



IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

On the Rényi Divergence and the Joint Range of Relative Entropies

Igal Sason

CCIT Report #878
February 2015

■ ■ ■ ■ Electronics
■ ■ ■ ■ Computers
■ ■ ■ ■ Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL



On the Rényi Divergence and the Joint Range of Relative Entropies

Igal Sason

Abstract

This paper starts with a study of the minimum of the Rényi divergence subject to a fixed (or minimal) value of the total variation distance. Relying on the solution of this minimization problem, we determine the exact region of the points $(D(Q||P_1), D(Q||P_2))$ where P_1 and P_2 are any probability distributions whose total variation distance is not below a fixed value, and the probability distribution Q is arbitrary (none of these three distributions is assumed to be fixed). It is further shown that all the points of this convex region are attained by a triple of 2-element probability distributions. As a byproduct of this characterization, we provide a geometric interpretation of the minimal Chernoff information subject to a minimal total variation distance.

Keywords: Chernoff information, Lagrange duality, relative entropy, Rényi divergence, total variation distance.

I. INTRODUCTION

The Rényi divergence, introduced in [22], has been studied so far in various information-theoretic contexts (and it has been actually used before it had a name [24]). These include generalized cutoff rates for hypothesis testing ([1], [7]), generalized guessing moments [9], strong converse theorems for classes of networks [11], channel coding error exponents ([14], [19], [24]), strong data processing theorems for discrete memoryless channels [20], two-sensor composite hypothesis testing [25], and one-shot bounds for various information-theoretic problems [29].

This work starts with a study of the minimum of the Rényi divergence subject to a fixed (or minimal) value of the total variation distance. The derivation of an exact expression for this minimum is initialized by adapting some arguments that have been used by Fedotov *et al.* [10] for the minimization of the relative entropy (a.k.a. Kullback-Leibler divergence), subject to a fixed value of the total variation distance. Our analysis further relies on the Lagrange duality and a solution of the Karush-Kuhn-Tucker (KKT) equations, while asserting strong duality for the studied problem. The use of Lagrange duality significantly simplifies the computational task of the studied minimization problem. The exact expression for the Rényi divergence generalizes, in a non-trivial way, previous studies of the minimization of the relative entropy under the same constraint on the total variation distance (see [10], [15], [21]). The exact expression for this minimum is also compared with known Pinsker-type lower bounds on the Rényi divergence [16] when the total variation distance is fixed. It should be noted that the studied problem minimizes the Rényi divergence w.r.t. all pairs of probability distributions with a total variation distance which is not below a given value; this differs from the type of problems studied in [3] and [18], in connection to the minimization of the relative entropy $D(P||Q)$ with a minimal allowed value of the total variation distance, where the probability distribution (PD) Q was fixed.

I. Sason is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: sason@ee.technion.ac.il). This work has been supported by the Israeli Science Foundation, grant 12/12. The paper was submitted to the IEEE Transactions on Information Theory in January 15, 2015. This work is submitted in part to the 2015 IEEE International Symposium on Information Theory (Hong-Kong, China).

The motivation for this study is provided in the second part of this work. It is used to obtain an exact characterization of the joint range of the relative entropies $(D(Q||P_1), D(Q||P_2))$ where P_1 and P_2 are arbitrary PDs whose total variation distance is not below a fixed value, and Q is an arbitrary PD (note that none of these three PDs is set to be fixed). The use of such a characterization can be exemplified, due to the significance of the relative entropy in various fundamental problems in information theory and statistics. These include, e.g., the characterization of the gap of the compression rate, in lossless source coding, to the entropy of a memoryless and stationary source when there exists a mismatch between the assumed PD of the source and its true PD. Such a characterization implies that if Q is an arbitrary PD of a memoryless and stationary source, and two lossless source encoders assume arbitrary PDs of P_1 and P_2 whose total variation distance is not below a fixed value (the total variation distance is used here as a quantitative measure for the distinction between P_1 and P_2), then the simultaneous gaps of their compression rates to the entropy of the source is an interior or a boundary point of the region which is specified by the joint range of the relative entropies $(D(Q||P_1), D(Q||P_2))$ as above. Every point in this region is shown to be achievable by a triple of 2-element PDs P_1 , P_2 and Q , and these PDs are determined by relying on the solution to the problem of minimizing the Rényi divergence subject to a minimal total variation distance. Furthermore, no point outside this region is achievable, irrespectively of the PDs P_1 , P_2 and Q as above. This region can get some other interpretations since the relative entropy plays a fundamental role in the method of types, the large deviation theory (Sanov's theorem), the conditional limit theorem, and the characterization of the best achievable error exponent for a Bayesian probability of error in binary hypothesis testing [6, Chapter 11]. Note that the studied problem here differs from the study in [13] which considered the joint range of f -divergences for a single pair of PDs.

This paper is focused on the Rényi divergence. To avoid confusion, recall the distinction between the Rényi divergence and the α -divergence that has been introduced by Sundaresan [27], [28]. Unlike the Rényi divergence, it does not satisfy the data processing inequality; this property is used in this paper for the minimization of the Rényi divergence for a fixed (or minimal) total variation distance. Nevertheless, the α -divergence satisfies the Pythagorean property on convex and closed sets of general PDs [2], and it has found some information-theoretic interpretations in the context of guessing under uncertainty [28], and mismatched encoding of tasks [5].

This paper is structured as follows: Section II solves the minimization problem for the Rényi divergence under a fixed total variation distance, Section III provides an exact characterization of the considered joint range of the relative entropies. Proofs are mostly relegated to the appendices.

Definitions and Notation

We end this section by providing the definitions, and setting the notation used in this paper.

Definition 1 (Rényi divergence): Let P and Q be two PDs defined on a countable set \mathcal{X} , and let $\alpha \in (0, 1) \cup (1, \infty)$. The Rényi divergence of order α of P from Q is given by

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \left(\sum_{x \in \mathcal{X}} P^\alpha(x) Q^{1-\alpha}(x) \right) \quad (1)$$

with the convention that if $\alpha > 1$ and $Q(x) = 0$ then $P^\alpha(x) Q^{1-\alpha}(x)$ equals 0 or ∞ if $P(x) = 0$ or $P(x) > 0$, respectively. The extreme cases of $\alpha = 0, 1, \infty$ are defined as follows:

- If $\alpha = 0$ then $D_0(P||Q) = -\log Q(\text{Support}(P))$ where $\text{Support}(P) = \{x \in \mathcal{X} : P(x) > 0\}$ denotes the support of P ,
- If $\alpha = +\infty$ then $D_\infty(P||Q) = \log \left(\text{ess sup } \frac{P}{Q} \right)$ where $\text{ess sup } f$ denotes the essential supremum of a function f ,

- If $\alpha = 1$, it is defined to be the relative entropy $D(P||Q) = \sum P(x) \log \frac{P(x)}{Q(x)}$.

If $D(P||Q) < \infty$, it can be verified by L'Hôpital's rule that $D(P||Q) = \lim_{\alpha \rightarrow 1^-} D_\alpha(P||Q)$. Properties of the Rényi divergence are provided in [8] (with a summary in [8, p. 3799]).

Another measure used in this paper is the total variation distance, defined as follows:

Definition 2 (Total variation distance): Let P and Q be two PDs defined on a set \mathcal{X} . The total variation distance between P and Q is defined by

$$d_{\text{TV}}(P, Q) \triangleq \sup_{A \subseteq \mathcal{X}} |P(A) - Q(A)|. \quad (2)$$

If \mathcal{X} is a countable set, the total variation distance in (2) can be simplified to

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \sum_x |P(x) - Q(x)| = \frac{\|P - Q\|_1}{2} \quad (3)$$

so, the total variation distance is equal to one-half the l_1 -distance between P and Q . In the continuous setting, PDs are replaced by *pdfs*, and the sum in (3) is replaced by an integral.

Throughout this paper, \mathbb{N} denotes the set of positive integers and logarithms are to the base e .

II. MINIMIZATION OF THE RÉNYI DIVERGENCE SUBJECT TO A FIXED TOTAL VARIATION DISTANCE

The task of minimizing an arbitrary *symmetric* f -divergence for a fixed total variation distance has been studied in [15], leading to a closed-form solution of this optimization problem. Although the Rényi divergence is not an f -divergence, it is a function of an f -divergence; however, this f -divergence is *asymmetric*, except for the case where $\alpha = \frac{1}{2}$, so the closed-form expression in [15] cannot be utilized to obtain a tight lower bound on the Rényi divergence subject to a fixed total variation distance.

In this section, we derive a tight lower bound on the Rényi divergence $D_\alpha(P_1||P_2)$ subject to a fixed total variation distance between P_1 and P_2 . We further show that this lower bound is attained with equality for a pair of 2-element probability distributions P_1 and P_2 , and both distributions are obtained explicitly in terms of the order α and the fixed total variation distance $d_{\text{TV}}(P_1, P_2) = \varepsilon \in [0, 1)$ (note that if $\varepsilon = 1$ then $\text{Supp}(P_1) \cap \text{Supp}(P_2) = \emptyset$, and consequently $D_\alpha(P_1||P_2) = \infty$). For orders $\alpha \in (0, 1)$, the new tight lower bound is compared with existing Pinsker-type lower bounds on the Rényi divergence [16]. The special case where $\alpha = 1$, which is particularized to the minimization of the relative entropy subject to a fixed total variation distance, has been studied extensively, and three equivalent forms of the solution to this optimization problem have been derived in [10], [15] and [21].

In [16, Corollaries 6 and 9], Gilardoni derived two Pinsker-type lower bounds on the Rényi divergence of order $\alpha \in (0, 1)$ in terms of the total variation distance. Among these two bounds, the improved lower bound is

$$D_\alpha(P||Q) \geq 2\alpha\varepsilon^2 + \frac{4}{9}\alpha(1 + 5\alpha - 5\alpha^2)\varepsilon^4, \quad \forall \alpha \in (0, 1) \quad (4)$$

where $\varepsilon \triangleq d_{\text{TV}}(P, Q)$ denotes the total variation distance between P and Q (see Definition 2). Note that in the limit where $\varepsilon \rightarrow 1$, this lower bound converges to a finite limit that is at most $\frac{22}{9}$. This, however, is an artifact of the lower bound, as it is stated in the following:

Lemma 1:

$$\lim_{\varepsilon \rightarrow 1^-} \inf_{P, Q: d_{\text{TV}}(P, Q) = \varepsilon} D_\alpha(P||Q) = \infty, \quad \forall \alpha > 0. \quad (5)$$

Proof: See Appendix I-A. ■

Lemma 1 motivates a study of the exact characterization of the infimum (or minimum) of the Rényi divergence for a fixed total variation distance. In the following, we derive a tight lower bound which is shown to be achievable by pairs of 2-element PDs for any fixed value $\varepsilon \in [0, 1)$ of the total variation distance. For $\alpha > 0$, let

$$g_\alpha(\varepsilon) \triangleq \inf_{P_1, P_2: d_{\text{TV}}(P_1, P_2) = \varepsilon} D_\alpha(P_1 \| P_2), \quad \forall \varepsilon \in [0, 1). \quad (6)$$

Since $g_\alpha(\varepsilon)$ is monotonic non-decreasing in $\varepsilon \in [0, 1)$, it can be expressed as

$$g_\alpha(\varepsilon) = \inf_{P_1, P_2: d_{\text{TV}}(P_1, P_2) \geq \varepsilon} D_\alpha(P_1 \| P_2), \quad \forall \varepsilon \in [0, 1). \quad (7)$$

Remark 1: For $\alpha \in [0, 1]$, since $D_\alpha(P \| Q)$ is jointly convex in (P, Q) , the same arguments by Fedotov *et al.* [10] yield that g_α is convex, and the infimum in (6) and (7) is a minimum.

In the following, we evaluate the function g_α in (6) and (7). Following [10, Section 2] that characterizes the minimum of the relative entropy in terms of the total variation distance, we first extend their argument to prove this lemma:

Lemma 2: There is no loss of generality by restricting the minimization in (6) or (7) to pairs of 2-element PDs.

Proof: See Appendix I-B. ■

The following proposition provides an expression for g_α .

Proposition 1: Let $\alpha \in (0, 1) \cup (1, \infty)$ and $\varepsilon \in [0, 1)$. The function g_α in (6) satisfies

$$g_\alpha(\varepsilon) = \min_{p, q \in [0, 1]: |p - q| \geq \varepsilon} d_\alpha(p \| q) \quad (8)$$

where

$$d_\alpha(p \| q) \triangleq \frac{\log(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha})}{\alpha - 1} \quad (9)$$

is the Rényi divergence $D_\alpha(P \| Q)$ between the 2-element PDs $P = (p, 1-p)$ and $Q = (q, 1-q)$.

Proof: Eq. (8) follows from Lemma 2 where $D_\alpha(P_1 \| P_2)$ is minimized over all pairs of 2-element PDs $P_1 = (p, 1-p)$, $P_2 = (q, 1-q)$ with $|p - q| = d_{\text{TV}}(P_1, P_2) \geq \varepsilon$. ■

Corollary 1: For $\alpha \in (0, 1)$ and $\varepsilon \in [0, 1)$

$$g_\alpha(\varepsilon) = \left(\frac{\alpha}{1 - \alpha} \right) g_{1-\alpha}(\varepsilon), \quad (10)$$

and

$$g_\alpha(\varepsilon) \geq c_1(\alpha) \log \left(\frac{1}{1 - \varepsilon} \right) + c_2(\alpha), \quad (11)$$

where $c_1(\alpha) \triangleq \min \left\{ 1, \frac{\alpha}{1-\alpha} \right\}$, and $c_2(\alpha) \triangleq -\frac{\log(2)}{1-\alpha}$.

For $\alpha = \frac{1}{2}$ and $\alpha = 2$, the function g_α admits the following closed-form expressions:

$$g_{\frac{1}{2}}(\varepsilon) = -\log(1 - \varepsilon^2), \quad (12)$$

and

$$g_2(\varepsilon) = \begin{cases} \log(1 + 4\varepsilon^2), & \text{if } \varepsilon \in [0, \frac{1}{2}], \\ \log \left(\frac{1}{1-\varepsilon} \right), & \text{if } \varepsilon \in (\frac{1}{2}, 1). \end{cases} \quad (13)$$

Proof: See Appendix II. ■

Remark 2: The lower bound on g_α in (11) provides another proof of Lemma 1, showing that

$$\lim_{\varepsilon \rightarrow 1^-} g_\alpha(\varepsilon) = \infty$$

for $\alpha \in (0, 1)$; this lemma also holds for $\alpha \geq 1$ since D_α is monotonic non-decreasing in $\alpha \in (0, \infty)$, and due to the definition of g_α in (6).

Solving the Optimization Problem in Proposition 1 for $\alpha \in (0, 1)$

In the following, we use Lagrange duality to obtain an alternative expression for a solution of the minimization problem. This simplifies considerably the computational task of the solution to this problem, as explained below.

Lemma 3: Let $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1)$. The function

$$f_{\alpha, \varepsilon}(q) \triangleq \frac{\left(1 - \frac{\varepsilon}{1-q}\right)^{\alpha-1} - \left(1 + \frac{\varepsilon}{q}\right)^{\alpha-1}}{\left(1 + \frac{\varepsilon}{q}\right)^{\alpha} - \left(1 - \frac{\varepsilon}{1-q}\right)^{\alpha}}, \quad \forall q \in (0, 1 - \varepsilon). \quad (14)$$

is strictly monotonic increasing, positive, continuous, and

$$\lim_{q \rightarrow 0^+} f_{\alpha, \varepsilon}(q) = 0, \quad \lim_{q \rightarrow (1-\varepsilon)^-} f_{\alpha, \varepsilon}(q) = +\infty. \quad (15)$$

Proof: The proof of the following lemma is tricky, and it is given in Appendix III. ■

Corollary 2: For $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1)$, the equation

$$f_{\alpha, \varepsilon}(q) = \frac{1 - \alpha}{\alpha} \quad (16)$$

has a unique solution $q \in (0, 1 - \varepsilon)$.

Proof: It is a direct consequence of Lemma 3, and the mean value theorem for continuous functions. ■

Remark 3: Since $f_{\alpha, \varepsilon}: (0, 1 - \varepsilon) \rightarrow (0, \infty)$ and this function is strictly monotonic increasing (see Lemma 3), the task of numerically solving equation (16) and finding its unique solution is easy.

A solution of the optimization problem in Proposition 1 is provided in the following for $\alpha \in (0, 1)$.

Proposition 2: Let $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1)$ denote, respectively, the order of the Rényi divergence and the fixed value of the total variation distance. A solution of the minimization problem for g_α in Proposition 1 is obtained by calculating the binary Rényi divergence $d_\alpha(p||q)$ in (9) while taking the unique solution $q \in (0, 1 - \varepsilon)$ of equation (16), and setting $p = q + \varepsilon$.

Proof: The proof of this proposition relies on the Lagrange duality and KKT conditions, while strong duality is first asserted by verifying the satisfiability of Slater's condition. The proof is given in Appendix IV. ■

Remark 4: In light of Remark 3, Proposition 2 enables a quick and high-precision computation of $g_\alpha(\varepsilon)$. The running time of our computer program for a numerical calculation of $g_\alpha(\varepsilon)$ with Proposition 2 has been considerably reduced (by a factor of 100) in comparison to its direct computation with Proposition 1. This significant reduction has been very helpful, especially in the context of the computations that are performed in Section III.

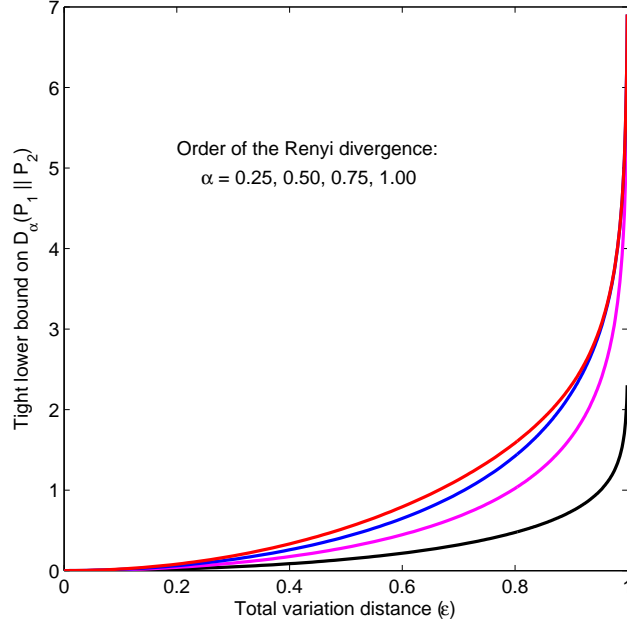


Fig. 1. A plot of the minimum of the Rényi divergence $D_\alpha(P_1||P_2)$ of order $\alpha = 0.25, 0.50, 0.75, 1.00$ (the special case of $\alpha = 1$ gives the Kullback-Leibler divergence) as a function of the total variation distance ε between the PDs P_1 and P_2 .

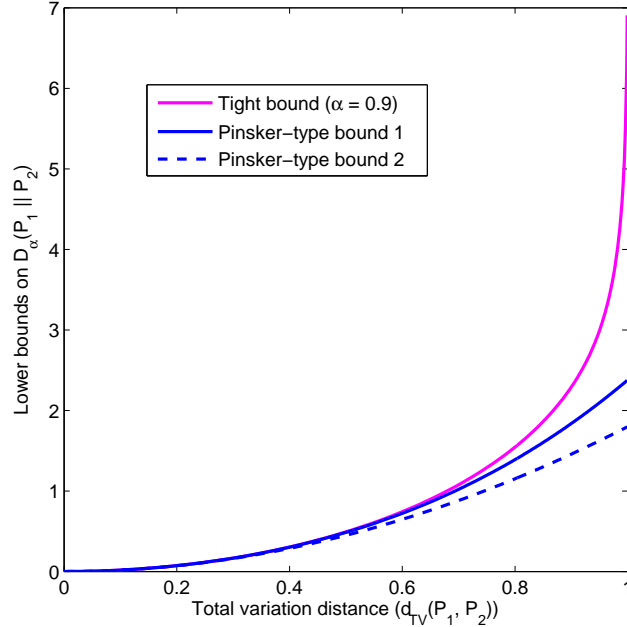


Fig. 2. A plot of the minimum of the Rényi divergence $D_\alpha(P_1||P_2)$ of order $\alpha = 0.90$ subject to a fixed total variation distance between P_1 and P_2 where $d_{TV}(P_1||P_2) = \varepsilon \in [0, 1)$. This tight lower bound is compared with the two Pinsker-type lower bounds in [16, Corollaries 6 and 9] (the improved lower bound from [16, Corollary 9] appears in Eq. (4)).

III. THE ACHIEVABLE REGION OF $(D(Q||P_1), D(Q||P_2))$ FOR ARBITRARY Q, P_1, P_2 SUBJECT TO A MINIMAL TOTAL VARIATION DISTANCE BETWEEN P_1 AND P_2

In this section, we address the following question:

Question 1: What is the achievable region of $(D(Q||P_1), D(Q||P_2))$ when P_1 and P_2 are arbitrary PDs whose total variation distance is at least $\varepsilon \in (0, 1)$, and Q is any PD that is absolutely continuous w.r.t. P_1 and P_2 ? (note that none of these three distributions is fixed).

The present section characterizes this achievable region exactly by relying on the results of Section II, and by using the following lemma which expresses the Rényi divergence as a linear combination of relative entropies.

Lemma 4: Let P_1 and P_2 be mutually absolutely continuous probability measures, and let Q be a third probability measure such that $Q \ll P_1$. Then, for an arbitrary $\alpha > 0$,

$$D_\alpha(P_1||P_2) = D(Q||P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_1) + \frac{1}{\alpha-1} \cdot D(Q||Q_\alpha) \quad (17)$$

where Q_α is given by

$$Q_\alpha(x) \triangleq \frac{P_1^\alpha(x) P_2^{1-\alpha}(x)}{\sum_u P_1^\alpha(u) P_2^{1-\alpha}(u)}, \quad \forall x \in \text{Supp}(P_1). \quad (18)$$

Proof: See Appendix V. ■

As a corollary of Lemma 4, the following tight inequality holds, which is attributed to Shayeitz (see [26, Section IV.B.8]). It will be useful for the continuation of this section, jointly with the results in Section II.

Corollary 3: If $\alpha \in (0, 1)$ then

$$\frac{\alpha}{1-\alpha} \cdot D(Q||P_1) + D(Q||P_2) \geq D_\alpha(P_1||P_2) \quad (19)$$

with equality if and only if $Q = Q_\alpha$ (see (18)). For $\alpha > 1$, inequality (19) is reversed with the same necessary and sufficient condition for an equality.

Remark 5: Corollary 3 with the optimizing PD Q_α in (18) strengthens Eq. (6) in [25] in the sense that it was stated there that, for $\alpha > 1$,

$$D_\alpha(P_1||P_2) = \max_{Q \ll P_1} \left\{ D(Q||P_2) + \frac{\alpha}{\alpha-1} \cdot D(Q||P_1) \right\} \quad (20)$$

where the max is replaced by min for $\alpha \in (0, 1)$. Equality (20) was proved in [25] by the method of types, and the optimizing PD $Q = Q_\alpha$ was stated in [26, Section IV.B.8]. The identity in Lemma 4, which to the best of our knowledge was not explicitly mentioned earlier, leads directly to the maximizing/ minimizing distribution $Q = Q_\alpha$ (due to the non-negativity of the relative entropy). The knowledge of the maximizing distribution in (18) plays an important role in the characterization of the achievable region studied in this section.

The region that includes all the achievable points of $(D(Q||P_1), D(Q||P_2))$ is determined as follows: let $d_{TV}(P_1, P_2) \geq \varepsilon$ for a fixed $\varepsilon \in (0, 1)$, and let $\alpha \in (0, 1)$ be chosen arbitrarily. By the tight lower bound in Section II, we have

$$D_\alpha(P_1||P_2) \geq g_\alpha(\varepsilon) \quad (21)$$

where g_α is expressed in (8) or by the efficient algorithm in Proposition 2. For $\alpha \in (0, 1)$ and for a fixed value of $\varepsilon \in (0, 1)$, let $p = p_\alpha$ and $q = q_\alpha$ in $(0, 1)$ be set to achieve the global minimum in (8) (note that, without loss of generality, one can assume that $p \geq q$ since if (p, q)

achieves the minimum in (8) then also $(1-p, 1-q)$ achieves the same minimum). Consequently, the lower bound in (21) is attained by the pair of 2-element PDs

$$P_1 = (p_\alpha, 1 - p_\alpha), \quad P_2 = (q_\alpha, 1 - q_\alpha). \quad (22)$$

From Corollary 3, and Eqs. (21) and (22), it follows that for every $\alpha \in (0, 1)$

$$g_\alpha(\varepsilon) \leq D(Q||P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_1) \quad (23)$$

where equality in (23) holds if P_1 and P_2 are the 2-element PDs in (22), and Q is the respective PD in (18) for P_1 and P_2 in (22). Hence, there exists a triple of 2-element PDs P_1, P_2, Q that satisfy (23) with equality, and they are easy to calculate for every $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1)$.

Remark 6: Similarly to (23), since $d_{\text{TV}}(P_1, P_2) = d_{\text{TV}}(P_2, P_1)$, it follows from (23) that

$$g_\alpha(\varepsilon) \leq D(Q||P_1) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_2). \quad (24)$$

By multiplying both sides of inequality (24) by $\frac{1-\alpha}{\alpha}$ and relying on the skew-symmetry property in (10), it follows that (24) is equivalent to

$$g_{1-\alpha}(\varepsilon) \leq D(Q||P_2) + \frac{1-\alpha}{\alpha} \cdot D(Q||P_1)$$

which is inequality (23) when $\alpha \in (0, 1)$ is replaced by $1 - \alpha$. Hence, since (23) holds for every $\alpha \in (0, 1)$, there is no additional information in (24).

Proposition 3: The intersection of the half spaces that are given in (23), where the parameter α varies continuously in $(0, 1)$, determines the joint range of $(D(Q||P_1), D(Q||P_2))$ that is addressed in Question 1. Furthermore, all the points in this region are achievable by triples of 2-element PDs P_1, P_2 and Q .

Proof: The boundary of this region is determined by letting α increase continuously in $(0, 1)$, and by drawing the following straight lines in the plane of $(D(Q||P_1), D(Q||P_2))$:

$$D(Q||P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_1) = g_\alpha(\varepsilon), \quad \forall \alpha \in (0, 1). \quad (25)$$

Once the boundary of this region is determined (see Figure 3), every point on the boundary of this region is a tangent point to one of the straight lines in (25). Furthermore, the triple of 2-element PDs P_1, P_2 and Q that achieves an arbitrary point on the boundary of this region is determined as follows:

- Find the slope s of the tangent line ($s < 0$), and let $\alpha = -\frac{s}{1-s}$; this implies that $\alpha \in (0, 1)$ satisfies $-\frac{\alpha}{1-\alpha} = s$ (see (25)).
- Determine the 2-element PDs $P_1 = (p, 1 - p)$, $P_2 = (q, 1 - q)$ such that $d_\alpha(p||q) = g_\alpha(\varepsilon)$. This is done by Proposition 2 where one first obtains the unique solution $q \in (0, 1 - \varepsilon)$ of equation (16) (recall that, from Lemma 3, the function on the right-hand side of (16) is monotonic increasing in q , and it maps the interval $(0, 1 - \varepsilon)$ to $(0, \infty)$), and set $p = q + \varepsilon$.
- Calculate the PD $Q = Q_\alpha$ in (18) for α , P_1 and P_2 .

Every point on the plane $(D(Q||P_1), D(Q||P_2))$, which is to the left of the boundary (i.e., the colored regions in Figures 3 and 4) is not achievable by any triple of PDs P_1, P_2 and Q with $d_{\text{TV}}(P_1, P_2) \geq \varepsilon$. This is because every such a point violates at least one of the inequality constraints in (23). On the other hand, every point which is to the right of this boundary is achievable by a triple of 2-element PDs P_1, P_2, Q . To verify the last claim, first note that it has been demonstrated to hold for all the points on the boundary. Furthermore, based on the set of inequalities in (23) for $\alpha \in (0, 1)$ and $\varepsilon \in [0, 1)$, choose an arbitrary interior point in the convex

region which is to the right of the boundary. Note that $g_\alpha(\cdot)$ is strictly monotonic increasing and continuous in $(0, 1)$; it also tends to infinity as we let ε tend to 1 (see Lemma 1). This implies that the achievable region of $(D(Q||P_1), D(Q||P_2))$, subject to the constraint where $D(P_1||P_2) \geq \varepsilon$, shrinks continuously as the value of $\varepsilon \in (0, 1)$ is increased, and it therefore lies on the boundary of the respective achievable region for some $\varepsilon' > \varepsilon$. One can find, accordingly, the 2-element probability distributions P_1, P_2 and Q in a similar way to the 3-item procedure outlined above (earlier in this proof) where ε is replaced by ε' . This therefore shows that all points on the boundary of this region, as well as all the interior points to the right of this boundary, are all achievable by 2-element PDs; furthermore, none of the points to the left of this boundary is achievable. This concludes the proof of Proposition 3. ■

Note that, from Figure 4, the boundaries of these achievable regions for different values of $\varepsilon \in (0, 1)$ do not form parallel lines; they become less curvy as the value of ε gets closer to 1.

On the Chernoff information and the point on the boundary with equal coordinates

We consider in the following the point in Figure 4 which is specified, in the plane of $(D(Q||P_1), D(Q||P_2))$, by the intersection of the straight line $D(Q||P_1) = D(Q||P_2)$ with the boundary of the achievable region for a fixed value of $\varepsilon \in (0, 1)$. Based on the above explanation (see, e.g., the third item after equation (25)), this intersection point satisfies the equality $D(Q_\alpha||P_1) = D(Q_\alpha||P_2)$ for some $\alpha \in (0, 1)$, 2-element PDs P_1, P_2 with $d_{TV}(P_1, P_2) = \varepsilon$, and Q_α in (18). The two equal coordinates of this intersection point are therefore equal to the Chernoff information $C(P_1, P_2)$ (see [6, Section 11.9]). In this case, due to the symmetry of the achievable region w.r.t. the line $D(Q||P_1) = D(Q||P_2)$ (this symmetry follows from the symmetry of the total variation distance $d_{TV}(P_1, P_2)$), the slope of the tangent line to the boundary at this intersection point is $s = -1$ (see Figure 4). This implies that $\alpha = -\frac{s}{1-s} = \frac{1}{2}$, and from Corollary 1 we have $g_\alpha(\varepsilon) = -\log(1-\varepsilon^2)$ for $\varepsilon \in [0, 1]$. Hence, from (25) with $\alpha = \frac{1}{2}$, the equal coordinates of this intersection point are $D(Q||P_1) = D(Q||P_2) = -\frac{1}{2} \log(1-\varepsilon^2)$. Based on [23, Proposition 2], this value is equal to the minimum of the Chernoff information subject to a fixed total variation distance $\varepsilon \in [0, 1]$. In the following, we also calculate the three PDs P_1, P_2 and Q that achieve this intersection point. Eq. (8) with $\alpha = \frac{1}{2}$ gives that

$$-2 \log(\sqrt{pq} + \sqrt{(1-p)(1-q)}) = -\log(1-\varepsilon^2)$$

subject to the inequality constraints $p, q \in [0, 1]$ and $|p-q| \geq \varepsilon$. A possible solution of this equation is $p = \frac{1+\varepsilon}{2}$ and $q = \frac{1-\varepsilon}{2}$, so the respective 2-element PDs are given by $P_1 = \left(\frac{1+\varepsilon}{2}, \frac{1-\varepsilon}{2}\right)$, $P_2 = \left(\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}\right)$ and, from (18), $Q = \left(\frac{1}{2}, \frac{1}{2}\right)$. As a byproduct of the characterization of this achievable region, we therefore provide a geometric interpretation of the minimal Chernoff information subject to a minimal total variation distance.

The straight line $D(Q||P_1) = D(Q||P_2)$, in the plane of Figure 4, intersects the boundaries of the respective regions at points whose coordinates are equal to the minimum Chernoff information for the fixed total variation distance (ε). The equal coordinates of each of these 4 intersection points in Figure 4, referring to $\varepsilon = 0.50, 0.70, 0.90, 0.99$, are equal to $-\frac{1}{2} \log(1-\varepsilon^2) = 0.144, 0.337, 0.830, 1.959$ nats, respectively.

The reader is also referred to [17] where a geometric interpretation of the Chernoff distribution (achieving the Chernoff information) has been provided. As a concluding remark, recall that the Chernoff information is related to the Rényi divergence by the equality

$$C(P_1, P_2) = \max_{\alpha \in [0, 1]} \left\{ (1-\alpha) D_\alpha(P_1||P_2) \right\}.$$

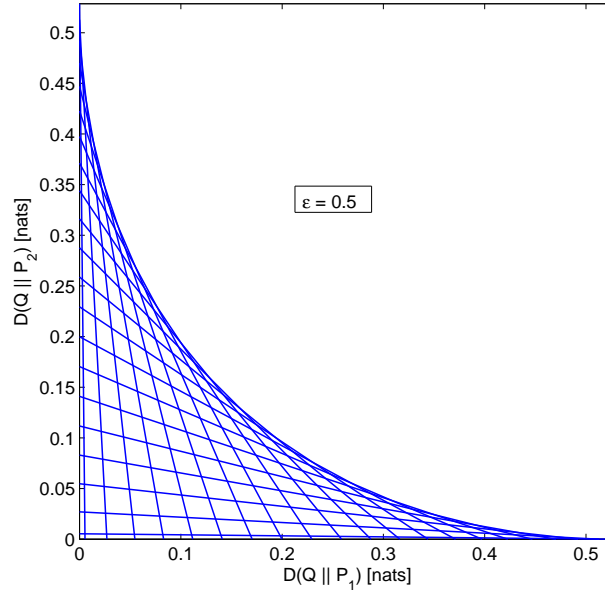


Fig. 3. The boundary of the achievable region of $(D(Q||P_1), D(Q||P_2))$ where Q goes over all possible PDs, and P_1 and P_2 go over all possible pairs of PDs whose total variation distance is at least $\varepsilon = 0.5$. The achievable region is the white region (i.e., it is to the right of its boundary), which is the intersection of all the inequality constraints in (23) where the parameter α varies continuously in $(0,1)$; in this plot, α gets values between 0.05 and 0.95 with increments of 0.05.

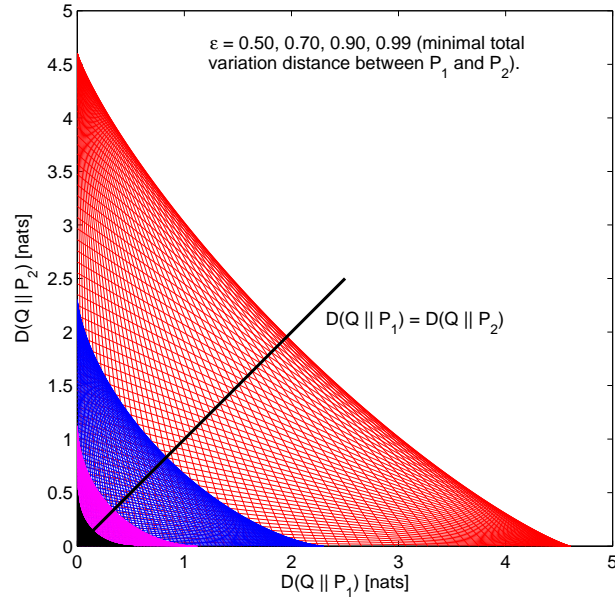


Fig. 4. This plot shows the boundaries of the 4 achievable regions of $(D(Q||P_1), D(Q||P_2))$ where Q goes over all possible PDs, and P_1 and P_2 go over all possible pairs of PDs whose total variation distance is at least $\varepsilon = 0.50, 0.70, 0.90, 0.99$. The respective achievable region for a fixed ε is to the right of its boundary, and it shrinks as the value of ε is increased.

APPENDIX I
PROOFS OF LEMMAS 1 AND 2

A. Proof of Lemma 1

For $\alpha = \frac{1}{2}$, $D_{\frac{1}{2}}(P||Q) = -2 \log Z(P, Q)$ where $Z(P, Q) \triangleq \sum_x \sqrt{P(x)Q(x)}$ denotes the Bhattacharyya coefficient between the two PDs P, Q . From [23, Proposition 1], it follows that $D_{\frac{1}{2}}(P||Q) \geq -\log(1 - \varepsilon^2)$ when $d_{\text{TV}}(P, Q) = \varepsilon$, so (5) holds for $\alpha = \frac{1}{2}$. Since D_α is non-decreasing in α (see [8, Theorem 3]), it follows that (5) holds for $\alpha \geq \frac{1}{2}$. Finally, due to the skew-symmetry property of D_α (see [8, Proposition 2]) where $D_\alpha(P||Q) = \left(\frac{\alpha}{1-\alpha}\right) D_{1-\alpha}(Q||P)$ for $\alpha \in (0, 1)$, and since the total variation distance is a symmetric measure and $\frac{\alpha}{1-\alpha} > 0$ for $\alpha \in (0, 1)$, the satisfiability of (5) for $\alpha \in (\frac{1}{2}, 1)$ yields that it also holds for $\alpha \in (0, \frac{1}{2})$.

B. Proof of Lemma 2

Let P_1 and P_2 be PDs that are defined on an arbitrary set \mathcal{A} of $k \geq 2$ elements. Denote by $\phi: \mathcal{A} \rightarrow \{1, 2\}$ the map given by

$$\phi(x) = \begin{cases} 1, & \text{if } P_1(x) \geq P_2(x), \\ 2, & \text{if } P_1(x) < P_2(x) \end{cases}$$

and define $\phi(P_i) = Q_i$ for $i \in \{1, 2\}$ where

$$Q_i(j) \triangleq \sum_{x \in \mathcal{A}: \phi(x)=j} P_i(x), \quad \forall i, j \in \{1, 2\}. \quad (\text{I.1})$$

We have

$$\begin{aligned} d_{\text{TV}}(P_1, P_2) &= \frac{1}{2} \sum_{x \in \mathcal{A}} |P_1(x) - P_2(x)| \\ &= \frac{1}{2} \sum_{x \in \mathcal{A}: \phi(x)=1} (P_1(x) - P_2(x)) + \frac{1}{2} \sum_{x \in \mathcal{A}: \phi(x)=2} (P_2(x) - P_1(x)) \\ &= \frac{1}{2} (Q_1(1) - Q_2(1)) + \frac{1}{2} (Q_2(2) - Q_1(2)) \\ &= \frac{1}{2} \sum_{j \in \{1, 2\}} |Q_1(j) - Q_2(j)| \\ &= d_{\text{TV}}(Q_1, Q_2). \end{aligned}$$

Furthermore, from the data processing theorem for the Rényi divergence (see [8, Theorem 9]),

$$D_\alpha(P_1||P_2) \geq D_\alpha(Q_1||Q_2) \quad (\text{I.2})$$

where Q_1 and Q_2 are the 2-element PDs defined in (I.1). This completes the proof of this lemma where it has been proved that for every pair of PDs P_1 and P_2 , there exists a pair of 2-element PDs Q_1 and Q_2 whose total variation distance is preserved, and they satisfy inequality (I.2).

APPENDIX II
PROOF OF COROLLARY 1

Eq. (10) in Corollary 1 holds since

$$\begin{aligned} g_\alpha(\varepsilon) &= \min_{p,q \in [0,1]: |p-q| \geq \varepsilon} \frac{\log(p^{1-\alpha}q^\alpha + (1-p)^{1-\alpha}(1-q)^\alpha)}{\alpha - 1} \\ &= \left(\frac{\alpha}{1-\alpha} \right) g_{1-\alpha}(\varepsilon) \end{aligned}$$

where the first equality holds by switching between p and q in (8), and the second equality also follows from (8). Alternatively, (10) follows from (6) and the skew-symmetry property of the Rényi divergence (see [8, Proposition 2]).

The lower bound on g_α in (11) follows from (8), which implies that for $\alpha \in (0, 1)$ and $\varepsilon \in [0, 1)$

$$g_\alpha(\varepsilon) = \frac{\log\left(\max_{p,q \in [0,1]: |p-q| \geq \varepsilon} (p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha})\right)}{\alpha - 1} \quad (\text{II.1})$$

and, we have

$$\begin{aligned} 0 &\leq \max_{p,q \in [0,1]: |p-q| \geq \varepsilon} (p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}) \\ &\leq \max_{p,q \in [0,1]: |p-q| \geq \varepsilon} p^\alpha q^{1-\alpha} + \max_{p,q \in [0,1]: |p-q| \geq \varepsilon} (1-p)^\alpha (1-q)^{1-\alpha} \\ &= 2 \max_{p,q \in [0,1]: |p-q| \geq \varepsilon} p^\alpha q^{1-\alpha} \\ &= 2 \max\{(1-\varepsilon)^\alpha, (1-\varepsilon)^{1-\alpha}\}. \end{aligned} \quad (\text{II.2})$$

The lower bound on g_α in (11) follows from the combination of (II.1) and (II.2).

Eq. (12) follows from the equality $D_{\frac{1}{2}}(P||Q) = -2 \log Z(P, Q)$ where $Z(P, Q)$ is the Bhattacharyya coefficient between P, Q , and since (see [23, Proposition 1])

$$\max_{P, Q: d_{\text{TV}}(P, Q) = \varepsilon} Z(P, Q) = \sqrt{1 - \varepsilon^2}, \quad \forall \varepsilon \in [0, 1).$$

Eq. (13) follows from (8), which gives

$$g_2(\varepsilon) = \min_{p,q \in [0,1]: |p-q| \geq \varepsilon} \log\left(\frac{p^2}{q} + \frac{(1-p)^2}{1-q}\right).$$

The solution of this minimization problem is $q = \frac{1}{2}$ and $p = \frac{1}{2} \pm \varepsilon$ if $\varepsilon \in [0, \frac{1}{2}]$, and its solution is $p = 1$ and $q = 1 - \varepsilon$ if $\varepsilon \in (\frac{1}{2}, 1)$.

Note that $D_2(P_1||P_2) = \log(1 + \chi^2(P_1, P_2))$ where

$$\chi^2(P_1, P_2) \triangleq \sum_x \frac{(P_1(x) - P_2(x))^2}{P_2(x)} = \sum_x \frac{P_1^2(x)}{P_2(x)} - 1$$

is the χ^2 -divergence (a.k.a. the quadratic divergence or Pearson divergence) between the two PDs P_1 and P_2 . An alternative way to derive g_2 in (13) is by relying on the closed-form solution of a minimization of the χ^2 -divergence, subject to a fixed value of the total variation distance $\varepsilon \in [0, 1)$, which is given by (see, e.g., [21, Eq. (58)])

$$\min_{P_1, P_2: d_{\text{TV}}(P_1, P_2) = \varepsilon} \chi^2(P_1, P_2) = \begin{cases} 4\varepsilon^2, & \text{if } \varepsilon \in [0, \frac{1}{2}], \\ \frac{\varepsilon}{1-\varepsilon}, & \text{if } \varepsilon \in (\frac{1}{2}, 1). \end{cases}$$

APPENDIX III
PROOF OF LEMMA 3

For $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1)$, we have

$$\begin{aligned} \lim_{q \rightarrow 0^+} \left(1 + \frac{\varepsilon}{q}\right)^{\alpha-1} &= 0, & \lim_{q \rightarrow 0^+} \left(1 + \frac{\varepsilon}{q}\right)^{\alpha} &= +\infty, \\ \implies \lim_{q \rightarrow 0^+} f_{\alpha, \varepsilon}(q) &= \lim_{q \rightarrow 0^+} \frac{(1 - \varepsilon)^{\alpha-1}}{\left(1 + \frac{\varepsilon}{q}\right)^{\alpha} - (1 - \varepsilon)^{\alpha}} = 0, \end{aligned}$$

and

$$\begin{aligned} \lim_{q \rightarrow (1-\varepsilon)^-} \left(1 - \frac{\varepsilon}{1-q}\right)^{\alpha-1} &= +\infty, & \lim_{q \rightarrow (1-\varepsilon)^-} \left(1 - \frac{\varepsilon}{1-q}\right)^{\alpha} &= 0, \\ \implies \lim_{q \rightarrow (1-\varepsilon)^-} f_{\alpha, \varepsilon}(q) &= \lim_{q \rightarrow (1-\varepsilon)^-} \frac{\left(1 - \frac{\varepsilon}{1-q}\right)^{\alpha-1} - (1 - \varepsilon)^{1-\alpha}}{(1 - \varepsilon)^{-\alpha} - \left(1 - \frac{\varepsilon}{1-q}\right)^{\alpha}} = +\infty. \end{aligned}$$

This proves the two limits in (15).

We prove in the following that $f_{\alpha, \varepsilon}(\cdot)$ is strictly increasing on the interval $[\frac{1-\varepsilon}{2}, 1 - \varepsilon)$, and we also prove later in this appendix that this function is monotonic increasing on the interval $(0, \frac{1-\varepsilon}{2}]$. These two parts of the proof yield that $f_{\alpha, \varepsilon}(\cdot)$ is strictly monotonic increasing on the interval $(0, 1 - \varepsilon)$. The positivity of $f_{\alpha, \varepsilon}$ on $(0, 1 - \varepsilon)$ follows from the first limit in (15), jointly with the monotonicity of this function which is proved in the following.

For a proof that $f_{\alpha, \varepsilon}(\cdot)$ is strictly monotonic increasing on $[\frac{1-\varepsilon}{2}, 1 - \varepsilon)$, this function (see (14)) is expressed as follows:

$$\begin{aligned} f_{\alpha, \varepsilon}(q) &= \frac{1}{1 + \frac{\varepsilon}{q}} \frac{\left(\frac{1 - \frac{\varepsilon}{1-q}}{1 + \frac{\varepsilon}{q}}\right)^{\alpha-1} - 1}{1 - \left(\frac{1 - \frac{\varepsilon}{1-q}}{1 + \frac{\varepsilon}{q}}\right)^{\alpha}} \\ &= \left(1 + \frac{\varepsilon}{q}\right)^{-1} u_{\alpha}(z_{\varepsilon}(q)) \end{aligned} \tag{III.1}$$

where

$$z_{\varepsilon}(q) \triangleq \frac{1 - \frac{\varepsilon}{1-q}}{1 + \frac{\varepsilon}{q}}, \tag{III.2}$$

$$u_{\alpha}(t) \triangleq \begin{cases} \frac{t^{\alpha-1} - 1}{1 - t^{\alpha}}, & \text{if } t \in (0, \infty) \setminus \{1\}, \\ \frac{1-\alpha}{\alpha}, & \text{if } t = 1. \end{cases} \tag{III.3}$$

Note that u_{α} in (III.3) was defined to be continuous at $t = 1$. In order to proceed, we need the following two lemmas:

Lemma III.1: Let $\varepsilon \in (0, 1)$. The function z_{ε} in (III.2) is strictly monotonic increasing on $(0, \frac{1-\varepsilon}{2}]$, and it is strictly monotonic decreasing on $[\frac{1-\varepsilon}{2}, 1 - \varepsilon)$. This function is also positive on $(0, 1 - \varepsilon)$.

Proof: $z_{\varepsilon}(q) > 0$ for $q \in (0, 1 - \varepsilon)$ since $1 - \frac{\varepsilon}{1-q} > 0$, and $1 + \frac{\varepsilon}{q} > 0$. In order to prove the monotonicity properties of z_{ε} , note that its derivative satisfies the equality

$$z'_{\varepsilon}(q) = \varepsilon z_{\varepsilon}(q) \left(\frac{1}{q(\varepsilon + q)} - \frac{1}{(1 - q)(1 - \varepsilon - q)} \right) \tag{III.4}$$

which is derived by taking logarithms on both sides of (III.2), followed by their differentiation. By setting the derivative $z'_\varepsilon(q)$ to zero, we have $q = \frac{1-\varepsilon}{2}$. Since $z_\varepsilon(q) > 0$ for $q \in (0, 1-\varepsilon)$, it follows from (III.4) that $z'_\varepsilon(q) > 0$ for $q \in (0, \frac{1-\varepsilon}{2})$, and $z'_\varepsilon(q) < 0$ for $q \in (\frac{1-\varepsilon}{2}, 1-\varepsilon)$. Hence, z_ε is strictly monotonic increasing on $(0, \frac{1-\varepsilon}{2}]$, and it is strictly monotonic decreasing on $[\frac{1-\varepsilon}{2}, 1-\varepsilon)$. ■

Lemma III.2: Let $\alpha \in (0, 1)$. The function u_α in (III.3) is strictly monotonic decreasing and positive on $(0, \infty)$.

Proof: Differentiation of u_α in (III.3) gives that for $t > 0$

$$u'_\alpha(t) = \frac{t^{\alpha-2}(t^\alpha - \alpha t + \alpha - 1)}{(t^\alpha - 1)^2}. \quad (\text{III.5})$$

Note that $\frac{d}{dt}(t^\alpha - \alpha t + \alpha - 1) = \alpha(t^{\alpha-1} - 1)$, so the derivative is zero at $t = 1$, it is positive if $t \in (0, 1)$, and it is negative if $t \in (1, \infty)$. This implies that $t^\alpha - \alpha t + \alpha - 1 \leq 0$ for every $t \in (0, \infty)$, and it is satisfied with equality if and only if $t = 1$. From (III.5), it follows that u_α is strictly monotonic decreasing on $(0, \infty)$. Since $\lim_{t \rightarrow \infty} u_\alpha(t) = 0$ (see (III.3)) and u_α is strictly monotonic decreasing on $(0, \infty)$ then it is positive on this interval. ■

From Lemmas III.1 and III.2, it follows that z_ε is strictly monotonic decreasing and positive on $[\frac{1-\varepsilon}{2}, 1-\varepsilon)$, and u_α is strictly monotonic decreasing and positive on $(0, \infty)$. This therefore implies that the composition $u_\alpha(z_\varepsilon(\cdot))$ is strictly monotonic increasing and positive on the interval $[\frac{1-\varepsilon}{2}, 1-\varepsilon)$. Hence, from (III.1), since $f_{\alpha,\varepsilon}(\cdot)$ is expressed as a product of two positive and strictly monotonic increasing functions on $[\frac{1-\varepsilon}{2}, 1-\varepsilon)$, also $f_{\alpha,\varepsilon}$ has these properties on this interval. This completes the first part of the proof where we show that $f_{\alpha,\varepsilon}(\cdot)$ is strictly monotonic increasing and positive on $[\frac{1-\varepsilon}{2}, 1-\varepsilon)$.

We prove in the following that $f_{\alpha,\varepsilon}(\cdot)$ is also strictly monotonic increasing and positive on $(0, \frac{1-\varepsilon}{2}]$. For this purpose, the function $f_{\alpha,\varepsilon}$ is expressed in the following alternative way:

$$\begin{aligned} f_{\alpha,\varepsilon}(q) &= \frac{1}{1 - \frac{\varepsilon}{q-1}} \left(\frac{1 - \frac{\varepsilon}{q-1}}{1 + \frac{\varepsilon}{q}} \right)^\alpha \frac{1 - \left(\frac{1 + \frac{\varepsilon}{q}}{1 - \frac{\varepsilon}{q-1}} \right)^{\alpha-1}}{1 - \left(\frac{1 - \frac{\varepsilon}{q}}{1 + \frac{\varepsilon}{q}} \right)^\alpha} \\ &= \left(1 - \frac{\varepsilon}{1-q} \right)^{-1} r_\alpha(z_\varepsilon(q)) \end{aligned} \quad (\text{III.6})$$

where z_ε is defined in (III.2), and

$$r_\alpha(t) \triangleq \begin{cases} \frac{t^\alpha(1-t^{1-\alpha})}{1-t^\alpha}, & \text{if } t \in (0, \infty) \setminus \{1\}, \\ \frac{1-\alpha}{\alpha}, & \text{if } t = 1. \end{cases} \quad (\text{III.7})$$

Note that it follows from Lemma III.1 and (III.2) that

$$z_\varepsilon(q) \leq z_\varepsilon\left(\frac{1-\varepsilon}{2}\right) = \left(\frac{1-\varepsilon}{1+\varepsilon}\right)^2 < 1$$

so the composition $r_\alpha(z_\varepsilon(\cdot))$ in (III.6) is independent of $r_\alpha(1)$; the value of $r_\alpha(1)$ is defined in (III.7) to obtain the continuity of r_α , which leads to the following lemma:

Lemma III.3: For $\alpha \in (0, 1)$, the function r_α in (III.7) is strictly monotonic increasing and positive on $(0, \infty)$.

Proof: A differentiation of r_α in (III.7) gives

$$r'_\alpha(t) = \frac{(1-\alpha)t^\alpha + \alpha t^{\alpha-1} - 1}{(t^\alpha - 1)^2} \quad (\text{III.8})$$

so the sign of r'_α is the same as of $(1 - \alpha)t^\alpha + \alpha t^{\alpha-1} - 1$. Since $\alpha \in (0, 1)$, and

$$\frac{d}{dt}((1 - \alpha)t^\alpha + \alpha t^{\alpha-1} - 1) = \alpha(1 - \alpha)t^{\alpha-2}(t - 1)$$

it follows that the last derivative is negative for $t \in (0, 1)$, zero at $t = 1$, and positive for $t \in (1, \infty)$. This implies that $t = 1$ is a global minimum of the numerator of r'_α (see (III.8)), so

$$(1 - \alpha)t^\alpha + \alpha t^{\alpha-1} - 1 \geq 0, \quad \forall t \in (0, \infty)$$

and equality holds if and only if $t = 1$. It therefore follows from (III.8) that $r'_\alpha(t) > 0$ for $t \in (0, \infty) \setminus \{1\}$, so $r_\alpha(\cdot)$ is strictly monotonic increasing on $(0, \infty)$. Since $\lim_{t \rightarrow 0} r_\alpha(t) = 0$, the monotonicity of $r_\alpha(\cdot)$ on $(0, \infty)$ yields that it is positive on this interval. ■

From Lemmas III.1 and III.3, z_ε is strictly monotonic increasing and positive on $(0, \frac{1-\varepsilon}{2}]$, and r_α is strictly monotonic increasing and positive on $(0, \infty)$. This implies that the composition $r_\alpha(z_\varepsilon(\cdot))$ is strictly monotonic increasing and positive on the interval $(0, \frac{1-\varepsilon}{2}]$. From (III.6), $f_{\alpha,\varepsilon}$ is expressed as a product of two strictly increasing and positive functions on the interval $(0, \frac{1-\varepsilon}{2}]$, which implies that $f_{\alpha,\varepsilon}(\cdot)$ also has these properties on this interval. This completes the second part of the proof where we show that $f_{\alpha,\varepsilon}(\cdot)$ is strictly monotonic increasing and positive on $(0, \frac{1-\varepsilon}{2}]$. The combination of the two parts of this proof completes the proof of Lemma 3.

APPENDIX IV PROOF OF PROPOSITION 2

For $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1)$ are fixed parameters, solving (8) is equivalent to solving the optimization problem

$$\begin{aligned} & \text{maximize} \quad p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha} \\ & \text{subject to} \end{aligned} \tag{IV.1}$$

$$\begin{cases} p, q \in [0, 1], \\ |p - q| \geq \varepsilon \end{cases}$$

where p, q are the optimization variables. The objective function of the optimization problem (IV.1) is concave for $\alpha \in (0, 1)$, so this maximization problem is a convex optimization problem. Since the problem is also strictly feasible at an interior point of the domain in (IV.1), Slater's condition yields that strong duality holds for this optimization problem (see [4, Section 5.2.3]). Note that the replacement of p, q with $1 - p$ and $1 - q$, respectively, does not affect the value of the objective function and the satisfiability of the constraints in (IV.1). Consequently, it can be assumed with loss of generality that $p \geq q$; together with the inequality constraint $|p - q| \geq \varepsilon$, it gives that $p - q \geq \varepsilon$. The Lagrangian of the dual problem is given by

$$L(p, q, \lambda) = p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha} + \lambda(q - p + \varepsilon)$$

and the KKT conditions lead to the following set of equations:

$$\begin{cases} \frac{\partial L}{\partial p} = \alpha[p^{\alpha-1}q^{1-\alpha} - (1-p)^{\alpha-1}(1-q)^{1-\alpha}] - \lambda = 0, \\ \frac{\partial L}{\partial q} = (1-\alpha)[p^\alpha q^{-\alpha} - (1-p)^\alpha(1-q)^{-\alpha}] + \lambda = 0, \\ \frac{\partial L}{\partial \varepsilon} = q - p + \varepsilon = 0. \end{cases} \tag{IV.2}$$

Eliminating λ from the first equation in (IV.2), and substituting it into the second equation gives

$$(1 - \alpha) \left[\left(\frac{p}{q} \right)^\alpha - \left(\frac{1-p}{1-q} \right)^\alpha \right] + \alpha \left[\left(\frac{p}{q} \right)^{\alpha-1} - \left(\frac{1-p}{1-q} \right)^{\alpha-1} \right] = 0. \tag{IV.3}$$

From the third equation of (IV.2), Substituting $p = q + \varepsilon$ into (IV.3), and re-arranging terms gives the equation $f_{\alpha,\varepsilon}(q) = \frac{1-\alpha}{\alpha}$, where $f_{\alpha,\varepsilon}$ is the function in (14).

APPENDIX V
PROOF OF LEMMA 4

For $\alpha \in (0, \infty) \setminus \{1\}$, the following equalities hold:

$$\begin{aligned}
& D(Q||P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_1) + \frac{1}{\alpha-1} D(Q||Q_\alpha) \\
&= \sum_x Q(x) \log \left(\frac{Q(x)}{P_2(x)} \right) + \frac{\alpha}{1-\alpha} \sum_x Q(x) \log \left(\frac{Q(x)}{P_1(x)} \right) + \frac{1}{\alpha-1} \sum_x Q(x) \log \left(\frac{Q(x)}{Q_\alpha(x)} \right) \\
&= - \sum_x Q(x) \log P_2(x) - \frac{\alpha}{1-\alpha} \sum_x Q(x) \log P_1(x) - \frac{1}{\alpha-1} \sum_x Q(x) \log Q_\alpha(x) \\
&= \frac{1}{\alpha-1} \sum_x Q(x) \log \left(\frac{P_1^\alpha(x) P_2^{1-\alpha}(x)}{Q_\alpha(x)} \right) \\
&\stackrel{(a)}{=} \frac{1}{\alpha-1} \sum_x Q(x) \log \left(\sum_u P_1^\alpha(u) P_2^{1-\alpha}(u) \right) \\
&= \frac{1}{\alpha-1} \log \left(\sum_u P_1^\alpha(u) P_2^{1-\alpha}(u) \right) \\
&= D_\alpha(P_1||P_2)
\end{aligned}$$

where equality (a) follows from the expression for Q_α in (18). This proves the identity in (17).

Acknowledgment

Sergio Verdú is acknowledged for a stimulating discussion.

REFERENCES

- [1] F. Alajaji, P. N. Chen and Z. Rached, "Csiszár's cutoff rates for the general hypothesis testing problem," *IEEE Trans. on Information Theory*, vol. 50, no. 4, pp. 663–678, April 2004.
- [2] M. A. Kumar and R. Sundaresan, "Minimization problems based on a parametric family of relative entropies I: forward projection," October 2014. [Online]. Available: <http://arxiv.org/abs/1410.2346>.
- [3] D. Berend, P. Harremoës and A. Kontorovich, "Minimum KL-divergence on complements of L_1 balls," *IEEE Trans. on Information Theory*, vol. 60, no. 6, pp. 3172–3177, June 2014.
- [4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [5] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *IEEE Trans. on Information Theory*, vol. 60, no. 9, pp. 5065–5076, September 2014.
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, second edition, 2006.
- [7] I. Csiszár, "Generalized cutoff rates and Rényi information measures," *IEEE Trans. on Information Theory*, vol. 41, no. 1, pp. 26–34, January 1995.
- [8] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.
- [9] T. van Erven and P. Harremoës, "Rényi divergence and majorization," *Proceedings of the 2010 IEEE International Symposium on Information Theory*, pp. 1335–1339, Austin, Texas, USA, June 2010.
- [10] A. A. Fedotov, P. Harremoës and F. Topsøe, "Refinements of Pinsker's inequality," *IEEE Trans. on Information Theory*, vol. 49, no. 6, pp. 1491–1498, June 2003.
- [11] S. L. Fong and V. Y. F. Tan, "Strong converse theorems for classes of multmessage multicast networks: a Rényi divergence approach," July 2014. [Online]. Available: <http://arxiv.org/abs/1407.2417>.
- [12] P. Harremoës, "Interpretations of Rényi entropies and divergences," *Physica A*, vol. 365, pp. 57–62, 2006.
- [13] P. Harremoës and I. Vajda, "On pairs of f -divergences and their joint range," *IEEE Trans. on Information Theory*, vol. 57, no. 6, pp. 3230–3235, June 2011.
- [14] R. G. Gallager, *Information Theory and Reliable Communications*, John Wiley, 1968.

- [15] G. L. Gilardoni, "On the minimum f -divergence for given total variation," *Comptes Rendus Mathematique*, vol. 343, no. 11–12, pp. 763–766, 2006.
- [16] G. L. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszár's f -divergences," *IEEE Trans. on Information Theory*, vol. 56, no. 11, pp. 5377–5386, November 2010.
- [17] F. Nielsen, "An information-geometric characterization of Chernoff information," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 269–272, March 2013.
- [18] E. Ordentlich and M. J. Weinberger, "A distribution dependent refinement of Pinsker's inequality," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1836–1840, May 2005.
- [19] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," *Proceedings of the Forty-Eighth Annual Allerton Conference*, pp. 1327–1333, Illinois, USA, October 2010.
- [20] M. Raginsky, "Logarithmic Sobolev inequalities and strong data processing theorems for discrete channels," *Proceedings of the 2013 IEEE International Symposium on Information Theory*, pp. 419–423, Istanbul, Turkey, July 2013.
- [21] M. D. Reid and R. C. Williamson, "Information, divergence and risk for binary experiments," *Journal of Machine Learning Research*, vol. 12, no. 3, pp. 731–817, March 2011.
- [22] A. Rényi, "On measures of entropy and information," *Proceedings of the 4th Berkeley Symposium on Probability Theory and Mathematical Statistics*, pp. 547–561, Berkeley, California, USA, 1961.
- [23] I. Sason, "Tight bounds for symmetric divergence measures and a refined bound for lossless source coding," *IEEE Trans. on Information Theory*, vol. 61, no. 2, pp. 701–707, February 2015.
- [24] C. Shannon, R. Gallager and E. Berlekamp, "Lower bounds to error probability for decoding on discrete memoryless channels - part I," *Information and Control*, vol. 10, pp. 65–103, February 1967.
- [25] O. Shayevitz, "On Rényi measures and hypothesis testing," *Proceedings of the 2011 IEEE International Symposium on Information Theory*, pp. 800–804, Saint Petersburg, Russia, August 2011.
- [26] O. Shayevitz, "A note on a characterization of Rényi measures, and its relation to composite hypothesis testing," December 2010. [Online]. Available at <http://arxiv.org/abs/1012.4401>.
- [27] R. Sundaresan, "A measure of discrimination and its geometric properties," *Proceedings of the IEEE 2002 International Symposium on Information Theory*, p. 264, Laussane, Switzerland, June 2002.
- [28] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. on Information Theory*, vol. 53, no. 1, pp. 269–287, January 2007.
- [29] N. A. Warsi, "One-shot bounds for various information-theoretic problems using smooth min and max Rényi divergences," *Proceedings of the 2013 IEEE Information Theory Workshop*, pp. 434–438, Seville, Spain, September 2013.