

A new approach to distributed hypothesis testing

Gil Katz, Pablo Piantanida, Merouane Debbah

▶ **To cite this version:**

Gil Katz, Pablo Piantanida, Merouane Debbah. A new approach to distributed hypothesis testing. 50th Asilomar Conference on Signals, Systems and Computers, Oct 2016, Pacific Grove, United States. 10.1109/acssc.2016.7869599 . hal-01436813

HAL Id: hal-01436813
https://hal.science/hal-01436813

Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A New Approach to Distributed Hypothesis Testing

Gil Katz*, Pablo Piantanida† and Mérouane Debbah*‡
*Large Systems and Networks Group (LANEAS), CentraleSupélec-CNRS-Université Paris-Sud, Gif-sur-Yvette, France
†Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec-CNRS-Université Paris-Sud, Gif-sur-Yvette, France
‡Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France
{gil.katz,pablo.piantanida,merouane.debbah}@centralesupelec.fr

***Abstract*—A distributed hypothesis testing problem is considered, where the goal is to declare the distribution of two random variables, based on their observations. Defining two error events, the error exponent of Type II is studied under a fixed constraint over the error of type I. A novel approach is presented, based on random binning. The benefits of this approach are demonstrated through an example, compared to a more traditional approach, as well as to a different binned decoding method. These performance gains are then generalized to a large set of probability distributions.**

I. Introduction

The field of hypothesis testing (HT) focuses on the ability to declare the distribution of one or more random variables (RVs) by using n available observations. In single-variable HT, a statistician observes n realizations of a single RV X, before declaring if the governing distribution is $P_{0,X}(x)$ (hypothesis H_0) or $P_{1,X}(x)$ (hypothesis H_1). Two error events can occur: An error of the first type, with probability α_n, happens when the statistician declares H_1 while $P_{0,X}$ is the distribution that governs X, while the opposite event, with probability β_n, is defined to be an error of the second type. *Stein's Lemma* (see e.g. [1]) determines the optimal exponential rate of decay for the probability of error of the second type, under a fixed constraint over the probability of error of the first type ($\alpha_n \leq \epsilon$, $\epsilon > 0$), to be

$$\theta_n \triangleq -\lim_{n\to\infty} \frac{1}{n} \log \beta_n^\star = \mathcal{D}(P_0||P_1) \ , \tag{1}$$

where $\mathcal{D}(P_0||P_1)$ is the Kullback-Leiber (KL) divergence between the two probability distributions. Note that the optimal exponential rate of decay of β_n *does not depend* on the specific constraint over the probability of error of the first type α_n, in this case.

In this paper we consider a distributed HT scenario, where two nodes in a system are required to reach a decision. While a "helper" sees realizations of the RV X at node A, the statistician, at node B, observes realizations of Y (see Fig. 1 for a visual representation of the model). We assume that the helper is allowed to send information about its observations to the statistician with rate $R\left[\frac{\text{bits}}{\text{symbol}}\right]$. The statistician is then required to declare the probability distribution that governs the RVs, out of two possible options, namely $P_{0,XY}$ (under hypothesis H_0) and $P_{1,XY}$ (under hypothesis H_1).

A similar model for distributed HT problems was first presented and analyzed in [2]. An achievable error-exponent was proposed for an error of the second type, under a fixed constraint for the error of the first type, as discussed above. The problem of *testing against independence*, where $P_{1,XY}(x,y) = P_{0,X}(x)P_{0,Y}(y)$, $\forall (x,y) \in \mathcal{X} \times \mathcal{Y}$, was studied as a special case. Here, $P_{0,X}(x) = \sum_{y\in\mathcal{Y}} P_{0,XY}(x,y)$ is the marginal distribution of X according to hypothesis H_0, and equivalently for $P_{0,Y}(y)$. For this special case, the achievable error-exponent was proven to also be *optimal*. In [3] the case of testing against independence was revisited, focusing on joint problems of detection and source estimation. While optimality results remain allusive for the case of general hypotheses (see e.g., [2]), [4] presents an interesting achievable region through type-by-type analysis of the observed sequences at the two nodes (see [5] for a tutorial on the method of types). The idea of using a coding technique called "random binning" (see e.g., [6]) in order to improve these results was first briefly mentioned in [7], and then analyzed thoroughly in [8]. In [9] it was shown that binning can be optimal for a special case called *testing against conditional independence*.

The method of random binning is attractive for HT problems, as it allows the transmitter to send more accurate information while using less resources. However, as this method involves randomly assigning code sequences into groups ('bins') and allowing the encoder to only send the *group number* to the decoder, using it for problems of HT entails a risk: The value of "side information" used at the statistician's side is unknown before a decision has been made. In [8] this method results in a trade-off between the event where the wrong sequence is chosen from the bin and the event of an erroneous decision while using the right sequence. In this paper we propose a new method, by which HT is performed *by using the entire bin*, without sequence retrieval at the statistician's side. We show through an example that for HT problems, where the statistician is only interested in making a decision, this method can be very beneficial. We continue to show that these gains are in fact general, for an important group of HT problems.

The rest of this paper is organized as follows: In section II we present the system model, as well as the state of the art. Section III presents our main result, while an example of the performance gain is presented in Section IV through binary symmetric sources. Section V generalizes this result

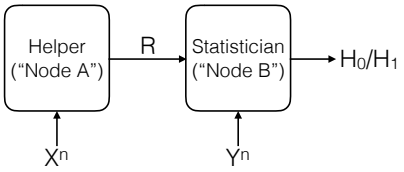


Fig. 1: Two-node detection model, with a unidirectional communication link

by showing that the gain in performance is significant in a large array of problems, before concluding remarks are given in Section VI.

II. System Model and Preliminaries

A. Notation

We use upper-case letters to denote RVs and lower-case letters to denote realizations of RVs. Vectors are denoted by bold-face letters. $Q_{\mathbf{x}}$ denotes the empirical distribution, or "type", of the vector $\mathbf{x}$. That is, for a vector $\mathbf{x}^n$ and a letter $a \in \mathcal{X}$ in the alphabet of X, $Q_{\mathbf{x}}(a) = \frac{N(a|\mathbf{x}^n)}{n}$, where $N(a|\mathbf{x}^n)$ is the number of times the letter a appears in the vector $\mathbf{x}^n$. The set of all vectors of length n in $\mathbf{x} \in \mathcal{X}^n$ with a specific type Q is denoted by $\mathcal{T}^n(Q)$, while the set of all vectors that are δ-typical is denoted by $\mathcal{T}^n_\delta(Q)$.

Using Csiszár's notation [5], we let $H(P) = \mathbb{E}\left[-\log P\right]$ denote the entropy of a RV distributed according to P, $\mathcal{D}(P_X||P'_X) = \sum\limits_{x\in\mathcal{X}} P_X(x)\log\left(\frac{P_X(x)}{P'_X(x)}\right)$ the KL divergence between the distributions P and P', and $I(P_X;P_{Y|X}) = I(X;Y) = \mathcal{D}(P_{XY}||P_XP_Y)$ the mutual information between X and Y while assuming that $P_XP_{Y|X}$ governs the pair. All logarithms and exponentials in this paper are base 2. Finally, $X \multimap Y \multimap Z$ signifies that the RVs X, Y and Z form a Markov chain.

B. Two-Node System Model

We consider a two-node system model, by which a helper and a statistician see n samples of the RVs X and Y, respectively. We assume that the RVs are jointly distributed according to one of two possible probability distributions

$$\begin{cases} H_0: & P_{0,XY}(x,y)\ , \\ H_1: & P_{1,XY}(x,y)\ . \end{cases} \tag{2}$$

Under each of these distributions, the variables are assumed to be jointly independent and identically distributed (i.i.d) in time. Furthermore, we assume that under each of these hypotheses, the marginal distribution of each of the RVs is identical. This assumption makes impossible for any node to make an independent decision, without communication with the other side.

Remark 1. *For the remainder of this paper, we often choose to view the system from the point of view of the helper. Thus, the system can be fully characterized by three RVs, $X \sim P_X(x)$, $Y_0 \sim P_{0,Y}(y)$ and $Y_1 \sim P_{1,Y}(y)$.*

Communication is allowed through a perfect link with rate R, measured in bits per symbol, from the helper to the statistician. Having received the sent information from the helper, the statistician makes a decision about the probability distribution that controls the RVs X and Y. In order to do so, he uses the received information, as well as his own observations $\mathbf{Y}^n$. In accordance to many previous works on the subject [2]–[4], [8], we consider two error probabilities, $\alpha_n \triangleq \Pr(H_1|XY \sim P_0(x,y))$ and $\beta_n \triangleq \Pr(H_0|XY \sim P_1(x,y))$, for errors of Type I and II, respectively. The performance of the system, for a given rate R, is measured by the exponential rate of decay of the probability of error of the second type β_n, when a fixed constraint is set over the probability of error of the first type $\alpha_n \leq \epsilon$ ($\epsilon > 0$). We denote this error exponent by E:

$$E(R,\epsilon) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log\beta_n(\epsilon)\ . \tag{3}$$

C. State of the Art

We bring forth two approaches, which would form the basis for comparison for our proposed approach, presented subsequently. An achievable scheme for the distributed HT system with two nodes and a unidirectional communication link, studied first in [2], was presented in [4] for the general case, which we will use extensively in this paper:

Proposition 1 ([4])**.** *In the case of distributed HT with two nodes, as seen in Fig. 1, and general hypotheses, the following is an achievable error exponent, as defined in* (3)*, as a function of R, and for any constraint ϵ:*

$$E(R) = \sup_{U\in\mathscr{S}(R)} \inf_{\tilde{U}\tilde{X}\tilde{Y}\in\mathscr{L}(U)} \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}||P_{UXY_1})\ , \tag{4}$$

where $\mathscr{S}(R) = \{U : I(U;X) \leq R,\quad U \multimap X \multimap Y\}$ and $\mathscr{L}(U) = \{\tilde{U}\tilde{X}\tilde{Y} : P(\tilde{U}\tilde{X}) = P(UX), P(\tilde{U}\tilde{Y}) = P(UY_0)\}$.

Intuitively, given a link with limited rate R, the transmitter is free to choose any strategy U such that its message is only dependent on X and can be sent over the given link. This is assured through limiting the choice of strategy U to the set $\mathscr{S}(R)$. Once a strategy has been chosen, the result is dependent on a *worst case* over the set $\mathscr{L}(U)$.

In [8] a new approach was offered, by which a binned codebook is used:

Proposition 2 (Performance improvement through binning [8])**.** *The following error exponent, as a function of R, is achievable for the two-node HT problem, for any two hypotheses, and any constraint ϵ:*

$$\begin{aligned} E(R) = \inf_{Q_X} \sup_{Q^\star_{U|X}} \inf_{Q_Y} \inf_{\substack{Q_{UXY} \\ Q_{U|X}=Q^\star_{U|X}}} \\ \Big\{ \min\big\{G[Q_{UXY},R], \\ \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathscr{L}(U)} \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}||Q^\star_{UX}P_{Y_1|X})\big\}\Big\}, \end{aligned} \tag{5}$$

where $\mathscr{L}(U)$ is as defined above and

$$\begin{aligned} G[Q_{UXY},R] = \min_{i\in\{0,1\}} \{\mathcal{D}(Q_{UXY}||P_{i,XY}Q^\star_{U|X}\} \\ + \left[R - I(Q_X;Q_{U|X}) + I(Q_Y;Q_{U|Y})\right]^+ \end{aligned} \tag{6}$$

if $I(Q_X;Q_{U|X}) > R$, and $G[Q_{UXY},R] = \infty$ otherwise. Here, $Q^\star_{U|X}$ is the chosen coding strategy of node A, and $Q^\star_{UX}$ is its joint probability with the source X, $Q^\star_{UX} = Q^\star_{U|X}P_X$.

The trade-off induced by the use of a binning approach can be identified in (5) through the minimum operator, imposed over $G[Q_{UXY},R]$ and a second function, which resembles the achievable result in Proposition 1. The function G represents the exponential probability of error in identifying the correct sequence in the bin, while the second function in (5) represents the error event where a faulty detection is made over the correct sequence from the bin.

III. Main Result

We now improve upon the result of [8] by proposing a new decoding strategy. While binning is still performed at the helper's side according to this new strategy, the receiver does not try to recover the original sequence from the bin. Instead, it uses the entire bin in order to make its decision, as explained subsequently.

Proposition 3 (New achievable error exponent)**.** *The following error exponent, as a function of R, is achievable for the two-node HT problem for any two hypotheses, and any constraint ϵ, when only node A is allowed to transmit:*

$$E(R) = \sup_{Q^\star_{U|X}} \Big\{ \min\big\{\hat{G}[Q_{UXY},R], \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathscr{L}(U)} \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}||Q^\star_{UX}P_{Y_1|X})\big\}\Big\} , \tag{7}$$

where $\hat{G}[Q_{UXY},R] = R - I(P_X;Q^\star_{U|X}) + I(P_{Y_0};Q^\star_{U|Y_0})$ and $\mathscr{L}(U)$ is as defined in Proposition 1.

In the following we give a proof outline for Proposition 3. We start by describing the codebook generation, as well as the encoding and decoding strategies. We then explain the general idea behind the calculation of the error exponent, and calculate $\hat{G}[Q_{UXY},R]$ explicitly.

Proof. Codebook Generation: We build the codebook only for types of $\mathbf{X}^n$, Q_X, such that $\mathcal{T}^n(Q_X) \subset \mathcal{T}^n_\delta(X)$, for some arbitrarily small $\delta > 0$. For each of these types set a unique index $k(Q_X)$. Then, fix a conditional type $Q^\star_{U|X}(Q_X)$. Randomly and uniformly choose a set of codewords, $\mathcal{C}^n_U(Q_X)$, from the resulting marginal type class $\mathcal{T}^n_{Q^\star_U}$, induced by Q_X and $Q^\star_{U|X}(Q_X)$. The size of $\mathcal{C}^n_U(Q_X)$ is an integer satisfying $\mathcal{C}^n_U(Q_X) \doteq \exp^{nI(Q_X;Q^\star_{U|X}(Q_X))}$, where $\doteq$ signifies equality in the exponent. Define $f_U : \mathcal{T}^n(Q_X) \to \mathcal{C}^n_U(Q_X)$. The function $f_U(\mathbf{x})$ determines the codeword that would be chosen by the encoder ("node A") in order to describe $\mathbf{x}$, and sent to the decoder ("node B"), as subsequently explained. We define $\mathbf{U}^n \triangleq f_U(\mathbf{X}^n)$.

Encoding: Given a sequence $\mathbf{x} \in \mathcal{T}^n(Q_X)$ node A first checks if $\mathcal{T}^n(Q_X) \subset \mathcal{T}^n_\delta(X)$. If not, it sends an error message. Otherwise, it searches for a sequence $\mathbf{u}_i$ in the codebook that belongs to the type of $\mathbf{x}$, such that $(\mathbf{u}_i,\mathbf{x}) \in \mathcal{T}^n_\mu(UX)$, with μ being some arbitrarily small value that satisfies $\mu > \delta$. The encoder sends the type of $\mathbf{x}$, $k(Q_\mathbf{x})$, with zero rate, as well as an index, that is determined by the chosen codeword $\mathbf{u}$, $F(f_U(\mathbf{x}))$. The nature of this index is dependent on the type of the observed sequence $\mathbf{x}$. There are two cases to consider:

1. $\log|\mathcal{C}^n_U(Q_X)| < nR$, in which case we can assign a unique index for each member of the codebook.
2. $\log|\mathcal{C}^n_U(Q_X)| \geq nR$, in which case we assign each distinct member of $\mathcal{C}^n_U(Q_X)$ to a message index uniformly at random. This step is called the binning process.

Let $F(f_U(\mathbf{x}))$ denote the element to which $f_U(\mathbf{x})$ is mapped. The encoder can be expressed mathematically as

$$\Psi(\mathbf{x}) = (F(f_U(\mathbf{x})), k(Q_\mathbf{x})) \ . \tag{8}$$

Decoding: In case an error message is received, the decoder declares H_1. The probability of this event, however, goes to zero when $n \to \infty$ thanks to the asymptotic equipartition property (AEP, see e.g., [6]) and the size of the codebook. When the encoder does not send an error message, the decoder operates on the entire bin in order to make a decision. Going over the sequences in the bin one by one, the decoder checks for each $\mathbf{u}_i$ if $(\mathbf{u}_i,\mathbf{y}) \in \mathcal{T}^n_\mu(UY_0)$. If a sequence in the bin is found, which is jointly typical with $\mathbf{y}$, the decoder declares H_0. If no such sequence is found, the decoder declares H_1.

Analysis of the Probability of Error: We now give a brief outline for the analysis of the probability of error. As showing that $\alpha_n \to 0$ when $n \to \infty$ is straight-forward, we neglect this part of the proof and concentrate on analyzing the exponential rate with wich $\beta_n \to 0$. When analyzing β_n, it is assumed throughout that the probability distribution that controls X and Y is the one implied by H_1. Node B makes an error of the Type II if it declares H_0. This can be the result of one of two events:

$$\begin{aligned} \mathcal{B}_1 &= \big\{(f(\mathbf{x}),\mathbf{y}) \in \mathcal{T}^n_\delta(UY_0)\big\} \ , \\ \mathcal{B}_2 &= \big\{(\exists i \in F(f(\mathbf{x})) : \mathbf{u}_i \neq f(\mathbf{x}), \\ &\qquad (\mathbf{u}_i,\mathbf{y}) \in \mathcal{T}^n_\delta(UY_0)\big\} \ . \end{aligned} \tag{9}$$

Here, the first event represents the case where the actual sequence node A intended to send is jointly typical with the observed sequence $\mathbf{y}$ according to H_0, despite the fact that we assume the probability distribution implied by H_1 controls the RVs. The second event represents a case where a different sequence in the bin "confuses" node B. The first event was analyzed in [8], showing that the result of [4] is applicable also when using random codebooks, and thus, for a given strategy $Q^\star_{U|X}$,

$$\lim_{n\to\infty} -\frac{1}{n}\log\Pr(\mathcal{B}_1) \geq \min D(P_{\tilde{U}\tilde{X}\tilde{Y}}||Q^\star_{UX}P_{Y_1|X}) \ , \tag{10}$$

where the minimum is over $\tilde{U}\tilde{X}\tilde{Y} \in \mathscr{L}(U)$. We thus concentrate on the analysis of the error exponent of event $\mathcal{B}_2$.

We start by presenting the following lemma,without proof:

Lemma 1. *Let $\mathcal{A}$ be the set of triplets, such that a binned codebook is necessary:*

$$\mathcal{A} = \{(\mathbf{u},\mathbf{x},\mathbf{y}) : \mathbf{u} \in \mathcal{T}^n_{Q^\star_{\mathbf{U}|\mathbf{X}}}, \log|\mathcal{C}^n_U(Q_{\mathbf{X}})| \geq nR\} \ . \tag{11}$$

Let $(\mathbf{u},\mathbf{x},\mathbf{y}) \in \mathcal{A}$. Let $\mathcal{B}_2$ be the event that $(\mathbf{u},\mathbf{y}) \in \mathcal{T}^n_\mu(UY_0)$, for some $\mathbf{u} \neq f(\mathbf{x})$ in the bin. Then

$$\begin{aligned}&\Pr\left(\mathcal{B}_2 | \mathbf{U^n} = \mathbf{u}, \mathbf{X}^n = \mathbf{x}, \mathbf{Y}^n = \mathbf{y}\right) \\ &\quad \leq \exp^{-n\left(R - \hat{J}(Q_{\mathbf{uxy}}) - \delta_n\right)} \ ,\end{aligned} \tag{12}$$

with

$$\hat{J}(Q_{\mathbf{uxy}}) \triangleq I(Q_{\mathbf{x}}; Q^\star_{U|X}(Q_{\mathbf{x}})) - H(Q_{\mathbf{u}}) + H(Q_{U|Y_0}|P_{Y_0}) \tag{13}$$

and δ_n goes to 0 with n. The probability in (12) is taken over the choice of the codebook in use.

Using Lemma 1 and summing over types and over sequences within each type, the probability of the event where an unintended sequence in the bin causes an error can be bounded by

$$\begin{aligned}&\lim_{n\to\infty} -\frac{1}{n}\log\Pr(\mathcal{B}_2) \geq \\ &= \min_{Q_X}\max_{Q^\star_{U|X}}\min_{Q_Y}\min_{Q_{UXY}}\left\{\mathcal{D}(Q_{UXY}||P_{1,XY}Q_{U|X})\right. \\ &\quad \left. + R - I(Q_X; Q^\star_{U|X}) + I(Q_{U|Y_0}; P_{Y_0})\right\} \ .\end{aligned} \tag{14}$$

As we only work with δ-typical $\mathbf{x}$ sequences, we may choose δ to be any value, as long as it is strictly positive. Thus, we may force Q_X to be arbitrarily close to P_X by taking $\delta \to 0^+$. The error exponent in question thus becomes

$$\begin{aligned}&\lim_{n\to\infty} -\frac{1}{n}\log\Pr(\mathcal{B}_2) \\ &\quad \geq \max_{Q^\star_{U|X}}\left\{R - I(P_X; Q^\star_{U|X}) + I(Q_{U|Y_0}; P_{Y_0})\right. \\ &\qquad \left. + \min_{Q_Y}\min_{Q_{UXY}}\left(\mathcal{D}(Q_{UXY}||P_{1,XY}Q^\star_{U|X})\right)\right\} + \hat{\epsilon} \\ &\quad = \max_{Q^\star_{U|X}}\left\{R - I(P_X; Q^\star_{U|X}) + I(Q_{U|Y_0}; P_{Y_0})\right\} + \hat{\epsilon} \ ,\end{aligned} \tag{15}$$

with $\hat{\epsilon} \to 0$ as $\delta \to 0$. Here, the last equality is true since we are now free to minimize the KL divergence independently of the maximized expression before it, which is set once $Q^\star_{U|X}$ is chosen. Thus, it can always be brought to its minimal value, which is zero. This, along with an analysis of the complementary error event similar to the one given in [8], completes the proof. The maximization over $Q^\star_{U|X}$ in Proposition 3 stems from the fact that the encoder can choose its strategy freely. □

IV. Binary Symmetric Sources

As an example, consider the following sources:

$$X = \text{Bern}\left(\frac{1}{2}\right), \quad \begin{cases} Y_0 = X + Z_0 \ , \\ Y_1 = X + Z_1 \ , \end{cases} \tag{16}$$

where $Z_0 = \text{Bern}(p)$ and $Z_1 = \text{Bern}(q)$, and we assume $0 \leq p < q \leq \frac{1}{2}$. For this case, we draw results under the assumptions that $Q^\star_{U|X}$ is chosen to be a binary symmetric channel (BSC) from X to the auxiliary RV U, with crossover probability δ. This of course may be a suboptimal choice. However, we will see that performance gains may be significant, even under this limiting choice. Performance results for this case are presented in Fig. 2. The curve entitled 'hypothesis testing error exponent' depicts the exponential probability of error under the assumptions detailed above, while using the right sequence at the decoder. Using the approach offered by [4], as summarized in Proposition 1, performance is limited by the available rate. This is depicted in Fig. 2 by the dashed line entitled 'hypothesis testing without binning'. It is important to emphasize that this line is presented as a lower bound over the performance of the binned approaches, and is not a function of δ.

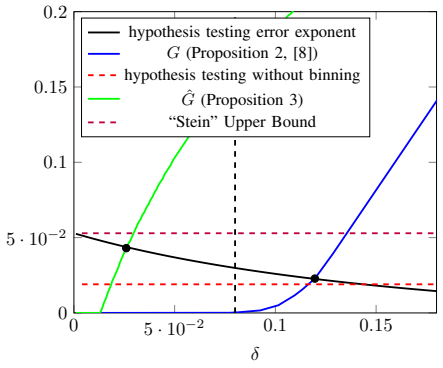


Fig. 2: Error exponents for both error events in the BSC case with $p = 0.1$, $q = 0.2$ and $R = 0.4$, under the strategies implied by Propositions 2 and 3.

As was explained in previous sections, the results attained by each of the binned approaches in Proposition 2 and Proposition 3 are each a result of a trade-off between two error events. The curve representing the function $G[Q_{UXY}, R]$, in blue, refers to the event that, under the approach of Proposition 2, the wrong sequence is chosen from the bin by the decoder. For this case, choosing $\delta \simeq 0.12$ results in the best performance offered by this approach, under our presumptions, detailed above. The curve representing $\hat{G}[Q_{UXY}, R]$, in green, refers to the event where, while going through all the sequences in the bin, as implied by the approach of Proposition 3, the decoder is confused by a sequence, which is not the one the encoder intended to send. Optimal results under this approach are attained by choosing $\delta \simeq 0.035$. It can be seen that, at lest for this case, Proposition 3 offers superior performance to the approach presented in [8].

V. Assessing the Gain in Performance

In this section we show that the performance gain shown for the specific example of binary symmetric sources in Section IV is in fact general for many cases. In order to do so, we choose to examine a "cross-section" of the performance gain, at the point where $R = I(U;X|Y_1)$. This cross-section is illustrated for the BSC example in Fig. 2 by a black dashed line. We choose this point because of its

importance to problems where both detection and source-estimation are required at the receiver (see e.g., [3], [10]). It can be seen in Fig. 2, that it is at this point that the curve of $G[Q_{UXY}, R]$ leaves 0. This is in fact general for all cases, and is implied by the decoding approach of Proposition 2, where a single sequence *must* first be chosen, before detection is performed.

In the example presented above, at the same point, the curve for $\hat{G}[Q_{UXY}, R]$ is above the one representing the 'hypothesis testing error exponent' while using the intended sequence (seen in black in Fig. 2). This implies that at this point, performance is *not limited* by the binning approach. We now check if this observation is true in general:

$$\begin{aligned}
&\left[\hat{G}[Q_{UXY}, R] - \mathcal{D}(P_{UY_0}||P_{UY_1})\right]\Big|_{R=I(U;X|Y_1)} \\
&= [R - I(U;X) + I(U;Y_0) - \mathcal{D}(P_{UY_0}||P_{UY_1})]|_{R=I(U;X|Y_1)} \\
&\overset{(a)}{=} [R - I(U;X|Y_0)\ \mathcal{D}(P_{UY_0}||P_{UY_1})]|_{R=I(U;X|Y_1)} \\
&= I(U;X|Y_1) - I(U;X|Y_0) - \mathcal{D}(P_{UY_0}||P_{UY_1}) \\
&= I(U;XY_1) - I(U;Y_1) - I(U;XY_0) \\
&\qquad + I(U;Y_0) - \mathcal{D}(P_{UY_0}||P_{UY_1}) \\
&\overset{(b)}{=} I(U;Y_0) - I(U;Y_1) - \mathcal{D}(P_{UY_0}||P_{UY_1}) \triangleq (*)\ .
\end{aligned} \tag{17}$$

Here, (a) stems from the Markov chain $U \multimap X \multimap Y_0$, while (b) stems from the same Markov chain, as well as $U \multimap X \multimap Y_1$. Through the chain rule for KL divergence we get that $(*) = H(Y_1|U) - H(Y_0|U) - \mathcal{D}(P_{Y_0|U}||P_{Y_1|U}|P_U)$ where $\mathcal{D}(P_{Y_0|U}||P_{Y_1|U}|P_U)$ is the *conditional* KL-divergence. In order to check if the performance at this point is not limited by the binning approach of Proposition 3, we would like to check if this expression is positive, or equivalently if $H(Y_1|U) - H(Y_0|U) \geq \mathcal{D}(P_{Y_0|U}||P_{Y_1|U}|P_U)$. This is a conditional version of Theorem 3 in [11]. A sufficient (but not necessary) condition for this inequality to hold is thus that Y_1 is *majorized* by Y_0, for any choice of U:

Definition 1 ([11])**.** *Consider discrete probability distributions $P = \{p_i\}$ and $Q = \{q_i\}$ defined on the positive integers labeled in decreasing probabilities, i.e.,*

$$p_i \geq p_{i+1}, \quad q_i \geq q_{i+1}\ . \tag{18}$$

Q is majorized by P if for all $k = 1, 2, \ldots$

$$\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i\ . \tag{19}$$

Lemma 2 ([11])**.** *If Q is majorized by P, then*

$$H(Q) - H(P) \geq D(P||Q)\ . \tag{20}$$

When considering the conditional case, as we are required to do here, it is enough to verify the majorization condition in Lemma 2 for the average of Y_0 (respectively, Y_1) over U. Nevertheless, we will restrict ourselves further by demanding that $(Y_1|X = x)$ is majorized by $(Y_0|X = x)$ for any $x \in \mathcal{X}$. In such a case, there will always be a $U \in \mathscr{S}(R)$ (not necessarily unique) that achieves the maximum over $\mathscr{S}(R)$ in Proposition 3, and such that the majorization constraint holds. Thus, performance is not limited by our proposed binning approach of Proposition 3, for any setting that complies with this condition, at our chosen reference point $R = I(U;X|Y_1)$. Comparing this to the approach in Proposition 2, where at the same reference point binning reduces the error exponent of interest to zero, the benefits of the approach presented in Proposition 3 are clear.

Remark 2. *While at a first glance enforcing the majorization condition for each $x \in \mathcal{X}$ might seem unnecessarily strict, in fact it still includes many interesting problems, including settings in which H_0 and H_1 imply the same channel from X to Y, with the difference that the channel implied by H_1 is noisier. This is in fact the case of the BSC example above.*

VI. Concluding Remarks

In this paper, a novel approach to the distributed HT problem was presented through the use of binned codebooks, without sequence retrieval at the decoder. Comparing our results to the state of the art, it was shown that performance gains may be significant. Moreover, it was shown that significant gains are assured in a wide array of interesting problems. This was achieved by showing that at a point where previous methods are highly restricted, the proposed approach is not limited by its binned codebook.

References

[1] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, ser. Springer Texts in Statistics.

[2] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *Information Theory, IEEE Transactions on*, vol. 32, no. 4, pp. 533–542, Jul 1986.

[3] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "Joint estimation and detection against independence," in *Proc. of the 52-nd Annual Allerton Conference on Coomunication, Control and Computing*, 2014.

[4] T. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, Nov 1987.

[5] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct 1998.

[6] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley & Sons, 1991.

[7] H. Shimokawa, T. Han, and S.-I. Amari, "Error bound of hypothesis testing with data compression," in *Inf. Theory, 1994 IEEE International Symposium on (ISIT)*, Jun 1994, p. 114.

[8] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "On the necessity of binning for the distributed ht problem," in *Inf. Theory, 2015 IEEE International Symposium on (ISIT)*, Jun 2015.

[9] S. Rahman and A. Wagner, "On the optimality of binning for distributed hypothesis testing," *Information Theory, IEEE Transactions on*, vol. 58, no. 10, pp. 6282–6303, Oct 2012.

[10] G. Katz, P. Piantanida, and M. Debbah, "Distributed binary detection with lossy data compression," *arXiv preprint arXiv:1601.01152*, 2016.

[11] S.-W. Ho and S. Verdu, "On the interplay between conditional entropy and error probability," *Information Theory, IEEE Transactions on*, vol. 56, no. 12, pp. 5930–5942, Dec 2010.