

On the Necessity of Binning for the Distributed Hypothesis Testing Problem

Gil Katz, Pablo Piantanida, Romain Couillet, Merouane Debbah

► **To cite this version:**

Gil Katz, Pablo Piantanida, Romain Couillet, Merouane Debbah. On the Necessity of Binning for the Distributed Hypothesis Testing Problem. IEEE International Symposium on Information Theory, Jun 2015, hong kong, Hong Kong SAR China. 10.1109/isit.2015.7282966 . hal-01242524

HAL Id: hal-01242524

<https://hal.archives-ouvertes.fr/hal-01242524>

Submitted on 14 Dec 2015

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.

On the Necessity of Binning for the Distributed Hypothesis Testing Problem

Gil Katz, Pablo Piantanida, Romain Couillet and Mérouane Debbah

Laboratoire des Signaux et Systèmes (L2S)

CentraleSupélec-CNRS-Université Paris-Sud

Gif sur Yvette, France

Email: {gil.katz, pablo.piantanida, romain.couillet, merouane.debbah}@centralesupelec.fr

Abstract—A distributed hypothesis testing (HT) problem is considered, comprising two nodes and a unidirectional communication link. The receiving node is required to make a decision as to the probability distribution in effect. A binning process is used in order to minimize the probability of error, resulting in a new achievable error-exponent. A sub-class of HT problems with general hypotheses is defined, which contains many interesting and relevant problems. The advantage of the binning strategy in comparison to the non-binning approach is demonstrated by means of a binary symmetric example.

I. INTRODUCTION

The subject of hypothesis testing (HT) is a familiar one in statistics. In HT problems, the interest lies in the ability to make a decision as to the probability distribution (or “law”) of a certain random variable (RV) X , based on n available realizations. Typically, two possible “options” are known, named H_0 and H_1 , and implying probability distributions $P_0(x)$ and $P_1(x)$, respectively. Two error events can be defined. We name the event that P_1 is declared the probability distribution in effect while P_0 is true an error of the first type (with probability α_n), while the opposite event is named an error of the second type (with probability β_n). *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

$$\theta_n \triangleq - \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^* = D(P_0 || P_1). \quad (1)$$

Notice 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 distributed HT problems (see e.g. [2], [3]) different RVs are assumed to be sampled in different locations (“nodes”) of the system. Different communication constraints can be imposed on the system, implying that the entirety of the information can not be obtained in a single place, and thus that Stein’s Lemma can not be directly applied. Throughout this paper we will focus on the system depicted in Figure 1, comprising two nodes and a single one-directional rate-limited link. In [4], the same system was studied under the assumption of *testing against independence*, where the

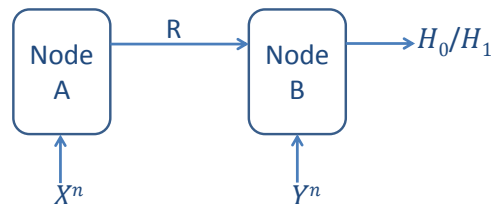


Fig. 1: Two-node detection model, with a unidirectional communication link

alternative hypothesis H_1 consists of X and Y being independent, while maintaining the same marginal distributions implied by H_0 (i.e., $P_{1,XY}(x,y) = P_{0,X}(x)P_{0,Y}(y)$, $\forall x \in \mathcal{X}, y \in \mathcal{Y}$). An optimal region was proved for the *joint detection and lossy compression* problem, by which when node B detects that the sources are dependent (i.e. H_0 is the true hypothesis) it also attempts to estimate the realizations of X with maximal average distortion D .

The authors in [5] investigate HT in a multi-node system. It is shown that in the case that under H_1 the sources are *conditionally independent* given some additional information Z , optimality can be achieved through the use of a *binned codebook*. The achievable error exponent (which then turns out to also be optimal) is the result of a trade-off between the exponential rates of decay of the probabilities of two different error events. This tension between different error events will be key in this paper as well.

In this paper, we concentrate on HT with general hypotheses. An achievable error exponent was proposed for this setting in [3]. Naturally, the performance is limited by the available rate. Our approach improves upon the achievable error-exponent proposed in [3] by using a binned codebook. Although binning for distributed HT problems with general hypotheses was first mentioned previously in [6], no proof of the claimed result is available in literature. Binning allows for a codebook containing more sequences, making a “closer” description of the observed source sequence more likely. However, as the binning process introduces to the system a new type of error event which is nonexistent in the non-binned approach (namely, the event where the wrong sequence is picked from the bin at the decoder’s side), it is still unclear whether this approach induces any performance gain. After establishing the achievable error-exponent in the

This research has been supported by the ERC Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering).

system by using the binning approach and technical tools inspired by [7], we present a specific class of problems we call “HT with degraded hypotheses”, and show through a specific example, that the binning approach does lead to performance gain in distributed HT. This is, however, a general result, as explained subsequently.

The rest of this paper is organized as follows: In section II the system model is presented. Section III gives our main result - an achievable error exponent for the two-node HT problem with general hypotheses. Section IV gives the codebook construction, coding and decoding strategies, as well as an outline of the calculation of the probability of error of the second type, under the suggested achievable scheme. In Section V the class of HT problems “with degraded hypotheses” is introduced. An example for a problem in this class is given, where the source is assumed to be binary and symmetric. Finally, Section VI gives our concluding remarks.

II. SYSTEM MODEL AND PRELIMINARIES

A. Notation

We use upper-case letters to denote random variables (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} . The set of all vectors of length n in $\mathbf{x} \in \mathcal{X}^n$ with a specific type Q is denoted by $T^n(Q)$, while the set of all vectors that are δ -typical (in the usual entropy sense) is denoted by $T_\delta^n(Q)$. Using Csiszár’s notation [8], we let $H(P)$ denote the entropy of a RV distributed according to P , $I(P_X; P_{Y|X})$ the mutual information between X and Y while assuming that $P_X P_{Y|X}$ governs the pair, and $D(P||P')$ the KL divergence between the distributions P and P' .

B. Two-Node System Model

We consider a distributed system comprised of two nodes, as depicted in Fig. 1. Node A and node B see n realizations of the RVs X and Y , respectively. X and Y are jointly distributed in one of two possible ways, $P_{0,XY}(x, y)$ and $P_{1,XY}(x, y)$, and are assumed to be discrete and i.i.d under each of these options. We further assume that the marginal distributions of both RVs are identical under each of the hypotheses, i.e.,

$$\begin{aligned} P_{0X}(x) &= P_{1X}(x) \quad \forall x \in \mathcal{X} , \\ P_{0Y}(y) &= P_{1Y}(y) \quad \forall y \in \mathcal{Y} . \end{aligned} \quad (2)$$

As the encoder only sees X , for the remainder of this paper we often take the encoder’s point of view and define three RVs in the system - X , Y_0 and Y_1 , with probability distributions $P_X(x)$, $\sum_{x \in \mathcal{X}} P_{Y_0|X}(y|x)P_X(x)$ and $\sum_{x \in \mathcal{X}} P_{Y_1|X}(y|x)P_X(x)$, respectively. After having received information from node A with rate $R[\frac{\text{bits}}{\text{symbol}}]$, node B is expected to make a decision about the joint distribution of the two RVs.

In a similar fashion to Stein’s Lemma and to previous work on distributed HT ([2]–[4]), we consider two types of error probabilities

$$\begin{aligned} \alpha_n &\triangleq \Pr(H_1|XY \sim P_0(x, y)) , \\ \beta_n &\triangleq \Pr(H_0|XY \sim P_1(x, y)) . \end{aligned} \quad (3)$$

The performance of the system, for a given rate of communication from node A to node B , 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 .

III. MAIN RESULT

Our main result is a new achievable error exponent for the general distributed HT problem.

Proposition 1 (Achievable error exponent). *The following error exponent, as a function of R , is achievable for the two-node HT problem, for any two hypotheses:*

$$\begin{aligned} E &= \inf_{Q_X} \sup_{Q_{U|X}^*} \inf_{Q_Y} \inf_{\substack{Q_{U,XY} \\ Q_{U|X}=Q_{U|X}^*}} \\ &\left\{ \min \left\{ G[Q_{U,XY}, R], \min_{\tilde{U}, \tilde{X}, \tilde{Y} \in \mathcal{L}(U)} D(P_{\tilde{U}\tilde{X}\tilde{Y}} || P_{U_1 X_1 Y_1}) \right\} \right\}, \end{aligned} \quad (4)$$

with U being an auxiliary RV such that the Markov chain $U \dashv X \dashv (Y_0, Y_1)$ is respected. The function $G[Q_{U,XY}, R]$ and the set $\mathcal{L}(U)$ appear in equations (5) and (6), respectively, at the top of the next page.

The expression for E encapsulates the innate tension between two possible error events, when a binning approach is taken. Choosing a strategy such that the description of \mathbf{x} is fine, the decoder is able to perform HT with a very large probability of success, when the correct description of the vector \mathbf{x} , \mathbf{u} , is available. However, such a choice would also induce a very large codebook, which, for a given R , would cause each bin to be very large in order to satisfy the rate constraint, making likely a mistake in choosing the sequence from the bin. When a crude description is chosen, the codebook is smaller, and thus so is each bin (if binning is at all necessary). The binning process is thus not likely to hurt performance significantly. However, the retrieved sequence is much less valuable for the sake of HT, because of the crude nature of the description it supplies about the source.

In order to ensure the achievability of the error exponent proposed in Proposition 1, a “worst-case” approach is taken. The first line in the expression for the error exponent in (4) can thus be read as follows: For every possible type of the observed vector \mathbf{x} , the encoder is allowed to choose its strategy of transmission (this is achieved by taking the supremum over $Q_{U|X}^*$). Having chosen the law for U , the proposed approach should apply for any type of the observed vector \mathbf{y} , as well as for any joint type of \mathbf{u} , \mathbf{x} and \mathbf{y} , as long as $Q_{U|X}^*$ is respected.

$$G[Q_{UXY}, R] = \begin{cases} \min_{i \in \{0,1\}} \{D(Q_{UXY} \| P_i Q_{U|X}) + [R - I(Q_X; Q_{U|X}) + I(Q_Y; Q_{U|Y})]^+\} & I(Q_X; Q_{U|X}) > R \\ \infty & \text{else} \end{cases} \quad (5)$$

$$\mathcal{L}(U) = \{P_{\bar{U}\bar{X}\bar{Y}} \in \mathcal{P}(U \times \mathcal{X} \times \mathcal{Y}) : P_{\bar{U}\bar{X}}(u, x) = P_{UX}(u, x), P_{\bar{U}\bar{Y}}(u, y) = P_{UY}(u, y), \forall u, x, y\}, \quad (6)$$

IV. PROOF OUTLINE FOR PROPOSITION 1

In this section we first describe the codebook construction, as well as coding and decoding strategies, that lead to the achievable region of Proposition 1. We then offer an outline to the analysis of error probability of the second type. Our approach uses some of the tools offered in [7] to calculate the error-exponent for the Wyner & Ziv problem [9], as well as the method of types in [8].

A. Encoder and Decoder Strategies

Codebook Construction: For a given block-length n we operate on a type-by-type basis. For each type Q_X set a unique index $k(Q_X)$. Then, fix a conditional type $Q_{U|X}^*(Q_X)$. Randomly and uniformly choose a set of codewords, $\mathcal{C}_U^n(Q_X)$, from the resulting marginal type class $T_{Q_U^*}^n$, induced by Q_X and $Q_{U|X}^*(Q_X)$. The size of $\mathcal{C}_U^n(Q_X)$ is an integer satisfying:

$$\begin{aligned} \exp^{nI(Q_X; Q_{U|X}^*(Q_X))} + (|\mathcal{U}||\mathcal{X}| + 2) \log(n+1) \\ \leq \mathcal{C}_U^n(Q_X) \leq \\ \exp^{nI(Q_X; Q_{U|X}^*(Q_X))} + (|\mathcal{U}||\mathcal{X}| + 4) \log(n+1). \end{aligned} \quad (7)$$

Define $f_U : T^n(Q_X) \rightarrow \mathcal{C}_U^n(Q_X)$. 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 $\mathbf{x} \in T^n(Q_X)$, search for a sequence \mathbf{u}_i in the codebook that belongs to the type of \mathbf{x} , such that $(\mathbf{u}_i, \mathbf{x}) \in T_\mu^n(UX)$. The encoder’s message then consists of two parts:

$$\begin{aligned} \mathcal{M}_1 &= \{1, 2, \dots, M_1 \triangleq \exp[nR]\}, \\ \mathcal{M}_2 &= \{1, 2, \dots, M_2 \triangleq (n+1)^{|\mathcal{X}|}\}. \\ \mathcal{M} &= \mathcal{M}_1 \times \mathcal{M}_2. \end{aligned} \quad (8)$$

The encoder sends the type of \mathbf{x} , $k(Q_{\mathbf{x}})$, with zero rate (“ \mathcal{M}_2 ”), 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 $\log |\mathcal{C}_U^n(Q_X)| < nR$, in which case we can map each member of $\mathcal{C}_U^n(Q_X)$ to an element of \mathcal{M}_1 in a one-to-one manner.
- 2 $\log |\mathcal{C}_U^n(Q_X)| \geq nR$, in which case we assign each distinct member of $\mathcal{C}_U^n(Q_X)$ to \mathcal{M}_1 uniformly at random.

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

$$\Psi(\mathbf{x}) = (F(f_U(\mathbf{x})), k(Q_{\mathbf{x}})). \quad (9)$$

Decoding: The decoder is aware of the type of the observed sequences at both nodes, and is free to optimize its strategy accordingly. It attempts to discover the word \mathbf{u} , by using the information sent to it by the encoder, as well as the vector of realizations \mathbf{y} , available at node B.

- 1 If $\log |\mathcal{C}_U^n(Q_X)| < nR$ the codeword can be decoded without error.
- 2 If $\log |\mathcal{C}_U^n(Q_X)| \geq nR$ the decoder receives a bin index and uses the side information to pick the best \mathbf{u} in the bin. Given the bin number, the type of \mathbf{x} and the side information \mathbf{y} , the decoder’s strategy is

$$\phi(F(f_U(\mathbf{x})), k(Q_{\mathbf{x}}), \mathbf{y}) = \hat{\mathbf{u}}, \quad (10)$$

if $\hat{\mathbf{u}} \in F(f_U(\mathbf{x}))$ and $\forall \tilde{\mathbf{u}} \in F(f_U(\mathbf{x})), \tilde{\mathbf{u}} \neq \hat{\mathbf{u}}, H(\tilde{\mathbf{u}}|\mathbf{y}) > H(\hat{\mathbf{u}}|\mathbf{y})$. $H(\mathbf{u}|\mathbf{y})$ denotes the empirical entropy

$$\begin{aligned} H(\mathbf{u}|\mathbf{y}) &\triangleq \mathbb{E}[-\log Q_{\mathbf{u}|\mathbf{y}}] \\ &= -\sum_{\substack{u \in \mathcal{U} \\ y \in \mathcal{Y}}} Q_{\mathbf{u}\mathbf{y}}(u, y) \log Q_{\mathbf{u}|\mathbf{y}}(u|y). \end{aligned} \quad (11)$$

Note that since our chosen test is over empirical entropies, it does not matter at this stage which hypothesis is the true one, for the sake of choosing the sequence from the bin. The true joint distribution of the sources is important for the sake of analyzing the error, but imperatively not important for the strategy of the decoder. After having retrieved a single sequence from the bin, the decoder can continue to perform HT by using a similar strategy to [3].

B. Analysis of the Probability of Error

Using the union bound, the probability of error in detection can be bounded by

$$P_e^{(n)} \leq P_r^{(n)} + P_d^{(n)}, \quad (12)$$

where $P_r^{(n)}$ is the probability of retrieving the wrong sequence from the bin, and $P_d^{(n)}$ is the probability of an erroneous detection despite using the correct sequence. We start by evaluating $P_r^{(n)}$ for a finite block-length n and then use a continuity argument to show that in the limit of $n \rightarrow \infty$, $P_r \leq G[Q_{UXY}, R]$. Since choosing the wrong sequence can only happen in case binning occurs, we are

only interested in the following subspace of the sequence space:

$$\mathcal{A} = \{(\mathbf{u}, \mathbf{x}, \mathbf{y}) : \mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n, \mathbf{u} \in T_{Q_{\mathbf{u}|\mathbf{x}}}^n, \log |\mathcal{C}_U^n(Q_{\mathbf{x}})| \geq nR\} . \quad (13)$$

We first evaluate the probability of choosing the wrong sequence within the set \mathcal{A} by using the following lemma, brought here without proof, which resembles the proof given in [7]:

Lemma 1. *Let $(\mathbf{u}, \mathbf{x}, \mathbf{y}) \in \mathcal{A}$. Let \mathcal{B} be the event that $\mathbf{u} \neq \phi(\psi(\mathbf{x}), \mathbf{y})$. If $\log |\mathcal{C}_U^n(Q_{\mathbf{x}})| \geq nR$ then*

$$\Pr(\mathcal{B} | \mathbf{U}^n = \mathbf{u}, \mathbf{X}^n = \mathbf{x}, \mathbf{Y}^n = \mathbf{y}) \leq \exp^{-n(R - J(Q_{\mathbf{u}|\mathbf{x}}) - \delta_n)} , \quad (14)$$

with

$$J(Q_{\mathbf{u}|\mathbf{x}}) = I(Q_{\mathbf{x}}; Q_{U|X}^*(Q_{\mathbf{x}})) - I(Q_{\mathbf{u}|\mathbf{y}}; Q_{\mathbf{y}}) \quad (15)$$

and

$$\delta_n = \frac{1}{n} \log(n+1)^{|U|(1+|\mathcal{X}|+|\mathcal{Y}|)+4} . \quad (16)$$

In order to bound the probability of choosing the wrong sequence from the bin, we use Lemma 1 while counting the sequences in the set \mathcal{A} . Changing the counting order to count types first and the sequences within each type second allows for the bounding of $P_r^{(n)}$. As a final step, continuity arguments are used in order to demonstrate the result of Proposition 1 when $n \rightarrow \infty$.

The second aspect of the error exponent analysis is the investigation of the detection error, under the assumption that the right sequence has been chosen from the bin. The probability that, given the right sequence \mathbf{u} , node B makes a wrong decision was investigated in detail in [3]. That result, however, is dependent on a specific codebook, conceived specifically to allow detection with high probability. As we use a random codebook in our scheme, it is essential to adapt the method of [3].

We propose here a slight modification to [3]. Intuitively, since we investigate the exponential decay of β_n while only enforcing a fixed upper bound on α_n , the ‘‘penalty’’ of replacing the codebook construction in [3] with random coding can be fully absorbed into α_n , leaving the error exponent result of β_n unmodified. From the construction of the codebook (specifically the size of the set $\mathcal{C}_U^n(Q_X)$), the number of sequences in the code book *per type of X* can be expressed by

$$M \triangleq \exp^{n[I(Q_X; Q_{U|X}^*(Q_X)) + \eta]} . \quad (17)$$

Assuming without loss of generality that $\mathbf{u}_1 = f_U(\mathbf{x})$, the decoder declares H_0 if $(\hat{\mathbf{u}}, \mathbf{y}) \in T_\mu^n(UY)$ and H_1 otherwise. Note that typicality is checked relative to the distribution implied by H_0 only. α_n can thus be expressed as follows:

$$\alpha_n \leq \Pr\{\mathbf{u}_1 \neq \phi(F(f_U(\mathbf{x})), k(Q_X), \mathbf{y})\} + \Pr\{(\mathbf{u}_1, \mathbf{y}) \notin T_\mu^n(UY) | H_0\} . \quad (18)$$

The first term here is not dependent on the hypothesis (assumed to be H_0 when analyzing α_n) and goes to zero

as explained above. The second term goes to zero as $n \rightarrow \infty$, as was the case in [4], thanks to the asymptotic equipartition property, the size of the codebook, the Markov chain $U \text{---} X \text{---} (Y_0, Y_1)$ and the Markov lemma. Thus, for any $\epsilon > 0$, $\alpha_n \leq \epsilon$ is fulfilled, for n large enough.

With this analysis of α_n , the analysis of the probability of error of the second type can be carried over unchanged from [3], and be applicable to our random codebook. Thus, for a fixed constraint over the error probability of the first type and a random codebook, one can conclude that

$$\theta_L(R) \triangleq \min_{\tilde{U}, \tilde{X}, \tilde{Y} \in \mathcal{L}(U)} D(P_{\tilde{U}\tilde{X}\tilde{Y}} || P_{U_1 X_1 Y_1}) \quad (19)$$

is an inner bound over the exponential rate of decay of the error probability of the second type, when the correct sequence is used at the decoder.

V. DOES BINNING IMPROVE PERFORMANCE?

Having proposed an approach to general HT, it is still not clear that this approach improves upon the known non-binning one [3]. In order to demonstrate the benefits of binning, we define a sub-class of HT problems, called ‘‘HT with degraded hypotheses’’. An example for such a problem is given subsequently, with numerical results.

Definition 1. *Let the marginal distribution at each side of a two-node HT problem be equal under both hypotheses. The system can thus be described by the probability distributions P_X , $P_{Y_0|X}$ and $P_{Y_1|X}$. Let H_1 be a degraded hypothesis, with relation to H_0 , if the channel from X to Y_1 is a stochastically degraded version of the channel from X to Y_0 . That is, there exists a RV \hat{Y}_1 , such that*

$$P_{Y_1|X}(y_1|x) = \sum_{y_0 \in \mathcal{Y}} P_{Y_0|X}(y_0|x) P_{\hat{Y}_1|Y_0}(y_1|y_0) . \quad (20)$$

This subclass includes many interesting HT problems, such as problems with additive channels. For the rest of this paper, we assume that the channel from X to Y_1 is *physically degraded* with relation to the channel from X to Y_0 , in order to improve readability. This assumption does not change the results, as they are only dependent on the marginal distributions $P_{Y_0|X}$ and $P_{Y_1|X}$. In the next section we give an example of a problem in this class, for which we show that the results attained by applying Proposition 1 beat the traditional non-binning approach.

As an example for a HT testing problem with degraded hypotheses, we take the following, scenario, composed of binary symmetric channels (BSCs):

$$X \sim \text{Bern}\left(\frac{1}{2}\right), \quad \begin{cases} H_0 : & Y = X \oplus Z_0, \\ H_1 : & Y = X \oplus Z_1, \end{cases} \quad (21)$$

where $Z_0 \sim \text{Bern}(p)$, $Z_1 \sim \text{Bern}(q)$ and $q > p$. Note that the marginal distribution of Y is similar for both hypotheses, making a decision without cooperation impossible.

Applying Proposition 1 to this setup, we choose to consider only distributions in which Q_X is a binary symmetric source ($\{0, 1\}$ with probability 0.5), and U is the result

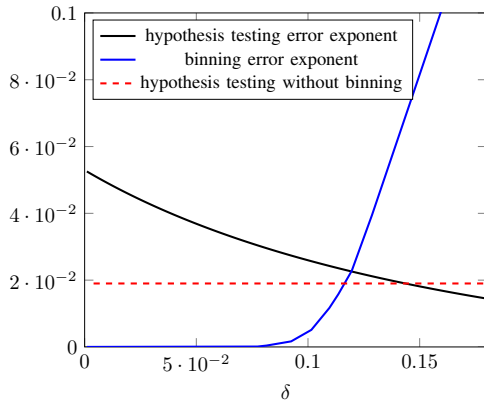


Fig. 2: Error exponents for both error events in the BSC case with $p = 0.1$, $q = 0.2$, $R = 0.4$. The resulting error exponent for each δ is the minimum between the two. Inner and outer bounds are dashed.

of passing X through a BSC with transition probability δ . While this is not necessarily an optimal choice, it can be justified as an optimal approach for the asymptotic regime [9]. In order to calculate the resulting error exponent, we need to calculate two values:

$$\inf_{Q_Y} \inf_{Q_{U|X}=Q_{U|X}^*} G[Q_{UXY}, R], \quad (22)$$

as a function of $Q_{U|X}^*$ (i.e., under our assumptions, as a function of δ), encapsulates the error exponent of the event where the wrong sequence is chosen from the bin.

$$\min_{\tilde{U}, \tilde{X}, \tilde{Y} \in \mathcal{L}(U)} D(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{U_1 X_1 Y_1}), \quad (23)$$

also as a function of $Q_{U|X}^*$, is the error exponent of the event where, while using the right sequence, an error occurs in detection. Having calculated these two functions, we can choose $Q_{U|X}^*$ such that the *minimum* between the two is *maximized*.

The tension between the two possible error events is shown in Figure 2, where the results for a binary symmetric source are depicted, for $p = 0.1$, $q = 0.2$ and $R = 0.4$. When δ is very small, a sequence \mathbf{u} can be found with high probability, such that \mathbf{x} is very well described, but the codebook contains many sequences \mathbf{u} . Thus, given the right sequence \mathbf{u} , an error in HT is not likely, and the error exponent of the event where HT fails is high. However, since the rate of communication is fixed, each bin has to contain many sequences when δ is small, enlarging the probability of error in choosing the sequence. When δ grows, the accuracy of the description of \mathbf{x} by \mathbf{u} is lower, making the probability of error in HT, while using the correct sequence, higher. The codebook, however, is smaller, making the task of choosing the right sequence in the bin easier. Note that the error exponent for choosing the sequence from within the bin has a threshold, under which it is zero. This threshold in this case is roughly $\delta \approx 0.08$, which is the value implied by

[9] as the minimal value for the binning approach, in the asymptotic regime.

As the system is free to choose its strategy (which boils down to choosing δ under our assumptions), it is clear that the chosen value should be $\delta \approx 0.12$. This value produces two error events with equal error exponents, and thus maximizes the minimum between the two. Clearly, this choice allows the binned approach to beat the non-binned one, as depicted in Figure 2 by a dashed line. We emphasize that *this bound is not drawn as a function of δ* , but rather depicts the best possible performance under the assumptions of [3]. Since this approach does not include binning, the size of the codebook, and thus the quality of the descriptions \mathbf{u} , is limited by the given rate of communication R .

Remark 1. A bigger improvement in performance is also possible, through a different approach by which the decoder does not attempt to retract a single sequence from the bin. Instead, HT is performed by checking all of the sequences in the bin. For the case of HT with degraded hypotheses, this approach can also be used to show that binning is always beneficial, except when testing against independence.

VI. CONCLUDING REMARKS

In this paper, an achievable region for the two-node HT problem with unidirectional communication was offered for general hypotheses. A class of HT problems, called “degraded hypotheses” was introduced. Using the example of a binary symmetric source, the merit of the binning approach taken in this paper was demonstrated, by showing that it beats the traditional non-binning approach. One possible suboptimal step in the offered approach could be the separation of the decision process into two steps, by first looking for the best sequence in the received bin and then making the decision. Future work will concern further improving performance in distributed HT problems with general hypotheses.

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] T. Han, “Hypothesis testing with multiterminal data compression,” *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, Nov 1987.
- [4] 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 Communication, Control and Computing*, 2014.
- [5] 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.
- [6] 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.
- [7] B. Kelly and A. Wagner, “Reliability in source coding with side information,” *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 5086–5111, Aug 2012.
- [8] I. Csiszár, “The method of types,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct 1998.
- [9] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 1–10, Jan 1976.