One-Shot Achievability Region for Hypothesis Testing with Communication Constraint

Yuanxin Guo, Sadaf Salehkalaibar, Stark C. Draper and Wei Yu
University of Toronto, Toronto, Canada
E-mails: yuanxin.guo@mail.utoronto.ca, sadafs@ece.utoronto.ca, stark.draper@utoronto.ca, weiyu@ece.utoronto.ca

***Abstract*—The paper considers a communication constrained distributed hypothesis testing problem in which the transmitter sends a message about its local observation to the receiver, and the receiver tries to decide whether or not its own observation is independent of the observation at the transmitter. We analyze the problem in the one-shot setting and derive an achievability region under both the fixed-length and the variable-length communication constraints. Novel information-theoretic tools, including the generalized Poisson matching lemma and the strong functional representation lemma, are applied. It is shown that the proposed one-shot schemes, when applied to the asymptotic case, recover the optimal fixed-length and variable-length type-II error exponents for testing against independence.**

I. Introduction

This paper considers a distributed hypothesis testing problem in which a sensor node and a decision node are located in two different locations and are linked by a noiseless channel of finite capacity. The sensor and decision nodes each make local observations. The sensor node transmits a message to the decision node to help it decide between two hypotheses: H_0 or H_1. Our aim is to quantify the effect of communication constraint on the probability of error.

We model the aforementioned problem by the system depicted in Fig. 1. The sensor node (i.e., the transmitter) observes n independent and identically distributed (i.i.d.) samples of a random variable X and the decision node (i.e., the receiver) observes n i.i.d. samples of another random variable Y. We assume that the marginal distributions of each of the two random variables are the same under both hypotheses, but that their joint distribution depends on the hypothesis. We also assume the transmitter and the receiver share unlimited common randomness. This paper focuses on an important special case, often referred to as *testing against independence*, where $(X,Y) \sim P_{XY}$ under the null hypothesis H_0, and $(X,Y) \sim P_X P_Y$ under the alternative hypothesis H_1. Here, P_{XY} is a given distribution over $\mathcal{X} \times \mathcal{Y}$, and P_X, P_Y are the corresponding marginals.

This hypothesis testing problem is formulated under a communication constraint between the transmitter and the receiver. The transmitter sends a message M based on X^n. We impose one of the two following types of communication constraints. The first is when the message takes values from a finite set and we restrict the log of the cardinality of the finite set to be less than nL; the second is when the message is a variable-length prefix-free binary string and we restrict the expected length of the message to be less than nL. Upon receiving M noiselessly and observing Y^n, the receiver makes a decision $\hat{H} \in \{H_0, H_1\}$. A detection error is classified as either a type-I error ($H = H_0$ while $\hat{H} = H_1$) or a type-II error ($H = H_1$ while $\hat{H} = H_0$). We are interested in the *achievable region* of this problem, which consists of all triples (L, α, β) where α, β are respective upper bounds on the type-I and type-II error probabilities.

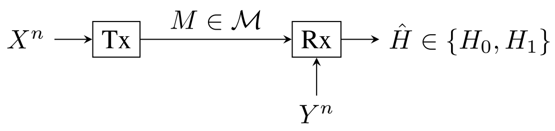

Fig. 1. Distributed hypothesis testing with communication constraint.

When the number of observed samples n tends to infinity, one would expect the two types of error probabilities to approach zero exponentially. Characterizing the error exponents of both types of error is difficult. Instead, a major line of work focuses on determining the largest exponential rate of decay of the type-II error probability while fixing some bound on the type-I error probability. Ahlswede and Csiszár [1] construct a quantization-based fixed-length testing scheme and prove that their scheme achieves the optimal type-II error exponent among all fixed-length schemes for testing against independence. Han [2] and Shimokawa, Han and Amari [3] improve upon Ahlswede and Csiszár's scheme and establish achievability results for testing between more general distributions. For a more general case termed "testing against conditional independence", Rahman and Wagner [4] establish tight achievability and converse results. A recent paper [5] demonstrates that a slightly larger type-II error exponent can be attained if variable-length schemes are allowed. There have also been studies on the trade-off between the type-I and the type-II error exponents, for example, [6], [7]. However, tight converse bounds for most cases are not known, except for the special cases of testing against (possibly conditional) independence.

As opposed to the asymptotic regime of $n \to \infty$, much less is known about the finite block-length regime. A Neyman-Pearson type hypothesis testing scheme is proposed in [8] and its non-asymptotic performance is analyzed. However, [8] is restricted to zero-rate settings in which the communication constraint scales sublinearly with block-length.

In this paper, we focus on the one-shot setting of $n = 1$.

We show how to apply the recently developed information theoretic tools of the generalized Poisson matching lemma (GPML) [9] and the strong functional representation lemma (SFRL) [10], developed to prove one-shot source and channel coding theorems, to the hypothesis testing problem. We construct both fixed-length and variable-length one-shot hypothesis testing schemes based on GPML and SFRL and obtain the corresponding achievable regions.

These results can be readily generalized to arbitrary n, i.e., the finite block-length setting. When the block-length tends to infinity, the optimal exponents in [1] and [5] are recovered.

Notation: We denote random variables by uppercase letters, e.g., X, Y, and their realizations by lowercase letters, e.g., x, y. The alphabets of random variables are denoted by calligraphic letters, e.g., $\mathcal{X}, \mathcal{Y}$. The set of all finite-length sequences over an alphabet $\mathcal{X}$ is denoted by $\mathcal{X}^*$. We use $\ell : \mathcal{X}^* \to \mathbb{N}$ to denote the length of the sequence. We denote the set of positive integers by $\mathbb{N} = \{1, 2, \cdots\}$ and the set of nonnegative numbers by $\mathbb{R}_+ = [0, \infty)$. For event $\mathcal{E} \subset \mathcal{X}$, the indicator function of $\mathcal{E}$ is denoted by $\mathbf{1}\{\cdot \in \mathcal{E}\} : \mathcal{X} \to \{0, 1\}$. Throughout the paper, we assume that $\log$ is base 2.

For a pair of random variables (X, Y), their joint distribution is denoted by P_{XY}, and the marginal distributions are denoted by P_X and P_Y respectively. The conditional distribution of Y given X is denoted by $P_{Y|X}$. For two distributions P, Q defined over the same alphabet $\mathcal{X}$ such that P is absolutely continuous with respect to Q (denoted $P \ll Q$), the Radon-Nikodym derivative is denoted by $\frac{dP}{dQ} : \mathcal{X} \to \mathbb{R}_+$. Given joint distribution P_{XY} (such that $P_{XY} \ll P_X P_Y$), we denote the *information density* as

$$i(x; y) = \log \frac{dP_{XY}}{d(P_X P_Y)}(x, y). \tag{1}$$

II. Technical tools

In this section, we briefly introduce the main tools needed to establish the one-shot achievability results. The exposition mainly follow the works [9], [10].

A. Poisson Functional Representation

Definition 1 (Poisson Function Representation (PFR))**.** Let $Z = \{\bar{U}_k, T_k\}_{k\in\mathbb{N}}$ be a Poisson point process with intensity $Q \times \lambda_{\mathbb{R}_+}$, where Q is a probability measure over $\mathcal{U}$ and $\lambda_{\mathbb{R}_+}$ is the Lebesgue measure on $[0, \infty)$. Without loss of generality we assume $\{T_k\}_{k\in\mathbb{N}}$ is monotonically non-decreasing in k. Equivalently we have $\bar{U}_k \overset{\text{i.i.d.}}{\sim} Q$ and $T_k - T_{k-1} \overset{\text{i.i.d.}}{\sim} \text{Exp}(1)$ (set $T_0 := 0$). For probability measure $P \ll Q$, we define

$$K_P(Z) = \operatorname*{arg\,min}_{k: \frac{dP}{dQ}(\bar{U}_k) > 0} \frac{T_k}{\frac{dP}{dQ}(\bar{U}_k)}, \quad \tilde{U}_P(Z) = \bar{U}_{K_P(Z)}, \tag{2}$$

where ties are broken arbitrarily. When the Poisson process is clear from the context, we usually omit Z and directly write K_P and $\tilde{U}_P$.

By the mapping theorem [11], we have that $\tilde{U}_P$ is a random variable such that $\tilde{U}_P \sim P$.

B. Generalized Poisson Matching Lemma

Lemma 1 (GPML [9])**.** Let $Z = \{\bar{U}_k, T_k\}_{k\in\mathbb{N}}$ be defined as in Definition 1. For any $P \ll Q$ and any $K \in \mathbb{N}$, we have

$$\Pr[K_P(Z) > K | \tilde{U}_P(Z)] \leq \left(1 - \left(1 + \frac{dP}{dQ}(\tilde{U}_P(Z))\right)^{-1}\right)^K. \tag{3}$$

C. Strong Functional Representation Lemma

Lemma 2 (SFRL [9], [10])**.** Let $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ be random variables such that $(X, Y) \sim P_{XY}$ with $I(X; Y) < \infty$. It is possible to construct a random variable $Z \perp\!\!\!\perp X$ and a function $g : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ such that $Y = g(X, Z)$ and

$$H(Y|Z) \leq I(X; Y) + \log(I(X; Y) + 1) + 4. \tag{4}$$

In [10], the authors propose a construction using the Poisson functional representation given in Definition 1. The random variable Z is the Poisson process $\{\bar{Y}_k, T_k\}$ where we let $Q = P_Y$, the marginal of P_{XY}. The function g is then given by

$$(X, \{\bar{Y}_k, T_k\}) \overset{g}{\mapsto} \tilde{Y}_{P_{Y|X}(\cdot|X)}.$$

III. Problem Formulation

Consider the system model depicted in Fig. 1 with block-length $n \in \mathbb{N}$. We study the hypothesis testing problem referred to as testing against independence:

$$H_0 : (X^n, Y^n) \overset{\text{i.i.d.}}{\sim} P_{XY}, \quad H_1 : (X^n, Y^n) \overset{\text{i.i.d.}}{\sim} P_X P_Y.$$

We assume that $P_{XY} \ll P_X P_Y$. Upon observing X^n, the transmitter sends a message $M \in \mathcal{M}_n$ to the receiver through a noiseless link, where $\mathcal{M}_n$ is the message set that will be specified later. The transmitter generates the message M using a possibly randomized encoding function:

$$\phi^{(n)} : \mathcal{X}^n \to \mathcal{M}_n, \quad M = \phi^{(n)}(X^n).$$

We consider two types of encoding schemes:

1) Variable-length scheme: Here $\mathcal{M}_n = \{0, 1\}^*$, i.e., the message set is the set of all finite-length binary sequences. We also require M to be prefix-free.
2) Fixed-length scheme: Here $\mathcal{M}_n$ is a finite index set. The cardinality of $\mathcal{M}_n$ is denoted by $|\mathcal{M}_n|$.

The receiver makes a decision between the hypotheses using the decision function

$$\psi^{(n)} : \mathcal{M}_n \times \mathcal{Y}^n \to \{H_0, H_1\},$$

i.e., $\hat{H}(\phi^{(n)}, \psi^{(n)}) = \psi^{(n)}(M, Y^n)$. When it is clear from context, we just write $\hat{H}$ for the decision of the receiver. We also assume the transmitter and the receiver have access to unlimited amount of common randomness.

For a hypothesis testing scheme specified by $(\phi^{(n)}, \psi^{(n)})$, the type-I and the type-II error probabilities are defined to be

$$\begin{aligned} p_{\text{I}}(\phi^{(n)}, \psi^{(n)}) &= \Pr[\hat{H} = H_1 | H = H_0], \\ p_{\text{II}}(\phi^{(n)}, \psi^{(n)}) &= \Pr[\hat{H} = H_0 | H = H_1], \end{aligned} \tag{5}$$

where H denotes the true hypothesis. We also omit $(\phi^{(n)}, \psi^{(n)})$ when it is clear from context.

A. One-Shot Setting

In the one-shot setting, we consider a single hypothesis testing scheme with block-length $n = 1$. We omit the superscripts and the subscripts n for this case and aim to characterize the achievable region defined below.

Definition 2 (One-shot achievable regions)**.** A triple (L, α, β) is said to be *fixed-length one-shot achievable* if there exists a fixed-length distributed hypothesis testing scheme (ϕ, ψ) satisfying the error probability constraint

$$p_{\mathrm{I}}(\phi, \psi) \le \alpha,\ p_{\mathrm{II}}(\phi, \psi) \le \beta, \tag{6}$$

and the message set size constraint $\log|\mathcal{M}| \le L$.

A triple (L, α, β) is said to be *variable-length one-shot achievable* if there exists a variable-length distributed hypothesis testing scheme (ϕ, ψ) satisfying the error probability constraint (6) and the expected message length constraint

$$\mathbb{E}[\ell(M)] = \mathbb{E}[\ell(\phi(X))] \le L. \tag{7}$$

The randomness of M comes from the source X and (possibly) the randomized encoding function ϕ.

We denote the set of all variable-length (resp. fixed-length) achievable triples by $\mathcal{R}_{\mathrm{VL}}$ (resp. $\mathcal{R}_{\mathrm{FL}}$).

The goal of this paper is to characterize the regions $\mathcal{R}_{\mathrm{FL}}$ and $\mathcal{R}_{\mathrm{VL}}$ and to understand the trade-off between the amount of communication and the two types of the error probabilities.

B. Asymptotic Setting

In contrast to the one-shot setting, we consider a sequence of hypothesis testing schemes $(\phi^{(n)}, \psi^{(n)})$ indexed by the block-length n. The classic results focus mainly on the following definition of optimal exponent.

Definition 3 (Optimal exponent)**.** Fix rate R and $\epsilon \in (0, 1)$, an exponent E is said to be *(ϵ, R)-fixed-length achievable* if there exists a sequence of distributed hypothesis testing scheme $\{(\phi^{(n)}, \psi^{(n)})\}_{n\in\mathbb{N}}$ satisfying

$$\begin{aligned}&\limsup_{n\to\infty} p_{\mathrm{I}}(\phi^{(n)}, \psi^{(n)}) \le \epsilon,\\ &\liminf_{n\to\infty} -\frac{1}{n}\log p_{\mathrm{II}}(\phi^{(n)}, \psi^{(n)}) \ge E.\end{aligned} \tag{8}$$

and

$$\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{M}_n| \le R. \tag{9}$$

Similarly, an exponent E is said to be (ϵ, R)*-variable-length achievable* if there exists a sequence of distributed hypothesis testing scheme $\{(\phi^{(n)}, \psi^{(n)})\}_{n\in\mathbb{N}}$ satisfying (8) and

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[\ell(\phi^{(n)}(X^n))] \le R. \tag{10}$$

The *optimal (ϵ, R)-fixed-length exponent $E_{FL}(\epsilon, R)$* (resp. *optimal (ϵ, R)-variable-length exponent $E_{\mathrm{VL}}(\epsilon, R)$*) is the supremum of all (ϵ, R)-fixed-length achievable (resp. (ϵ, R)-variable-length achievable) exponents.

Ahlswede and Csiszár provided the following single letter characterization of optimal (ϵ, R)-fixed-length exponent [1]:

$$E_{\mathrm{FL}}(\epsilon, R) = \sup_{P_{U|X}: R \ge I(U;X)} I(U;Y). \tag{11}$$

In their scheme, the transmitter sends a quantized version of its observed sequence, where quantization is performed via a test channel $P_{U|X}$ optimized subject to the rate constraint as in (11). The receiver performs a typicality test between the quantized sequence and its own observed sequence. Note that the right hand side of (11) does not depend on ϵ, i.e., this can be seen as a strong converse result.

The corresponding variable-length result is established in [5]. Somewhat surprisingly, the strong converse result fails to hold in this case:

$$E_{\mathrm{VL}}(\epsilon, R) = \sup_{P_{U|X}: R \ge (1-\epsilon) I(U;X)} I(U;Y). \tag{12}$$

Their proposed scheme makes use of the variability of the message length to modify the aforementioned scheme in [1]. Namely, whenever the transmitter's observed sequence falls into a predetermined subset with probability close to ϵ, it transmits a special message instructing the receiver to declare the alternative hypothesis.

IV. Fixed-Length Achievability

This section presents the main one-shot achievability result. We use $\mathcal{P}^\star$ to denote the set of all kernels $P_{U|X}$ such that $P_{U|X}(\cdot|X), P_{U|Y}(\cdot|Y) \ll P_U$ almost surely (with respect to P_X and P_Y respectively). For any $P_{U|X} \in \mathcal{P}^\star$, define the following region for real numbers $\gamma > 0$ and $\theta \in [0, 1]$:

$$\begin{aligned}&\mathcal{R}^{\mathrm{in}}_{\mathrm{FL}}(P_{U|X}, \gamma, \theta)\\ &\quad= \left\{(L, \alpha, \beta) : \begin{matrix}\alpha \ge P_0[i(U;Y) < \log\gamma] + (1-\theta)p_0\\ \beta \ge P_1[i(U;Y) \ge \log\gamma] + \theta p_0\end{matrix}\right\},\end{aligned} \tag{13}$$

where we use P_0 and P_1 as shorthand notation of $P_{U|X}P_{XY}$ and $P_{U|X}P_X P_Y$ respectively (in accordance with the hypotheses). The quantity p_0 is defined as

$$p_0 = \mathbb{E}\left[\left(1 - \left(1 + 2^{i(U;X)}\right)^{-1}\right)^{\lfloor 2^L \rfloor - 1}\right], \tag{14}$$

where the expectation is taken with respect to $P_{U|X}P_X$. We also define

$$\mathcal{R}^{\mathrm{in}}_{\mathrm{FL}} = \bigcup_{\substack{\gamma > 0,\, 0 \le \theta \le 1\\ P_{U|X} \in \mathcal{P}^\star}} \mathcal{R}^{\mathrm{in}}_{\mathrm{FL}}(P_{U|X}, \gamma, \theta). \tag{15}$$

The superscript "in" indicates that $\mathcal{R}^{\mathrm{in}}_{\mathrm{FL}}$ is an inner bound.

Theorem 1. For testing against independence, all triples in $\mathcal{R}_{\mathrm{FL}}$ are fixed-length achievable, i.e.,

$$\mathcal{R}^{\mathrm{in}}_{\mathrm{FL}} \subset \mathcal{R}_{\mathrm{FL}}.$$

Proof. It suffices to show that for a fixed $P_{U|X} \in \mathcal{P}^\star$ and $\gamma > 0$, there exists a fixed-length scheme (ϕ, ψ) with $|\mathcal{M}| = \lfloor 2^L \rfloor$ such that

$$\begin{aligned} p_{\mathrm{I}} &\le P_0[i(U;Y) < \log\gamma] + (1-\theta)p_0, \\ p_{\mathrm{II}} &\le P_1[i(U;Y) \ge \log\gamma] + \theta p_0. \end{aligned} \tag{16}$$

Hypothesis Testing Scheme: We apply GPML (cf. Lemma 1) in a similar fashion to channel simulation [12], [9, Remark 14]. Using the unlimited common randomness, the transmitter and the receiver generate the same samples of a Poisson point process $Z = \{\bar{U}_k, T_k\}_{k\in\mathbb{Z}}$ with intensity measure $P_U \times \lambda_{\mathbb{R}_+}$. The transmitter observes X and then transmits $M = \min\{K_{P_{U|X}(\cdot|X)}, \lfloor 2^L \rfloor\}$.

At the receiver, we consider two scenarios: if $M = \lfloor 2^L \rfloor$, the receiver flips a biased coin, i.e., it generates a sample $V \sim \mathrm{Ber}(1-\theta)$ where V is independent of (U, X, Y, Z) conditioning on the event $\{M = \lfloor 2^L \rfloor\}$, and declares $\hat{H} = H_i$ if $V = i$ for $i \in \{0, 1\}$. If $M < \lfloor 2^L \rfloor$, the receiver sets $\hat{U} = \bar{U}_M$ and performs a likelihood ratio test: namely, it declares $\hat{H} = H_0$ if $\frac{dP_{U|Y}(\cdot|Y)}{dP_U}(\hat{U}) \ge \gamma$ and $\hat{H} = H_1$ otherwise.

Error analysis: Define $U = \tilde{U}_{P_{U|X}(\cdot|X)}$. From the discussion in Section II-A, we already know that $U = \bar{U}_{K_{P_{U|X}(\cdot|X)}} \sim P_{U|X}(\cdot|X)$. Conditioning on the event $\{M < \lfloor 2^L \rfloor\}$, we have $M = K_{P_{U|X}(\cdot|X)}$. Therefore $\hat{U} = \bar{U}_M = \bar{U}_{K_{P_{U|X}(\cdot|X)}} = U$.

On the other hand, by GPML, we have

$$\begin{aligned} \Pr\left[M = \lfloor 2^L \rfloor\right] &= \mathbb{E}\left[\Pr\left[K_{P_{U|X}(\cdot|X)} > \lfloor 2^L \rfloor - 1 \,\middle|\, U\right]\right] \\ &\le \mathbb{E}\left[\left(1 - \left(1 + \frac{dP_{U|X}(\cdot|X)}{dP_U}(U)\right)^{-1}\right)^{\lfloor 2^L \rfloor - 1}\right] \\ &= \mathbb{E}\left[\left(1 - \left(1 + 2^{i(U;X)}\right)^{-1}\right)^{\lfloor 2^L \rfloor - 1}\right] = p_0. \end{aligned}$$

From the description of the scheme, we have that the distribution of M does not depend on the true hypothesis H, thus $\Pr[M = \lfloor 2^L \rfloor | H = H_0] = \Pr[M = \lfloor 2^L \rfloor | H = H_0] \le p_0$. We also have that $\Pr[V = 1 | H = H_0] = \Pr[V = 1 | H = H_1] = 1 - \theta$. The type-I error can now be upper bounded as

$$\begin{aligned} p_{\mathrm{I}} &\le \Pr[M < \lfloor 2^L \rfloor | H = H_0] \cdot \Pr[\hat{H} = H_1 | M < \lfloor 2^L \rfloor, H = H_0] \\ &\quad + \Pr[M = \lfloor 2^L \rfloor | H = H_0] \cdot \Pr[\hat{H} = H_1 | M = \lfloor 2^L \rfloor, H = H_0] \\ &\le 1 \cdot \Pr\left[\frac{dP_{U|Y}(\cdot|Y)}{dP_U}(U) < \gamma \,\middle|\, H = H_0\right] + p_0 \cdot \Pr[V = 1] \\ &\le P_0[i(U;Y) < \log\gamma] + (1-\theta)p_0, \end{aligned} \tag{17}$$

and the type-II error can be bounded similarly as

$$\begin{aligned} p_{\mathrm{II}} &\le \Pr[M < \lfloor 2^L \rfloor | H = H_1] \cdot \Pr[\hat{H} = H_0 | M < \lfloor 2^L \rfloor, H = H_1] \\ &\quad + \Pr[M = \lfloor 2^L \rfloor | H = H_1] \cdot \Pr[\hat{H} = H_0 | M = \lfloor 2^L \rfloor, H = H_1] \\ &\le 1 \cdot \Pr\left[\frac{dP_{U|Y}(\cdot|Y)}{dP_U}(U) < \gamma \,\middle|\, H = H_1\right] + p_0 \cdot \Pr[V = 0] \\ &\le P_1[i(U;Y) \ge \log\gamma] + \theta p_0. \end{aligned} \tag{18}$$

■

We provide the following lemma to upper bound the term $P_1[i(U;Y) \ge \log\gamma]$ which will be useful subsequently.

Lemma 3. $P_1[i(U;Y) \ge \log\gamma] \le \gamma^{-1} P_0[i(U;Y) \ge \log\gamma]$.

Proof. By calculation:

$$\begin{aligned} &P_1[i(U;Y) \ge \log\gamma] \\ &= \int \mathbf{1}\{i(u;y) \ge \log\gamma\} P_U(du) P_Y(dy) \\ &= \int \mathbf{1}\{i(u;y) \ge \log\gamma\} \frac{dP_U P_Y}{dP_{UY}}(u,y) P_{UY}(dudy) \\ &= \int \mathbf{1}\{i(u;y) \ge \log\gamma\} 2^{-i(u;y)} P_{UY}(dudy) \\ &\le \gamma^{-1} \int \mathbf{1}\{i(u;y) \ge \log\gamma\} P_{UY}(dudy) \\ &= \gamma^{-1} P_0[i(U;Y) \ge \log\gamma]. \end{aligned}$$

■

In the remaining part of the section, we show that the one-shot scheme of Theorem 1, when applied to the infinite block-length setting, can recover the optimal exponent given in (11).

Fix rate $R > 0$, type-I error threshold $\epsilon \in (0,1)$, and an arbitrarily small constant $\delta > 0$. Equation (11) implies the existence of a kernel $P_{U|X}$ such that for $(U, X, Y) \sim P_{U|X} P_{XY}$, we have $I(U;X) < R$ and $I(U;Y) > E_{\mathrm{FL}}(\epsilon, R) - \delta$. Consider the auxiliary random variable U^n where $P_{U^n|X^n} = P_{U|X}^n$. By Theorem 1 and Lemma 3, there exists a fixed-length scheme $(\phi^{(n)}, \psi^{(n)})$ with $|\mathcal{M}_n| = 2^{nR} + 1$ such that

$$\begin{aligned} p_{\mathrm{I}} &\le P_0^n[i(U^n;Y^n) < \log\gamma] + p_0, \\ p_{\mathrm{II}} &\le \gamma^{-1} P_0^n[i(U^n;Y^n) \ge \log\gamma], \end{aligned}$$

which is obtained by choosing $\theta = 0$. Note that $\limsup_{n\to\infty} \frac{1}{n} \log|\mathcal{M}_n| = R$. Let $\gamma = 2^{n(I(U;Y)-\delta)}$. We now analyze the type-I error. Under P^n, $\mathbb{E}[i(U^n;Y^n)] = nI(U;Y)$ and hence by law of large numbers, we have $P^n[i(U^n;Y^n) < \log\gamma] < \frac{\epsilon}{2}$ for sufficiently large $n \in \mathbb{N}$. Also

$$\begin{aligned} p_0 &= \mathbb{E}\left[\left(1 - \left(1 + 2^{i(U^n;X^n)}\right)^{-1}\right)^{2^{nR}}\right] \\ &\le \mathbb{E}\left[\left(1 - \left(1 + 2^{i(U^n;X^n) - nR}\right)^{-1}\right)\right] \\ &\le 1 - \left(1 + 2^{\mathbb{E}[i(U^n;X^n) - nR]}\right)^{-1} \\ &= 1 - \left(1 + 2^{-n(R - I(U;X))}\right)^{-1}, \end{aligned}$$

where the first inequality is by Taylor expansion, the second inequality is by Jensen's inequality (observe that $x \mapsto 1 - (1+2^x)^{-1}$ is concave on $\mathbb{R}_+$), and the last equality is because $\mathbb{E}[i(U^n;X^n)] = nI(U;X)$. Since $R - I(U;X) > 0$, we have $2^{-n(R-I(U;X))} \to 0$. Hence for sufficiently large n, $p_0 < \frac{\epsilon}{2}$. Combining, we have $p_1 < \epsilon$ whenever n is sufficiently large, and hence $\limsup_{n\to\infty} p_1 < \epsilon$. For the type-II error, we have

$$\begin{aligned} p_{\mathrm{II}} &\le 2^{-n(I(U;Y)-\delta)} \cdot P_0^n[i(U^n;Y^n) \ge n(I(U;Y) - \delta)] \\ &\le 2^{-n(E_{\mathrm{FL}}(\epsilon,R) - 2\delta)}. \end{aligned}$$

Hence, the exponent $E_{\mathrm{FL}}(\epsilon, R) - 2\delta$ is (ϵ, R)-fixed-length achievable using one-shot testing. The infinite block-length result follows, since δ can be made arbitrarily small.

V. Variable-Length Achievability

We now present the one-shot achievability result for the variable-length case. Recall that $\mathcal{P}^\star$ is the set of all kernels $P_{U|X}$ such that $P_{U|X}(\cdot|X), P_{U|Y}(\cdot|Y) \ll P_U$ almost surely (with respect to P_X and P_Y respectively). For $P_{U|X} \in \mathcal{P}^\star$ and a positive real number $\gamma > 0$, define the following region:

$$\begin{aligned}&\mathcal{R}^{\text{in}}_{\text{VL}}(P_{U|X},\gamma)\\ &= \left\{ (L,\alpha,\beta) : \begin{array}{l} L \ge I(U;X) + \log(I(U;X)+1)+5 \\ \alpha \ge P_0[i(U;Y) < \log\gamma] \\ \beta \ge P_1[i(U;Y) \ge \log\gamma] \end{array} \right\},\end{aligned}$$

where we use P_0 and P_1 as shorthand notation of $P_{U|X}P_{XY}$ and $P_{U|X}P_XP_Y$ respectively. We also define

$$\mathcal{R}^{\text{in}}_{\text{VL}} = \bigcup_{\substack{\gamma>0\\ P_{U|X}\in\mathcal{P}^\star}} \mathcal{R}^{\text{in}}_{\text{VL}}(P_{U|X},\gamma).$$

Theorem 2. For testing against independence, all triples in $\mathcal{R}_a$ are variable-length achievable, i.e.,

$$\mathcal{R}^{\text{in}}_{\text{VL}} \subset \mathcal{R}_{\text{VL}}.$$

Proof: It suffices to show that for a fixed $P_{U|X}$ and $\gamma > 0$, there exists a variable-length scheme (ϕ, ψ) such that

$$\begin{aligned}\mathbb{E}[\ell(M)] &\le I(U;X) + \log(I(U;X)+1) + 5,\\ p_{\text{I}} &\le P_0[i(U;Y) < \log\gamma], \\ p_{\text{II}} &\le P_1[i(U;Y) \ge \log\gamma].\end{aligned} \tag{19}$$

Hypothesis Testing Scheme. We apply the SFRL. By Theorem 2 and the discussion after, there exists a function g such that for the Poisson point process $Z = \{\bar{U}_k, T_k\}_{k\in\mathbb{Z}}$ with intensity $P_U \times \lambda_{\mathbb{R}_+}$, we have $U = g(X, Z)$ and

$$H(U|Z) \le I(U;X) + \log(I(U;X)+1) + 4.$$

Using the unlimited shared randomness, the transmitter and the receiver can generate the same sample Z at both terminals. Upon observing X, the transmitter generates a sample U from $P_{U|X}(\cdot|X)$ and then uses a Huffman code corresponding to distribution $P_{U|Z}(\cdot|Z)$ to encode U. In other words, the encoding function ϕ maps a sample X to the Huffman codeword of U. Averaging over the realization of Z, the expected message length is upper bounded by

$$\begin{aligned}\mathbb{E}[\ell(M)] &\le H(U|Z) + 1\\ &\le I(U;X) + \log(I(U;X)+1) + 5.\end{aligned} \tag{20}$$

At the decoder, the receiver can decode U losslessly upon receipt of the message M and knowledge of Z. Now upon observing a sample Y, the decoder performs a likelihood ratio test. Namely, the decoder declares $\hat{H} = H_0$ if $\frac{dP_{U|Y}(\cdot|Y)}{dP_U}(U) \ge \gamma$ and $\hat{H} = H_1$ otherwise.

Error analysis. Note that $U|X \sim P_{U|X}$. By a similar analysis as in (17) and (18), we obtain the last two inequalities in (19). Combining with (20) finishes the proof. ■

For the remaining part of the section, we slightly modify the one-shot scheme in Theorem 2 then recover the optimal exponent for infinite block-length case given in (12).

Fix rate $R > 0$, type-I error threshold $\epsilon \in (0,1)$ and an arbitrarily small number $\delta \in (0,\epsilon)$. Equation (12) implies the existence of a kernel $P_{U|X}$ such that for $(U,X,Y) \sim P_{U|X}P_{XY}$, we have $(1-\epsilon)I(U;X) < R$ and $I(U;Y) > E_{\text{VL}}(\epsilon,R) - \delta$. Consider the auxiliary random variable U^n where $P_{U^n|X^n} = P^n_{U|X}$. Denote the testing scheme presented in the proof of Theorem 2 by $(\phi^{(n)}, \psi^{(n)})$. We define a new testing scheme $(\tilde{\phi}^{(n)}, \tilde{\psi}^{(n)})$ as follows. Let V_n be an independent $\text{Ber}(\epsilon-\delta)$ random variable. Further, let

$$\tilde{\phi}^{(n)}(x^n) = \begin{cases} 0, & V_n = 1,\\ 1 \circ \phi^{(n)}(x^n), & V_n = 0,\end{cases}$$

where $\circ$ denotes string concatenation, and

$$\tilde{\psi}^{(n)}(m_n, y^n) = \begin{cases} H_1, & m_n = 0,\\ \psi^{(n)}(\tilde{m}_n, y^n), & m_n = 1 \circ \tilde{m}_n.\end{cases}$$

The expected message length of this scheme is given by

$$\begin{aligned}&\mathbb{E}[\ell(\tilde{\phi}(X))]\\ &= \Pr[V_n = 1]\cdot 1 + \Pr[V_n = 0]\cdot(\mathbb{E}[\ell(\phi(X))] + 1)\\ &\le 1 + (1-\epsilon+\delta)(nI(U;X) + \log(nI(U;X)+1) + 5).\end{aligned}$$

Hence $\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[\ell(\phi^{(n)}(X^n))] \le (1-\epsilon+\delta)I(U;X)$.

Let $\gamma = 2^{n(I(U;Y)-\delta)}$. We upper bound the type-I error as:

$$\begin{aligned}p_{\text{I}}(\tilde{\phi}^{(n)}, \tilde{\psi}^{(n)}) &\le p_{\text{I}}(\phi^{(n)}, \psi^{(n)}) + \Pr[V = 1]\\ &\le P_0^n[i(U^n;Y^n) < n(I(U;Y)-\delta)] + \epsilon - \delta.\end{aligned}$$

Since $\mathbb{E}[i(U^n;Y^n)] = nI(U;Y)$, by law of large numbers, $P^n[i(U^n;Y^n) < n(I(U;Y)-\delta)] < \delta$ for sufficiently large n, hence $\limsup_{n\to\infty} p_{\text{I}}(\tilde{\phi}^{(n)}, \tilde{\psi}^{(n)}) < \epsilon$. It is straightforward that the type-II error for the modified scheme is upper bounded by that of the original scheme:

$$\begin{aligned}&p_{\text{II}}(\tilde{\phi}^{(n)}, \tilde{\psi}^{(n)})\\ &\le p_{\text{II}}(\phi^{(n)}, \psi^{(n)})\\ &\le 2^{-n(I(U;Y)-\delta)} \cdot P_0^n[i(U^n;Y^n) \ge n(I(U;Y)-\delta)]\\ &\le 2^{-n(E_{\text{VL}}(\epsilon,R)-2\delta)},\end{aligned}$$

where the second inequality is by Theorem 2 and Lemma 3. From this, we have that $E_{\text{VL}}(\epsilon,R)-2\delta$ is (ϵ,R)-variable length achievable using the modified scheme. The infinite block-length result follows since δ can be arbitrarily small.

VI. Conclusion

This paper provides one-shot achievable regions for a distributed hypothesis testing problem under communication constraints for both the fixed-length and variable length cases. The results are stated for testing against independence, but can also be generalized for testing between arbitrary distributions. When extended to asymptotic i.i.d. case, these results recover the previously known asymptotic optimal error exponents for testing against independence.

References

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

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

[3] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *Proc. IEEE Int. Symp. Inf. Theory*, 1994, p. 114.

[4] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6282–6303, 2012.

[5] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 3, pp. 681–694, 2020.

[6] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 2–14, 2006.

[7] N. Weinberger and Y. Kochman, "On the reliability function of distributed hypothesis testing under optimal detection," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4940–4965, 2019.

[8] S. Watanabe, "Neyman–pearson test for zero-rate multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 4923–4939, 2017.

[9] C. T. Li and V. Anantharam, "A unified framework for one-shot achievability via the poisson matching lemma," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2624–2651, 2021.

[10] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

[11] J. F. C. Kingman, *Poisson processes*. Clarendon Press, 1992, vol. 3.

[12] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.