (4) and $H(Q \| P)$ denotes the relative entropy (or Kullback-Leibler divergence) between two distributions P and Q.

Remarks.

1. A slightly different error-exponent. Alternatively, we can define a version of the optimal error-exponent by considering only D-coverings $\{C_n\}$ with mass bounded by 2^{nR} for *all* n:

$$E'(R, D) \triangleq \liminf_{n \to \infty} -\frac{1}{n} \log \left\{ \min_{C_n\ :\ M^n(C_n) \leq 2^{nR}} \left[1 - P^n([C_n]_D) \right] \right\}.$$

From the theorem it easily follows that $E'(R, D)$ is also equal to $E^*(R, D)$ at all points R where $E^*(R, D)$ is continuous and, since it is nondecreasing in R, $E^*(R, D)$ is indeed continuous at all except countably many values of R. But in general it may fail to be continuous everywhere, as illustrated in the discussions by Marton [7] and Ahlswede [1] for the special case of lossy data compression (which corresponds to taking $M(x) \equiv 1$; see Example 2 below).

2. Proof. The proof of the theorem is a modification of Marton's [7] original argument for the case of error-exponents in lossy data compression. The optimal sets $\{C_n\}$ achieving $E^*(R, D)$ are randomly generated, and they are *universal* in

that their construction only depends on R, D, and M. Therefore, they achieve the optimal error-exponent simultaneously for all distributions P.

Example 1: Hypothesis Testing.

Let P_0 and P_1 be two probability distributions A with all positive probabilities. Suppose that the null hypothesis that a sample $X_1^n = (X_1, X_2, \ldots, X_n)$ of n independent observations comes from P_0 is to be tested against the simple alternative that X_1^n comes from P_1. Any test between these two hypotheses is simply a decision region $C_n \subset A^n$: If $X_1^n \in C_n$ we declare that $X_1^n \sim P_1^n$, otherwise we declare $X_1^n \sim P_0^n$. The set C_n is called the *critical region*, and the type-I and type-II probabilities of error associated with the test are, respectively,

$$\alpha_n = P_0^n(C_n) \quad \text{and} \quad \beta_n = P_1^n(C_n^c).$$

Clearly we wish to have α_n and β_n both decrease to zero as fast as possible. In particular, we may ask how quickly β_n can decay to zero if we require that α_n decays exponentially at some rate $r > 0$, i.e., $\alpha_n \approx 2^{-nr}$. In statistical terminology, we are asking for the fastest rate of decay of the type-II error probability among all tests with significance level $\alpha_n \leq 2^{-nr}$.

Formally, we want to identify the best exponent of the error probability $\beta_n = 1 - P_1^n(C_n)$ among all C_n with $P_0^n(C_n) \leq 2^{-nr}$. Taking $P = P_1$, $M = P_0$, $R = -r$, and allowing *no* distortion, this question reduces exactly to the our earlier sphere-covering problem. [To be precise, allowing no distortion means we take $D = 0$ with $\rho(x, y)$ being Hamming distortion as in (2).] Accordingly, $R(D; P, M) = R(0; P_1, P_0)$ turns out to be equal to $-H(P_1 \| P_0)$, and from the theorem we immediately obtain the following classical result of Hoeffding. Also see [2, Thms. 9, 10] and [3, Ex.12, p.43] for versions of this result in the information theory literature.

Corollary 1. (Hypothesis Testing) [5] Let $\{C_n\}$ be an arbitrary sequence of tests with associated error probabilities α_n and β_n as above. Among all tests with

$$\limsup_{n \to \infty} \frac{1}{n} \log \alpha_n \leq -r$$

for some $r \in (0, H(P_1 \| P_0))$, the fastest achievable asymptotic rate of decay of β_n is

$$\lim_{n \to \infty} -\frac{1}{n} \log \beta_n = \inf_{Q : H(Q \| P_0) < r} H(Q \| P_1).$$

As mentioned earlier, the decision regions C_n in the Corollary are randomly generated. Therefore, although they do achieve asymptotically optimal performance, they are not optimal for finite n in the Neyman-Pearson sense.

Example 2: Lossy Data Compression.

Suppose data $X_1^n = (X_1, X_2, \ldots, X_n)$ is generated by a stationary, memoryless source, i.e., X_1^n are i.i.d. (independent and identically distributed) random variables, with distribution P on the finite alphabet A. The objective of lossy data compression is to find efficient representations $y_1^n \in A^n$ for all source strings

$x_1^n \in A^n$. In particular, suppose that the maximum amount of distortion $\rho_n(x_1^n, y_1^n)$ that we are willing to tolerate between a source string x_1^n and its representation y_1^n is some $D \geq 0$, where $\{\rho_n\}$ is a family of single-letter distortion measures as in (1). Then the problem is to find an efficient codebook $C_n \subset A^n$ such that for most of the source strings x_1^n there is a $y_1^n \in C_n$ with $\rho_n(x_1^n, y_1^n) \leq D$.

Here, an efficient codebook C_n is one that leads to good compression, i.e., one whose size is as small as possible. And, on the other hand, we also want to make sure that the probability that a source string cannot be represented by any element of C_n with distortion D or less, is small. Taking M to be counting measure ($M(x) = 1$ for all $x \in A$), the mass $M^n(C_n)$ of the codebook becomes its size $|C_n|$, and the problem of finding a good codebook reduces to the earlier sphere-covering question. Accordingly, the rate-function $R(D; P; M)$ reduces to Shannon's rate-distortion function $R(D; P)$, and the theorem yields Marton's error-exponents result.

Corollary 2. (Lossy Data Compression) [7] Let $D \geq 0$ be a given distortion level, and $R(D; P) < R < \log |A|$. Among all sequences of codebooks $\{C_n\}$ with asymptotic rate no greater than R bits/symbol,

$$\limsup_{n \to \infty} \frac{1}{n} \log |C_n| \leq R,$$

the fastest achievable asymptotic rate of decay of the probability of error is

$$\lim_{n \to \infty} -\frac{1}{n} \log \left[1 - P^n([C_n]_D) \right] = \inf_{Q \,:\, R(D;Q) > R} H(Q\|P).$$

Example 3: Measure Concentration on the Binary Cube.

Consider again the setting of the example described in the introduction. There we asked for the *best* achievable error exponent among all D-coverings $\{C_n\}$ with probability no greater that $\approx e^{-nr}$. Taking $M = P$ in the theorem, we obtain the answer to this question in the following Corollary. Let $H_e(P\|Q)$ denote the relative entropy expressed in nats rather than bits, $H_e(P\|Q) = (\log_e 2) H(P\|Q)$, and similarly write $R_e(D; P, M) = (\log_e 2) R(D; P, M)$.

Corollary 3. (Converse Measure Concentration) Let $D \geq 0$ and $0 < r < -R_e(D; P, P)$. Among all D-coverings $\{C_n\}$ with

$$\limsup_{n \to \infty} \frac{1}{n} \log_e P^n(C_n) \leq -r,$$

the fastest achievable asymptotic rate of decay of the probability of error is

$$\lim_{n \to \infty} -\frac{1}{n} \log_e \left[1 - P^n([C_n]_D) \right] = \mathcal{E}^*(r, D),$$

where

$$\mathcal{E}^*(r, D) \triangleq \inf_{Q \,:\, R_e(D;Q,P) > -r} H_e(Q\|P).$$

Although the exponent $\mathcal{E}^*(r, D)$ above is not as explicit as $(\frac{D^2}{2} - r)$ in (3), it is easy to evaluate numerically and it contains much more useful information. For example, Figure 1 shows the graph of $\mathcal{E}^*(r, D)$ as a function of r, for $D = 0.3$, P being the Bernoulli(0.4) distribution, and r running over the range $r \in (0.6109, 0.6393)$ where $\mathcal{E}^*(r, D)$ is nontrivial (i.e., finite and nonzero). In this case, (3) is only useful when $(\frac{D^2}{2} - r)$ is positive, i.e., for $r \in (0, 0.045)$: There (3) says that, whenever $P^n(C_n) \approx e^{-nr}$ for some $r \in (0, 0.045)$, the probability of error decays exponentially fast. But in that range, and in fact for all r up to ≈ 0.61, we have $\mathcal{E}^*(r, D) = \infty$ so there are sets C_n with $P^n(C_n) \approx e^{-nr}$ and probability of error decaying *super*-exponentially fast. Moreover, in the range $r \in (0.6109, 0.6393)$ where $\mathcal{E}^*(r, D)$ is nontrivial, we can choose C_n with $P^n(C_n) \approx e^{-nr}$ and $\Pr\{\text{"error"}\} \approx e^{-n\mathcal{E}^*(r,D)}$.
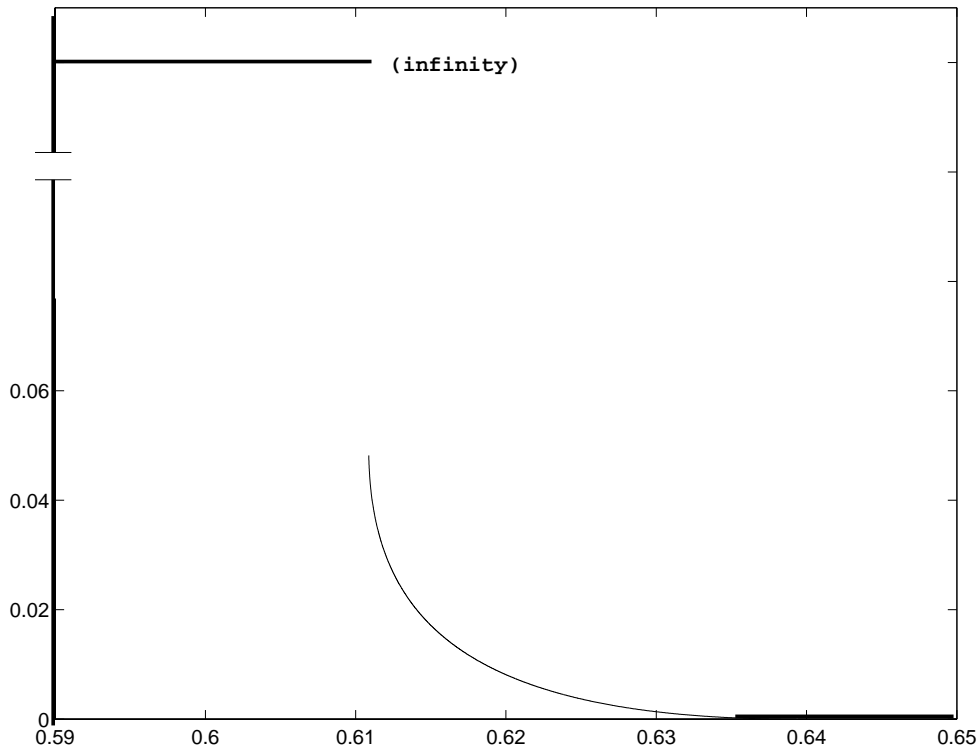

FIGURE 1. Graph of the error-exponent function $\mathcal{E}^*(r, D)$ in Corollary 3 as a function of r, for $D = 0.3$ and $P(1) = 0.4$. Note that $\mathcal{E}^*(r, D)$ is infinite for all $r \in (0, 0.6109)$, and that it is zero for $r > 0.6393$.

Finally we remark that the "extremal" sets in the classical isoperimetric problem, namely, those C_n that achieve equality in (3), are very different from the extremal sets in Corollary 3. The former are well-known to be Hamming balls B_n centered at $0^n = (0, 0, \ldots, 0) \in A^n$, $B_n = \{x_1^n \; : \; \rho_n(x_1^n, 0^n) \leq \delta\}$ (see [4][8, p. 174][10, Sec. 2.3]), while the latter are collections of strings y_1^n randomly selected from a collection of suitable strings.

Extensions.

1. Different alphabets. Although we assumed from the start that $\rho(x, y)$ is a distortion measure on $A \times A$, it is straightforward to generalize the main result as well as the subsequent discussion above to the case when $\rho(x, y)$ is a distortion measure between the "source" alphabet A and a different ("reproduction") alphabet $\hat{A}$, as long as it is still the case that for each $a \in A$ there exists a $b \in \hat{A}$ with $\rho(a, b) = 0$. The necessary modifications to the statements and proofs follow exactly as in the case of Marton's result; see [3, Sec. 2.4].

2. Strong converse. As mentioned earlier, the theorem is stated only for values of R above $R(D; P, M)$ since we trivially have $E(R, D) = 0$ for $R < R(D; P, M)$; see [6, Thm. 1]. In that range it is also possible to prove a "strong converse" showing that, not only $E(R, D) = 0$, but in fact the probability of error goes to one exponentially fast with a certain rate.

3. Proof

First we prove the *converse* part of the theorem, i.e., that $E(R, D) \leq E^*(R, D)$.

Note that the rate-function $R(D; P, M)$ defined in (4) is jointly uniformly continuous in $D \geq 0$ and P; this can be easily seen to be the case by arguing along the lines of the proof of [3, Lemma 2.2.2] for the rate-distortion function $R(D; P)$. Now let $\{C_n\}$ be an arbitrary D-covering with $\{C_n\} \in \mathcal{C}(R)$. Take any Q on A such that $R(D; Q, M) > R$ (if no such Q exists then the claim is trivially true), and let $\delta > 0$ be such that $R(D; Q, M) > R + \delta$. Since $\{C_n\} \in \mathcal{C}(R)$, we have $\log M^n(C_n) < n(R + \delta/2)$, eventually, and by the continuity of $R(D; Q, M)$ in D we can find an $\eta > 0$ small enough so that

$$\log M^n(C_n) < n(R + \delta/2) < nR(D + \eta; Q, M), \quad \text{eventually.}$$

Therefore, by the "weak converse" in [6, Thm. 1], we must also have

$$(5) \qquad E_{Q^n} \left[\min_{y_1^n \in C_n} \rho_n(X_1^n, y_1^n) \right] > D + \eta, \quad \text{eventually,}$$

where X_1^n denote n i.i.d. random variables with distribution Q^n. Writing

$$Z_n \stackrel{\triangle}{=} \min_{y_1^n \in C_n} \rho_n(X_1^n, y_1^n),$$

the bound in equation (5) implies that

$$D + \eta < E[Z_n] \leq D \, Q^n(Z_n \leq D) + D_{\max} Q^n(Z_n > D)$$

i.e.,

$$Q^n(Z_n > D) > \frac{\eta}{D_{\max} - D}.$$

From Stein's lemma [3, Cor. 1.1.2] we also know that, for any P and any $\epsilon > 0$,

$$\lim_{n \to \infty} \frac{1}{n} \log \left[\min_{B_n \subset A^n \, : \, Q^n(B_n) > \epsilon} P^n(B_n) \right] = -D(Q\|P).$$

Taking $\epsilon = \eta/(D_{\max} - D) > 0$ and applying this to the events

$$B_n \triangleq \{Z_n > D\} = [C_n]_D^c,$$

yields

$$\liminf_{n \to \infty} \frac{1}{n} \log \left[1 - P^n([C_n]_D) \right] \geq -D(Q\|P),$$

and since this holds for all Q with $R(D; Q, M) > R$, we obtain

$$\limsup_{n \to \infty} -\frac{1}{n} \log \left[1 - P^n([C_n]_D) \right] \leq E^*(R, D).$$

Finally, since the sequence $\{C_n\} \in \mathcal{C}(R)$ was arbitrary, this establishes that $E(R, D) \leq E^*(R, D)$, as required.

To prove the *direct* part of the theorem, asserting the existence of a D-covering $\{C_n\} \in \mathcal{C}(R)$ such that

$$\liminf_{n \to \infty} -\frac{1}{n} \log \left[1 - P^n([C_n]_D) \right] \geq E^*(R, D),$$

we follow the same outline as in the proof of the direct part of [3, Thm. 2.4.5].

Using the joint uniform continuity of $R(D; P, M)$ in $D \geq 0$ and P, the proof of the type-covering lemma [3, Lemma 2.4.1] can be generalized to the corresponding statement with $R(D; P, M)$ in place of $R(D; P)$. The main new observation here is that, since all the elements y_1^n of the covering set B are drawn from the set $T_{[Y^*]}^n$ of Y^*-typical strings, where (X^*, Y^*) achieve the infimum in the definition (4) of $R(D; P, M)$, their mass $M^n(y_1^n)$ satisfies

$$\frac{1}{n} \log M^n(y_1^n) \leq E[\log M(Y^*)] + \delta_n \left[\sum_y \log M(y) \right],$$

where the sequence $\delta_n \to 0$ as $n \to \infty$.

Finally, following the same steps as in the proof of the direct part of [3, Thm. 2.4.5] and replacing $R(D; P)$ by $R(D; P, M)$, we obtain the existence of a D-covering $\{C_n\} \in \mathcal{C}(R)$ with error exponent no worse than $E^*(R, D) - \delta$, where $\delta > 0$ is an arbitrary constant. This proves that $E(R, D) \geq E^*(R, D)$, and completes the proof. $\qquad \square$

Acknowledgments

References

[1] R. Ahlswede. Extremal properties of rate-distortion functions. *IEEE Trans. Inform. Theory*, 36(1):166–171, 1990.

[2] R.E. Blahut. Hypothesis testing and information theory. *IEEE Trans. Inform. Theory*, 20(4):405–417, 1974.

[3] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.

[4] L.H. Harper. Optimal numberings and isoperimetric problems on graphs. *J. Combinatorial Theory*, 1:385–393, 1966.

[5] W. Hoeffding. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.*, 36:369–408, 1965.

[6] I. Kontoyiannis. Sphere-covering, measure concentration, and source coding. *IEEE Trans. Inform. Theory*, 47:1544–1552, May 2001.

[7] K. Marton. Error exponent for source coding with a fidelity criterion. *IEEE Trans. Inform. Theory*, 20:197–199, 1974.

[8] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics (Norwich, 1989)*, pages 148–188. London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press, Cambridge, 1989.

[9] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Algorithms Combin., 16, Springer, Berlin, 1998.

[10] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, No. 81:73–205, 1995.

Division of Applied Mathematics, and Department of Computer Science, Brown University, 182 George Street, Providence, RI 02912, USA
E-mail address: `yiannis@dam.brown.edu`

Division of Applied Mathematics, Brown University, 182 George Street, Providence, RI 02912, USA
E-mail address: `ali_sezer@brown.edu`