Testing Against Independence and a Rényi Information Measure

Amos Lapidoth and Christoph Pfister
Signal and Information Processing Laboratory
ETH Zurich, 8092 Zurich, Switzerland
Email: {lapidoth,pfister}@isi.ee.ethz.ch

***Abstract*—The achievable error-exponent pairs for the type I and type II errors are characterized in a hypothesis testing setup where the observation consists of independent and identically distributed samples from either a known joint probability distribution or an unknown product distribution. The empirical mutual information test, the Hoeffding test, and the generalized likelihood-ratio test are all shown to be asymptotically optimal. An expression based on a Rényi measure of dependence is shown to be the Fenchel biconjugate of the error-exponent function obtained by fixing one error exponent and optimizing the other. An example is provided where the error-exponent function is not convex and thus not equal to its Fenchel biconjugate.**

I. Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets and P_{XY} a probability mass function (PMF) over $\mathcal{X} \times \mathcal{Y}$. Based on a sequence of pairs of random variables $\{(X_i, Y_i)\}_{i=1}^n$, we want to distinguish between two hypotheses:

0) Under the null hypothesis, $(X_1, Y_1), \ldots, (X_n, Y_n)$ are IID according to P_{XY}.
1) Under the alternative hypothesis, $(X_1, Y_1), \ldots, (X_n, Y_n)$ are IID according to some unknown PMF of the form $Q_{XY} = Q_X Q_Y$, where $Q_X \in \mathcal{P}(\mathcal{X})$ and $Q_Y \in \mathcal{P}(\mathcal{Y})$ are arbitrary PMFs over $\mathcal{X}$ and $\mathcal{Y}$, respectively.

An error-exponent pair $(\mathsf{E_P}, \mathsf{E_Q}) \in \mathbb{R}^2$ is achievable if there exists a sequence of deterministic tests $\{T_n\}_{n=1}^\infty$ satisfying the following two conditions:

$$\liminf_{n\to\infty} -\frac{1}{n} \log P_{XY}^{\times n}[T_n(X^n, Y^n) = 1] > \mathsf{E_P}, \tag{1}$$

$$\liminf_{n\to\infty} \inf_{Q_X, Q_Y} -\frac{1}{n} \log (Q_X Q_Y)^{\times n}[T_n(X^n, Y^n) = 0] > \mathsf{E_Q}, \tag{2}$$

where a deterministic test T_n is a function from $\mathcal{X}^n \times \mathcal{Y}^n$ to $\{0, 1\}$; we denote by $R_{XY}^{\times n}[\mathcal{A}]$ the probability of an event $\mathcal{A}$ when $\{(X_i, Y_i)\}_{i=1}^n$ are IID according to R_{XY}; the infimum is over all $Q_X \in \mathcal{P}(\mathcal{X})$ and all $Q_Y \in \mathcal{P}(\mathcal{Y})$; and all logarithms in this paper are natural logarithms. If an error-exponent pair $(\mathsf{E_P}, \mathsf{E_Q})$ is achievable, then, since the inequalities in (1) and (2) are strict, there exists a sequence of tests $\{T_n\}_{n=1}^\infty$ such that for sufficiently large n and for all $(Q_X, Q_Y) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$,

$$P_{XY}^{\times n}[T_n(X^n, Y^n) = 1] \le e^{-n\mathsf{E_P}}, \tag{3}$$

$$(Q_X Q_Y)^{\times n}[T_n(X^n, Y^n) = 0] \le e^{-n\mathsf{E_Q}}. \tag{4}$$

(The reverse is not true: (3) and (4) are not sufficient for the achievability of the pair $(\mathsf{E_P}, \mathsf{E_Q})$; see Section II for more motivation for our definition.)

Our first result characterizes the achievable error-exponent pairs.

Theorem 1. *An error-exponent pair $(\mathsf{E_P}, \mathsf{E_Q})$ is achievable if, and only if, for all $R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$,*

$$\big(D(R_{XY} \| P_{XY}) > \mathsf{E_P}\big) \vee \big(D(R_{XY} \| R_X R_Y) > \mathsf{E_Q}\big). \tag{5}$$

This characterization is also valid when randomized tests are allowed in (1) *and* (2).

In Lemmas 8–10 we show that the empirical mutual information test, the Hoeffding test, and the generalized likelihood-ratio test can achieve every achievable error-exponent pair. Defining the error-exponent functions $\mathsf{E_P} \colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ and $\mathsf{E_Q} \colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ as

$$\mathsf{E_P}(\mathsf{E_Q}) \triangleq \sup\{\mathsf{E_P} \in \mathbb{R} : (\mathsf{E_P}, \mathsf{E_Q}) \text{ is achievable}\}, \tag{6}$$

$$\mathsf{E_Q}(\mathsf{E_P}) \triangleq \sup\{\mathsf{E_Q} \in \mathbb{R} : (\mathsf{E_P}, \mathsf{E_Q}) \text{ is achievable}\}, \tag{7}$$

we obtain

Corollary 2. *For all $\mathsf{E_Q} \in \mathbb{R}$,*

$$\mathsf{E_P}(\mathsf{E_Q}) = \inf_{\substack{R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}): \\ D(R_{XY} \| R_X R_Y) \le \mathsf{E_Q}}} D(R_{XY} \| P_{XY}), \tag{8}$$

and for all $\mathsf{E_P} \in \mathbb{R}$,

$$\mathsf{E_Q}(\mathsf{E_P}) = \inf_{\substack{R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}): \\ D(R_{XY} \| P_{XY}) \le \mathsf{E_P}}} D(R_{XY} \| R_X R_Y). \tag{9}$$

Our next result relates the Rényi measure of dependence $J_\alpha(P_{XY})$ to $\mathsf{E}_\mathsf{P}^{**}(\cdot)$, the Fenchel biconjugate of $\mathsf{E_P}(\cdot)$. Both $J_\alpha(P_{XY})$ and $\mathsf{E}_\mathsf{P}^{**}(\cdot)$ are discussed in Section II. (The analogous result for $\mathsf{E}_\mathsf{Q}^{**}(\cdot)$ is Theorem 13.)

Theorem 3. *For all $\mathsf{E_Q} \in \mathbb{R}$,*

$$\sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} \big(J_\alpha(P_{XY}) - \mathsf{E_Q}\big) = \mathsf{E}_\mathsf{P}^{**}(\mathsf{E_Q}). \tag{10}$$

Furthermore, $\mathsf{E}_\mathsf{P}^{**}(\mathsf{E_Q}) = \mathsf{E_P}(\mathsf{E_Q})$ for all $\mathsf{E_Q} \in \mathbb{R}$ if, and only if, $\mathsf{E_P}(\cdot)$ is convex on $\mathbb{R}$.

Our last contribution is Example 14, where $\mathsf{E_P}(\cdot)$ is not convex and thus for some $\mathsf{E_Q} \in \mathbb{R}$,

$$\sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} \big(J_\alpha(P_{XY}) - \mathsf{E_Q}\big) \ne \mathsf{E_P}(\mathsf{E_Q}). \tag{11}$$

The rest of this paper is organized as follows: in Section II, we review the Rényi divergence and the Fenchel conjugation; in Section III, we review results on simple and composite hypothesis testing; in Section IV, we prove Theorem 1 and provide asymptotically optimal tests; in Section V, we relate $J_\alpha(P_{XY})$ to the Fenchel biconjugates $\mathsf{E}_\mathsf{P}^{**}$ and $\mathsf{E}_\mathsf{Q}^{**}$; and in Section VI, we discuss Example 14, where $\mathsf{E}_\mathsf{P}(\cdot)$ is not convex. Additional proofs can be found in [1].

II. Preliminaries

Let P and Q be PMFs over a finite set $\mathcal{Z}$. The relative entropy (or Kullback–Leibler divergence) is defined as

$$D(P\|Q) \triangleq \sum_{z\in\mathcal{Z}} P(z)\log\frac{P(z)}{Q(z)} \tag{12}$$

with the conventions that $0\log(0/q) = 0$ for all $q \geq 0$ and $p\log(p/0) = +\infty$ for all $p > 0$. The Rényi divergence of order α [2], [3] is defined for all positive α other than 1 as

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha-1}\log\sum_{z\in\mathcal{Z}} P(z)^\alpha\, Q(z)^{1-\alpha} \tag{13}$$

with the conventions that $\log 0 = -\infty$ and that for $\alpha > 1$, we read $P(z)^\alpha\, Q(z)^{1-\alpha}$ as $P(z)^\alpha/Q(z)^{\alpha-1}$ and use $0/0 = 0$ and $p/0 = +\infty$ for all $p > 0$. By continuous extension [3, Theorem 5], $D_1(P\|Q) \triangleq D(P\|Q)$.

The measure of dependence $J_\alpha(P_{XY})$ [4] is defined as

$$J_\alpha(P_{XY}) \triangleq \min_{Q_X\in\mathcal{P}(\mathcal{X}),\, Q_Y\in\mathcal{P}(\mathcal{Y})} D_\alpha(P_{XY}\|Q_XQ_Y) \tag{14}$$

for all positive α and as zero when α is zero.

The convex conjugate (or Fenchel conjugate) of a function $f\colon \mathbb{R}\to[-\infty,+\infty]$ is the function $f^*\colon\mathbb{R}\to[-\infty,+\infty]$,

$$f^*(\lambda) \triangleq \sup_{x\in\mathbb{R}}\,[\lambda\, x - f(x)]. \tag{15}$$

It is lower semicontinuous and convex [5, Section 7.1 and Proposition 1.2.2].

The (Fenchel) biconjugate of a function $f\colon\mathbb{R}\to[-\infty,+\infty]$ is $f^{**}\colon\mathbb{R}\to[-\infty,+\infty]$, the convex conjugate of f^*. For every f and for every $x\in\mathbb{R}$, $f^{**}(x)\leq f(x)$ [6, Section 4.2].

We next motivate the strict inequalities in (1) and (2). Let $\tilde{\mathsf{E}}_\mathsf{P}(\mathsf{E}_\mathsf{Q})$ denote the error-exponent function that would have resulted had we replaced the strict inequalities in (1) and (2) with weak inequalities. Then, $\tilde{\mathsf{E}}_\mathsf{P}(\cdot)$ and $\tilde{\mathsf{E}}_\mathsf{P}^{**}(\cdot)$ cannot be equal because, unlike $\mathsf{E}_\mathsf{P}(\cdot)$, $\tilde{\mathsf{E}}_\mathsf{P}(\cdot)$ is not lower semicontinuous. The difference between $\mathsf{E}_\mathsf{P}(\cdot)$ and $\tilde{\mathsf{E}}_\mathsf{P}(\cdot)$ is best seen at zero: While $\tilde{\mathsf{E}}_\mathsf{P}(0)$ is $+\infty$, it turns out that $\mathsf{E}_\mathsf{P}(0)$ is the optimal error exponent if for a fixed $\epsilon\in(0,1)$, we require the tests to satisfy $(Q_XQ_Y)^{\times n}[T_n(X^n,Y^n)=0]\leq\epsilon$ for all n and all (Q_X,Q_Y). (This setup is similar to the one in Stein's lemma [7, Corollary 1.2]; we do not explore it further in this paper.)

To see that (3) and (4) are not sufficient for the achievability of an error-exponent pair, observe that (3) and (4) hold for every $\mathsf{E}_\mathsf{P}\in\mathbb{R}$ if $\mathsf{E}_\mathsf{Q}=0$ and $T_n(X^n,Y^n)=0$ irrespective of X^n and Y^n. Yet, (8) implies that $\mathsf{E}_\mathsf{P}(0)$ is finite, so $(\mathsf{E}_\mathsf{P},0)$ is not achievable for every E_P.

We conclude this section with two lemmas.

Lemma 4. *For all R_{XY}, Q_X, and Q_Y,*

$$D(R_{XY}\|Q_XQ_Y)\geq D(R_{XY}\|R_XR_Y). \tag{16}$$

Consequently,

$$\inf_{Q_X,Q_Y} D(R_{XY}\|Q_XQ_Y) = D(R_{XY}\|R_XR_Y). \tag{17}$$

Proof. We have

$$\begin{aligned}
&D(R_{XY}\|Q_XQ_Y)\\
&\quad= D(R_{XY}\|R_XR_Y)+D(R_X\|Q_X)+D(R_Y\|Q_Y) &(18)\\
&\quad\geq D(R_{XY}\|R_XR_Y), &(19)
\end{aligned}$$

where (19) holds because $D(P\|Q)\geq 0$. Equality is achieved for $Q_X=R_X$ and $Q_Y=R_Y$, which proves (17). ∎

Lemma 5. *Let P and Q be PMFs over a finite set $\mathcal{Z}$. Then, for all $\mathsf{E}_\mathsf{Q}\in\mathbb{R}$,*

$$\inf_{\substack{R\in\mathcal{P}(\mathcal{Z}):\\ D(R\|Q)\leq\mathsf{E}_\mathsf{Q}}} D(R\|P) = \sup_{\alpha\in(0,1]}\frac{1-\alpha}{\alpha}\big(D_\alpha(P\|Q)-\mathsf{E}_\mathsf{Q}\big). \tag{20}$$

Proof. Omitted. ∎

III. Related Work

Let P and Q be PMFs over a finite set $\mathcal{Z}$. In the simple hypothesis testing setup where one has to guess whether $\{Z_i\}_{i=1}^n$ are IID according to P or Q, Hoeffding [8] and Csiszár and Longo [9] essentially showed that

$$\tilde{\mathsf{E}}_\mathsf{P}(\mathsf{E}_\mathsf{Q}) = \inf_{R\in\mathcal{P}(\mathcal{Z}):\, D(R\|Q)\leq\mathsf{E}_\mathsf{Q}} D(R\|P), \tag{21}$$

where $\tilde{\mathsf{E}}_\mathsf{P}(\cdot)$ is the error-exponent function for the simple hypothesis testing setup. More properties of $\tilde{\mathsf{E}}_\mathsf{P}(\cdot)$ were studied by Blahut [10]; relevant for us is

$$\tilde{\mathsf{E}}_\mathsf{P}(\mathsf{E}_\mathsf{Q}) = \sup_{\alpha\in(0,1]}\frac{1-\alpha}{\alpha}\big(D_\alpha(P\|Q)-\mathsf{E}_\mathsf{Q}\big), \tag{22}$$

which follows from [10, Theorem 7] by substituting $\alpha=\frac{1}{1+s}$ and identifying the Rényi divergence.

In the composite hypothesis testing setup where P is tested against an unknown Q from some set $\mathcal{Q}$, Hoeffding [8] showed that his likelihood-ratio test is asymptotically optimal against all $Q\in\mathcal{Q}$; see also [7, Problem 2.13(b)]. This test statistic is used in Lemma 9.

For the hypothesis testing setup of this paper, Tomamichel and Hayashi [11, first part of (57)] showed that for sufficiently large E_Q,

$$\sup_{\alpha\in(\frac{1}{2},1)}\frac{1-\alpha}{\alpha}\big(J_\alpha(P_{XY})-\mathsf{E}_\mathsf{Q}\big) = \mathsf{E}_\mathsf{P}(\mathsf{E}_\mathsf{Q}). \tag{23}$$

We provide a negative answer to the question at the end of the paragraph in [11]: an equality of the form (23) does not hold in general because the LHS of (23) is always convex in E_Q, but $\mathsf{E}_\mathsf{P}(\cdot)$ from Example 14 is not convex.

Conditions for which the generalized likelihood-ratio test is asymptotically optimal in a Neyman–Pearson sense are studied in [12]. A different approach to composite hypothesis testing has been proposed in [13].

Independence testing is a related setup where one wants to know whether or not the PMF generating $\{(X_i, Y_i)\}_{i=1}^n$ has a product form (whereas here, we test a fixed P_{XY} against an unknown product distribution). Since the empirical mutual information in Lemma 8 does not depend on P_{XY}, it can also be used for independence testing; see for example [14, "G-test of independence"], where G is $2n$ times the empirical mutual information.

IV. Achievable Error-Exponent Pairs

After two preparatory lemmas, we present in Lemmas 8–10 three tests that achieve any error-exponent pair $(\mathsf{E_P}, \mathsf{E_Q})$ for which

$$\big(D(R_{XY}\|P_{XY}) > \mathsf{E_P}\big) \vee \big(D(R_{XY}\|R_X R_Y) > \mathsf{E_Q}\big) \tag{24}$$

holds for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$. These tests are all based on the type [7] $\hat{R}_{XY}$ of the sequence $\{(X_i, Y_i)\}_{i=1}^n$. The asymptotic optimality of these tests follows from the converse proved in Lemma 11, which establishes Theorem 1 and Corollary 2.

Lemma 6. *If* (24) *holds for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, then there exists an $\epsilon > 0$ such that for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$,*

$$\big(D(R_{XY}\|P_{XY}) \geq \mathsf{E_P} + \epsilon\big) \vee \big(D(R_{XY}\|R_X R_Y) \geq \mathsf{E_Q} + \epsilon\big). \tag{25}$$

Proof. Define the function $f\colon \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \to \mathbb{R}\cup\{+\infty\}$,

$$R_{XY} \mapsto \max\big\{D(R_{XY}\|P_{XY}) - \mathsf{E_P}, \\ D(R_{XY}\|R_X R_Y) - \mathsf{E_Q}\big\}. \tag{26}$$

Suppose that (24) holds for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, and consider

$$\eta \triangleq \inf_{R_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})} f(R_{XY}). \tag{27}$$

If $\eta > 0$, then (25) holds with $\epsilon = \eta$. We show by contradiction that $\eta \leq 0$ is impossible. Assume $\eta \leq 0$. Observe that f is lower semicontinuous on $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ and that $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ is a compact set. By the extreme value theorem, there would exist an $R^*_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ with $f(R^*_{XY}) = \eta \leq 0$. This leads to a contradiction because then (24) would not hold for R^*_{XY}. ■

Lemma 7. *Let $\mathsf{E_P}$, $\mathsf{E_Q}$, and $\epsilon > 0$ be such that* (25) *holds for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$. Define $\tau \triangleq (n+1)^{|\mathcal{X}\times\mathcal{Y}|}$. Then, for all $Q_X \in \mathcal{P}(\mathcal{X})$ and all $Q_Y \in \mathcal{P}(\mathcal{Y})$,*

$$P_{XY}^{\times n}[D(\hat{R}_{XY}\|P_{XY}) \geq \mathsf{E_P} + \epsilon] \leq \tau\, e^{-n(\mathsf{E_P}+\epsilon)}, \tag{28}$$

$$P_{XY}^{\times n}[D(\hat{R}_{XY}\|\hat{R}_X\hat{R}_Y) < \mathsf{E_Q} + \epsilon] \leq \tau\, e^{-n(\mathsf{E_P}+\epsilon)}, \tag{29}$$

$$(Q_X Q_Y)^{\times n}[D(\hat{R}_{XY}\|P_{XY}) < \mathsf{E_P} + \epsilon] \leq \tau\, e^{-n(\mathsf{E_Q}+\epsilon)}, \tag{30}$$

$$(Q_X Q_Y)^{\times n}[D(\hat{R}_{XY}\|\hat{R}_X\hat{R}_Y) \geq \mathsf{E_Q} + \epsilon] \leq \tau\, e^{-n(\mathsf{E_Q}+\epsilon)}. \tag{31}$$

Proof. Omitted. (It is based on Sanov's theorem.) ■

Lemma 8 (Empirical Mutual Information Test)**.** *If* (24) *is satisfied for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, then there exists an $\epsilon > 0$ such that the error-exponent pair* $(\mathsf{E_P}, \mathsf{E_Q})$ *is achieved by the sequence of tests*

$$T_n(\hat{R}_{XY}) \triangleq \begin{cases} 1 & \text{if } D(\hat{R}_{XY}\|\hat{R}_X\hat{R}_Y) < \mathsf{E_Q} + \epsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

Proof. Use the $\epsilon > 0$ from Lemma 6. Then, the sequence of tests $\{T_n\}_{n=1}^\infty$ satisfies (1) because

$$\begin{aligned} &P_{XY}^{\times n}[T_n(X^n, Y^n) = 1] \\ &\quad = P_{XY}^{\times n}[D(\hat{R}_{XY}\|\hat{R}_X\hat{R}_Y) < \mathsf{E_Q} + \epsilon] &(33)\\ &\quad \leq (n+1)^{|\mathcal{X}\times\mathcal{Y}|}\cdot e^{-n(\mathsf{E_P}+\epsilon)}, &(34)\end{aligned}$$

where (34) follows from Lemma 7. Similarly, the sequence of tests $\{T_n\}_{n=1}^\infty$ satisfies (2). ■

Lemma 9 (Hoeffding's Test [8])**.** *If* (24) *is satisfied for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, then there exists an $\epsilon > 0$ such that the error-exponent pair* $(\mathsf{E_P}, \mathsf{E_Q})$ *is achieved by the sequence of tests*

$$T_n(\hat{R}_{XY}) \triangleq \begin{cases} 0 & \text{if } D(\hat{R}_{XY}\|P_{XY}) < \mathsf{E_P} + \epsilon, \\ 1 & \text{otherwise.} \end{cases} \tag{35}$$

Proof. The proof is very similar to the proof of Lemma 8. ■

Lemma 10 (Generalized Likelihood-Ratio Test)**.** *The logarithm of the generalized likelihood ratio, divided by n, is*

$$\begin{aligned}\Gamma &\triangleq \frac{1}{n}\log\frac{P_{XY}^{\times n}(X^n, Y^n)}{\sup\limits_{Q_X\in\mathcal{P}(\mathcal{X}),\, Q_Y\in\mathcal{P}(\mathcal{Y})}(Q_X Q_Y)^{\times n}(X^n, Y^n)} &(36)\\ &= D(\hat{R}_{XY}\|\hat{R}_X\hat{R}_Y) - D(\hat{R}_{XY}\|P_{XY}). &(37)\end{aligned}$$

If (24) *is satisfied for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, then the error-exponent pair* $(\mathsf{E_P}, \mathsf{E_Q})$ *is achieved by the sequence of tests*

$$T_n(\hat{R}_{XY}) \triangleq \begin{cases} 1 & \text{if } \Gamma \leq \mathsf{E_Q} - \mathsf{E_P}, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

Proof. The proof of (37) is omitted. Using the $\epsilon > 0$ from Lemma 6, the sequence of tests $\{T_n\}_{n=1}^\infty$ satisfies (1) because

$$\begin{aligned} &P_{XY}^{\times n}[T_n(X^n, Y^n) = 1] \\ &\quad \leq P_{XY}^{\times n}[D(\hat{R}_{XY}\|\hat{R}_X\hat{R}_Y) < \mathsf{E_Q} + \epsilon] \\ &\qquad + P_{XY}^{\times n}[D(\hat{R}_{XY}\|P_{XY}) \geq \mathsf{E_P} + \epsilon] &(39)\\ &\quad \leq 2\,(n+1)^{|\mathcal{X}\times\mathcal{Y}|}\cdot e^{-n(\mathsf{E_P}+\epsilon)}, &(40)\end{aligned}$$

where (39) follows from the union bound because the events $D(\hat{R}_{XY}\|\hat{R}_X\hat{R}_Y) \geq \mathsf{E_Q} + \epsilon$ and $D(\hat{R}_{XY}\|P_{XY}) < \mathsf{E_P} + \epsilon$ imply $\Gamma > \mathsf{E_Q} - \mathsf{E_P}$; and (40) follows from Lemma 7. In the same way, the sequence of tests $\{T_n\}_{n=1}^\infty$ satisfies (2). ■

Lemma 11. *If* (24) *does not hold for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, i.e., if there exists an $R^*_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ satisfying*

$$\big(D(R^*_{XY}\|P_{XY}) \leq \mathsf{E_P}\big) \wedge \big(D(R^*_{XY}\|R^*_X R^*_Y) \leq \mathsf{E_Q}\big), \tag{41}$$

then the error-exponent pair $(\mathsf{E_P}, \mathsf{E_Q})$ *is not achievable. (Not even if randomized tests are allowed.)*

Proof. Omitted. ■

Proof of Theorem 1. The theorem follows from Lemma 8 and from Lemma 11. ■

Proof of Corollary 2. For a fixed $\mathsf{E_Q} \in \mathbb{R}$, define

$$C \triangleq \inf_{\substack{R_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}):\\ D(R_{XY}\|R_X R_Y)\leq\mathsf{E_Q}}} D(R_{XY}\|P_{XY}). \tag{42}$$

By Theorem 1, all error-exponent pairs $(\mathsf{E}_\mathsf{P}, \mathsf{E}_\mathsf{Q})$ with $\mathsf{E}_\mathsf{P} < C$ are achievable, while those with $\mathsf{E}_\mathsf{P} > C$ are not. Therefore, $\mathsf{E}_\mathsf{P}(\mathsf{E}_\mathsf{Q}) = C$. An analogous argument proves (9). ■

V. Error-Exponent Functions and $J_\alpha(P_{XY})$

After a preparatory lemma, we prove Theorem 3 and state Theorem 13, the analog of Theorem 3 for $\mathsf{E}_\mathsf{Q}^{**}(\cdot)$.

Lemma 12. *The convex conjugate of* $\mathsf{E}_\mathsf{P}(\cdot)$ *is*

$$\mathsf{E}_\mathsf{P}^*(\lambda) = \begin{cases} +\infty & \text{if } \lambda > 0, \\ \lambda \, J_{\frac{1}{1-\lambda}}(P_{XY}) & \text{otherwise.} \end{cases} \tag{43}$$

Proof. By the definition of the convex conjugate,

$$\mathsf{E}_\mathsf{P}^*(\lambda) = \sup_{\mathsf{E}_\mathsf{Q} \in \mathbb{R}} [\lambda \, \mathsf{E}_\mathsf{Q} - \mathsf{E}_\mathsf{P}(\mathsf{E}_\mathsf{Q})]. \tag{44}$$

For $\lambda > 0$, the RHS of (44) is $+\infty$, since we can lower-bound the supremum over E_Q with the limit as E_Q tends to infinity and since $\mathsf{E}_\mathsf{P}(\mathsf{E}_\mathsf{Q})$ is zero for all $\mathsf{E}_\mathsf{Q} \geq D(P_{XY} \| P_X P_Y)$, which can be verified by choosing $R_{XY} = P_{XY}$ in the RHS of (8).

Now assume $\lambda \leq 0$. Then,

$$\begin{aligned}
&\sup_{\mathsf{E}_\mathsf{Q} \in \mathbb{R}} [\lambda \, \mathsf{E}_\mathsf{Q} - \mathsf{E}_\mathsf{P}(\mathsf{E}_\mathsf{Q})] \\
&= \sup_{\mathsf{E}_\mathsf{Q} \in \mathbb{R}} \Bigl[\lambda \, \mathsf{E}_\mathsf{Q} - \inf_{\substack{R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}): \\ D(R_{XY} \| R_X R_Y) \leq \mathsf{E}_\mathsf{Q}}} D(R_{XY} \| P_{XY}) \Bigr] && (45) \\
&= \sup_{\mathsf{E}_\mathsf{Q} \in \mathbb{R}} \sup_{\substack{R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}): \\ D(R_{XY} \| R_X R_Y) \leq \mathsf{E}_\mathsf{Q}}} [\lambda \, \mathsf{E}_\mathsf{Q} - D(R_{XY} \| P_{XY})] && (46) \\
&= \sup_{R_{XY}} \sup_{\mathsf{E}_\mathsf{Q} \colon \mathsf{E}_\mathsf{Q} \geq D(R_{XY} \| R_X R_Y)} [\lambda \, \mathsf{E}_\mathsf{Q} - D(R_{XY} \| P_{XY})] && (47) \\
&= \sup_{R_{XY}} [\lambda \, D(R_{XY} \| R_X R_Y) - D(R_{XY} \| P_{XY})] && (48) \\
&= -(1-\lambda) \inf_{Q_X, Q_Y} \inf_{R_{XY}} \Bigl[\frac{-\lambda}{1-\lambda} D(R_{XY} \| Q_X Q_Y) \\
&\qquad + \frac{1}{1-\lambda} D(R_{XY} \| P_{XY}) \Bigr] && (49) \\
&= \lambda \inf_{Q_X, Q_Y} D_{\frac{1}{1-\lambda}}(P_{XY} \| Q_X Q_Y) && (50) \\
&= \lambda \, J_{\frac{1}{1-\lambda}}(P_{XY}), && (51)
\end{aligned}$$

where (45) follows from (8); (48) holds because $\lambda \leq 0$, so $\mathsf{E}_\mathsf{Q} = D(R_{XY} \| R_X R_Y)$ achieves the maximum; (49) follows from Lemma 4 because $\frac{-\lambda}{1-\lambda} \geq 0$ and $1 - \lambda \geq 1$; (50) follows from [3, Theorem 30] with $\alpha = \frac{1}{1-\lambda} \in (0, 1]$; and (51) follows from the definition of $J_\alpha(P_{XY})$. (Technically, the case $\alpha = 1$ is not covered by [3, Theorem 30], but it is easy to see that (50) also holds if $\alpha = 1$, i.e., if $\lambda = 0$.) ■

Proof of Theorem 3. Using Lemma 12, we have

$$\begin{aligned}
\mathsf{E}_\mathsf{P}^{**}(\mathsf{E}_\mathsf{Q}) &= \sup_{\lambda \in \mathbb{R}} [\lambda \, \mathsf{E}_\mathsf{Q} - \mathsf{E}_\mathsf{P}^*(\lambda)] && (52) \\
&= \sup_{\lambda \leq 0} [\lambda \, \mathsf{E}_\mathsf{Q} - \mathsf{E}_\mathsf{P}^*(\lambda)] && (53) \\
&= \sup_{\lambda \leq 0} [\lambda \, \mathsf{E}_\mathsf{Q} - \lambda \, J_{\frac{1}{1-\lambda}}(P_{XY})] && (54) \\
&= \sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} \bigl(J_\alpha(P_{XY}) - \mathsf{E}_\mathsf{Q} \bigr), && (55)
\end{aligned}$$

where (53) holds because $\mathsf{E}_\mathsf{P}^*(\lambda) = +\infty$ for all $\lambda > 0$, and (55) follows from the substitution $\alpha = \frac{1}{1-\lambda} \in (0, 1]$.

By [6, Theorem 4.2.1], a function $h \colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is equal to its biconjugate if, and only if, it is lower semicontinuous and convex. The function $\mathsf{E}_\mathsf{P} \colon \mathbb{R} \to [0, +\infty]$ is always lower semicontinuous. (This follows from a topological argument, which is omitted here.) Thus, $\mathsf{E}_\mathsf{P}(\cdot)$ is equal to its biconjugate if, and only if, it is convex. ■

Theorem 13. *For all* $\mathsf{E}_\mathsf{P} \in \mathbb{R}$,

$$\sup_{\alpha \in [0,1)} \Bigl[J_\alpha(P_{XY}) - \frac{\alpha}{1-\alpha} \mathsf{E}_\mathsf{P} \Bigr] = \mathsf{E}_\mathsf{Q}^{**}(\mathsf{E}_\mathsf{P}). \tag{56}$$

Furthermore, $\mathsf{E}_\mathsf{Q}^{**}(\mathsf{E}_\mathsf{P}) = \mathsf{E}_\mathsf{Q}(\mathsf{E}_\mathsf{P})$ *for all* $\mathsf{E}_\mathsf{P} \in \mathbb{R}$ *if, and only if,* $\mathsf{E}_\mathsf{Q}(\cdot)$ *is convex on* $\mathbb{R}$.

Proof. Omitted; the proof is similar to the proofs of Lemma 12 and Theorem 3. ■

VI. An Example Where $\mathsf{E}_\mathsf{P}(\cdot)$ Is Not Convex

Example 14. *Consider* $\mathcal{X} = \mathcal{Y} = \{1, 2, 3\}$ *and* P_{XY} *given by*

$P_{XY}(x,y)$	$y=1$	$y=2$	$y=3$
$x=1$	10^{-4}	γ	γ
$x=2$	γ	10^{-4}	γ
$x=3$	γ	γ	10^{-4},

where $\gamma = \frac{9997}{60000} \approx 0.167$. *Then,*

$$\mathsf{E}_\mathsf{P}\bigl(3898 \,/\, 2^{17}\bigr) \leq 58593464420737815 \,/\, 2^{56}, \tag{57}$$

$$\mathsf{E}_\mathsf{P}\bigl(3984 \,/\, 2^{17}\bigr) \leq 58382556630811219 \,/\, 2^{56}, \tag{58}$$

$$\mathsf{E}_\mathsf{P}\bigl(3941 \,/\, 2^{17}\bigr) \geq 58488010525784883 \,/\, 2^{56}. \tag{59}$$

This implies

$$\begin{aligned}
&\mathsf{E}_\mathsf{P}\Bigl(\frac{3898 + 3984}{2 \cdot 2^{17}}\Bigr) - \frac{1}{2} \mathsf{E}_\mathsf{P}\Bigl(\frac{3898}{2^{17}}\Bigr) - \frac{1}{2} \mathsf{E}_\mathsf{P}\Bigl(\frac{3984}{2^{17}}\Bigr) \\
&\quad = \mathsf{E}_\mathsf{P}\Bigl(\frac{3941}{2^{17}}\Bigr) - \frac{1}{2} \mathsf{E}_\mathsf{P}\Bigl(\frac{3898}{2^{17}}\Bigr) - \frac{1}{2} \mathsf{E}_\mathsf{P}\Bigl(\frac{3984}{2^{17}}\Bigr) && (60) \\
&\quad \geq 10366 \,/\, 2^{56} \approx 1.44 \cdot 10^{-13}, && (61)
\end{aligned}$$

so $\mathsf{E}_\mathsf{P}(\cdot)$ *is not convex. (We estimate the LHS of* (60) *to be in the order of* 10^{-7}*.)*

To verify (57), we use (8) and check (see Remark 17 below) that a specific $R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ satisfies

$$D(R_{XY} \| R_X R_Y) \leq 3898 \,/\, 2^{17}, \tag{62}$$

$$D(R_{XY} \| P_{XY}) \leq 58593464420737815 \,/\, 2^{56}. \tag{63}$$

Similarly, (58) can be verified. Establishing (59) is much more involved and is the topic of the rest of this section.

Let $\mathcal{Q}$ denote the set of all product distributions on $\mathcal{X} \times \mathcal{Y}$,

$$\mathcal{Q} \triangleq \{Q_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : Q_{XY} = Q_X Q_Y\}. \tag{64}$$

We express $\mathcal{Q}$ as a finite union, i.e.,

$$\mathcal{Q} = \bigcup_{i=1}^{k} \mathcal{Q}_i, \tag{65}$$

where for each $i \in \{1, \ldots, k\}$,

$$\mathcal{Q}_i \triangleq \{Q_X Q_Y : (Q_X \in \mathcal{Q}_{X,i}) \wedge (Q_Y \in \mathcal{Q}_{Y,i})\}, \quad (66)$$

$$\begin{aligned}\mathcal{Q}_{X,i} \triangleq \{Q_X \in \mathcal{P}(\mathcal{X}) : \quad &(l_{i,1} \le Q_X(1) \le u_{i,1}) \\ \wedge &(l_{i,2} \le Q_X(2) \le u_{i,2}) \\ \wedge &(l_{i,3} \le Q_X(3) \le u_{i,3})\}, \quad (67)\end{aligned}$$

$$\begin{aligned}\mathcal{Q}_{Y,i} \triangleq \{Q_Y \in \mathcal{P}(\mathcal{Y}) : \quad &(l_{i,4} \le Q_Y(1) \le u_{i,4}) \\ \wedge &(l_{i,5} \le Q_Y(2) \le u_{i,5}) \\ \wedge &(l_{i,6} \le Q_Y(3) \le u_{i,6})\}, \quad (68)\end{aligned}$$

where $l_{i,1}, \ldots, l_{i,6}$ and $u_{i,1}, \ldots, u_{i,6}$ are nonnegative numbers. With the help of Lemma 16 below, we can verify that for specific $\mathsf{E_Q} \in \mathbb{R}$ and $\Gamma \in \mathbb{R}$ and for all $i \in \{1, \ldots, k\}$,

$$\inf_{Q_{XY} \in \mathcal{Q}_i} \sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} \big(D_\alpha(P_{XY} \| Q_{XY}) - \mathsf{E_Q}\big) \ge \Gamma, \quad (69)$$

which by Lemma 15 below and (65) implies

$$\mathsf{E_P}(\mathsf{E_Q}) \ge \Gamma. \quad (70)$$

More details are given in Remark 17.

Lemma 15. *For all* $\mathsf{E_Q} \in \mathbb{R}$,

$$\mathsf{E_P}(\mathsf{E_Q}) = \inf_{Q_{XY} \in \mathcal{Q}} \sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} \big(D_\alpha(P_{XY} \| Q_{XY}) - \mathsf{E_Q}\big). \quad (71)$$

Proof. We have

$$\begin{aligned}\mathsf{E_P}(\mathsf{E_Q}) &= \inf_{\substack{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}): \\ D(R_{XY} \| R_X R_Y) \le \mathsf{E_Q}}} D(R_{XY} \| P_{XY}) \quad (72)\\ &= \inf_{Q_{XY} \in \mathcal{Q}} \inf_{\substack{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}): \\ D(R_{XY} \| Q_{XY}) \le \mathsf{E_Q}}} D(R_{XY} \| P_{XY}) \quad (73)\\ &= \inf_{Q_{XY} \in \mathcal{Q}} \sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} \big(D_\alpha(P_{XY} \| Q_{XY}) - \mathsf{E_Q}\big), \quad (74)\end{aligned}$$

where (72) follows from (8); (73) follows from Lemma 4; and (74) follows from Lemma 5. ■

Lemma 16. *Let* $\mathcal{Q}_i$ *be defined as in* (66)*, let* $\alpha \in (0, 1)$*, and let* $\beta \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_{\ge 0}$*. Define*

$$D \triangleq \inf_{Q_{XY} \in \mathcal{Q}_i} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \beta(x, y)\, Q_{XY}(x, y)^{1-\alpha}. \quad (75)$$

Then, for all $\mathsf{E_Q} \in \mathbb{R}$,

$$\begin{aligned}&\inf_{Q_{XY} \in \mathcal{Q}_i} \sup_{\tilde{\alpha} \in (0,1]} \frac{1-\tilde{\alpha}}{\tilde{\alpha}} \big(D_{\tilde{\alpha}}(P_{XY} \| Q_{XY}) - \mathsf{E_Q}\big) \\ &\quad \ge \frac{-1}{\alpha} \log\Bigg\{ \Bigg[\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \big(P_{XY}(x, y)^\alpha + \beta(x, y)\big)^{\frac{1}{\alpha}} \Bigg]^\alpha \\ &\qquad\qquad - D \Bigg\} - \frac{1-\alpha}{\alpha} \mathsf{E_Q}. \quad (76)\end{aligned}$$

Proof. Omitted. ■

Remark 17. *We finish with a few comments about the verification of Example 14.*

- *Computing* D *in Lemma 16 for fixed* $\mathcal{Q}_i$*,* α*, and* β *is easy: One can show that there exist an extreme point* Q_X^* *of* $\mathcal{Q}_{X,i}$ *and an extreme point* Q_Y^* *of* $\mathcal{Q}_{Y,i}$ *such that* $Q_X^* Q_Y^*$ *achieves the infimum in the RHS of* (75)*. (This basically holds because* $\mathcal{Q}_{X,i}$ *and* $\mathcal{Q}_{Y,i}$ *are bounded convex polytopes and because the objective function is concave in* Q_X *for fixed* Q_Y *and concave in* Q_Y *for fixed* Q_X*.) Since* $\mathcal{Q}_{X,i}$ *and* $\mathcal{Q}_{Y,i}$ *have at most six extreme points, there are at most 36 candidate points. One can evaluate the objective function at the candidate points; the minimum function value among these is equal to* D*.*
- *To establish* (59)*, we use* (65) *with* $k = 1\,323\,238$*. To ensure that* (65) *holds, we start with a collection* $\mathcal{C}$ *of sets that initially contains only* $\mathcal{Q}$*; we iteratively remove a* $\mathcal{Q}_i$ *from* $\mathcal{C}$*, split it into two parts, and add each part to* $\mathcal{C}$*; and we stop when* $\mathcal{C}$ *has the desired structure.*
- *We use interval arithmetic [15] to obtain exact bounds.*
- *The splits to obtain* $\mathcal{C}$*, the* α *and* β *needed in Lemma 16 for every* $i \in \{1, \ldots, k\}$*, and the code that allows for a mathematically rigorous verification of our bounds can be found in [16].*

References

[1] A. Lapidoth and C. Pfister, “Testing against independence and a Rényi information measure,” preprint, arXiv:1805.11059.

[2] A. Rényi, “On measures of entropy and information,” in *Proc. 4th Berkeley Symp. Math. Statist. and Probability*, vol. 1, 1961, pp. 547–561.

[3] T. van Erven and P. Harremoës, “Rényi divergence and Kullback–Leibler divergence,” *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

[4] A. Lapidoth and C. Pfister, “Two measures of dependence,” in *Proc. 2016 IEEE Int. Conf. Sci. Elect. Eng.*, pp. 1–5.

[5] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*, Belmont, MA: Athena Scientific, 2003.

[6] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, 2nd ed. New York: Springer, 2006.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge: Cambridge University Press, 2011.

[8] W. Hoeffding, “Asymptotically optimal tests for multinomial distributions,” *Ann. Math. Statist.*, vol. 36, no. 2, pp. 369–401, Apr. 1965.

[9] I. Csiszár and G. Longo, “On the error exponent for source coding and for testing simple statistical hypotheses,” *Stud. Sci. Math. Hungarica*, vol. 6, pp. 181–191, 1971.

[10] R. E. Blahut, “Hypothesis testing and information theory,” *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, July 1974.

[11] M. Tomamichel and M. Hayashi, “Operational interpretation of Rényi information measures via composite hypothesis testing against product and Markov distributions,” *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1064–1082, Feb. 2018.

[12] O. Zeitouni, J. Ziv, and N. Merhav, “When is the generalized likelihood ratio test optimal?,” *IEEE Trans. Inf. Theory*, vol. 38, no. 5, pp. 1597–1602, Sept. 1992.

[13] M. Feder and N. Merhav, “Universal composite hypothesis testing: a competitive minimax approach,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1504–1517, June 2002.

[14] R. R. Sokal and F. J. Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research*, 4th ed. New York: W. H. Freeman and Company, 2012.

[15] N. Revol and F. Rouillier, “Motivations for an arbitrary precision interval arithmetic and the MPFI library,” *Reliable Computing*, vol. 11, no. 4, pp. 275–290, Aug. 2005.

[16] A. Lapidoth and C. Pfister. *Supplementary Material*. (2018) [Online]. Available: https://github.com/pfistchr/arxiv-1805.11059