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

\[
\liminf_{n \to \infty} -\frac{1}{n} \log P^n_{XY}[T_n(X^n, Y^n) = 1] > E_P,
\]

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

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

\[
P^n_{XY}[T_n(X^n, Y^n) = 1] \leq e^{-n E_P},
\]

\[
(Q_X Q_Y)^{\times n}[T_n(X^n, Y^n) = 0] \leq e^{-n E_Q}.
\]

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

Our first result characterizes the achievable error-exponent pairs.

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

\[
\left(D(R_{XY}||P_{XY}) > E_P \right) \lor \left(D(R_{XY}||R_X R_Y) > E_Q\right). (5)
\]

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 $E_P : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ and $E_Q : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ as

\[
E_P(E_Q) \triangleq \sup \{E_P \in \mathbb{R} : (E_P, E_Q) \text{ is achievable}\},
\]

\[
E_Q(E_P) \triangleq \sup \{E_Q \in \mathbb{R} : (E_P, E_Q) \text{ is achievable}\},
\]

we obtain

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

\[
E_P(E_Q) = \inf_{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : D(R_{XY}||P_{XY}) \leq E_Q} D(R_{XY}||R_X R_Y), (8)
\]

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

\[
E_Q(E_P) = \inf_{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : D(R_{XY}||P_{XY}) \leq E_P} D(R_{XY}||R_X R_Y). (9)
\]

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

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

\[
\sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} (J_\alpha(P_{XY}) - E_Q) = E_P^\alpha(E_Q). (10)
\]

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

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

\[
\sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha} (J_\alpha(P_{XY}) - E_Q) \neq E_P(E_Q). (11)
\]
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 \( J_0(P_{XY}) \) to the Fenchel biconjugates \( E^*_P \) and \( E^*_Q \); and in Section VI, we discuss Example 14, where \( E_P(\cdot) \) is not convex. Additional proofs can be found in [1].

II. PRELIMINARIES

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

\[
D(P||Q) \triangleq \sum_{z \in Z} P(z) \log \frac{P(z)}{Q(z)}
\]

(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 \( \alpha \), [2], [3] is defined for all positive \( \alpha \) other than 1 as

\[
D_\alpha(P||Q) \triangleq \frac{1}{\alpha - 1} \log \sum_{z \in Z} P(z)^\alpha Q(z)^{1-\alpha}
\]

(13)

with the conventions that \( \log 0 = -\infty \) and that for \( \alpha > 1 \), we read \( P(z)^\alpha /Q(z)^{\alpha-1} \) 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], \( D_1(P||Q) = D(P||Q) \).

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

\[
J_\alpha(P_{XY}) \triangleq \min_{Q_X \in \mathcal{P}(X), Q_Y \in \mathcal{P}(Y)} D_\alpha(P_{XY}||Q_X Q_Y)
\]

(14)

for all positive \( \alpha \) and as zero when \( \alpha = 0 \).

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

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

(15)

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

The (Fenchel) biconjugate of a function \( f : \mathbb{R} \to [-\infty, +\infty] \) is \( f^{**} : \mathbb{R} \to [-\infty, +\infty] \), 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 \( \tilde{E}_P(E_Q) \) denote the error-exponent function that would have resulted had we replaced the strict inequalities in (1) and (2) with weak inequalities. Then, \( \tilde{E}_P(\cdot) \) and \( \tilde{E}_P^*(\cdot) \) cannot be equal because, unlike \( E_P(\cdot) \), \( \tilde{E}_P(\cdot) \) is not lower semicontinuous. The difference between \( E_P(\cdot) \) and \( \tilde{E}_P(\cdot) \) is best seen at zero: While \( \tilde{E}_P(0) \) is +\( \infty \), it turns out that \( E_P(0) \) is the optimal error exponent if for a fixed \( \epsilon \in (0, 1) \), we require the tests to satisfy \( (Q_X Q_Y)^{xy[T_\epsilon(X^n, Y^n) = 0]} \leq \epsilon \) 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 \( E_P \in \mathbb{R} \) if \( E_Q = 0 \) and \( T_\epsilon(X^n, Y^n) \) \( = 0 \) irrespective of \( X^n \) and \( Y^n \). Yet, (8) implies that \( E_P(0) \) is finite, so \( (E_P, 0) \) is not achievable for every \( E_P \).

We conclude this section with two lemmas.

Lemma 4. For all \( R_{XY} \), \( Q_X \), and \( Q_Y \),

\[
D(R_{XY}||Q_X Q_Y) \geq D(R_{XY}||R_X R_Y).
\]

(16)

Consequently,

\[
\inf_{Q_X, Q_Y} D(R_{XY}||Q_X Q_Y) = D(R_{XY}||R_X R_Y).
\]

(17)

Proof. We have

\[
D(R_{XY}||Q_X Q_Y) = D(R_{XY}||R_X R_Y) + D(R_X||Q_X) + D(R_Y||Q_Y)
\]

(18)

\[
\geq D(R_{XY}||R_X R_Y)
\]

(19)

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 \( Z \). Then, for all \( E_Q \in \mathbb{R} \),

\[
\inf_{R \in \mathcal{P}(Z)} D(R||P) = \sup_{\alpha \in (0,1]} \frac{1}{\alpha} \left( D_\alpha(P||Q) - E_Q \right).
\]

(20)

Proof. Omitted.

III. RELATED WORK

Let \( P \) and \( Q \) be PMFs over a finite set \( Z \). 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

\[
\tilde{E}_P(E_Q) = \inf_{R \in \mathcal{P}(Z) : D(R||Q) \leq E_Q} D(R||P),
\]

(21)

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

\[
\tilde{E}_P(E_Q) = \sup_{\alpha \in (0,1]} \frac{1}{\alpha} (D_\alpha(P||Q) - E_Q),
\]

(22)

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

In the composite hypothesis testing setup where \( P \) is tested against an unknown \( Q \) from some set \( Q \), Hoeffding [8] showed that his likelihood-ratio test is asymptotically optimal against all \( Q \in Q \); 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 \( E_Q \),

\[
\sup_{\alpha \in (0,1]} \frac{1}{\alpha} \left( J_\alpha(P_{XY}) - E_Q \right) = \tilde{E}_P(E_Q).
\]

(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 \( E_P(\cdot) \) 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 $(E, E_Q)$ for which

$$(D(R_{XY}||P_{XY}) > E) \vee (D(R_{XY}||R_X R_Y) > E_Q)$$

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

$$\mathbb{P}(R_{XY}||P_{XY}) \geq E + \epsilon \quad \text{and} \quad \mathbb{P}(R_{XY}||R_X R_Y) \geq E_Q + \epsilon.$$  

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

$$R_{XY} \mapsto \min\left\{ D(R_{XY}||P_{XY}) - E, D(R_{XY}||R_X R_Y) - E_Q \right\}.$$  

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

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

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 $P(\mathcal{X} \times \mathcal{Y})$ and that $P(\mathcal{X} \times \mathcal{Y})$ is a compact set. By the extreme value theorem, there would exist an $\hat{R}_{XY} \in P(\mathcal{X} \times \mathcal{Y})$ with $f(\hat{R}_{XY}) = \eta \leq 0$. This leads to a contradiction because then (24) would not hold for $\hat{R}_{XY}$.

Lemma 7. Let $E, E_Q,$ and $\epsilon > 0$ be such that (25) holds for all $R_{XY} \in P(\mathcal{X} \times \mathcal{Y})$. Define $\tau \triangleq (n + 1)^{1/2}$, Then, for all $Q_{XY} \in P(X)$ and all $Q_Y \in P(Y)$,

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

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

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

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

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

$$P_{XY}^n[T_n(X^n, Y^n) = 1] = P_{XY}^n[D(\hat{R}_{XY}||R_X \hat{R}_Y) < E_Q + \epsilon] \leq (n + 1)^{1/2} e^{-n(E_P + \epsilon)},$$

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 $R_{XY} \in P(\mathcal{X} \times \mathcal{Y})$, then there exists an $\epsilon > 0$ such that the error-exponent pair $(E, E_Q)$ is achieved by the sequence of tests

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

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

$$\Gamma \triangleq \frac{1}{n} \log \frac{P_{XY}^n(X^n, Y^n)}{Q_X^n Q_Y^n} \leq D(\hat{R}_{XY}||R_X \hat{R}_Y) - D(\hat{R}_{XY}||P_{XY})$$

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

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

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

$$P_{XY}^n[T_n(X^n, Y^n) = 1] \leq P_{XY}^n[D(\hat{R}_{XY}||R_X \hat{R}_Y) < E_Q + \epsilon] + P_{XY}^n[D(\hat{R}_{XY}||P_{XY}) \geq E_P + \epsilon] \leq 2(n + 1)^{1/2} e^{-n(E_P + \epsilon)},$$

where (39) follows from the union bound because the events $D(\hat{R}_{XY}||R_X \hat{R}_Y) \geq E_Q + \epsilon$ and $D(\hat{R}_{XY}||P_{XY}) \leq E_P + \epsilon$ imply $\Gamma > E_Q - E_P$, 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 $R_{XY} \in P(\mathcal{X} \times \mathcal{Y})$, i.e., if there exists an $\hat{R}_{XY} \in P(\mathcal{X} \times \mathcal{Y})$ satisfying

$$D(\hat{R}_{XY}^*||P_{XY}) \leq E_P$$

then the error-exponent pair $(E_P, E_Q)$ 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 $E_Q \in \mathbb{R}$, define

$$C \triangleq \inf_{R_{XY} \in P(\mathcal{X} \times \mathcal{Y})} D(R_{XY}||P_{XY}).$$

(42)
By Theorem 1, all error-exponent pairs \( (E_P, E_Q) \) with \( E_P < C \) are achievable, while those with \( E_P > C \) are not. Therefore, \( E_P(E_Q) = C \). An analogous argument proves (9).

V. ERROR-EXEMPLARY FUNCTIONS AND J\(_\alpha\)(P\(_{XY}\))

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

**Lemma 12.** The convex conjugate of \( E_P(\cdot) \) is

\[
E_P^*(\lambda) = \begin{cases} 
\infty & \text{if } \lambda > 0, \\
\lambda J_{\frac{1}{1-\lambda}}(P_{XY}) & \text{otherwise.}
\end{cases}
\]

(43)

**Proof.** By the definition of the convex conjugate,

\[
E_P^*(\lambda) = \sup_{E_Q \in \mathbb{R}} [\lambda E_Q - E_P(E_Q)].
\]

(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 \( E_P(E_Q) \) is zero for all \( E_Q \geq D(P_{XY}||P_X Y) \), which can be verified by checking \( R_{XY} = P_{XY} \) in the RHS of (8).

Now assume \( \lambda \leq 0 \). Then,

\[
\sup_{E_Q \in \mathbb{R}} [\lambda E_Q - E_P(E_Q)]
= \sup_{E_Q \in \mathbb{R}} [\lambda E_Q - \inf_{R_{XY} \in \mathcal{P}(X \times Y)} D(R_{XY}||P_{XY})]
= \sup_{E_Q \in \mathbb{R}} \sup_{R_{XY} \in \mathcal{P}(X \times Y)} [\lambda E_Q - D(R_{XY}||P_{XY})]
= \sup_{R_{XY}} [\lambda D(R_{XY}||P_X Y) - D(R_{XY}||P_{XY})]
= -(1 - \lambda) \inf_{Q_{XY} \in \mathcal{P}(X \times Y)} \frac{1}{1 - \lambda} D(R_{XY}||P_{XY})
= \lambda J_{\frac{1}{1-\lambda}}(P_{XY}),
\]

(45)

(46)

(47)

(48)

(49)

(50)

(51)

where (45) follows from (8); (48) holds because \( \lambda \leq 0 \), so \( E_Q = D(R_{XY}||P_X Y) \) achieves the maximum; (49) follows from Lemma 4 because \( \frac{1}{1-\lambda} \geq 0 \) 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 \( J_{\frac{1}{1-\lambda}}(P_{XY}) \). (Technically, the case \( \alpha = 1 \) is not covered by [3, Theorem 30], but it is clear to see that (50) also holds if \( \alpha = 1 \), i.e., if \( \lambda = 0 \).)

**Proof of Theorem 3.** Using Lemma 12, we have

\[
E_P^*(E_Q) = \sup_{\lambda \in \mathbb{R}} [\lambda E_Q - E^*_P(\lambda)]
= \sup_{\lambda \leq 0} [\lambda E_Q - E^*_P(\lambda)]
= \sup_{\lambda \leq 0} [\lambda E_Q - \lambda J_{\frac{1}{1-\lambda}}(P_{XY})]
= \sup_{\alpha \in [0, 1]} \frac{1-\alpha}{\alpha} (J_{\alpha}(P_{XY}) - E_Q),
\]

(52)

(53)

(54)

(55)

where (53) holds because \( E^*_P(\lambda) = +\infty \) 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 \( h : \mathbb{R} \to \mathbb{R} \cup \{+\infty\} \) is equal to its biconjugate if, and only if, it is lower semi-continuous and convex. The function \( E_P : \mathbb{R} \to [0, +\infty] \) is always lower semi-continuous. (This follows from a topological argument, which is omitted here.) Thus, \( E_P(\cdot) \) is equal to its biconjugate if, and only if, it is convex.

**Theorem 13.** For all \( E_P \in \mathbb{R} \),

\[
\sup_{\alpha \in [0, 1]} \left[ J_{\alpha}(P_{XY}) - \frac{\alpha}{1-\alpha} E_P \right] = E_Q^{**}(E_P).
\]

(56)

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

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

VI. AN EXAMPLE WHERE \( E_P(\cdot) \) IS NOT CONVEX

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

\[
\begin{array}{cccc}
P_{XY}(x, y) & y = 1 & y = 2 & y = 3 \\
\quad x = 1 & 10^{-4} & \gamma & \gamma \\
\quad x = 2 & 10^{-4} & \gamma & \gamma \\
\quad x = 3 & \gamma & \gamma & 10^{-4},
\end{array}
\]

where \( \gamma = \frac{9997}{9999} \approx 0.167 \). Then,

\[
E_P(\frac{3988}{2^{17}}) \leq 58593464420737815 / 2^{56},
\]

(57)

\[
E_P(\frac{3984}{2^{17}}) \leq 5838255630811219 / 2^{56},
\]

(58)

\[
E_P(\frac{3941}{2^{17}}) \geq 58488201525784883 / 2^{56}.
\]

(59)

This implies

\[
E_P^*(\frac{3988 + 3984}{2 \cdot 2^{17}} - \frac{1}{2} E_P(\frac{3988}{2^{17}}) - \frac{1}{2} E_P(\frac{3984}{2^{17}}))
= \frac{3941}{2^{17}} - \frac{1}{2} E_P(\frac{3988}{2^{17}}) - \frac{1}{2} E_P(\frac{3984}{2^{17}})
\geq 10366 / 2^{56} \approx 1.44 \cdot 10^{-13},
\]

(60)

(61)

so \( E_P(\cdot) \) 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

\[
D(R_{XY}||P_X Y) \leq \frac{3988}{2^{17}},
\]

(62)

\[
D(R_{XY}||P_{XY}) \leq \frac{3988}{2^{17}}.
\]

(63)

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

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

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

(64)

We express \( Q \) as a finite union, i.e.,

\[
Q = \bigcup_{i=1}^k Q_i.
\]

(65)
where for each $i \in \{1, \ldots, k\}$,
\[
\begin{align*}
Q_{X,i} &= \{ Q_X Q_Y: (Q_X \in Q_{X,i}) \land (Q_Y \in Q_{Y,i}) \}, \quad (66) \\
Q_{X,Y,i} &= \{ Q_X \in \mathcal{P}(X): \quad (l_{i,1} \leq Q_X(1) \leq u_{i,1}) \\
&\quad \land (l_{i,2} \leq Q_X(2) \leq u_{i,2}) \\
&\quad \land (l_{i,3} \leq Q_X(3) \leq u_{i,3}) \}, \quad (67) \\
Q_{Y,i} &= \{ Q_Y \in \mathcal{P}(Y): \quad (l_{i,4} \leq Q_Y(1) \leq u_{i,4}) \\
&\quad \land (l_{i,5} \leq Q_Y(2) \leq u_{i,5}) \\
&\quad \land (l_{i,6} \leq Q_Y(3) \leq u_{i,6}) \}, \quad (68)
\end{align*}
\]

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 $E_Q \in \mathbb{R}$ and $\Gamma \in \mathbb{R}$ and for all $i \in \{1, \ldots, k\}$,
\[
\inf_{Q_X \in Q_i} \sup_{a \in [0,1]} \frac{1 - a}{a} (D(a(P_X || Q_X) - E_Q) \geq \Gamma, \quad (69)
\]
which by Lemma 15 below and (65) implies
\[
E_P(E_Q) \geq \Gamma. \quad (70)
\]

More details are given in Remark 17.

**Lemma 15.**\textit{ For all }$E_Q \in \mathbb{R}$,\textit{ }\[
E_P(E_Q) = \inf_{Q_X \in Q} \sup_{a \in [0,1]} \frac{1 - a}{a} (D(a(P_X || Q_X) - E_Q). \quad (71)
\]

**Proof.**\textit{ We have }\[
E_P(E_Q) = \left\{ \begin{array}{ll}
\inf_{R_{X,Y} \in \mathcal{P}(X \times Y)} & D(R_{X,Y} || P_{X,Y}) \land D(R_{X,Y} || R_{X,Y} \leq E_Q) \\
\inf_{Q_X \in Q} & \inf_{R_{X,Y} \in \mathcal{P}(X \times Y)} D(R_{X,Y} || P_{X,Y}) \land D(R_{X,Y} || Q_{X,Y} \leq E_Q) \\
\inf_{Q_X \in Q} & \sup_{a \in [0,1]} \frac{1 - a}{a} (D(a(P_X || Q_X) - E_Q), \quad (74)
\end{array} \right.
\]

where (72) follows from (8); (73) follows from Lemma 4; and (74) follows from Lemma 5. \hfill \[ \square \]

**Lemma 16.**\textit{ Let }$Q_i$\textit{ be defined as in }$(66)$, \textit{let }$a \in (0, 1)$, \textit{and let }$\beta: X \times Y \rightarrow \mathbb{R}_{\geq 0}$. \textit{Define }\[
D \triangleq \inf_{Q_X \in Q_i, (x,y) \in X \times Y} \beta(x, y) Q_X(x, y)^{1-a}. \quad (75)
\]

\textit{Then, for all }$E_Q \in \mathbb{R}$,\textit{ }\[
\inf_{Q_X \in Q_i} \sup_{a \in [0,1]} \frac{1 - a}{a} (D(a(P_X || Q_X) - E_Q) \geq \frac{1}{a} \log \left\{ \sum_{(x,y) \in X \times Y} (P_X(x, y)^a + \beta(x, y) \right)^a - D \right\} - \frac{1 - a}{a} E_Q. \quad (76)
\]

**Proof.**\textit{ Omitted.} \hfill \[ \square \]

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

- \textit{Computing }$D$\textit{ in Lemma 16 for fixed }$Q_i$, $\alpha$, and $\beta$ \textit{is easy: One can show that there exist an extreme point }$Q_X^* \subseteq Q_{X,i}$ \textit{and an extreme point }$Q_Y^* \subseteq Q_{Y,i}$ \textit{such that }$Q_X^* Q_Y^*$ \textit{achieves the infimum in the RHS of (75). (This basically holds because }$Q_{X,i}$ \textit{and }$Q_{Y,i}$ \textit{are bounded convex polytopes and because the objective function is concave in }$Q_X$\textit{ for fixed }$Q_Y$\textit{ and concave in }$Q_Y$\textit{ for fixed }$Q_X$, \textit{since }$Q_{X,i}$ \textit{and }$Q_{Y,i}$ \textit{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 }$D$.\textit{ To establish (59), we use (65) with }$k = 1323238$. \textit{To ensure that (65) holds, we start with a collection }$C$ \textit{of sets that initially contains only }$Q$; \textit{we iteratively remove a }$Q_i$ \textit{from }$C$\textit{, split it into two parts, and add each part to }$C$; \textit{and we stop when }$C$ \textit{has the desired structure.}\textit{ We use interval arithmetic [15] to obtain exact bounds.}\textit{ The splits to obtain }$C$, \textit{the }$\alpha$ \textit{and }$\beta$ \textit{needed in Lemma 16 for every }$i \in \{1, \ldots, k\}$, \textit{and the code that allows for a mathematically rigorous verification of our bounds can be found in [16].}

**REFERENCES**