Extra Samples can Reduce the Communication for Independence Testing

K. R. Sahasranand* Himanshu Tyagi*

***Abstract*—Two parties observing sequences of bits want to determine if their bits were generated independently or not. To that end, the first party communicates to the second. A simple communication scheme involves taking as few sample bits as determined by the sample complexity of independence testing and sending it to the second party. But is there a scheme that uses fewer bits of communication than the sample complexity, perhaps by observing more sample bits? We show that the answer to this question is in the affirmative when the joint distribution is a binary symmetric source. More generally, for any given joint distribution, we present a distributed independence test that uses linear correlation between functions of the observed random variables. Furthermore, we provide lower bounds for the general setting that use hypercontractivity and reverse hypercontractivity to obtain a measure change bound between the joint and the independent distributions. The resulting bounds are tight for both a binary symmetric source and a Gaussian symmetric source.**

I. Introduction

Parties $\mathcal{P}_1$ and $\mathcal{P}_2$ observe sequences X^n and Y^n comprising finite-valued, independent and identically distributed samples (X_i, Y_i). They seek to determine if the samples are generated from P_{XY} or $\mathrm{P}_X\mathrm{P}_Y$. To that end, $\mathcal{P}_1$ communicates to $\mathcal{P}_2$, and the latter declares the output of the hypothesis test. A simple scheme entails taking recourse to the standard collocated version of the problem whereby $\mathcal{P}_1$ sends a subset of its samples to $\mathcal{P}_2$ who then applies an optimal likelihood ratio test. The number of samples sent is dictated by the reliability requirements: $\mathcal{P}_1$ sends the minimum number of samples needed to attain the required reliability. Can we reduce the communication by using a more carefully designed scheme, possibly by observing more samples?

Our interest in this question is motivated by applications arising in the Internet of Things (IoT) where we need to enable distributed inference and testing by communicating over a low bandwidth link. In such applications, the nodes are not restricted by the number of samples they collect, but by the permissible amount of communication. An easily implementable solution entails collecting all the samples at a single location and applying a standard statistical procedure. Restricting to the problem of independence testing, we seek to explore if it is even possible to communicate less by using a more sophisticated scheme, without paying heed to the number of samples collected.

In contrast, the available information theory literature on distributed independence testing (see [12] for a survey) primarily focuses on characterizing the exponential decay of probability of error under fixed rate communication, as the number of samples grows to infinity. In their seminal work [2], Ahlswede and Csiszár initiated the study of the tradeoff between the error exponent and the rate of communication for the general distributed hypothesis testing problem. In the specific case of independence testing, they provided a complete single-letter characterization of the error exponent. However, these results cannot be used directly to address the question we raise. In particular, the error exponent can be shown to be 0 when we restrict to communication of rate 0 ($cf.$ [19]). In our question, we are not interested in the dependence of error on the number of samples, and hence, even this zero-rate regime must be investigated.

In this paper, we derive general bounds for communication requirements for independence testing and show that the answer to the question raised above is in the affirmative. In particular, we exhibit a reduction in communication requirement over the simple scheme when P_{XY} is a binary symmetric source $\mathtt{BSS}(\rho)$, $-1 \le \rho \le 1$, $i.e.$, $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $\mathrm{P}_{XY}(0,0) = \mathrm{P}_{XY}(1,1) = (1+\rho)/4$, and $\mathrm{P}_{XY}(0,1) = (1-\rho)/4$.

Denote by $C(\delta, \varepsilon)$ the minimum amount of communication required to ensure that the test declares independence erroneously with probability smaller than δ and correctly with probability greater than $1-\varepsilon$. Similarly, denote by $n(\delta, \varepsilon)$ the minimum number of samples required in the standard collocated independence testing problem where the observations of both parties are at one place. It can be shown[1] that for $\mathtt{BSS}(\rho)$,

$$n(\delta, \varepsilon) = \frac{1}{1 - h((1-\rho)/2)} \log \frac{1}{\varepsilon} + \Theta_\delta\left(\sqrt{\log \frac{1}{\varepsilon}}\right), \quad (1)$$

where $h(\cdot)$ denotes the binary entropy function[2]. On the other hand, we show in Corollary 6 that

$$C(\delta, \varepsilon) = \frac{1}{\rho^2} \log \frac{1}{\varepsilon} + O_\delta\left(\sqrt{\log \frac{1}{\varepsilon}}\right).$$

Note that $\rho^2 > 1 - h((1-\rho)/2)$ for all $\rho \notin \{-1, 0, 1\}$. Therefore, for a fixed δ and a sufficiently small ε, our proposed

*Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India. Email: {sahasranand, htyagi}@iisc.ac.in

[1]This result is essentially a finite sample version of Stein's lemma and is easy to derive using standard concentration bounds; we omit the simple proof due to lack of space.

[2]The notation Θ_x denotes that the constant implied by Θ depends on x; similarly for O_x.

scheme communicates less than the simple scheme that simply shares $n(\delta,\varepsilon)$ sample bits.

A similar gain can be shown for the other regime where ε is fixed and δ is sufficiently small. Indeed, in the manner of equation (1), we can show that the simple scheme requires roughly $\frac{2}{\log\frac{1}{1-\rho^2}}\cdot\log\frac{1}{\delta}$ bits of communication. On the other hand, here our proposed scheme requires roughly (see Corollary 6) $\frac{1-\rho^2}{\rho^2}\cdot\log\frac{1}{\delta}$ bits of communication, which is less than that for simple scheme for all $\rho\notin\{-1,0,1\}$.

In fact, we derive upper and lower bounds for $C(\delta,\varepsilon)$ for a general P_{XY}. Our upper bound is achieved by a scheme based on [11] where communication for common randomness generation ($cf.$ [3]) was considered. Drawing on the heuristic connection between independence testing and common randomness generation highlighted in [22], [21], we adapt the scheme of [11] to devise a distributed independence test.

Specifically, our scheme compares the linear correlation of appropriately chosen zero mean, unit variance functions $f(X)$ and $g(Y)$ under P_{XY} and $\mathrm{P}_X\mathrm{P}_Y$. To save on the communication, the vector $(f(X_1),...,f(X_n))$ is first quantized using a shared (randomly generated) codebook comprising $\{-1,+1\}^n$-valued codewords. Roughly, $\mathcal{P}_1$ identifies the codeword u^n such that the inner product $\sum_{i=1}^n f(X_i)u_i$ is above a threshold τ and sends its index to $\mathcal{P}_2$. Next, $\mathcal{P}_2$ checks if $\sum_{i=1}^n g(Y_i)u_i$ is greater than $\theta\tau$, in which case it declares P_{XY}; else, $\mathcal{P}_2$ declares independence. The specific choice of f and g is so that the $\mathbb{E}\left[f(X)g(Y)\right]$ is maximized. This maximum value is termed the *maximal correlation coefficient* of (X,Y) and denoted $\rho_m(X,Y)$. We choose θ to roughly equal $\rho_m(X,Y)$ in our scheme. Note that for $\mathtt{BSS}(\rho)$, $\rho_m(X,Y)$ equals ρ.

For the converse, we use a change of measure argument to relate the probabilities of sets under P_{XY} and $\mathrm{P}_X\mathrm{P}_Y$. The distributed nature of our independence test imposes a rectangle structure on the acceptance region for P_{XY}. Our lower bound is obtained by using the classic hypercontractivity and reverse-hypercontractivity bounds ($cf.$ [7], [10], [6], [8], [4], [14]) to relate the measures of rectangles under the joint and the product distributions.

While our upper bound involves ρ_m, the lower bound is expressed in terms of the so-called hypercontractivity ribbon of X and Y ($cf.$ [4], [5]). Indeed, the two are closely related (see [5, Theorem 4]). In the particular case of $\mathtt{BSS}(\rho)$, the entire hypercontractivity ribbon is characterized by ρ (see, for instance, [17]), and our bounds coincide. The same also applies for the case of a Gaussian symmetric source $\mathtt{GSS}(\rho)$, $-1\le\rho\le 1$, where (X,Y) are jointly Gaussian with zero mean, $\mathbb{E}\left[X^2\right]=\mathbb{E}\left[Y^2\right]=1$ and $\mathbb{E}\left[XY\right]=\rho$.

The paper is organized as follows: next section contains a formal description of our problem and the main results. Our scheme is given in Section III followed by the proof-sketch for the lower bounds in Section IV. We conclude with pointers to possible extensions, related literature, and open problems in the final section.

Notation. Random variables are denoted by capital letters such as X, Y, $etc.$; their specific realizations by the corresponding small letters such as x, y, $etc.$; and their ranges by the corresponding calligraphic forms such as $\mathcal{X}$, $\mathcal{Y}$, $etc.$. The distribution of a random variable X is denoted by P_X. Most of the information theory notations are borrowed from [9]. All the logarithms are to the base 2; when needed, we use $\ln a$ to denote the natural logarithm of a. For a $p>0$, we denote by p' the Hölder conjugate of p given by $p/(p-1)$, with the convention that $p'=\infty$ for $p=1$.

II. Minimum one-way communication for independence testing

We consider a simple binary hypothesis testing problem. The observation consists of n independent and identically distributed (i.i.d.) samples, generated with common distribution P_{XY} under the null hypothesis $\mathcal{H}_0$ and with common distribution $\mathrm{P}_X\times\mathrm{P}_Y$ under the alternative hypothesis $\mathcal{H}_1$. In our distributed setup, the party $\mathcal{P}_1$ observes X^n and the party $\mathcal{P}_2$ observes Y^n. In addition, $\mathcal{P}_1$ and $\mathcal{P}_2$ have access to a shared random variable U. A *distributed test* $T=(c,d)$ consists of a communication of length l described by a mapping $c:\mathcal{X}^n\times\mathcal{U}\to\{0,1\}^l$, where $\mathcal{P}_1$ transmits $c(X^n,U)$ upon observing X^n, and a decision mapping $d:\mathcal{Y}^n\times\{0,1\}^l\times\mathcal{U}\to\{0,1\}$, where $\mathcal{P}_2$, upon observing Y^n and receiving bits $B^l=c(X^n,U)$ from $\mathcal{P}_1$, declares the hypothesis $d(Y^n,B^l,U)$.

A distributed test $T=(c,d)$ constitutes an (l,δ,ε)-test with observation length n if c is a communication of length l,

$$\mathbb{P}_{\mathcal{H}_0}\left(d\big(Y^n,B^l,U\big)=1\right)\le\delta,\ \text{and}$$
$$\mathbb{P}_{\mathcal{H}_1}\left(d\big(Y^n,B^l,U\big)=0\right)\le\varepsilon.$$

Our goal is to design a distributed test that communicates as few bits as possible, while possessing desired probabilities of error. Note that in contrast to the classic Neyman-Pearson formulation where we choose a nominal value for one error and require the other error to be very small, here we study the more general problem of characterizing the communication requirement for the entire decision region. However, our results for this complete range are only partial.

Formally, we seek bounds for the minimum communication for independence testing, defined next.

Definition 1. Given $\delta,\varepsilon\in[0,1]$, the minimum communication for independence testing $C(\delta,\varepsilon)$ is the least l such that there exists an (l,δ,ε)-test, for some observation length n.

Remark 1. Clearly, the communication requirement does not increase with n. In defining $C(\delta,\varepsilon)$, we allow n to be arbitrarily large. However, it suffices to have an n that is polynomial in $(1/\delta\varepsilon)$ for our proposed scheme to work.

We derive general lower and upper bounds for $C(\delta,\varepsilon)$. For the cases of doubly symmetric binary and Gaussian sources, our bounds match in specific regimes for δ and ε.

A. Upper bounds

Our proposed scheme for independence testing relies on the linear correlation between appropriately chosen, zero mean

functions f and g of X and Y, respectively. The specific quantity that shows up is the Hirschfeld-Gebelein-Rényi correlation coefficient or the maximal correlation coefficient of (X,Y), denoted $\rho_m(X,Y)$ ($cf.$ [18]) and defined as

$$\rho_m(X,Y) = \sup_{f,g} \mathbb{E}\left[f(X)g(Y)\right],$$

where the supremum is over all $\mathbb{R}$-valued functions f of X and g of Y satisfying $\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[g(Y)\right] = 0$ and $\mathbb{E}\left[f^2(X)\right] = \mathbb{E}\left[g^2(Y)\right] = 1$. Note that when computed under the independent distribution, the expected value of $f(X)g(Y)$ is 0. Our scheme relies on this difference in the expected value of $f(X)g(Y)$ under the two hypotheses.

We now report the upper bound on $C(\delta,\varepsilon)$ achieved by our scheme. We will present the scheme in Section III; the analysis of the scheme and the proof of Theorems 1 and 2 are omitted – only a sketch is provided in Section III.

Theorem 1 (Upper bound for small ε,δ)**.** *For $\delta,\varepsilon \in (0,1/2)$ and P_{XY} with maximal correlation coefficient $\rho_m(X,Y)=\rho$,*

$$C(\delta,\varepsilon) \leq \frac{1}{\rho^2}\cdot\left(\sqrt{\log\frac{1}{\varepsilon}} + \sqrt{\left(1-\rho^2\right)\log\frac{1}{\delta}}\right)^2 + O\left(\sqrt{\log\frac{1}{\varepsilon\delta}}\right).$$

Remark 2. When (X,Y) correspond to $BSS(\rho)$ or $GSS(\rho)$, $-1\leq\rho\leq 1$, the maximal correlation coefficient $\rho_m(X,Y)$ equals ρ ($cf.$ [18]).

In Section II-C, we shall see that for $BSS(\rho)$ and $GSS(\rho)$ the bound above is tight up to the leading term in dependence on ε or δ, though not simultaneously for both. In fact, for the practically uninteresting regime of $\delta > 1/2$, the next result offers an improvement which will be seen to be tight for $BSS(\rho)$ and $GSS(\rho)$ up to the leading term, simultaneously for ε and δ.

Theorem 2 (Upper bound for small ε, large δ)**.** *For $\varepsilon \in (0,1/2)$, $\delta\in(1/2,1)$ and P_{XY} with maximal correlation coefficient $\rho_m(X,Y)=\rho$,*

$$C(\delta,\varepsilon) \leq \frac{1}{\rho^2}\cdot\left(\sqrt{\log\frac{1}{\varepsilon}} - \sqrt{\left(1-\rho^2\right)\log\frac{1}{1-\delta}}\right)^2 + O\left(\sqrt{\log\frac{1}{\varepsilon(1-\delta)}}\right).$$

B. Lower bounds

Our lower bounds involve the notions of *hypercontractivity* and *reverse hypercontractivity* ($cf.$ [4], [16]). For $1\leq q\leq p<\infty$, a pair of random variables (X,Y) is (p,q)-hypercontractive if for all $\mathbb{R}$-valued functions f of X and g of Y

$$\mathbb{E}\left[|f(X)g(Y)|\right] \leq \|f(X)\|_{p'}\|g(Y)\|_q,$$

where $p'=p/(p-1)$ is the Hölder conjugate of p. Similarly, for $1\geq q>p$, a pair of random variables (X,Y) is (p,q)-reverse hypercontractive if for all $\mathbb{R}$-valued functions f of X and g of Y

$$\mathbb{E}\left[|f(X)g(Y)|\right] \geq \|f(X)\|_{p'}\|g(Y)\|_q.$$

The set of all (p,q) for which (X,Y) is (p,q)-hypercontractive and (p,q)-reverse hypercontractive, respectively, are called the hypercontractivity ribbon and the reverse hypercontractivity ribbon of (X,Y). We use the notions of hypercontractivity and reverse hypercontractivity to obtain the change of measure bounds between the joint distribution and the independent distribution, which in turn lead to the following lower bounds for $C(\delta,\varepsilon)$.

Theorem 3 (Lower bound 1)**.** *Given $\delta,\varepsilon\in(0,1)$ and (p,q) such that $1\leq p'\leq q\leq p$ and (X,Y) is (p,q)-hypercontractive, the minimum communication for independence testing $C(\delta,\varepsilon)$ is bounded below as*

$$C(\delta,\varepsilon) \geq \frac{p}{q}\log\frac{1}{\varepsilon} - p\log\frac{1}{1-\delta}. \tag{2}$$

The proof of Theorem 3 is given in Section IV. The next result can be proved using similar manipulations with the reverse hypercontractivity bound replacing the hypercontractivity bound and is omitted.

Theorem 4 (Lower bound 2)**.** *Given $\delta,\varepsilon\in(0,1)$ and (p,q) such that $1\geq q\geq 0\geq q'\geq p$ and (X,Y) is (p,q)-reverse hypercontractive, the minimum communication for independence testing $C(\delta,\varepsilon)$ is bounded below as*

$$C(\delta,\varepsilon) \geq \frac{p}{q}\log\frac{1}{1-\varepsilon} - p\log\frac{1}{\delta}. \tag{3}$$

C. Special cases: Binary and Gaussian symmetric sources

To obtain tight lower bounds for specific distributions, we need to optimize our lower bounds over the entire hypercontractivity and reverse hypercontractivity ribbon. In general, an expression for this optimized lower bound is unavailable and its relation to the maximum correlation that appears in the upper bounds is unclear. However, for the special cases of $BSS(\rho)$ and $GSS(\rho)$ the optimized lower bounds will be seen to match the upper bounds. This is the content of the current section. Note that for these special cases the functions f and g attaining the maximum correlation are linear (for $BSS(\rho)$, we replace the alphabet with $\{-1,1\}$). Thus, our distributed independence test in these cases takes a linear form, too.

We rely on the following characterizations of the hypercontractivity and the reverse hypercontractivity ribbons.

Theorem 5 ([7], [10], [6], [8], [15])**.** *Let P_{XY} correspond to $BSS(\rho)$ or $GSS(\rho)$, $-1\leq\rho\leq 1$. For $1\leq q\leq p$, (X,Y) is (p,q)-hypercontractive if and only if*

$$\frac{q-1}{p-1} \geq \rho^2. \tag{4}$$

Furthermore, for $1\geq q\geq p$, (X,Y) is (p,q)-reverse hypercontractive if and only if

$$\frac{1-q}{1-p} \geq \rho^2. \tag{5}$$

The next corollary is obtained by maximizing the right-sides of (2) and (3), respectively, over the set of (p, q) satisfying (4) and (5); the upper bound is from Theorem 1.

Corollary 6. *Let P_{XY} correspond to $BSS(\rho)$ or $GSS(\rho)$, $-1 \le \rho \le 1$. Then,*

1) *for $\delta \in (0, 1/2)$ and ε such that $\delta + \varepsilon^{\frac{1-|\rho|}{1+|\rho|}} \le 1$,*

$$C(\delta, \varepsilon) = \frac{1}{\rho^2} \log \frac{1}{\varepsilon} + O_\delta\left(\sqrt{\log \frac{1}{\varepsilon}}\right);$$

2) *for $\varepsilon, \delta \in (0, 1/2)$,*

$$C(\delta, \varepsilon) = \frac{1-\rho^2}{\rho^2} \log \frac{1}{\delta} + O_\varepsilon\left(\sqrt{\log \frac{1}{\delta}}\right),$$

where the notation O_x denotes that the constant implied by O depends on x.

Note that the result above yields the leading asymptotic term for dependence on ε and δ, considered separately. However, it falls short of characterizing the joint-dependence on (δ, ε). Indeed, a characterization of such a joint-dependence is difficult to obtain even for sample complexity $n(\delta, \varepsilon)$. Nevertheless, when we allow a large δ and have ε sufficiently small, we can obtain a result characterizing simultaneous dependence. Interestingly, the amount of communication is below $(1/\rho^2) \log 1/\varepsilon$ in this case.

Corollary 7. *Let P_{XY} correspond to $BSS(\rho)$ or $GSS(\rho)$, $-1 \le \rho \le 1$. Then, for $\delta \in (1/2, 1)$ and ε such that $\delta + \varepsilon^{\frac{1-|\rho|}{1+|\rho|}} \le 1$,*

$$C(\delta, \varepsilon) = \frac{1}{\rho^2}\left(\sqrt{\log \frac{1}{\varepsilon}} - \sqrt{\left(1-\rho^2\right) \log \frac{1}{1-\delta}}\right)^2 + O\left(\sqrt{\log \frac{1}{\varepsilon(1-\delta)}}\right).$$

III. The scheme achieving our upper bounds

We describe a slightly restricted form of our scheme for the case when Y is a zero mean and unit variance random variable; the extension to the general case is straightforward and will be mentioned later. For this case, let $\rho^2 = \mathbb{E}\left[\mathbb{E}\left[Y|X\right]^2\right]$ and define $f(X) = \rho^{-1}\mathbb{E}\left[Y|X\right]$. Note that since Y has zero mean, $f(X)$ is a zero mean and unit variance random variable. Our distributed test for such (X, Y) is described below[3]:

Fix parameters $r > 0$, $\theta \in (0, 1]$, and $k \in \mathbb{N}$.

1) Using the shared randomness, parties generate a $n \times 2^k$ matrix $\mathbf{U}$ consisting of i.i.d. $\{-1, +1\}$-valued entries U_{ij}, $1 \le i \le n$, $0 \le j \le 2^k - 1$, drawn uniformly.
2) $\mathcal{P}_1$ finds the *least* index $j \in [2^k]$ such that $\sum_{i=1}^n U_{ij} \cdot f(X_i) \ge r\sqrt{n}$ holds and sends the k-bit representation of j to $\mathcal{P}_2$. If no such j is found, declare[4] $\mathcal{H}_1$.
3) $\mathcal{P}_2$, upon receiving j, declares $\mathcal{H}_0$ if $\sum_{i=1}^n U_{ij} \cdot Y_i \ge \theta \cdot r\sqrt{n}$ and $\mathcal{H}_1$ otherwise.

This distributed independence test extends to general distributions by replacing Y with any function $g(Y)$ such that $\mathbb{E}\left[g(Y)\right] = 0$ and $\mathbb{E}\left[g(Y)^2\right] = 1$ and apply the procedure above; the resulting ρ^2 will be given by $\mathbb{E}\left[\mathbb{E}\left[g(Y)|X\right]^2\right]$. As shall be seen below, the number of bits communicated by the distributed test above depends on ρ^2, and the latter must be chosen appropriately for optimality.

The analysis of the scheme relies on Gaussian approximation using the Berry-Esseen theorem. Note that the factor of $\sqrt{n}$ appearing in the thresholds in steps (2) and (3) is crucial for removing the dependence of Gaussian tails on n, and thereby that of k on n. In particular, we can show that for any fixed $\eta \in (0, 1)$ and a sufficiently large n

$$\mathbb{P}_{\mathcal{H}_1}\left(\text{Declare } \mathcal{H}_0 \mid \mathbf{U} = \mathbf{u}\right) \le 2^{k+1} \cdot Q(r) \cdot Q(\theta r), \tag{6}$$

and

$$\mathbb{P}_{\mathcal{H}_0}\left(\text{Declare } \mathcal{H}_0\right) \ge (1-\eta) Q\left(\frac{(\theta-\rho)r}{\sqrt{1-\rho^2}}\right)\left(1 - e^{-2^{k-1}Q(r)}\right), \tag{7}$$

where $Q(\cdot)$ is the complementary cumulative distribution function of a standard normal random variable. Thus, we obtain a (k, δ, ε)-test upon choosing θ, r, k, and η such that the right-side of (6) is bounded above by ε and the right-side of (7) is bounded below by $1 - \delta$. In particular, to prove Theorems 1 and 2, we find the minimum k for which these constraints can be satisfied for some θ, r, and η.

Sketch of proof of Theorems 1 and 2. It suffices to ensure that

$$2 \ln \frac{4}{\delta} \le 2^k Q(r) \le 4 \ln \frac{4}{\delta} \text{ and } Q\left(\frac{(\rho-\theta)r}{\sqrt{1-\rho^2}}\right) \le \frac{\delta}{4},$$

which for $\rho \ge \theta$ will hold if we choose $k \approx r^2/(2 \ln 2)$ for an r^2 satisfying

$$\frac{r^2}{2 \ln 2} \ge \min_{\theta \le \rho} \max\left\{\frac{\alpha}{(\rho-\theta)^2}, \frac{\beta}{\theta^2}\right\} = \frac{1}{\rho^2}(\sqrt{\alpha} + \sqrt{\beta})^2,$$

where $\alpha = (1-\rho^2) \log \frac{4}{\delta}$ and $\beta = \log \frac{1}{\varepsilon} + \log \log \frac{4}{\delta} + 2$.

A similar optimization yields Theorem 2; here, too, we set $k \approx r^2/(2 \ln 2)$, but r^2 now must satisfy

$$\frac{r^2}{2 \ln 2} \ge \min_{\theta \ge \rho} \max\left\{\frac{\alpha}{(\theta-\rho)^2}, \frac{\beta}{\theta^2}\right\} = \frac{1}{\rho^2}(\sqrt{\alpha} - \sqrt{\beta})^2,$$

where $\alpha \approx (1-\rho^2) \log \frac{1}{1-\delta}$ and $\beta \approx \log \frac{1}{\varepsilon}$.

IV. Proof of lower bound 1

For $1 \le q \le p$, suppose that (X, Y) is (p, q)-hypercontractive. Furthermore, assume that $p' \le q$ which is the same as $q' \le p$. Then, for any subset $\mathcal{A} \subset \mathcal{X}^n$ and $\mathcal{B} \subset \mathcal{Y}^n$, we have

$$\mathrm{P}_{X^n Y^n}\left(\mathcal{A} \times \mathcal{B}\right) \le \mathrm{P}_{X^n}\left(\mathcal{A}\right)^{\frac{1}{p'}} \mathrm{P}_{Y^n}\left(\mathcal{B}\right)^{\frac{1}{q}}. \tag{8}$$

[3] We provide an operational description of our distributed test; the mappings c and d can be identified readily from this.

[4] Formally, $\mathcal{P}_1$ will output 1 and communicate this outcome to $\mathcal{P}_2$ using additional 1-bit communication.

For brevity, we only consider a deterministic test where the shared randomness U is constant; randomness can be handled using Jensen's inequality. Specifically, given a deterministic (l, δ, ε)-test $T = (c, d)$, denoting $L = 2^l$, let $\mathcal{A}_i = c^{-1}(i)$ for $i = 1, ..., L$. Then, $\{\mathcal{A}_1, ..., \mathcal{A}_L\}$ constitutes a partition of $\mathcal{X}^n$. Further, let $\mathcal{B}_i$ denote the set $\{\mathbf{y} \in \mathcal{Y}^n : d(\mathbf{y}, i) = 0\}$, namely the set of $\mathbf{y}$ where $\mathcal{P}_2$ declares $\mathcal{H}_0$ upon receiving i from $\mathcal{P}_1$. Denoting $a_i = \mathrm{P}_{X^n}(\mathcal{A}_i)$ and $b_i = \mathrm{P}_{Y^n}(\mathcal{B}_i)$, it follows by (8) that $1 - \delta \leq \sum_{i=1}^{L} \mathrm{P}_{X^n Y^n}(\mathcal{A}_i \times \mathcal{B}_i) \leq \sum_{i=1}^{L} a_i^{\frac{1}{p'}} b_i^{\frac{1}{q}}$. Using Hölder's inequality, we can bound the right-side by

$$\Big(\sum_{i=1}^{L} a_i b_i\Big)^{\frac{1}{q}} \Big(\sum_{i=1}^{L} a_i^{q'(\frac{1}{p'} - \frac{1}{q})}\Big)^{\frac{1}{q'}} \leq \varepsilon^{\frac{1}{q}} \Big(\sum_{i=1}^{L} a_i^{q'(\frac{1}{p'} - \frac{1}{q})}\Big)^{\frac{1}{q'}},$$

where the previous inequality uses the requirement $\mathbb{P}_{\mathcal{H}_1}(\text{Declare } \mathcal{H}_0) \leq \varepsilon$. Upon noting that $q'(1/p' - 1/q) = 1 - q'/p$, the bound above together with the assumption $q' \leq p$ and Hölder's inequality, yields the desired bound $(1 - \delta) \leq \varepsilon^{\frac{1}{q}} L^{\frac{1}{p}}$.

V. Concluding remarks

For $\mathtt{BSS}(\rho)$ with $X, Y \in \{-1, 1\}$, the distributed independence test we have presented, entails identifying a sequence $u^n \in \{-1, 1\}^n$ such that x^n is close to u^n and checking if y^n, too, is appropriately close to u^n. Alternatively, we can send the sequence x^n and directly check if y^n is close to x^n. If we ignore the Berry-Esseen correction, this simple scheme attains the communication rate of Corollary 6. However, in a formal analysis, the number of samples needed is dominated by the correction term. An important idea in our distributed scheme, which we borrow from [11], is to appropriately choose the parameters so that only the communication cost k shows up in the Gaussian-tail bounds and not n.

A natural extension of the problem we consider, is to allow multiple rounds of interaction. In the error-exponent regime, such a formulation has been considered in [24], [25]. However, we are unable to handle interaction in our current treatment. A specific question of interest is whether multiple rounds of interaction can improve the communication warranted by Corollary 6 for $\mathtt{BSS}(\rho)$. A similar question was addressed for the related problem of secret key agreement in [20] ($cf.$ [13]).

Also, in the context of IoT, it is of interest to consider more involved communication topologies such as those studied in [26], [23]. Here too, our current techniques fall short. In particular, we have difficulty in handling a separate decision center using a random codebook generated independently of X and Y.

Finally, note that we have not addressed the problem in a universal setting where the distribution P_{XY} is not known, but only a separation of ε between P_{XY} and $\mathrm{P}_X \mathrm{P}_Y$ in total variation distance is assumed. Note that for $\mathtt{BSS}(\rho)$, the separation ε equals ρ. Therefore, our lower bound for $\mathtt{BSS}(\rho)$ yields an $\Omega(1/\varepsilon^2)$ lower bound for the worst-case communication cost over all distributions for binary X and Y. Using the known sample complexity results for the collocated case ($cf.$ [1]), this lower bound implies that the simple scheme is order-optimal. However, the case of larger alphabet-sizes remains an interesting open problem.

References

[1] J. Acharya, C. Daskalakis, and G. Kamath, "Optimal testing for properties of distributions," in *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, pp. 3591–3599.

[2] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, July 1986.

[3] ——, "Common randomness in information theory and cryptography–part II: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, January 1998.

[4] R. Ahlswede and P. Gacs, "Spreading of sets in product spaces and hypercontraction of the markov operator," *Ann. Probab.*, vol. 4, no. 6, pp. 925–939, December 1976.

[5] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover," 2013.

[6] W. Beckner, "Inequalities in Fourier analysis," *Ann. of Math.*, vol. 102, no. 1, pp. 159–182, July 1975.

[7] A. Bonami, "Etudes des coefficients Fourier des fonctiones de $L^p(G)$," *Ann. Inst. Fourier*, vol. 20, no. 2, pp. 335–402, 1970.

[8] C. Borell, "Positivity improving operators and hypercontractivity," *Mathematische Zeitschrift*, no. 180, pp. 225–234, 1982.

[9] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless channels*. Academic Press, 1981.

[10] L. Gross, "Logarithmic sobolev inequalities," *American Journal of Mathematics*, vol. 97, no. 4, pp. 1061–1083, 1975.

[11] V. Guruswami and J. Radhakrishnan, "Tight bounds for communication-assisted agreement distillation," in *Proceedings of the 31st Conference on Computational Complexity*, 2016, pp. 6:1–6:17.

[12] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, October 1998.

[13] J. Liu, P. Cuff, and S. Verdú, "Secret key generation with limited interaction," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7358–7381, November 2017.

[14] E. Mossel, K. Oleszkiewicz, and A. Sen, "On reverse hypercontractivity," *Geometric and Functional Analysis*, vol. 23, no. 3, pp. 1062–1097, June 2013.

[15] C. Nair and Y. N. Wang, "Reverse hypercontractivity region for the binary erasure channel," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 938–942.

[16] C. Nair, "Equivalent formulations of hypercontractivity using information measures," *Proceedings of International Zürich Seminar on Communications*, 2014.

[17] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[18] A. Rényi, "On measures of dependence," *Acta Mathematica Hungarica*, vol. 10, no. 3-4, pp. 441–451, 1959.

[19] H. M. H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 254–267, March 1992.

[20] H. Tyagi, "Common information and secret key capacity," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5627–5640, September 2013.

[21] H. Tyagi and S. Watanabe, "Converses for secret key agreement and secure computing," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4809–4827, September 2015.

[22] ——, "A bound for multiparty secret key agreement and implications for a problem of secure computing," in *EUROCRYPT*, 2014, pp. 369–386.

[23] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," in *International Conference on Signal Processing and Communications (SPCOM)*, June 2016, pp. 1–5.

[24] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing with communication constraints," in *50th Annual Allerton Conference on Communication, Control, and Computing*, October 2012, pp. 1065–1072.

[25] ——, "Interactive hypothesis testing against independence," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 2840–2844.

[26] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *52nd Annual Allerton Conference on Communication, Control, and Computing*, September 2014, pp. 1246–1251.