PAPER *Special Section on Information Theory and Its Applications*

The Error Exponent of Zero-Rate Multiterminal Hypothesis Testing for Sources with Common Information*

Makoto UEDA†, ***Nonmember and*** **Shigeaki KUZUOKA**††a), ***Senior Member***

SUMMARY The multiterminal hypothesis testing problem with zero-rate constraint is considered. For this problem, an upper bound on the optimal error exponent is given by Shalaby and Papamarcou, provided that the positivity condition holds. Our contribution is to prove that Shalaby and Papamarcou's upper bound is valid under a weaker condition: (i) two remote observations have a common random variable in the sense of Gács and Körner, and (ii) when the value of the common random variable is fixed, the conditional distribution of remaining random variables satisfies the positivity condition. Moreover, a generalization of the main result is also given.

key words: *common information, distributed detection, error exponent, multiterminal hypothesis testing, zero-rate compression*

1. Introduction

We consider the multiterminal hypothesis testing problem depicted in Fig. 1. A pair (x^n, y^n) of sequences of length n is emitted from one of the two distributions P_{XY} (null hypothesis) or $P_{\overline{XY}}$ (alternative hypothesis). At the site A, only the sequence x^n is observed and compressed to a codeword at the rate R_X. Similarly, at the site B, y^n is observed and compressed at the rate R_Y. The site C determines whether to accept or reject the null hypothesis based on received two codewords.

The main problem of the multiterminal hypothesis testing is to determine the *optimal error exponent* [1]. In the special case of the rate constraint called *zero-rate*, i.e., the case where $R_X = 0$ and/or $R_Y = 0$, Han [2] proved that the optimal error exponent is lower bounded by

$$\min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}} \| P_{\overline{XY}}) \tag{1}$$

where the minimum is taken over all possible distributions $P_{\tilde{X}\tilde{Y}}$ satisfying $P_{\tilde{X}} = P_X$ and $P_{\tilde{Y}} = P_Y$. On the other hand, Shalaby and Papamarcou [3] proved that Han's lower bound is tight under the *positivity condition*

$$P_{\overline{XY}}(x, y) > 0, \quad (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{2}$$

In this paper, the main argument is devoted to weakening the positivity condition. In the proof given by Shalaby and Papamarcou [3], the positivity condition is needed in order to invoke the blowing-up lemma [4]. Our key idea is that the blowing-up lemma holds for independent but *not* necessarily identical distributions, although Shalaby and Papamarcou use the lemma for i.i.d. distribution. Based on this idea, we prove that the error exponent is upper bounded by (1) if the alternative hypothesis $P_{\overline{XY}}$ satisfies following two conditions:

1. Two separate observations $\overline{X}$ and $\overline{Y}$ have a common random variable $\overline{U}$ in the sense of Gács and Köner; i.e., there are two random variables $\overline{U}_0$ and $\overline{U}_1$ such that $\overline{X} = (\overline{U}, \overline{U}_0)$, $\overline{Y} = (\overline{U}, \overline{U}_1)$.
2. When the value of the common random variable $\overline{U}$ is fixed, the conditional distribution of remaining satisfies the positivity condition; i.e., for all triplets (u, u_0, u_1), $P_{\overline{U}_0\overline{U}_1|\overline{U}}(u_0, u_1|u) > 0$.

Furthermore, we give a generalization of the result and show that the error exponent can be determined even though the above two conditions are not satisfied.

The rest of the paper is organized as follows. In Sect. 2, we introduce notations used in this paper. In Sect. 3, we formulate the multiterminal hypothesis testing problem and explain known results. Then, we state our main theorem in Sect. 4. The proof of the main theorem is given in Sect. 5. A generalization of the main theorem is given in Sect. 6. Section 7 concludes the paper.

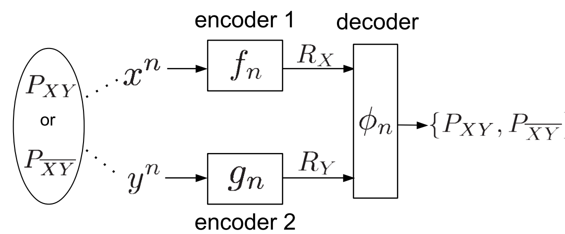

Fig. 1 Multiterminal hypothesis testing.

2. Notations

Throughout this paper, random variables (e.g. X) and their realizations (e.g. x) are denoted by capital and lower case letters respectively. All random variables take values in finite alphabets which are denoted by the respective calligraphic letters (e.g. $\mathcal{X}$). For a finite set $\mathcal{X}$, the nth Cartesian product is denoted by $\mathcal{X}^n$ and a sequence $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ of the length n is denoted by x^n. For a random variable

Manuscript received February 12, 2015.
Manuscript revised May 26, 2015.
†The author is with the Graduate School of Systems Engineering, Wakayama University, Wakayama-shi, 640-8510 Japan.
††The author is with the Faculty of System Engineering, Wakayama University, Wakayama-shi, 640-8510 Japan.
*This work was presented in part at 2014 International Symposium on Information Theory and its Applications (ISITA2014).
a) E-mail: kuzuoka@ieee.org
DOI: 10.1587/transfun.E98.A.2384

Copyright © 2015 The Institute of Electronics, Information and Communication Engineers

(e.g. X), the distribution is denoted by P with the subscript (e.g. P_X). For a distribution P_X, we define P_X^n by

$$P_X^n(x^n) \triangleq \prod_{i=1}^{n} P_X(x_i), \quad x^n \in \mathcal{X}^n.$$

Similarly, for a conditional distribution $V : \mathcal{X} \to \mathcal{Y}$, we define V^n by

$$V^n(y^n|x^n) \triangleq \prod_{i=1}^{n} V(y_i|x_i), \quad (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n.$$

For information quantities, we will use standard notations as in [5], [6]: the *entropy* is denoted by $H(X)$ or $H(P_X)$, the *divergence* is denoted by $D(P_X\|P_{\tilde{X}})$ or $D(X\|\tilde{X})$, and the *conditional divergence* is denoted by $D(V\|W|P_X)$ or $D(V\|W|X)$. Throughout this paper, log denotes the natural logarithm.

3. Problem and Known Results

Let us consider two distributions P_{XY} and $P_{\overline{XY}}$ on $\mathcal{X} \times \mathcal{Y}$. The distribution P_{XY} is called the *null hypothesis*, and $P_{\overline{XY}}$ is called the *alternative hypothesis*. A pair (x^n, y^n) of sequences is emitted from P_{XY}^n or $P_{\overline{XY}}^n$. We assume

$$D(P_{XY}\|P_{\overline{XY}}) < \infty. \tag{3}$$

There are two separate remote sites A and B. At the site A, only x^n is observed and compressed by the *encoder function*

$$f_n : \mathcal{X}^n \to \mathcal{M}_n \triangleq \{1, 2, 3, \ldots, M_n\}.$$

Similarly, at the site B, observed y^n is compressed by

$$g_n : \mathcal{Y}^n \to \mathcal{N}_n \triangleq \{1, 2, 3, \ldots, N_n\}.$$

The *rate* R_X (resp. R_Y) of the encoder function f_n (resp. g_n) is defined by

$$R_X \triangleq \limsup_{n\to\infty} \frac{1}{n} \log M_n, \quad R_Y \triangleq \limsup_{n\to\infty} \frac{1}{n} \log N_n. \tag{4}$$

We consider the hypothesis testing problem under compression at (asymptotically) *zero-rate*. In other words, we assume that the codebook sizes M_n, N_n satisfy constraints of the type $R_X = 0$ and/or $R_Y = 0$.

The compressed sequences are sent to the center C where the unknown distribution is estimated by the *decoder function*

$$\phi_n : \mathcal{M}_n \times \mathcal{N}_n \to \left\{P_{XY}, P_{\overline{XY}}\right\}.$$

A triplet (f_n, g_n, ϕ_n) of encoders f_n, g_n and a decoder ϕ_n is called a code. Given a code (f_n, g_n, ϕ_n), the *acceptance region* $\mathcal{A}_n$ is defined by

$$\mathcal{A}_n \triangleq \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \phi_n(f_n(x^n), g_n(y^n)) = P_{XY}\}.$$

Then, probabilities of the error of the first kind α_n and the second kind β_n are defined by

$$\alpha_n \triangleq P_{XY}^n(\mathcal{A}_n^c), \quad \beta_n \triangleq P_{\overline{XY}}^n(\mathcal{A}_n)$$

where $\mathcal{A}_n^c$ is the complementary set of $\mathcal{A}_n$. We impose the *constant-type* constraint on α_n as

$$\alpha_n \le \varepsilon \tag{5}$$

where $0 < \varepsilon < 1$ is an arbitrarily fixed constant. Let $\beta_n^*(\varepsilon, R_X, R_Y|XY, \overline{XY})$ denote the minimum of β_n over all possible codes (f_n, g_n, ϕ_n) satisfying conditions (4) and (5). Then, let us define

$$\sigma(\varepsilon, R_X, R_Y|XY, \overline{XY}) \triangleq \liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n^*(\varepsilon, R_X, R_Y|XY, \overline{XY})}$$

which is called the *optimal error exponent* for the multiterminal hypothesis testing. Instead of $\beta_n^*(\varepsilon, R_X, R_Y|XY, \overline{XY})$ (resp. $\sigma(\varepsilon, R_X, R_Y|XY, \overline{XY})$), we will use the notation $\beta_n^*(\varepsilon, R_X, R_Y)$ (resp. $\sigma(\varepsilon, R_X, R_Y)$), if sources XY and $\overline{XY}$ are clear from the context.

As mentioned in Sect. 1, a lower bound on $\sigma(\varepsilon, R_X, R_Y)$ for the case $R_X = 0$ was given by Han.

Proposition 1 (Direct part of Theorem 5 of [2]) Assume that (3) holds. For all $0 < \varepsilon < 1$ and $R \ge 0$,

$$\sigma(\varepsilon, 0, R) \ge \min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}).$$

Remark 1 In [2], Han gave an upper bound of the error exponent under the 0_2 rate (i.e., the codebook size $M_n = 2$): If (3) holds then there exists some constant $0 < \varepsilon_0 \le 1$ such that, for all $0 < \varepsilon < \varepsilon_0$ and $R \ge 0$,

$$\sigma(\varepsilon, 0_2, R) \le \min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}).$$

On the other hand, Shalaby and Papamarcou proved that Han's bound is tight under the positivity condition.

Proposition 2 (Theorem 1 of [3]) Assume that (2) and (3) hold. Then, for all $0 < \varepsilon < 1$ and $R \ge 0$,

$$\sigma(\varepsilon, 0, R) \le \min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}).$$

4. Main Theorem

Instead of the positivity condition (2), we consider the following condition.

Condition 1 (Conditional Positivity Condition) There are finite sets $\mathcal{U}$, $\mathcal{U}_0$, $\mathcal{U}_1$ and a triplet of random variables $(\overline{U}, \overline{U}_0, \overline{U}_1)$ on $\mathcal{U} \times \mathcal{U}_0 \times \mathcal{U}_1$ satisfying the following two conditions:

Table 1 Example of $P_{\overline{XY}}$.

$\overline{x} \setminus \overline{y}$	00	01	10	11
00	p_1	p_2	0	0
01	p_3	p_4	0	0
10	0	0	p_5	p_6
11	0	0	p_7	p_8

1. $(\overline{X}, \overline{Y})$ is written as
$$\overline{X} = (\overline{U}, \overline{U}_0), \quad \overline{Y} = (\overline{U}, \overline{U}_1). \tag{6}$$
2. $P_{\overline{U}_0\overline{U}_1|\overline{U}}$ satisfies
$$P_{\overline{U}_0\overline{U}_1|\overline{U}}(u_0, u_1|u) > 0, \quad (u, u_0, u_1) \in \mathcal{U} \times \mathcal{U}_0 \times \mathcal{U}_1. \tag{7}$$

Remark 2 The positivity condition (2) can be considered as a special case of Condition 1. Indeed, if (2) holds, we can choose so that $|\mathcal{U}| = 1$, $\mathcal{U}_0 = \mathcal{X}$, and $\mathcal{U}_1 = \mathcal{Y}$, where $|\cdot|$ denotes the cardinality of the set.

Example 1 We show an example of a distribution $P_{\overline{XY}}$ satisfying Condition 1 in Table 1, where $p_i > 0$ $(i = 1, \ldots, 8)$. Since $\{00, 01, 10, 11\} = \{0, 1\} \times \{0, 1\}$, it is easy to verify that $P_{\overline{XY}}$ satisfies Condition 1. It should be noted that the distribution $P_{\overline{XY}}$ does not satisfy (2).

Now, we state our main theorem.

Theorem 1 Assume that (3) holds and that the alternative hypothesis $P_{\overline{XY}}$ satisfies Condition 1. Then, for all $0 < \varepsilon < 1$ and $R \geq 0$,
$$\sigma(\varepsilon, 0, R) \leq \min_{\substack{P_{\tilde{X}} = P_X \\ P_{\tilde{Y}} = P_Y}} D(P_{\tilde{X}\tilde{Y}} \| P_{\overline{XY}}).$$

Remark 3 Let us consider a case where $p_6 = p_8 = 0$ in Table 1 of Example 1. In this case, $P_{\overline{XY}}$ does not satisfy (7) but satisfies
$$P_{\overline{U}_0\overline{U}_1|\overline{U}}(u_0, u_1|u) = 0 \Rightarrow P_{\overline{U}_0|\overline{U}}(u_0|u) P_{\overline{U}_1|\overline{U}}(u_1|u) = 0. \tag{8}$$
By modifying the proof given in the next section, we can show that Theorem 1 holds even if (7) is replaced with (8).

Remark 4 Suppose the assumptions of Theorem 1 hold. Then, we can show that there exists a triplet of the random variables (U, U_0, U_1) on $\mathcal{U} \times \mathcal{U}_0 \times \mathcal{U}_1$ such that
$$X = (U, U_0), \quad Y = (U, U_1). \tag{9}$$
Equations (6) and (9) mean that the observed sequences x^n and y^n can be written as
$$x^n = (u^n, u_0^n), \quad y^n = (u^n, u_1^n),$$
where $(u^n, u_0^n, u_1^n) \in \mathcal{U}^n \times \mathcal{U}_0^n \times \mathcal{U}_1^n$. In other words, u^n is observed at both A and B as the *common information*, while u_0^n (resp. u_1^n) is observed only at A (resp. B).

5. Proof of Theorem 1

In the proof, we will use the method of types and the blowing-up lemma. Before proving our main theorem, we introduce several lemmas used in the proof.

5.1 Preliminary for the Proof

The *type* P_{x^n} of x^n is defined by
$$P_{x^n}(a) \triangleq \frac{1}{n}|\{i : x_i = a\}|, \quad a \in \mathcal{X}.$$
The set of all types of sequences of length n is denoted by $\mathcal{P}_n(\mathcal{X})$.

For a number η_n and a random variable X, the typical set $\mathcal{T}_{\eta_n}^{(n)}(X)$ is defined by
$$\mathcal{T}_{\eta_n}^{(n)}(X) \triangleq \{x^n \in \mathcal{X}^n : D(P_{x^n} \| P_X) < \eta_n\}.$$
Throughout this paper, we fix η_n such that
$$\eta_n > 0, \quad \eta_n \to 0, \quad n\eta_n / \log n \to \infty, \quad (n \to \infty). \tag{10}$$

We say that y^n has *conditional type* $V : \mathcal{X} \to \mathcal{Y}$ given x^n if
$$|\{i : (x_i, y_i) = (a, b)\}| = V(b|a)|\{i : x_i = a\}|, \quad (a, b) \in \mathcal{X} \times \mathcal{Y}.$$
The set of all conditional types of given x^n is denoted by $\mathcal{V}_n(\mathcal{Y}|x^n)$.

Given x^n and a conditional type $V : \mathcal{X} \to \mathcal{Y}$, the *V-shell* given x^n is defined by
$$\mathcal{T}^{(n)}(V|x^n) \triangleq \{y^n \in \mathcal{Y}^n : |\{i : (x_i, y_i) = (a, b)\}| = V(b|a)|\{i : x_i = a\}|\}.$$
Given a conditional distribution $W : \mathcal{X} \to \mathcal{Y}$ and x^n, the conditional typical set $\mathcal{T}_{\eta_n}^{(n)}(W|x^n)$ is defined by
$$\mathcal{T}_{\eta_n}^{(n)}(W|x^n) = \bigcup_{\substack{V \in \mathcal{V}_n(\mathcal{Y}|x^n) \\ D(V \| W | P_{x^n}) \leq \eta_n}} \mathcal{T}^{(n)}(V|x^n) \tag{11}$$
where η_n satisfies (10).

In the proof, we will use the following properties (For proofs of lemmas, see [5], [7]).

Lemma 1 For any sequence $x^n \in \mathcal{X}^n$, we have
$$P_X^n(x^n) = \exp(-n(H(P_{x^n}) + D(P_{x^n} \| P_X))) \tag{12}$$
$$\leq \exp(-nH(P_{x^n})). \tag{13}$$

Lemma 2 For any distribution P_X, there exists $\epsilon_n \to 0$ such that
$$P_X^n(\mathcal{T}_{\eta_n}^{(n)}(X)) \geq 1 - \epsilon_n.$$

Lemma 3 For two conditional distributions $V : \mathcal{X} \to \mathcal{Y}$,

$W : \mathcal{X} \to \mathcal{Y}$ and a sequence x^n such that $V \in \mathcal{V}_n(\mathcal{Y}|x^n)$, we have

$$W^n(\mathcal{T}^{(n)}(V|x^n)|x^n) \geq \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \exp\left(-nD(V\|W|P_{x^n})\right).$$

Lemma 4 For any sequence $x^n \in \mathcal{X}^n$ and any conditional distribution $V : \mathcal{X} \to \mathcal{Y}$, there exists $\epsilon_n \to 0$ such that

$$V^n(\mathcal{T}^{(n)}_{\eta_n}(V|x^n)|x^n) \geq 1 - \epsilon_n.$$

A key idea in our proof is that the blowing-up lemma holds for independent but *not* necessarily identical distributions. Indeed, the following lemma holds [4] (see also [5, Lemma 5.4]).

Lemma 5 (Blowing-up Lemma) Fix a conditional distribution $V : \mathcal{X} \to \mathcal{Y}$ and $x^n \in \mathcal{X}^n$. For any set $C_n \subseteq \mathcal{Y}^n$ and sequence $\delta_n \to 0$ $(n \to \infty)$, there exists a sequence of positive integers k_n with $k_n/n \to 0$ $(n \to \infty)$ and a sequence $\gamma_n \to 0$ $(n \to \infty)$ such that

$$V^n(C|x^n) \geq \exp(-n\delta_n) \implies V^n(\Gamma^{k_n}C|x^n) \geq 1 - \gamma_n,$$

where

$$\Gamma^{k_n}C_n \triangleq \{u^n \in \mathcal{X}^n : d(x^n, u^n) \leq k_n \text{ for some } x^n \in C_n\}$$

and $d(x^n, u^n) \triangleq |\{i : x_i \neq u_i\}|$.

5.2 Proof of Theorem 1

We prove the theorem by 4 steps. The proof arguments presented in this paper exactly parallel those in the case with the positivity condition shown in [3]. In particular, in Step 3 of the proof, we use the technique developed in [3].

Step 1) At first, we introduce conditional distributions, which are used in our proof. Since $P_{\overline{XY}}$ satisfies Condition 1, there exists a triplet of random variables $(\overline{U}, \overline{U_0}, \overline{U}_1)$ satisfying

$$\overline{X} = (\overline{U}, \overline{U_0}), \quad \overline{Y} = (\overline{U}, \overline{U}_1). \tag{14}$$

From $(\overline{U}, \overline{U}_0, \overline{U}_1)$, we define $\overline{V}_0 : \mathcal{U} \to \mathcal{U}_0$, $\overline{V}_1 : \mathcal{U} \to \mathcal{U}_1$, and $\overline{W} : \mathcal{U} \to \mathcal{U}_0 \times \mathcal{U}_1$ as

$$\overline{V}_0 = P_{\overline{U}_0|\overline{U}}, \quad \overline{V}_1 = P_{\overline{U}_1|\overline{U}}, \quad \overline{W} = P_{\overline{U}_0\overline{U}_1|\overline{U}}.$$

It should be noted that $\overline{W}(u_0, u_1|u) \neq \overline{V}_0(u_0|u)\overline{V}_1(u_1|u)$ in general.

Similarly, as mentioned in Remark 4, there exists (U, U_0, U_1) satisfying

$$X = (U, U_0), \quad Y = (U, U_1). \tag{15}$$

From (U, U_0, U_1), we define $V_0 : \mathcal{U} \to \mathcal{U}_0$, $V_1 : \mathcal{U} \to \mathcal{U}_1$, and $W : \mathcal{U} \to \mathcal{U}_0 \times \mathcal{U}_1$ as

$$V_0 = P_{U_0|U}, \quad V_1 = P_{U_1|U}, \quad W = P_{U_0U_1|U}. \tag{16}$$

Fix a pair of random variables $(\tilde{X}, \tilde{Y})$ satisfying

$$P_{\tilde{X}} = P_X, \quad P_{\tilde{Y}} = P_Y \tag{17}$$

and $D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}) < \infty$. Then, in the same manner as the above, we can show that there exists $(\tilde{U}, \tilde{U}_0, \tilde{U}_1)$ such that

$$\tilde{X} = (\tilde{U}, \tilde{U}_0), \quad \tilde{Y} = (\tilde{U}, \tilde{U}_1). \tag{18}$$

From $(\tilde{U}, \tilde{U}_0, \tilde{U}_1)$, we define $\tilde{V}_0 : \mathcal{U} \to \mathcal{U}_0$, $\tilde{V}_1 : \mathcal{U} \to \mathcal{U}_1$ and $\tilde{W} : \mathcal{U} \to \mathcal{U}_0 \times \mathcal{U}_1$ as

$$\tilde{V}_0 = P_{\tilde{U}_0|\tilde{U}}, \quad \tilde{V}_1 = P_{\tilde{U}_1|\tilde{U}}, \quad \tilde{W} = P_{\tilde{U}_0\tilde{U}_1|\tilde{U}}. \tag{19}$$

Since (15), (16), (17), (18), and (19) hold, we have

$$\tilde{U} = U, \quad \tilde{V}_0 = V_0, \quad \tilde{V}_1 = V_1, \quad \tilde{W} = W. \tag{20}$$

Step 2) We bound the probabilities of the error by using random variables introduced in Step 1. From definitions of the rate and the optimal error exponent, $\sigma(\varepsilon, 0, R) \leq \sigma(\varepsilon, 0, \infty)$ holds. Hence, it is enough if we prove that

$$\sigma(\varepsilon, 0, \infty) \leq \min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}). \tag{21}$$

When the rate R_Y is sufficiently large and $R_Y \geq \log|\mathcal{Y}|$, we can choose the encoder g_n such that $g_n(y^n) = y^n$. Further, as mentioned in Remark 4, we can write

$$x^n = (u^n, u_0^n), \quad y^n = (u^n, u_1^n).$$

Thus, the acceptance region $\mathcal{A}_n$ can be written as

$$\mathcal{A}_n = \{((u^n, u_0^n),(u^n, u_1^n)) \in (\mathcal{U}^n \times \mathcal{U}_0^n) \times (\mathcal{U}^n \times \mathcal{U}_1^n) : \phi_n(f_n((u^n, u_0^n)), (u^n, u_1^n)) = P_{XY}\}.$$

Let the set $\mathcal{A}'_n \subseteq \mathcal{U}^n \times \mathcal{U}_0^n \times \mathcal{U}_1^n$ be

$$\mathcal{A}'_n = \{(u^n, u_0^n, u_1^n) : ((u^n, u_0^n), (u^n, u_1^n)) \in \mathcal{A}_n\}.$$

Then, we have

$$P^n_{XY}(\mathcal{A}_n) = P^n_{UU_0U_1}(\mathcal{A}'_n). \tag{22}$$

Farther, for each $u^n \in \mathcal{U}^n$, let the set $\mathcal{A}_n(u^n) \subseteq \mathcal{U}_0^n \times \mathcal{U}_1^n$ be

$$\begin{aligned}\mathcal{A}_n(u^n) &= \{(u_0^n, u_1^n) : \phi_n(f_n((u^n, u_0^n)), (u^n, u_1^n)) = P_{XY}\} \\ &= \left\{(u_0^n, u_1^n) : (u^n, u_0^n, u_1^n) \in \mathcal{A}'_n\right\}.\end{aligned}$$

Then $\mathcal{A}'_n$ can be written as

$$\mathcal{A}'_n = \bigcup_{u^n \in \mathcal{U}^n} \{u^n\} \times \mathcal{A}_n(u^n). \tag{23}$$

From (16), (22), and (23), we can rewrite (5) as

$$\varepsilon \geq P^n_{XY}(\mathcal{A}^c_n) = \sum_{u^n \in \mathcal{U}^n} P^n_U(u^n) W^n(\mathcal{A}'^c_n(u^n)|u^n).$$

Letting

$$\mathcal{U}^n_{[\varepsilon]} \triangleq \left\{u^n \in \mathcal{U}^n : W^n(\mathcal{A}'^c_n(u^n)|u^n) \le \sqrt{\varepsilon}\right\},$$

and using Markov's inequality, we can show that

$$1 - \sqrt{\varepsilon} < P^n_U(\mathcal{U}^n_{[\varepsilon]}). \tag{24}$$

On the other hand, we have

$$\begin{aligned}\beta_n =& P^n_{\overline{UU_0U_1}}(\mathcal{A}'_n)\\ \ge& \sum_{u^n \in \mathcal{U}^n_{[\varepsilon]}} P^n_{\overline{UU_0U_1}}(u^n, \mathcal{A}'_n(u^n)).\end{aligned} \tag{25}$$

Step 3) We apply similar argument as [3] for a fixed $u^n \in \mathcal{U}^n_{[\varepsilon]}$. Define $C_i(u^n)$ and $F_i(u^n)$ by

$$C_i(u^n) = \left\{u^n_0 \in \mathcal{U}^n_0 : f_n((u^n, u^n_0)) = i\right\}$$
$$F_i(u^n) = \left\{u^n_1 \in \mathcal{U}^n_1 : \phi_n(i, (u^n, u^n_1)) = P_{XY}\right\}$$

where $C_i(u^n) \cap C_j(u^n) = \emptyset$ for all $i \ne j$. We can write

$$\mathcal{A}_n(u^n) = \bigcup_{i=1}^{M_n} C_i(u^n) \times F_i(u^n).$$

Since $u^n \in \mathcal{U}^n_{[\varepsilon]}$, there exists an index i_0 such that

$$W^n(C_{i_0}(u^n) \times F_{i_0}(u^n)|u^n) \ge \frac{1-\sqrt{\varepsilon}}{M_n}.$$

Letting $C(u^n) = C_{i_0}(u^n)$ and $F(u^n) = F_{i_0}(u^n)$, we can rewrite it as

$$W^n(C(u^n) \times F(u^n)|u^n) \ge \exp(-n\delta_n) \tag{26}$$

where $\delta_n \triangleq -\frac{1}{n}\log(1-\sqrt{\varepsilon}) + \frac{1}{n}\log M_n$ and $\delta_n \to 0$. Equation (26) clearly implies that

$$V^n_0(C(u^n)|u^n) \ge \exp(-n\delta_n).$$

Thus, from Lemma 5, there exists k_n and γ_n such that

$$V^n_0(\Gamma^{k_n}C(u^n)|u^n) \ge 1 - \gamma_n. \tag{27}$$

From (20) and (27), we have

$$\tilde{V}^n_0(\Gamma^{k_n}C(u^n)|u^n) \ge 1 - \gamma_n. \tag{28}$$

Similarly, we have

$$\tilde{V}^n_1(\Gamma^{k_n}F(u^n)|u^n) \ge 1 - \gamma_n. \tag{29}$$

Equations (28) and (29) imply that

$$\tilde{W}^n(\Gamma^{k_n}C(u^n) \times \Gamma^{k_n}F(u^n)|u^n) \ge 1 - 2\gamma_n.$$

Hence, from Lemma 4, for all sufficiently large n, we have

$$\tilde{W}^n\left((\Gamma^{k_n}C(u^n) \times \Gamma^{k_n}F(u^n)) \cap \mathcal{T}^{(n)}_{\eta_n^2}(\tilde{W})\middle| u^n\right) \ge \frac{1}{2}. \tag{30}$$

From (11) and (30), there exists a conditional type $\hat{W}$ such that

$$D(\hat{W}\|\tilde{W}|P_{u^n}) \le \eta_n^2, \quad \hat{W} \in \mathcal{V}_n(\mathcal{Y}|u^n) \tag{31}$$

and

$$\frac{|(\Gamma^{k_n}C(u^n) \times \Gamma^{k_n}F(u^n)) \cap \mathcal{T}^{(n)}(\hat{W}|u^n)|}{|\mathcal{T}^{(n)}(\hat{W}|u^n)|} \ge \frac{1}{2}. \tag{32}$$

From (32), we have

$$\overline{W}^n(\Gamma^{k_n}C(u^n) \times \Gamma^{k_n}F(u^n)|u^n) \ge \frac{1}{2}\overline{W}^n(\mathcal{T}^{(n)}(\hat{W}|u^n)|u^n). \tag{33}$$

On the other hand, from the definition of Γ^{k_n}, there exists at least one element $(u^n_0, u^n_1) \in C(u^n) \times F(u^n)$ such that (u^n_0, u^n_1) differs from $(s^n, t^n) \in \Gamma^{k_n}C(u^n) \times \Gamma^{k_n}F(u^n)$ for at most $2k_n$. Moreover, from Conditional 1, we have $\overline{W}(u_{0;i}, u_{1;i}|u_i) > 0$, where $u_{0;i}$ (resp. $u_{1;i}$) is the ith symbol of u^n_0 (resp. u^n_1). Hence, we have

$$\overline{W}^n(s^n, t^n|u^n) = \prod_{i=1}^{n}\overline{W}(s_i, t_i|u_i) \le \rho^{-2k_n}\overline{W}^n(u^n_0, u^n_1|u^n) \tag{34}$$

where $\rho \triangleq \min\left\{\overline{W}(u_0, u_1|u) > 0 : (u, u_0, u_1) \in \mathcal{U} \times \mathcal{U}_0 \times \mathcal{U}_1\right\}$. As (s^n, t^n) ranges over $\Gamma^{k_n}C(u^n) \times \Gamma^{k_n}F(u^n)$, each element $(u^n_0, u^n_1) \in C(u^n) \times F(u^n)$ will be selected at most $|\Gamma^{k_n}\{u^n_0\}||\Gamma^{k_n}\{u^n_1\}|$ times.[†] Thus, we have

$$\begin{aligned}&\overline{W}^n(\Gamma^{k_n}C(u^n) \times \Gamma^{k_n}F(u^n)|u^n)\\ &\quad\le \sum_{(s^n,t^n)\in\Gamma^{k_n}C(u^n)\times\Gamma^{k_n}F(u^n)} \rho^{-2k_n}\overline{W}^n(u^n_0, u^n_1|u^n) &(35)\\ &\quad\le |\Gamma^{k_n}\{u^n_0\}||\Gamma^{k_n}\{u^n_1\}|\rho^{-2k_n}\overline{W}^n(C(u^n) \times F(u^n)|u^n)\\ &\quad\le \exp(n\varepsilon_1(n))\overline{W}^n(C(u^n) \times F(u^n)|u^n) &(36)\end{aligned}$$

where (35) follows from (34), and $\varepsilon_1(n) \triangleq 2h\left(\frac{k_n}{n}\right) + \frac{k_n}{n}(\log|\mathcal{U}_0| + \log|\mathcal{U}_1|) - \frac{2k_n}{n}\log\rho \to 0$ and $h(p)$ is the binary entropy. Combining (33) and (36), we have

$$\begin{aligned}&\overline{W}^n(C(u^n) \times F(u^n)|u^n)\\ &\quad\ge \frac{1}{2}\exp(-n\varepsilon_1(n))\overline{W}^n(\mathcal{T}^{(n)}(\hat{W}|u^n))\\ &\quad\ge \exp(-n(D(\hat{W}\|\overline{W}|P_{u^n}) + \varepsilon_2(n))) &(37)\end{aligned}$$

where (37) follows from Lemma 3 and $\varepsilon_2(n) \triangleq \varepsilon_1(n) + \frac{\log 2}{n} + \frac{|\mathcal{U}||\mathcal{U}_0||\mathcal{U}_1|}{n}\log(n+1) \to 0$.

Hence, we have

$$\begin{aligned}&P^n_{\overline{UU_0U_1}}(u^n, C(u^n) \times F(u^n))\\ &= \overline{W}^n(C(u^n) \times F(u^n)|u^n)P^n_{\overline{U}}(u^n)\\ &\ge P^n_{\overline{U}}(u^n)\exp\left(-n(D(\hat{W}\|\overline{W}|P_{u^n})) + \varepsilon_2(n)\right) &(38)\\ &= \exp(-nH(P_{u^n}))\\ &\quad\times\exp\left(-n(D(\hat{W}\|\overline{W}|P_{u^n}) + D(P_{u^n}\|P_{\overline{U}})) + \varepsilon_2(n)\right) &(39)\end{aligned}$$

[†]Note that, by the definition of Γ^{k_n}, we have $\Gamma^{k_n}\{u^n_j\} = \{\tilde{u}^n_j \in \mathcal{U}^n_j : d(\tilde{u}^n_j, u^n_j) \le k_n\}$ for $j = 0, 1$.

$$= \exp(-nH(P_{u^n})) \times \exp\left(-n(D(\hat{U}\hat{U}_0\hat{U}_1\|\overline{U}\,\overline{U}_0\overline{U}_1)) + \varepsilon_2(n)\right) \quad (40)$$

where (38) follows from (37), (39) follows from (12), and (40) follows from the chain rule for divergence [5]:

$$D(\hat{W}\|\overline{W}|P_{u^n}) + D(P_{u^n}\|P_{\overline{U}}) = D(\hat{U}\hat{U}_0\hat{U}_1\|\overline{U}\,\overline{U}_0\overline{U}_1)$$

where $\hat{U}, \hat{U}_0, \hat{U}_1$ are random variables defined by

$$P_{\hat{U}\hat{U}_0\hat{U}_1}(u, u_0, u_1) = P_{u^n}(u)\hat{W}(u_0, u_1|u). \quad (41)$$

Step 4) Combining (25) and (40), we have

$$\beta_n \geq \sum_{u^n \in \mathcal{U}^n_{[\varepsilon]}} \exp\left(-n(D(\hat{U}\hat{U}_0\hat{U}_1\|\overline{UU_0U_1}) + \varepsilon_2(n))\right) \times \exp(-nH(P_{u^n}))$$
$$\geq \sum_{u^n \in \mathcal{U}^n_{[\varepsilon]} \cap \mathcal{T}^{(n)}_{\eta_n}(\tilde{U})} P_{\tilde{U}}(u^n) \times \exp\left(-n(D(\hat{U}\hat{U}_0\hat{U}_1\|\overline{UU_0U_1}) + \varepsilon_2(n))\right) \quad (42)$$

where (42) follows from (13). Recall that $(\hat{U}, \hat{U}_0, \hat{U}_1)$ is defined by (41), and $\hat{W}$ satisfies (31). Thus, for $u^n \in \mathcal{U}^n_{[\varepsilon]} \cap \mathcal{T}^{(n)}_{\eta_n}(\tilde{U})$, there exists $\xi_n \to 0$ such that

$$\left|D(\hat{U}\hat{U}_0\hat{U}_1\|\overline{UU_0U_1}) - D(\tilde{U}\tilde{U}_0\tilde{U}_1\|\overline{UU_0U_1})\right| \leq \xi_n. \quad (43)$$

Since $D(\tilde{U}\tilde{U}_0\tilde{U}_1\|\overline{UU_0U_1})$ does not depend on u^n, we have

$$\beta_n \geq \exp(-n(D(\tilde{U}\tilde{U}_0\tilde{U}_1\|\overline{UU_0U_1}) + \varepsilon_3(n))) \times \sum_{u^n \in \mathcal{U}^n_{[\varepsilon]} \cap \mathcal{T}^{(n)}_{\eta_n}(\tilde{U})} P_{\tilde{U}}(u^n) \quad (44)$$
$$\geq \exp(-n(D(\tilde{U}\tilde{U}_0\tilde{U}_1\|\overline{UU_0U_1}) + \varepsilon_3(n))) \times \left(P_{\tilde{U}}(\mathcal{U}^n_{[\varepsilon]}) + P_{\tilde{U}}(\mathcal{T}^{(n)}_{\eta_n}(\tilde{U})) - 1\right)$$
$$\geq \exp(-n(D(\tilde{U}\tilde{U}_0\tilde{U}_1\|\overline{UU_0U_1}) + \varepsilon_3(n))) \times \left(1 - \sqrt{\varepsilon} + 1 - \epsilon_n - 1\right) \quad (45)$$
$$\geq \exp(-n(D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}) + \varepsilon_4(n))) \quad (46)$$

where (44) follows from (43) and $\varepsilon_3(n) \triangleq \varepsilon_2(n) + \xi_n \to 0$, (45) follows from (24) and Lemma 2, (46) follows from (14) and (18) and $\varepsilon_4(n) \triangleq \varepsilon_3(n) - \frac{\log(1-\sqrt{\varepsilon}-\epsilon_n)}{n} \to 0$.

Hence, we have

$$\sigma(\varepsilon, 0, \infty) \leq D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}).$$

Moreover, since $P_{\tilde{X}\tilde{Y}}$ satisfies the appropriate marginal constraints $P_{\tilde{X}} = P_X, P_{\tilde{Y}} = P_Y$ as in (17), we have (21).

6. Generalization

In this section, we give a generalization of Theorem 1.

Theorem 2 Given P_{XY} and $P_{\overline{XY}}$, assume that there exist distributions $P_{X'Y'}, P_{\bar{X}'\bar{Y}'}$ and functions Λ_1, Λ_2 satisfying the following conditions:

1. $\Lambda_1 : \mathcal{X}' \to \mathcal{X}$ and $\Lambda_2 : \mathcal{Y}' \to \mathcal{Y}$ satisfy
$$(X, Y) = (\Lambda_1(X'), \Lambda_2(Y')) \quad (47)$$
$$(\overline{X}, \overline{Y}) = (\Lambda_1(\bar{X}'), \Lambda_2(\bar{Y}')). \quad (48)$$
2. $P_{X'Y'}$ and $P_{\bar{X}'\bar{Y}'}$ satisfy the assumptions of Theorem 1, i.e., $D(P_{X'Y'}\|P_{\bar{X}'\bar{Y}'}) < \infty$ and $P_{\bar{X}'\bar{Y}'}$ satisfies Condition 1.
3. $P_{X'Y'}$ and $P_{\bar{X}'\bar{Y}'}$ satisfy
$$\min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}) = \min_{\substack{P_{\tilde{X}'}=P_{X'} \\ P_{\tilde{Y}'}=P_{Y'}}} D(P_{\tilde{X}'\tilde{Y}'}\|P_{\bar{X}'\bar{Y}'}). \quad (49)$$

Then, we have

$$\sigma(\varepsilon, 0, R|XY, \overline{XY}) = \min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}). \quad (50)$$

Proof. From Proposition 1, we have

$$\min_{\substack{P_{\tilde{X}}=P_X \\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}) \leq \sigma(\varepsilon, 0, R|XY, \overline{XY}). \quad (51)$$

On the other hand, by using Λ_1 and Λ_2 satisfying (47) and (48), we can convert any code (f_n, g_n, ϕ_n) for the hypothesis testing XY versus $\overline{XY}$ into a code (f'_n, g'_n, ϕ'_n) for the hypothesis testing $X'Y'$ versus $\bar{X}'\bar{Y}'$ without increasing the error probabilities α_n and β_n. Indeed, we can construct (f'_n, g'_n, ϕ'_n) as

$$f'_n(x'^n) \triangleq f_n(\Lambda_1(x'^n))$$
$$g'_n(y'^n) \triangleq g_n(\Lambda_2(y'^n))$$

and for $c_1 \in \mathcal{M}_n$ and $c_2 \in \mathcal{N}_n$,

$$\phi'_n(c_1, c_2) \triangleq \begin{cases} P_{X'Y'} & (\phi_n(c_1, c_2) = P_{XY}) \\ P_{\bar{X}'\bar{Y}'} & (\phi_n(c_1, c_2) = P_{\overline{XY}}). \end{cases}$$

This fact implies that

$$\sigma(\varepsilon, 0, R|XY, \overline{XY}) \leq \sigma(\varepsilon, 0, R|X'Y', \bar{X}'\bar{Y}'). \quad (52)$$

Further, since $P_{X'Y'}$ and $P_{\bar{X}'\bar{Y}'}$ satisfy the assumptions in Theorem 1, we have

$$\sigma(\varepsilon, 0, R|X'Y', \bar{X}'\bar{Y}') \leq \min_{\substack{P_{\tilde{X}'}=P_{X'} \\ P_{\tilde{Y}'}=P_{Y'}}} D(P_{\tilde{X}'\tilde{Y}'}\|P_{\bar{X}'\bar{Y}'}). \quad (53)$$

Combining (51), (52) and (53) with the assumption (49), we obtain (50). □

In the rest of this section, we give an example for which the assumptions of Theorem 2 hold. To do this, we define some notations.

Let us define $\mathcal{X}_n$ and $\mathcal{Y}_n$ $(n = 1, 2 \ldots)$ as $\mathcal{X}_0 \triangleq \mathcal{X}$, $\mathcal{Y}_0 \triangleq \mathcal{Y}$, and

$$\mathcal{X}_{n+1} \triangleq \mathcal{X}_n \setminus \left\{x \in \mathcal{X}_n : \left|\left\{y \in \mathcal{Y}_n : P_{\overline{XY}}(x, y) > 0\right\}\right| = 1\right\}$$
$$\mathcal{Y}_{n+1} \triangleq \mathcal{Y}_n \setminus \left\{y \in \mathcal{Y}_n : \left|\left\{x \in \mathcal{X}_n : P_{\overline{XY}}(x, y) > 0\right\}\right| = 1\right\}.$$

Let n^* be the minimum of n such that $\mathcal{X}_{n+1} = \mathcal{X}_n$ and $\mathcal{Y}_{n+1} = \mathcal{Y}_n$.

For any $x \notin \mathcal{X}_{n^*}$, let us define $\mathsf{i}(x)$ as follows: we can choose the unique i $(0 \le i \le n^* - 1)$ such that $x \in \mathcal{X}_i$ and $x \notin \mathcal{X}_{i+1}$. We denote such i by $\mathsf{i}(x)$. For $x \in \mathcal{X}_{n^*}$, let $\mathsf{i}(x) \triangleq n^*$. For $y \in \mathcal{Y}$, $\mathsf{j}(y)$ is defined in a similar way.

For any x such that $\mathsf{i}(x) = i < n^*$, by the definition of $\mathsf{i}(x)$ and $\mathcal{X}_i$, we have

$$\begin{aligned}\mathsf{i}(x) = i &\Leftrightarrow x \in \mathcal{X}_i \text{ and } x \notin \mathcal{X}_{i+1}\\ &\Leftrightarrow \exists y \text{ s. t. } |\{y \in \mathcal{Y}_{\mathsf{i}(x)} : P_{XY}(x, y) > 0\}| = 1,\end{aligned}$$

and thus, we can choose the unique $y \in \mathcal{Y}_{\mathsf{i}(x)}$ satisfying $P_{\overline{XY}}(x, y) > 0$. For x such that $\mathsf{i}(x) < n^*$, let $\mathsf{s}(x) \in \mathcal{Y}_{\mathsf{i}(x)}$ be such y. For y such that $\mathsf{j}(y) < n^*$, $\mathsf{t}(y) \in \mathcal{X}_{\mathsf{j}(y)}$ is defined similarly.

Let us $\mathcal{X}' = \{0, \ldots, n^*\} \times \mathcal{X}$ and $\mathcal{Y}' = \{0, \ldots, n^*\} \times \mathcal{Y}$. From $P_{\overline{XY}}$ on $\mathcal{X} \times \mathcal{Y}$, we construct the distribution $P_{\bar{X}'\bar{Y}'}$ on $\mathcal{X}' \times \mathcal{Y}'$ as follows: Let $P_{\bar{X}'\bar{Y}'}((i, x), (j, y)) \triangleq P_{\overline{XY}}(x, y)$ if (i, x) and (j, y) satisfy one of the following three conditions

Case 1 $i = j = \mathsf{i}(x) = \mathsf{j}(y) = n^*$,
Case 2 $i = j = \mathsf{i}(x) \ne n^*$ and $y = \mathsf{s}(x)$,
Case 3 $i = j = \mathsf{j}(y) \ne n^*$ and $x = \mathsf{t}(y)$.

Otherwise, let $P_{\bar{X}'\bar{Y}'}((i, x), (j, y)) \triangleq 0$.

Example 2 Let us consider $P_{\overline{XY}}$ in Table 2. From the definition of $\mathcal{X}_0$ and $\mathcal{Y}_0$, we have

$$\mathcal{X}_0 = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}, \quad \mathcal{Y}_0 = \{\mathsf{d}, \mathsf{e}, \mathsf{f}\}.$$

Since $\left|\left\{y \in \mathcal{Y}_0 : P_{\overline{XY}}(\mathsf{c}, y) > 0\right\}\right| = |\{\mathsf{f}\}| = 1$, we have

$$\mathcal{X}_1 = \{\mathsf{a}, \mathsf{b}\}, \quad \mathcal{Y}_1 = \mathcal{Y}_0, \quad \mathsf{i}(\mathsf{c}) = 0, \quad \mathsf{s}(\mathsf{c}) = \mathsf{f}.$$

Further, since $\left|\left\{x \in \mathcal{X}_1 : P_{\overline{XY}}(x, \mathsf{f}) > 0\right\}\right| = |\{\mathsf{b}\}| = 1$, we have,

$$\mathcal{X}_2 = \{\mathsf{a}, \mathsf{b}\}, \quad \mathcal{Y}_2 = \{\mathsf{d}, \mathsf{e}\}, \quad \mathsf{j}(\mathsf{f}) = 1, \quad \mathsf{t}(\mathsf{f}) = \mathsf{b}.$$

Since $\mathcal{X}_3 = \mathcal{X}_2, \mathcal{Y}_3 = \mathcal{Y}_2$, we have $n^* = 2$. As a consequence, we have $P_{\bar{X}'\bar{Y}'}$ in Table 3.

Now, we give an example for which the assumptions of Theorem 2 hold.

Table 2 $\overline{XY}$.

$x \setminus y$	d	e	f
a	p_1	p_2	0
b	p_3	p_4	p_5
c	0	0	p_6

Table 3 $\bar{X}'\bar{Y}'$.

$\bar{x}' \setminus \bar{y}'$	(2, d)	(2, e)	(1, f)	(0, f)	others
(2, a)	p_1	p_2	0	0	0
(2, b)	p_3	p_4	0	0	0
(1, b)	0	0	p_5	0	0
(0, c)	0	0	0	p_6	0
others	0	0	0	0	0

Proposition 3 Assume that (3) holds and that $P_{\overline{XY}}$ restricted on $\mathcal{X}_{n^*} \times \mathcal{Y}_{n^*}$ satisfies Condition 1. Then, the assumptions of Theorem 2 hold.

Remark 5 In Proposition 3, "$P_{\overline{XY}}$ restricted on $\mathcal{X}_{n^*} \times \mathcal{Y}_{n^*}$ satisfies Condition 1" means that $P_{\bar{X}^*\bar{Y}^*}$ on $\mathcal{X}_{n^*} \times \mathcal{Y}_{n^*}$ defined by

$$P_{\bar{X}^*\bar{Y}^*}(x, y) \triangleq \frac{P_{\overline{XY}}(x, y)}{P_{\overline{XY}}(\mathcal{X}_{n^*} \times \mathcal{Y}_{n^*})}, \quad (x, y) \in \mathcal{X}_{n^*} \times \mathcal{Y}_{n^*}$$

satisfies Condition 1.

Proof. Let $P_{\bar{X}'\bar{Y}'}$ be the distribution constructed from $P_{\overline{XY}}$ according to the construction above Example 2. Similarly, we construct $P_{X'Y'}$ from P_{XY}; i.e., let $P_{X'Y'}((i, x), (j, y)) \triangleq P_{XY}(x, y)$ if (i, x) and (j, y) satisfy one of the three conditions (Case 1–Case 3) above Example 2, and let $P_{X'Y'}((i, x), (j, y)) \triangleq 0$ otherwise. It should be emphasized that, in the construction of $P_{X'Y'}$, we use n^*, $\mathsf{i}, \mathsf{j}, \mathsf{s}$, and t defined from $P_{\overline{XY}}$ (not P_{XY}).†

It is apparent that the functions $\Lambda_1 : \mathcal{X}' \to \mathcal{X}$ and $\Lambda_2 : \mathcal{Y}' \to \mathcal{Y}$ by

$$\Lambda_1((i, x)) \triangleq x, \quad \Lambda_2((j, y)) \triangleq y.$$

Further it is not hard to see that if $P_{\overline{XY}}$ restricted on $\mathcal{X}_{n^*} \times \mathcal{Y}_{n^*}$ satisfies Condition 1, then $P_{\bar{X}'\bar{Y}'}$ also satisfies Condition 1. Moreover, if (3) holds then $D(P_{X'Y'} \| P_{\bar{X}'\bar{Y}'}) < \infty$. Thus, $P_{X'Y'}$ and $P_{\bar{X}'\bar{Y}'}$ satisfy the assumptions of Theorem 1.

Hence, what we have to prove is that (49) holds.

From the definition of $\mathcal{X}_{n^*}$ and $\mathcal{Y}_{n^*}$, we can show that if $P_{\tilde{X}\tilde{Y}}$ satisfies that $P_{\tilde{X}} = P_X$, $P_{\tilde{Y}} = P_Y$, and $D(P_{\tilde{X}\tilde{Y}} \| P_{\overline{XY}}) < \infty$ then $P_{\tilde{X}\tilde{Y}}$ must satisfy (i) $P_{\tilde{X}}(x) = P_X(x)$ for all $x \notin \mathcal{X}_{n^*}$ and (ii) $P_{\tilde{Y}}(y) = P_Y(y)$ for all $y \notin \mathcal{Y}_{n^*}$. Hence, we have

$$\begin{aligned}&\min_{\substack{P_{\tilde{X}}=P_X\\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}} \| P_{\overline{XY}})\\ &= \sum_{\substack{(x,y):\\ x\notin\mathcal{X}_{n^*} \text{ or } y\notin\mathcal{Y}_{n^*}}} P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_{\overline{XY}}(x, y)}\\ &\quad + \min \sum_{\substack{(x,y):\\ x\in\mathcal{X}_{n^*} \text{ and } y\in\mathcal{Y}_{n^*}}} P_{\tilde{X}\tilde{Y}}(x, y) \log \frac{P_{\tilde{X}\tilde{Y}}(x, y)}{P_{\overline{XY}}(x, y)}. \end{aligned} \tag{54}$$

where the minimum in (54) is taken over all $P_{\tilde{X}\tilde{Y}}$ satisfying

$$\sum_{y\in\mathcal{Y}_{n^*}} P_{\tilde{X}\tilde{Y}}(x, y) = \sum_{y\in\mathcal{Y}_{n^*}} P_{XY}(x, y), \quad x \in \mathcal{X}_{n^*} \tag{55}$$

$$\sum_{x\in\mathcal{X}_{n^*}} P_{\tilde{X}\tilde{Y}}(x, y) = \sum_{x\in\mathcal{X}_{n^*}} P_{XY}(x, y), \quad y \in \mathcal{Y}_{n^*}. \tag{56}$$

Similarly, from the constructions of $P_{\bar{X}'\bar{Y}'}$ and $P_{X'Y'}$, we

†If $P_{\overline{XY}}(x, y) > 0 \Leftrightarrow P_{XY}(x, y) > 0$ for all x, y then n^*, $\mathsf{i}, \mathsf{j}, \mathsf{s}$, and t defined from P_{XY} are identical with those from $P_{\overline{XY}}$. However, if there exists (x, y) such that $P_{\overline{XY}}(x, y) > 0$ and $P_{XY}(x, y) = 0$ then n^*, $\mathsf{i}, \mathsf{j}, \mathsf{s}$, and t defined from P_{XY} may not be identical with those from $P_{\overline{XY}}$.

have

$$\begin{aligned}
&\min_{\substack{P_{\tilde{X}'}=P_{X'}\\ P_{\tilde{Y}'}=P_{Y'}}} D(P_{\tilde{X}'\tilde{Y}'}\|P_{\overline{XY}}) \\
&= \min_{\substack{P_{\tilde{X}'}=P_{X'}\\ P_{\tilde{Y}'}=P_{Y'}}} \Bigg[\sum_{\substack{(i,x),(j,y):\\ \text{Case 2 or 3}}} P_{\tilde{X}'\tilde{Y}'}((i,x),(j,y)) \log \frac{P_{\tilde{X}'\tilde{Y}'}((i,x),(j,y))}{P_{\bar{X}'\bar{Y}'}((i,x),(j,y))} \\
&\quad + \sum_{\substack{(x,y):\\ x\in\mathcal{X}_{n^*} \text{ and } y\in\mathcal{Y}_{n^*}}} P_{\tilde{X}'\tilde{Y}'}((n^*,x),(n^*,y)) \log \frac{P_{\tilde{X}'\tilde{Y}'}((n^*,x),(n^*,y))}{P_{\bar{X}'\bar{Y}'}((n^*,x),(n^*,y))} \Bigg] \\
&= \sum_{\substack{(i,x),(j,y):\\ \text{Case 2 or 3}}} P_{X'Y'}((i,x),(j,y)) \log \frac{P_{X'Y'}((i,x),(j,y))}{P_{\bar{X}'\bar{Y}'}((i,x),(j,y))} \\
&\quad + \min \sum_{\substack{(x,y):\\ x\in\mathcal{X}_{n^*} \text{ and } y\in\mathcal{Y}_{n^*}}} P_{\tilde{X}'\tilde{Y}'}((n^*,x),(n^*,y)) \log \frac{P_{\tilde{X}'\tilde{Y}'}((n^*,x),(n^*,y))}{P_{\bar{X}'\bar{Y}'}((n^*,x),(n^*,y))} \\
&= \sum_{\substack{(x,y):\\ x\notin\mathcal{X}_{n^*} \text{ or } y\notin\mathcal{Y}_{n^*}}} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_{\overline{XY}}(x,y)} \\
&\quad + \min \sum_{\substack{(x,y):\\ x\in\mathcal{X}_{n^*} \text{ and } y\in\mathcal{Y}_{n^*}}} P_{\tilde{X}\tilde{Y}}(x,y) \log \frac{P_{\tilde{X}\tilde{Y}}(x,y)}{P_{\overline{XY}}(x,y)} \qquad (57)
\end{aligned}$$

where the minimum in (57) is taken over all $P_{\tilde{X}\tilde{Y}}$ satisfying (55) and (56).

By combining (54) and (57), we have (49). □

It is not hard to see that if (3) holds and $|\mathcal{X}| \le 2$ or $|\mathcal{Y}| \le 2$ then

1. the positivity condition is satisfied, or
2. the assumptions of Theorem 1 are satisfied, or
3. the assumptions of Proposition 3 are satisfied.

Hence, we have the following corollary.

Corollary 1 Assume that (3) holds and that $|\mathcal{X}| \le 2$ or $|\mathcal{Y}| \le 2$. Then, we have

$$\sigma(\varepsilon, 0, R|XY, \overline{XY}) = \min_{\substack{P_{\tilde{X}}=P_X\\ P_{\tilde{Y}}=P_Y}} D(P_{\tilde{X}\tilde{Y}}\|P_{\overline{XY}}).$$

7. Conclusion

In this paper, we proved that Shalaby and Papamarcou's upper bound of the optimal error exponent for multiterminal hypothesis testing holds under a condition weaker than their condition.

However, there are some cases for which error exponent is not determined yet. For example, $P_{\overline{XY}}$ in Table 4 dose not satisfy the positivity condition nor conditions given in this paper. Although the authors conjecture that Shalaby and Papamarcou's bound holds if $D(P_{XY}\|P_{\overline{XY}}) < \infty$ (i.e., Han's bound is tight), it remains as a future work to prove this conjecture.

In this paper, we focus our investigation on the two encoders system. On the other hand, Shalaby and Papamarcou showed that their result with the positivity condition

Table 4 Alternative hypothesis $\overline{XY}$.

$\bar{x} \setminus \bar{y}$	1	2	3
1	q_1	q_2	q_3
2	q_4	q_5	q_6
3	q_7	q_8	0

can be extended to the systems employing arbitrary finite number of encoders [3, Section IV]. Based on the same argument as [3, Section IV], it is not so hard to extend our Theorem 1 to the case where the number of encoders is m. To do this, we need to generalize the conditional positivity condition as follows: Assume that the i-th encoder can observe $(\overline{U}, \overline{U}_{i-1})$ and the conditional distribution $P_{\overline{U}_0\ldots\overline{U}_{m-1}|\overline{U}}$ of $(\overline{U}_0, \ldots, \overline{U}_{m-1})$ given the common random variable U is positive (i.e., $P_{\overline{U}_0\ldots\overline{U}_{m-1}|\overline{U}}$ satisfies the positivity condition such as (7)). We can see that the statements given in Section IV of [3] holds not only under the positivity condition but also under the conditional positivity condition generalized as above. Although the above generalization is straightforward, we may be able to extend our result to more general case. For example, we may consider cross observations at the encoders (such as [8]) and assume an appropriate condition on the joint (and/or conditional) distribution of sources. It remain as a future work.

Furthermore, some important problems remain as future works:

- In the paper [9], Shalaby and Papamarcou generalized their result in [3] to the case where the source is a correlated Markov process. In [9], it is assumed that the transition matrix of the alternative hypothesis satisfies the positivity condition; this assumption corresponds to the positivity condition in [3]. It is an important future work to establish the result of [9] under a weaker condition.
- In this paper, we consider only the case where the alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite. It is also important future work to generalize the result to the case where alphabets are countably infinite or continuous.

Acknowledgment

The authors would like to thank the associate editor and anonymous reviewers for valuable comments which helped to improve the presentation.

The work of S. Kuzuoka was supported in part by JSPS KAKENHI Grant Number 26820145.

References

[1] T.S. Han and S. Amari, "Statistical inference under multiterminal data compression," IEEE Trans. Inform. Theory, vol.44, no.6, pp.2300–2324, Oct. 1998.
[2] T.S. Han, "Hypothesis testing with multiterminal data compression," IEEE Trans. Inform. Theory, vol.33, no.6, pp.759–772, Nov. 1987.
[3] H.M.H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," IEEE Trans. Inform. Theory, vol.38, no.2, pp.254–267, March 1992.

[4] K. Marton, “A simple proof of the blowing-up lemma,” IEEE Trans. Inform. Theory, vol.32, no.3, pp.445–446, May 1986.
[5] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, New York, 1982.
[6] T.S. Han, Information-Spectrum Methods in Information Theory, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
[7] T. Uyematsu, Modern Shannon Theory, Baifukan, 1998. (in Japanese)
[8] T.S. Han, “Source coding with cross observations at the encoders,” IEEE Trans. Inform. Theory, vol.25, no.3, pp.360–361, May 1979.
[9] H.M.H. Shalaby and A. Papamarcou, “Error exponents for distributed detection of Markov sources,” IEEE Trans. Inform. Theory, vol.40, no.2, pp.397–408, March 1994.

Makoto Ueda received the B.E. and M.E. degrees from Wakayama University in 2013 and 2015, respectively. Currently he is working for USE Co., Ltd.

Shigeaki Kuzuoka received the B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology in 2002, 2004, and 2007, respectively. He was an assistant professor from 2007 to 2009, a lecturer from 2009 to 2015, and has been an associate professor since 2015 in the Faculty of Systems Engineering at Wakayama University. His current research interests are in the areas of information theory, especially Shannon theory, source coding, and multiterminal information theory. He is a member of the IEEE.