On the Communication Exponent of Distributed Testing for Gaussian Correlations

Yuval Kochman
School of CSE, HUJI
Jerusalem, Israel
Email: yuvalko@cs.huji.ac.il

Ligong Wang
ETIS, CY Cergy Paris Université, ENSEA, CNRS
Cergy-Pontoise, France
Email: ligong.wang@ensea.fr

***Abstract*—This work addresses distributed binary hypothesis testing, where observations at two terminals are jointly Gaussian, each one standard, with two possible correlation coefficients. We assume that one of the terminals is colocated with the decision center, and focus on a single (Stein) error exponent. Rather than the traditional exponent that is defined with respect to the source blocklength, we assume the source data to be unlimited, and consider the error exponent as a function of the communication message length. We examine two different approaches, one by quantization and the other by sending the index of the maximum, and find them to yield the same exponent. We further find that binning improves upon both approaches in the same way. Finally we compare the obtained exponents to two upper bounds and determine the optimal exponent in some very special cases.**

I. Introduction

The goal of distributed hypothesis testing (DHT) is to distinguish between different possible joint distributions of data observed at several terminals, when the communication between them is constrained. We shall consider the popular side-information setting, where two terminals observe sequences X^k and Y^k, respectively, which are independent and identically distributed (i.i.d.) over the k time instances. The first ("encoder") creates an n-nat message $W = f(X^k)$, while the second ("decoder") makes a decision using some discriminant rule $g(W, Y^k)$. Further, we shall consider the Stein problem, where there are two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ regarding the joint distribution, where the error probability under $\mathcal{H}_0$ is only required to approach zero (possibly slowly), and one seeks the fastest decay of error probability under $\mathcal{H}_1$.

Let $p(\epsilon, n, k)$ denote the smallest attainable error probability under $\mathcal{H}_1$ when the source blocklength is k, the message contains no more than n nats, and the error probability under $\mathcal{H}_0$ is required to be no larger than ϵ. Traditionally, the error probability asymptotics are considered under fixed rates. That is, for some fixed $R > 0$, one is interested in

$$E(R) \triangleq \lim_{\epsilon \downarrow 0} \lim_{k \to \infty} -\frac{1}{k} \log p(\epsilon, \lfloor kR \rfloor, k). \tag{1}$$

Although this exponent is not known in general, many works derived bounds on it; we shall refer to some of them [1]–[5] in the sequel. The focus of this work is on a different exponent, inspired by the analysis of distributed parameter estimation by Hadar and Shayevitz [6]. We assume unlimited source data, and study the exponential decay of the error probability as a function of the message length, so the quantity of interest is the *communication exponent*:

$$E^{\mathrm{c}} \triangleq \lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \left(-\frac{1}{n} \lim_{k \to \infty} \log p(\epsilon, n, k) \right). \tag{2}$$

Since one can apply a fixed-rate scheme on unbounded source data, we have

$$E^{\mathrm{c}} \geq \lim_{R \downarrow 0} \frac{E(R)}{R}. \tag{3}$$

It is unclear whether equality should hold in the above or not.

In order for the communication exponent to be finite, the X- and Y-marginals of the joint distribution must not depend on the hypothesis. We consider in this work such a case, where each (X, Y) pair is zero-mean jointly Gaussian, and where the covariance matrix under $\mathcal{H}_i$ is given by

$$\Sigma_i = \begin{bmatrix} 1 & \rho_i \\ \rho_i & 1 \end{bmatrix}, \tag{4}$$

where $\rho_0 \in [-1, 1]$ and, without loss of generality, $\rho_1 \in [0, 1]$.

Deriving exponents for the Gaussian case requires more than substitution in known single-letter expressions for discrete-alphabet sources.[1] In Section II we present necessary preliminaries and a technical lemma, which is used extensively in this work.

In Section III we derive achievable exponents using two very different approaches. In the first we derive fixed-rate exponents for this source by extending well-known schemes for discrete-alphabet sources: the simple Ahlswede-Csiszár (AC) quantization scheme [1], and Han's improved quantization scheme [2]; in extending Han's scheme, we apply the above-mentioned technical lemma. Applying (3) to these bounds gives the following achievable communication exponents, respectively:

$$E_{\mathrm{AC}} = (\rho_1 - \rho_0)^2 \tag{5}$$

$$E_{\mathrm{Han}} = \frac{(\rho_1 - \rho_0)^2}{1 - \rho_1^2}. \tag{6}$$

In the second approach we apply to the DHT problem the technique used in [6] for correlation estimation, where an exponentially long source block is used, and the index of the

[1] This is similar to the case of error exponents for the Gaussian channel, which are not simple extensions of their discrete-channel counterparts.

maximum X within the block is sent. We find that the resulting exponent equals E_{Han}; we reason that the sharp concentration of the maximum mimics Han's use of improved ("fixed-type") quantization.

In Section IV, we add a binning element to the above approaches. We analyze the Gaussian version of the quantization-and-binning scheme of Shimokawa-Han-Amari (SHA) [3]. We also provide a (nontrivial) variant of the index-of-maximum scheme where the index is binned. Again both approaches yield the same communication exponent: for $\rho_0 \geq 0$,

$$E_{\text{SHA}} = \frac{(\rho_1 - \rho_0)^2}{(1 - \rho_1 \max(\rho_0, \rho_1))(1 - \rho_0 \min(\rho_0, \rho_1))}. \quad (7)$$

Finally, in Section V we compare E_{AC}, E_{Han}, and E_{SHA} to two upper bounds on E^{c}. The comparison reveals that beyond the case $\rho_1 = 0$, we have the exact value of E^{c} also for $\rho_0 \in \{0, 1\}$.

Due to the space limit, we shall focus on intuitive explanations to our approaches, and omit some technical details.

II. Preliminaries: Typicality of Gaussian Vectors

Typicality. Following [7], we define a spherical ϵ-type class of dimension k and power P as

$$\mathcal{T}_{k,P,\epsilon} = \left\{ x^k \in \mathbb{R}^k : \left| \frac{\|x^k\|^2}{kP} - 1 \right| \leq \epsilon \right\}, \quad (8)$$

where $\|\cdot\|$ denotes the Euclidean norm. We say that a vector $x^k \in \mathbb{R}^k$ is typical with respect to variance P if[2] $x^k \in \mathcal{T}_{k,P,\epsilon}$. We say that a vector pair (x^k, y^k) is jointly typical with respect to a covariance matrix Σ if any linear combination of the vectors is typical, i.e., if, for any $\boldsymbol{a} = (a_1, a_2)^T \in \mathbb{R}^2$, $a_1 x^k + a_2 y^k$ is typical with respect to variance $\boldsymbol{a}^T \Sigma \boldsymbol{a}$.[3]

We shall need the following large-deviation results on scalar amplitude asymptotics and dimension asymptotics.

Amplitude asymptotics. For parameters k, P, ϵ as above, define the interval

$$\mathcal{S}_{k,P,\epsilon} \triangleq [\sqrt{kP}(1-\epsilon), \sqrt{kP}(1+\epsilon)]. \quad (9)$$

Let X be a standard Gaussian variable. Let P be fixed, then

$$\lim_{\epsilon \downarrow 0} \lim_{k \to \infty} -\frac{1}{k} \log \Pr\{X \in \mathcal{S}_{k,P,\epsilon}\} = \frac{P}{2}. \quad (10)$$

This continues to hold when we replace X by $-X$.

Dimension asymptotics. Let the random vector X^k be i.i.d. standard Gaussian and fix $P > 0$, then

$$\lim_{\epsilon \downarrow 0} \lim_{k \to \infty} -\frac{1}{k} \log \Pr\{X^k \in \mathcal{T}_{k,P,\epsilon}\} = D_{\text{G}}(P), \quad (11)$$

where

$$D_{\text{G}}(P) \triangleq \frac{1}{2}[P - \log P - 1] \quad (12)$$

is the Kullback-Leibler divergence between zero-mean scalar Gaussian distributions of variances P and 1, respectively.

[2]This notion is of typicality is equivalent to the one in [8] with respect to a Gaussian density of mean zero and variance P.

[3]This definition is equivalent to requiring x^k and y^k be typical, and $\langle x^k, y^k \rangle$ be close to $k\text{E}[XY]$.

The results (10) and (11) are standard; we omit their derivations. The same exponents hold for the probability for $\|X^k\|^2$ to be above or below a threshold, e.g.,

$$\lim_{k \to \infty} -\frac{1}{k} \log \Pr\{\|X^k\|^2 \geq kP\} = D_{\text{G}}(P) \text{ if } P > 1. \quad (13)$$

We further need the following lemma concerning a mixture of an i.i.d. Gaussian vector and a deterministic vector. We call X^k a γ-*mixture* if

$$X^k = \sqrt{1-\gamma}\, Z^k + \sqrt{\gamma}\, v^k, \quad (14)$$

where $Z^k \in \mathbb{R}^k$ is i.i.d. standard Gaussian, and $v^k \in \mathcal{T}_{k,1,\epsilon}$.

Lemma 1: Fix $\gamma \in (0,1)$, and let $\{X^k, k \in \mathbb{Z}^+\}$ be γ-mixtures as in (14). Then

$$\lim_{\epsilon \downarrow 0} \lim_{k \to \infty} -\frac{1}{k} \log \Pr\{X^k \in \mathcal{T}_{k,P,\epsilon}\} = D(\gamma, P), \quad (15)$$

where

$$D(\gamma, P) \triangleq \frac{\gamma + P - A}{2\gamma} - \frac{1}{2} \log \frac{A - 1 + \gamma}{2\gamma}, \quad (16)$$

where $A = \sqrt{(1-\gamma)^2 + 4P\gamma}$. Furthermore,

$$\lim_{k \to \infty} -\frac{1}{k} \log \Pr\{\|X^k\|^2 \geq kP\} = D(\gamma, P) \text{ if } P > 1, \quad (17)$$

$$\lim_{k \to \infty} -\frac{1}{k} \log \Pr\{\|X^k\|^2 \leq kP\} = D(\gamma, P) \text{ if } P < 1. \quad (18)$$

Proof Method: The proof follows the technique used in [9] for deriving the Gaussian sphere-packing exponent. We decompose X^k into two parts: a component along the direction of v^k, and the remaining $k-1$ dimensions. Roughly speaking, we can then apply (10) to the former, and (11) to the latter. Finally, we find the (exponentially) most likely combination of the two components such that $\|X^k\|^2$ is around kP . ■

One may verify that $D(\gamma, P) \geq D_{\text{G}}(P)$,[4] with equality for $P = 1$ as well as in the limit $\gamma \to 0$. We shall need in the sequel the following fact: for P close to 1,

$$D(\gamma, P) = \frac{(P-1)^2}{4(1-\gamma^2)} + O((P-1)^3). \quad (19)$$

III. Quantization versus Index of Maximum

In this section and in the next one, we shall often use the following relationship between X and Y: under $\mathcal{H}_i$, $i \in \{0, 1\}$,

$$Y = \rho_i X + \sqrt{1 - \rho_i^2}\, Z, \quad (20)$$

where Z is standard Gaussian and independent of X.

A. Ahlswede-Csiszár-Style Exponent

In a classic work by Ahlswede and Csiszár [1] a simple scheme is proposed for DHT on discrete memoryless sources. The encoder's observation X^k is quantized at rate R to a sequence U^k. The decoder will declare $\mathcal{H}_0$ if and only if (U^k, Y^k) are jointly typical according to $\mathcal{H}_0$. It is assumed that the quantizer emulates in an i.i.d. manner the quantization test-channel between U and X.

[4]Intuitively, X^k as in (14) is "less random" than an i.i.d. Gaussian vector, thus its probability of atypical behavior is smaller than that of the latter.

A corresponding coding scheme in the Gaussian setting works as follows. The encoder generates quantization codewords $\{u^k(m), m = 1, \ldots, \lfloor e^{nR} \rfloor\}$ i.i.d. according to the zero-mean Gaussian distribution with variance

$$\sigma_R^2 \triangleq 1 - \exp\{-2R\}, \tag{21}$$

where R is the communication rate. Given x^k, it finds $u^k(m)$ which minimizes $x^k - u^k(m)$ and sends the index m. The decoder checks if $u^k(m)$ and y^k are jointly typical under $\mathcal{H}_0$. The obtained error exponent is (see also [5, Eq. (5)]):

$$E_{\text{AC}}(R) = \frac{1}{2} \log \left(\frac{1 - \rho_1^2 \sigma_R^2}{1 - \rho_0^2 \sigma_R^2} \right) - \frac{\rho_1(\rho_0 - \rho_1)\sigma_R^2}{1 - \rho_1^2 \sigma_R^2}. \tag{22}$$

We then obtain (5) by noting that, when R is small, $\sigma_R^2 \approx 2R$.

B. Han-Style Exponent

In [2], Han improves upon the Ahlswede-Csiszár scheme by noting that an i.i.d. quantization model is suboptimal, in the sense that it allows an atypical behavior of (U^k, X^k) to contribute to an error event in which, under $\mathcal{H}_1$, (U^k, Y^k) is typical with respect to $\mathcal{H}_0$. Thus, Han proposed to verify at the encoder that U^k, X^k are jointly typical. In the Gaussian case, we use the same idea to propose the following scheme. Consider the following joint distribution: U is zero-mean Gaussian with variance σ_R^2 as in (21); Q (the quantization noise) is standard Gaussian independent of U; and

$$X = U + \sqrt{1 - \sigma_R^2}\, Q. \tag{23}$$

- *Encoder:* Generate $\{u^k(m)\}$ i.i.d. according to the zero-mean Gaussian distribution of variance (21); discard those that are atypical. Given x^k, look for an index m such that x^k and $u^k(m)$ are jointly typical (in the sense of Section II) with respect to the covariance matrix of (U, X) as defined above. If successful, send m. If no such m can be found, send a special message to inform the decoder to declare $\mathcal{H}_1$.
- *Decoder:* If the special message is received, declare $\mathcal{H}_1$. Otherwise, check if $u^k(m)$ and y^k are jointly typical with respect to the joint distribution on (U, Y) computed from (U, X) as above, with Y given by (20) for $i = 0$, and with Z being independent of (U, X). If they are jointly typical, declare $\mathcal{H}_0$; otherwise declare $\mathcal{H}_1$.

One can verify that the error probability under $\mathcal{H}_0$ indeed vanishes. Under $\mathcal{H}_1$, the only error event is that the pair $(U^k(m), X^k)$ is "ρ_1-typical," but $(U^k(m), Y^k)$ is "ρ_0-typical." For this analysis, we consider a suboptimal rule that declares $\mathcal{H}_0$ whenever $S^k \triangleq Y^k - \eta\, U^k(m)$ is typical, where $\eta > 0$ will be specified later. By (20), under $\mathcal{H}_1$ we have:

$$S^k = \big(\rho_1 X^k - \eta\, U^k(m)\big) + \sqrt{1 - \rho_1^2}\, Z^k. \tag{24}$$

Recall that the encoder guarantees X^k and $U^k(m)$ to be jointly typical. Also note that, by (23),

$$\rho_1 X - \eta\, U = (\rho_1 - \eta)\, U + \rho_1 \sqrt{1 - \sigma_R^2}\, Q, \tag{25}$$

where U and Q are independent. We then obtain that $\|\rho_i X^k - \eta\, U^k(m)\|^2/k$ must be close to $\sigma^2_{\eta,R,1}$, where we define

$$\sigma^2_{\eta,R,i} \triangleq (\rho_i - \eta)^2 \sigma_R^2 + \rho_i^2(1 - \sigma_R^2), \quad i \in \{0, 1\}. \tag{26}$$

Under $\mathcal{H}_i$, S^k has a typical power of

$$P_{\eta,i}(R) = 1 - \rho_i^2 + \sigma^2_{\eta,R,i} = 1 + \eta(\eta - 2\rho_i)\sigma_R^2. \tag{27}$$

Thus, under $\mathcal{H}_1$, $\frac{S^k}{P_{\eta,1}(R)}$ is a $\frac{\sigma^2_{\eta,R,1}}{P_{\eta,1}(R)}$-mixture vector. We can then apply Lemma 1 to obtain that the probability of S^k being typical has exponent

$$E_\eta(R) = D\left(\frac{\sigma^2_{\eta,R,1}}{P_{\eta,1}(R)}, \frac{P_{\eta,0}(R)}{P_{\eta,1}(R)} \right). \tag{28}$$

The best error exponent achievable with this scheme is then $\sup_{\eta \geq 0} E_\eta(R)$. An explicit solution of the maximization is cumbersome. For a lower bound, we look at the regime where $\eta\sigma_R \gg 1$ while $\sigma_R^2 \ll 1$. In this regime,

$$\frac{\sigma^2_{\eta,R}}{P_{\eta,1}(R)} \approx 1 - \frac{1 - \rho_1^2}{\eta^2 \sigma_R^2} \tag{29a}$$

$$\frac{P_{\eta,0}(R)}{P_{\eta,1}(R)} \approx 1 + \frac{2}{\eta}(\rho_1 - \rho_0). \tag{29b}$$

Using (19) we then obtain (6).

Remark 1: Recall the decoder above examines typicality of $S^k = Y^k - \eta\, U^k(m)$. This is equivalent to examining whether $|\hat{\rho} - \rho_0| \leq \epsilon$, where

$$\hat{\rho} \triangleq \frac{1 + \eta^2 \sigma_R^2 - \|s^k(m)\|^2/k}{2\eta\sigma_R^2}. \tag{30}$$

This notation will allow easier comparison with the schemes that follow.

C. Exponent via Index of Maximum

Let X^k consist of i.i.d. standard Gaussian random variables, where $k = \exp n$, and let J be the index of the maximum component of X^k. The following is a well-known result in order statistics [10, Ex. 10.5.3]:

$$\mathrm{E}[X_J] = \sqrt{2n} + o(1) \tag{31a}$$

$$\mathrm{Var}[X_J] = o(1). \tag{31b}$$

Using (20), we have

$$Y_J = \rho_i(\sqrt{2n} + \tilde{X}) + \sqrt{1 - \rho_i^2}\, Z_J, \tag{32}$$

where $\tilde{X} \triangleq X_J - \sqrt{2n}$, and Z_J is standard Gaussian independent of $\tilde{X}$. In [6], Hadar and Shayevitz use this observation to study mean-squared error estimation of the correlation. It is almost immediate to extend the idea to the DHT problem, as we describe below.

The encoder finds the index J of the maximum in X^k and computes $\tilde{X}$. If $\tilde{X}$ is larger than a chosen positive constant ϵ, it sends a special message informing the decoder to declare $\mathcal{H}_1$; otherwise it conveys the index J. The decoder declares $\mathcal{H}_0$ if and only if it receives an index (i.e., not the special message) and $|\tilde{\rho} - \rho_0| \leq \epsilon$, where

$$\tilde{\rho} \triangleq \frac{Y_J}{\sqrt{2n}}. \tag{33}$$

One can verify that the error probability under $\mathcal{H}_0$ tends to zero as n grows large. Under $\mathcal{H}_1$,

$$\frac{Y_J - \sqrt{2n}\,\rho_1}{\sqrt{1-\rho_1^2}}$$

is "approximately" standard Gaussian. We can hence use (10) with

$$P = \frac{2(\rho_1 - \rho_0)^2}{1-\rho_1^2} \tag{34}$$

to recover the Han-style exponent (6). Notice that $\tilde{\rho}$ "plays the role" of $\hat{\rho}$ (30) of the quantization scheme.

One may ask why the index-of-maximum approach obtains E_{Han} rather than, say, the more straightforward E_{AC}. Intuitively, this is because the strong concentration of the maximal value (31) plays a similar role to the jointly typical (U^k, X^k) sequences in the Han-type scheme.

IV. Improvement by Binning

A. Shimokawa-Han-Amari-Style Exponent

In [3], Shimokawa et al. add a binning element to Han's scheme. We next analyze a Gaussian counterpart to this scheme. We claim without analysis that, whenever $\rho_0 < 0$, binning is not useful, so in the rest we assume $\rho_0 \geq 0$.

- *Encoder:* Fix some $\beta > 1$. Use a quantizer as in Han's scheme, but with rate βR instead of R. Accordingly, the sequences $\{u^k(m)\}$ are now generated with variance

$$\sigma_{\beta R}^2 = 1 - \exp\{-2\beta R\}. \tag{35}$$

  Randomly distribute the indices $\{1, \ldots, \lfloor e^{\beta R} \rfloor\}$ into $\exp\{(\beta-1)kR\}$ bins. If the encoder finds a unique m such that $u^k(m)$ and x^k are jointly typical, it sends the bin index (using rate R); otherwise it sends a special message informing the decoder to declare $\mathcal{H}_1$.
- *Decoder:* If the special message is received, declare $\mathcal{H}_1$. Otherwise, fix $\eta > 0$, and, for each m' in the bin, evaluate

$$\hat{\rho}(m') \triangleq \frac{1 + \eta^2\sigma_{\beta R}^2 - \|s^k(m')\|^2/k}{2\eta\sigma_{\beta R}^2}, \tag{36}$$

  where $s^k(m') \triangleq y^k - u^k(m')$. Declare $\mathcal{H}_0$ if, and only if, both the following hold.
  1) There exists a single $\hat{m}$ in the bin for which $|\hat{\rho}(\hat{m}) - \rho_0| < \epsilon$. This means that $s^k(\hat{m})$ is typical with respect to power

$$P_{\eta,0}(\beta R) = 1 + \eta(\eta - 2\rho_0)\sigma_{\beta R}^2. \tag{37}$$

  2) For all $m' \neq \hat{m}$ in the bin, $\hat{\rho}(m') < \rho_0$.

Under $\mathcal{H}_0$, with high probability the encoder will be able to find a proper m, and also with high probability $\hat{\rho}(m) \approx \rho_0$. It remains to analyze the probability that some $m' \neq m$ in the bin is such that $\hat{\rho}(m') \geq \rho_0$, i.e., that $\|s^k(m')\|^2 \geq kP_{\eta,0}(\beta R)$. Note that, for $m' \neq m$, $U^k(m')$ is generated independently of Y^k. Hence, by Lemma 1, the probability that a specific m' causes such an error has exponent

$$\bar{E}_\eta(\beta R) = D\left(\frac{\eta^2\sigma_{\beta R}^2}{1+\eta^2\sigma_{\beta R}^2}, \frac{P_{\eta,0}(\beta R)}{1+\eta^2\sigma_{\beta R}^2}\right). \tag{38}$$

Using the union bound, we can bound the exponent of the total probability of error of this kind by

$$\bar{E}_\eta(\beta R) - (\beta - 1)R. \tag{39}$$

In order to guarantee that the error probability under $\mathcal{H}_0$ be small, we need (39) to be positive, i.e., we need $\beta \leq \beta_{\max}$ where $\beta_{\max}$ is the (unique) root of (39).

We now evaluate the error exponent under $\mathcal{H}_1$. An error under $\mathcal{H}_1$ can occur only if at least one of the following two conditions holds:

1) The "correct" index m is such that $\hat{\rho}(m) \approx \rho_0$. As in Section III-B, the exponent is $E_\eta(\beta R)$, which is given by (28) with all R replaced by βR.
2) In the bin there exists $m' \neq m$ such that $\hat{\rho}(m') \approx \rho_0$ and $\hat{\rho}(m) < \rho_0$. The probability exponent of the former is the same as under $\mathcal{H}_0$ and is given by (39). The latter has probability close to one if $\rho_0 > \rho_1$, and has probability exponent $E_\eta(\beta R)$ if $\rho_0 < \rho_1$.

We conclude that the following exponent is achievable:

$$E_{\text{SHA}}(R) = \sup_{\eta \geq 0} \max_{0 \leq \beta \leq \beta_{\max}} \min\{E_\eta(\beta R),\, \bar{E}_\eta(\beta R) - (\beta - 1) + E_\eta(\beta R)\mathbb{1}_{\{\rho_0 < \rho_1\}}\}. \tag{40}$$

When $\rho_0 < \rho_1$, the minimum is always the first term, thus

$$E_{\text{SHA}}(R) = E_{\text{Han}}(\beta_{\max} R), \quad \rho_0 < \rho_1. \tag{41}$$

When $\rho_0 > \rho_1$, analytical expression of the optimal β is difficult to find.

Finally we consider the low-rate limit. It turns out that, in this regime, when η grows large, $\bar{E}_\eta(\beta R)$ as in (38) approaches its supremum, which is approximately $\rho_0^2 \beta R$. Accordingly,

$$\beta_{\max} \approx \frac{1}{1-\rho_0^2}.$$

When $\rho_0 < \rho_1$ we obtain

$$E_{\text{SHA}}(R) \approx \frac{(\rho_1 - \rho_0)^2}{1-\rho_1^2}\beta_{\max} R \approx \frac{(\rho_1 - \rho_0)^2}{(1-\rho_0^2)(1-\rho_1^2)} R. \tag{42}$$

When $\rho_0 > \rho_1$,

$$E_{\text{SHA}}(R) \approx \max_{0 \leq \beta \leq \beta_{\max}} \min\left\{\frac{(\rho_1 - \rho_0)^2}{1-\rho_1^2}\beta R, (\rho_0^2 - \beta + 1)R\right\}. \tag{43}$$

The maximum on the right-hand side is achieved by

$$\beta^* = \frac{1-\rho_1^2}{(1-\rho_0\rho_1)^2}, \tag{44}$$

thus

$$E_{\text{SHA}}(R) \approx \frac{(\rho_1 - \rho_0)^2}{(1-\rho_0\rho_1)^2} R. \tag{45}$$

This in turn yields (7).

B. Index of Maximum with Binning

It is natural to ask whether binning may help to improve upon the basic index-of-maximum approach of Section III-C. Indeed, in an estimation setting, Hadar et al. [11] apply a combination of index of maximum and binning. However, their scheme only works when the correlation coefficient is approximately known. We next present a scheme for DHT, where the correlation coefficient can take two different values.

Consider a source sequence X^k, where $k = e^{\beta n}$ for some $\beta > 1$. The indices $1, \ldots, e^{\beta n}$ are grouped into e^n bins each of length $e^{(\beta-1)n}$. Let m be the index of the maximum in X^k. The encoder checks if $X_m - \sqrt{2\beta n}$ is sufficiently small. If it is larger than a chosen positive constant, it sends a special message to inform the decoder to declare $\mathcal{H}_1$; otherwise it conveys the index of the bin containing m using n nats.

If the decoder receives a bin-index (i.e., not the special message), then, for each m' in the bin, it calculates

$$\tilde{\rho}(m') = \frac{Y_{m'}}{\sqrt{2\beta n}}. \tag{46}$$

It declares $\mathcal{H}_0$ if and only if the following two conditions hold.

1) There exists an index $\hat{m}$ with $|\tilde{\rho}(\hat{m}) - \rho_0| < \epsilon$.
2) For any $m' \neq \hat{m}$ in the bin, $\tilde{\rho}(m') < \rho_0$.

The analysis follows the same logic as for the SHA-style scheme where $\tilde{\rho}(m)$ replaces $\hat{\rho}(m)$. Notice that for the "correct" index m,

$$Y_m = \rho_i X_m + \sqrt{1-\rho_i^2} Z$$

is approximately Gaussian with mean X_m, while the other Ys are zero-mean Gaussian. Thus, one can use the amplitude asymptotics (9) on Y_m, which yield exponents for $\tilde{\rho}(m)$ that are exactly the same as those that $\hat{\rho}(m)$ satisfies in the low-rate limit. For example, by the union bound, the error probability under $\mathcal{H}_0$ is guaranteed to be small provided

$$\beta\rho_0^2 - (\beta - 1) \tag{47}$$

is positive, which is exactly the condition (39) in the low-rate limit, where $\bar{E}_\eta(\beta R) \approx \rho_0^2 \beta R$. Eventually, we obtain the same exponent (7) as in quantization with binning.

V. Upper Bounds and Comparison

We compare the above achievable communication exponents with two upper bounds. The first upper bound is adapted from Rahman and Wagner [4]. We note that the original upper bound of [4] is for a nonzero rate $R > 0$. It does not immediately yield an upper bound on the communication exponent by letting $R \downarrow 0$; recall that (3) does not necessarily hold with equality. Accordingly, our upper bound $\hat{E}_{\text{RW}}$ is obtained via a proof that is slightly different from [4]:

$$\hat{E}_{\text{RW}} = \frac{(\rho_1 - \rho_0)^2}{(1-\rho_1)^2}, \quad \frac{\rho_0 + 1}{2} \geq \rho_1 \geq 0. \tag{48}$$

The second upper bound is taken almost directly from Hadar et al. [5]:

$$\hat{E}_{\text{HLPS}} = \frac{(\rho_1 - \rho_0)^2}{(1 - \min(\rho_0, \rho_1))^2 - (\rho_0 - \rho_1)^2}, \quad \rho_0, \rho_1 \geq 0. \tag{49}$$

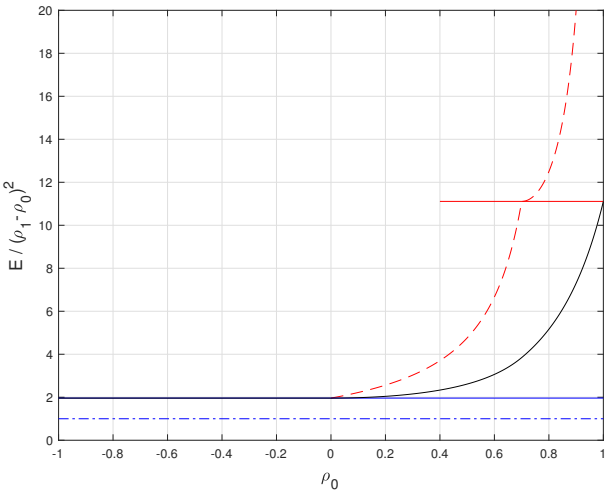

Fig. 1. Comparison of the different bounds as functions of ρ_0, for $\rho_1 = 0.7$. Notice that all bounds are normalized by $(\rho_1 - \rho_0)^2$. Lower bounds, starting from the lowest, are: E_{AC} in dash-dotted blue, E_{Han} in solid blue, and E_{SHA} in solid black. Upper bounds are: $\hat{E}_{\text{HLPS}}$ in dashed red (valid for $\rho_0 \geq 0$), and $\hat{E}_{\text{RW}}$ in solid red (valid for $\rho_0 \geq 0.4$).

Comparing the lower and upper bounds, we see that the communication exponent E^{c} is known for the following cases:

1) Testing against independence ($\rho_1 = 0$) : $E^{\text{c}} = \rho_0^2$;
2) Testing for independence ($\rho_0 = 0$): $E^{\text{c}} = \rho_1^2/(1-\rho_1^2)$;
3) Testing for equality ($\rho_0 = 1$): $E^{\text{c}} = \rho_1^2/(1-\rho_1)^2$.

In Figure 1 we plot the different lower and upper bounds on the communication exponent for a specific value of ρ_1. For better visibility, we normalize all bounds by $(\rho_1 - \rho_0)^2$. Notice that we have no upper bound for $\rho_0 < 0$.

References

[1] R. Ahlswede and I. Csiszar. Hypothesis testing with communication constraints. *IEEE Trans. Information Theory*, 32(4):533–542, Jul 1986.

[2] T. S. Han. Hypothesis testing with multiterminal data compression. *IEEE Trans. Information Theory*, 33(6):759–772, Nov 1987.

[3] H. Shimokawa, T. S. Han, and S. Amari. Error bound of hypothesis testing with data compression. In *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*, pages 114–, Jun 1994.

[4] M. Rahman and A. Wagner. On the optimality of binning for distributed hypothesis testing. *IEEE Trans. Information Theory*, 58(10):6282–6303, Oct 2012.

[5] U. Hadar, J. Liu, Y. Polyanskiy, and O. Shayevitz. Error exponents in distributed hypothesis testing of correlations. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2674–2678, 2019.

[6] U. Hadar and O. Shayevitz. Distributed estimation of gaussian correlations. *IEEE Transactions on Information Theory*, 65(9):5323–5338, 2019.

[7] E. Arikan and N. Merhav. Guessing subject to distortion. *IEEE Transactions on Information Theory*, 44(3):1041–1056, 1998.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[9] C. Swannack, U. Erez, and G. W Wornell. Reflecting on the AWGN error exponent. In *43rd Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois*, 2005.

[10] H. A. David and H. N. Nagaraja. *Order Statistics*. Wiley, Hoboken, NJ, 3rd edition, 2003.

[11] U. Hadar, J. Liu, Y. Polyanskiy, and O. Shayevitz. Communication complexity of estimating correlations. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 792–803, 2019.