2016 ICSEE International Conference on the Science of Electrical Engineering

Two Measures of Dependence

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

***Abstract*—Motivated by a distributed task-encoding problem, two closely related families of dependence measures are introduced. They are based on the Rényi divergence of order α and the relative α-entropy, respectively, and both reduce to the mutual information when the parameter α is one. Their properties are studied and it is shown that the first measure shares many properties with mutual information, including the data-processing inequality. The second measure does not satisfy the data-processing inequality, but it appears naturally in the context of distributed task encoding.**

I. Introduction

At the heart of information theory lies the Shannon entropy

$$H(X) = \sum_{x\in\mathcal{X}} P(x) \log \frac{1}{P(x)}, \tag{1}$$

which, together with relative entropy and mutual information, appears in numerous contexts. One of the more successful attempts to generalize Shannon entropy was performed by Rényi [1], who introduced the Rényi entropy of order α,

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x\in\mathcal{X}} P(x)^\alpha, \tag{2}$$

which is defined for $\alpha > 0$ and $\alpha \neq 1$ and has the desirable property that $\lim_{\alpha\to 1} H_\alpha(X) = H(X)$. But there does not seem to be a unique way to generalize relative entropy and mutual information to the Rényi setting.

The two classical generalizations of relative entropy are reviewed in Section II. In Section III, our proposed generalizations of mutual information, $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$, are introduced. Their properties are analyzed in Sections IV and V. Section VI provides an operational meaning to $K_\alpha(X;Y)$. Additional proofs can be found in [2].

The measure $J_\alpha(X;Y)$ was discovered independently by Tomamichel and Hayashi, who show its operational meaning in composite hypothesis testing [3].

Other generalizations of mutual information appeared in the past. Notable are those by Sibson [4], Arimoto [5], and Csiszár [6]. An overview and some properties of these proposals are provided by Verdú [7].

II. Generalizations of Relative Entropy

Throughout this section, P and Q are probability mass functions on a finite set $\mathcal{X}$. The relative entropy (or Kullback-Leibler divergence) of P with respect to Q is defined as

$$D(P||Q) = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} \tag{3}$$

with the convention $0 \log \frac{0}{q} = 0$ and $p \log \frac{p}{0} = \infty$ for $p > 0$.

The Rényi divergence of order α of Q from P, which was introduced by Rényi [1], is defined for $\alpha > 0$ and $\alpha \neq 1$ as

$$D_\alpha(P||Q) = \frac{1}{\alpha-1} \log \sum_{x\in\mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha} \tag{4}$$

with the convention that for $\alpha > 1$, we read $P(x)^\alpha Q(x)^{1-\alpha}$ as $\frac{P(x)^\alpha}{Q(x)^{\alpha-1}}$ and say that $\frac{0}{0} = 0$ and $\frac{p}{0} = \infty$ for $p > 0$. Its properties are studied in detail by van Erven and Harremoës [8]. By a continuity argument [8, Theorem 5], $D_1(P||Q)$ is defined as $D(P||Q)$.

The relative α-entropy of P with respect to Q is defined for $\alpha > 0$ and $\alpha \neq 1$ as

$$\Delta_\alpha(P||Q) = \frac{\alpha}{1-\alpha} \log \sum_{x\in\mathcal{X}} P(x) Q(x)^{\alpha-1} + \log \sum_{x\in\mathcal{X}} Q(x)^\alpha - \frac{1}{1-\alpha} \log \sum_{x\in\mathcal{X}} P(x)^\alpha \tag{5}$$

with the convention that for $\alpha < 1$, we read $P(x)Q(x)^{\alpha-1}$ as $\frac{P(x)}{Q(x)^{1-\alpha}}$ and say that $\frac{0}{0} = 0$ and $\frac{p}{0} = \infty$ for $p > 0$. It was first identified by Sundaresan [9] in the context of the Massey-Arikan guessing problem [10], [11] and it also plays a role in the context of mismatched task encoding as shown by Bunte and Lapidoth [12]. Further properties of relative α-entropy are studied by Kumar and Sundaresan [13], [14]. By a continuity argument [13, Lemma 2], $\Delta_1(P||Q)$ is defined as $D(P||Q)$.

The following lemma shows that $\Delta_\alpha(P||Q)$ and $D_\alpha(P||Q)$ are in fact closely related. (This relationship was first described in [9, Section V, Property 4].)

Lemma 1. *Let P and Q be PMFs over a finite set $\mathcal{X}$ and let $\alpha > 0$ be a constant. Define the PMFs*

$$\widetilde{P}(x) = \frac{P(x)^\alpha}{\sum_{x'\in\mathcal{X}} P(x')^\alpha}, \tag{6}$$

$$\widetilde{Q}(x) = \frac{Q(x)^\alpha}{\sum_{x'\in\mathcal{X}} Q(x')^\alpha}. \tag{7}$$

Then,

$$\Delta_\alpha(P||Q) = D_{\frac{1}{\alpha}}(\widetilde{P}||\widetilde{Q}), \tag{8}$$

where the LHS is ∞ if and only if the RHS is ∞.

Proof. Note that (6) and (7) are well-defined for every $\alpha > 0$. For $\alpha \in (0,1)$ and for $\alpha > 1$, (8) follows from the definitions (4) and (5) and from the transformations (6) and (7). Checking

978-1-5090-2152-9/16/$31.00

the conditions under which either side of (8) is ∞ establishes that the LHS is ∞ if and only if the RHS is ∞ because $\widetilde{P}(x)$ and $\widetilde{Q}(x)$ are zero if and only if $P(x)$ and $Q(x)$ are zero, respectively. For $\alpha = 1$, (8) is valid because we have $\widetilde{P} = P$, $\widetilde{Q} = Q$, and $\Delta_1(P||Q) = D_1(P||Q) = D(P||Q)$ by definition. ∎

III. Two Measures of Dependence

Throughout this section, X and Y are random variables taking values in finite sets according to the joint PMF P_{XY}. Based on the observation that mutual information can be characterized as

$$I(X;Y) = D(P_{XY}||P_X P_Y) \tag{9}$$

$$= \min_{Q_X, Q_Y} D(P_{XY}||Q_X Q_Y), \tag{10}$$

where the minimization is over all PMFs Q_X and Q_Y, two generalizations are proposed:

$$J_\alpha(X;Y) \triangleq \min_{Q_X, Q_Y} D_\alpha(P_{XY}||Q_X Q_Y), \tag{11}$$

$$K_\alpha(X;Y) \triangleq \min_{Q_X, Q_Y} \Delta_\alpha(P_{XY}||Q_X Q_Y). \tag{12}$$

Because $D_1(P||Q) = \Delta_1(P||Q) = D(P||Q)$ and because of (10), $J_1(X;Y)$ and $K_1(X;Y)$ are equal to $I(X;Y)$.

The measures $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ are well-defined for all $\alpha > 0$: Because $D_\alpha(P||Q)$ and $\Delta_\alpha(P||Q)$ are nonnegative and continuous in Q and because $D_\alpha(P_{XY}||Q_X Q_Y)$ and $\Delta_\alpha(P_{XY}||Q_X Q_Y)$ are finite for $Q_X = P_X$ and $Q_Y = P_Y$, the minima in the RHS of (11) and (12) exist. Note that, (10) notwithstanding, this choice of Q_X and Q_Y need not be optimal if $\alpha \neq 1$. For all $\alpha \geq \frac{1}{2}$, the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY}||Q_X Q_Y)$ is convex in the pair (Q_X, Q_Y), so (11) can be formulated as a convex optimization problem.[1]

In light of Lemma 1, $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ are related as follows:

Lemma 2. *Let P_{XY} be a joint PMF over the finite sets $\mathcal{X}$ and $\mathcal{Y}$ and let $\alpha > 0$ be a constant. Define the PMF*

$$\widetilde{P}_{XY}(x,y) = \frac{P_{XY}(x,y)^\alpha}{\sum_{x' \in \mathcal{X}} \sum_{y' \in \mathcal{Y}} P_{XY}(x',y')^\alpha}. \tag{13}$$

Then,

$$K_\alpha(X;Y) = J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}). \tag{14}$$

Proof. For every $\alpha > 0$,

$$K_\alpha(X;Y) = \min_{Q_X, Q_Y} \Delta_\alpha(P_{XY}||Q_X Q_Y) \tag{15}$$

$$= \min_{Q_X, Q_Y} D_{\frac{1}{\alpha}}(\widetilde{P}_{XY}||\widetilde{Q_X Q_Y}) \tag{16}$$

$$= \min_{Q_X, Q_Y} D_{\frac{1}{\alpha}}(\widetilde{P}_{XY}||\widetilde{Q}_X \widetilde{Q}_Y) \tag{17}$$

$$= \min_{Q_X, Q_Y} D_{\frac{1}{\alpha}}(\widetilde{P}_{XY}||Q_X Q_Y) \tag{18}$$

$$= J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}), \tag{19}$$

where (15) follows from the definition (12); (16) follows from Lemma 1; (17) follows because the transformation (7) of a product is the product of the transformations; (18) follows because the transformation (7) is bijective on the set of PMFs; and (19) follows from the definition (11). ∎

[1] The proof is omitted; for $\alpha \in (0, \frac{1}{2})$, convexity does not hold in general.

IV. Properties of $J_\alpha(X;Y)$

Theorem 1. *Let X, X_1, X_2, Y, Y_1, Y_2, and Z be random variables on finite sets. The following properties of the mutual information $I(X;Y)$ are also satisfied by $J_\alpha(X;Y)$ for all $\alpha > 0$:*

1) *$J_\alpha(X;Y) \geq 0$ with equality if and only if X and Y are independent (nonnegativity).*
2) *$J_\alpha(X;Y) = J_\alpha(Y;X)$ (symmetry).*
3) *$J_\alpha(X;Z) \leq J_\alpha(X;Y)$ if $X \multimap\!\!\!- Y \multimap\!\!\!- Z$, i.e., if X, Y, and Z form a Markov chain (data-processing inequality).*
4) *$J_\alpha(X_1, X_2; Y_1, Y_2) = J_\alpha(X_1;Y_1) + J_\alpha(X_2;Y_2)$ if the pairs (X_1, Y_1) and (X_2, Y_2) are independent (additivity).*
5) *$J_\alpha(X;Y) \leq \log|\mathcal{X}|$ and $J_\alpha(X;Y) \leq \log|\mathcal{Y}|$.*

In addition,

6) *$J_1(X;Y) = I(X;Y)$.*
7) *$J_\alpha(X;Y)$ is continuous and nondecreasing in α for all $\alpha > 0$.*
8) *For all $\alpha > 0$ and $\alpha \neq 1$, $J_\alpha(X;Y)$ is equal to*

$$\min_{Q_X} \frac{\alpha}{\alpha - 1} \log \sum_{y \in \mathcal{Y}} \left[\sum_{x \in \mathcal{X}} P_{XY}(x,y)^\alpha Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}}, \tag{20}$$

where the minimization is over all PMFs Q_X. This is a convex optimization problem if $\alpha \geq \frac{1}{2}$.

For all $\alpha \in (0,1)$:

9) *$J_\alpha(X;Y) = \min_{R_{XY}} \Big[I_{R_{XY}}(X;Y) + \frac{\alpha}{1-\alpha} D(R_{XY}||P_{XY}) \Big]$, where the minimization is over all joint PMFs R_{XY} and $I_{R_{XY}}(X;Y)$ denotes $D(R_{XY}||R_X R_Y)$.*

For all $\alpha > 1$:

10) *$J_\alpha(X;Y) = \max_{R_{XY}} \Big[I_{R_{XY}}(X;Y) + \frac{\alpha}{1-\alpha} D(R_{XY}||P_{XY}) \Big]$, where the maximization is over all joint PMFs R_{XY} and $I_{R_{XY}}(X;Y)$ denotes $D(R_{XY}||R_X R_Y)$. The expression in brackets is strictly concave in R_{XY}. It is maximized by R_{XY} if and only if it is equal to $D_\alpha(P_{XY}||R_X R_Y)$.*

Furthermore,

11) *$J_\alpha(X;Y)$ is concave in P_X for fixed $P_{Y|X}$ and $\alpha \geq 1$.*
12) *$J_\alpha(X;X) = \begin{cases} H_{\frac{\alpha}{2\alpha-1}}(X) & \text{if } \alpha > \frac{1}{2}, \\ \frac{\alpha}{1-\alpha} H_\infty(X) & \text{if } \alpha \in (0, \frac{1}{2}]. \end{cases}$*

Proof. It is well-known that Properties 1–5 are satisfied by the mutual information [15, Chapter 2]. We are left to show that $J_\alpha(X;Y)$ satisfies Properties 1–12:

1) We use the fact that for all $\alpha > 0$, $D_\alpha(P||Q) \geq 0$ with equality if and only if $P = Q$ [8, Theorem 8]. Then, the nonnegativity of $J_\alpha(X;Y)$ follows from (11) and from $D_\alpha(P||Q) \geq 0$. If X and Y are independent, i.e., if $P_{XY} = P_X P_Y$, the choice $Q_X = P_X$ and $Q_Y = P_Y$ in the RHS of (11) achieves $J_\alpha(X;Y) = 0$. Conversely,

$J_\alpha(X;Y) = 0$ implies that $P_{XY} = Q_X Q_Y$ for some PMFs Q_X and Q_Y, which in turn implies that X and Y are independent.

2) The symmetry of $J_\alpha(X;Y)$ in X and Y follows because (11) is symmetric in X and Y.
3) Assume that $X \multimap Y \multimap Z$, which is equivalent to

$$P_{Z|XY}(z|x,y) = P_{Z|Y}(z|y) \tag{21}$$

for all x, y, and z. Let Q_X and Q_Y be PMFs that achieve the minimum in the RHS of (11), so

$$J_\alpha(X;Y) = D_\alpha(P_{XY}||Q_X Q_Y). \tag{22}$$

Define the PMF Q_Z as follows:

$$Q_Z(z) = \sum_{y\in\mathcal{Y}} P_{Z|Y}(z|y) Q_Y(y). \tag{23}$$

We will show that for all $\alpha > 0$,

$$D_\alpha(P_{XZ}||Q_X Q_Z) \le D_\alpha(P_{XY}||Q_X Q_Y), \tag{24}$$

which implies the data-processing inequality because

$$\begin{aligned} J_\alpha(X;Z) &\le D_\alpha(P_{XZ}||Q_X Q_Z) && (25)\\ &\le D_\alpha(P_{XY}||Q_X Q_Y) && (26)\\ &= J_\alpha(X;Y), && (27)\end{aligned}$$

where (25) follows from (11); (26) follows from (24); and (27) follows from (22). In order to prove (24), we use the fact that $D_\alpha(P||Q)$ satisfies a data-processing inequality, namely, that for any conditional PMF $A(x|x')$,

$$D_\alpha((PA)||(QA)) \le D_\alpha(P||Q), \tag{28}$$

where $(PA)(x) = \sum_{x'} A(x|x')P(x')$ and (QA) is defined in the same way [8, Theorem 9]. We choose

$$A(x,z|x',y') = I\{x = x'\} P_{Z|XY}(z|x',y'), \tag{29}$$

where $I\{x = x'\}$ is the indicator function that is one if $x = x'$ and zero otherwise. Processing P_{XY} leads to

$$\begin{aligned} (PA)(x,z) &= \sum_{x',y'} A(x,z|x',y') P_{XY}(x',y') && (30)\\ &= \sum_{y} P_{Z|XY}(z|x,y) P_{XY}(x,y) && (31)\\ &= P_{XZ}(x,z), && (32)\end{aligned}$$

where (31) follows from (29). Processing $Q_X Q_Y$ leads to

$$\begin{aligned} (QA)(x,z) &= \sum_{x',y'} A(x,z|x',y') Q_X(x') Q_Y(y') && (33)\\ &= \sum_{y} P_{Z|XY}(z|x,y) Q_X(x) Q_Y(y) && (34)\\ &= \sum_{y} P_{Z|Y}(z|y) Q_X(x) Q_Y(y) && (35)\\ &= Q_X(x) Q_Z(z), && (36)\end{aligned}$$

where (34) follows from (29); (35) follows from (21); and (36) follows from (23). Combining (28), (32), and (36) now leads to (24).

4) The proof of this property is omitted.
5) For $\alpha > 1$,

$$\begin{aligned} J_\alpha(X;Y) &\le \max_{R_{XY}} I_{R_{XY}}(X;Y) && (37)\\ &\le \log|\mathcal{X}|, && (38)\end{aligned}$$

where (37) follows from Property 10 and (38) follows because $I_{R_{XY}}(X;Y) \le \log|\mathcal{X}|$ for all R_{XY}. The bound extends to all $\alpha > 0$ because $J_\alpha(X;Y)$ is nondecreasing in α. Because $J_\alpha(X;Y)$ is symmetric in X and Y, $J_\alpha(X;Y) \le \log|\mathcal{Y}|$ follows.

6) Because $D_1(P||Q) = D(P||Q)$ and because of (10), $J_1(X;Y) = I(X;Y)$.
7) Let $\alpha > 0$ and let Q_X^* and Q_Y^* be PMFs that achieve the minimum in the RHS of (11), so

$$J_\alpha(X;Y) = D_\alpha(P_{XY}||Q_X^* Q_Y^*). \tag{39}$$

The monotonicity of $J_\alpha(X;Y)$ in α follows because for every $0 < \alpha' \le \alpha$,

$$\begin{aligned} J_{\alpha'}(X;Y) &\le D_{\alpha'}(P_{XY}||Q_X^* Q_Y^*) && (40)\\ &\le D_\alpha(P_{XY}||Q_X^* Q_Y^*) && (41)\\ &= J_\alpha(X;Y), && (42)\end{aligned}$$

where (40) follows from (11); (41) follows because $D_\alpha(P||Q)$ is nondecreasing in α [8, Theorem 3]; and (42) follows from (39).

The continuity of $J_\alpha(X;Y)$ in α for $\alpha > 0$ and $\alpha \neq 1$ follows because the set of all PMFs is compact and because $D_\alpha(P_{XY}||Q_X Q_Y)$ is jointly continuous in α, Q_X, and Q_Y.[2] To establish the continuity of $J_\alpha(X;Y)$ at $\alpha = 1$, we first show $\limsup_{\alpha\to 1} J_\alpha(X;Y) \le I(X;Y)$. This follows because $J_\alpha(X;Y) \le D_\alpha(P_{XY}||P_X P_Y)$; because $D_\alpha(P_{XY}||P_X P_Y)$ is continuous in α [8, Theorem 7]; and because $D_1(P_{XY}||P_X P_Y) = I(X;Y)$. Next, we have $J_\alpha(X;Y) \ge I(X;Y)$ for $\alpha \ge 1$ because $J_\alpha(X;Y)$ is nondecreasing in α. To finish the proof, it remains to show $\liminf_{\alpha\uparrow 1} J_\alpha(X;Y) \ge I(X;Y)$. For convenience, set $\alpha = 1-\delta$ for $\delta \in (0,1)$, and observe that

$$\begin{aligned} &2^{-\delta D_{1-\delta}(P_{XY}||Q_X Q_Y)}\\ &= \sum_{x,y} P(x,y)\left[\frac{Q_X(x)Q_Y(y)}{P(x,y)}\right]^\delta && (43)\\ &= \sum_{x,y} P(x,y)\left[\frac{P_X(x)P_Y(y)}{P(x,y)}\right]^\delta \left[\frac{Q_X(x)Q_Y(y)}{P_X(x)P_Y(y)}\right]^\delta && (44)\\ &\le \left\{\sum_{x,y} P(x,y)\left[\frac{P_X(x)P_Y(y)}{P(x,y)}\right]^{2\delta}\right\}^{\frac{1}{2}}\\ &\quad\cdot \left\{\sum_{x,y} P(x,y)\left[\frac{Q_X(x)Q_Y(y)}{P_X(x)P_Y(y)}\right]^{2\delta}\right\}^{\frac{1}{2}}, && (45)\end{aligned}$$

[2]This requires a topological argument, which is omitted here.

where (43) follows from (4) and (45) follows from the Cauchy-Schwarz inequality. For $\delta \in (0, \frac{1}{4})$, the second factor in the RHS of (45) can be bounded as

$$\left\{ \sum_{x,y} P(x,y) \left[\frac{Q_X(x) Q_Y(y)}{P_X(x) P_Y(y)} \right]^{2\delta} \right\}^{\frac{1}{2}}$$

$$\leq \left\{ \sum_x P_X(x) \left[\frac{Q_X(x)}{P_X(x)} \right]^{4\delta} \right\}^{\frac{1}{4}} \cdot \left\{ \sum_y P_Y(y) \left[\frac{Q_Y(y)}{P_Y(y)} \right]^{4\delta} \right\}^{\frac{1}{4}} \quad (46)$$

$$= 2^{-\delta D_{1-4\delta}(P_X || Q_X)} \cdot 2^{-\delta D_{1-4\delta}(P_Y || Q_Y)} \quad (47)$$

$$\leq 1, \quad (48)$$

where (46) follows from the Cauchy-Schwarz inequality and from marginalization; (47) follows from (4); and (48) follows because the Rényi divergence is nonnegative. Combining (11), (45), and (48), we obtain

$$J_{1-\delta}(X;Y) \geq \frac{-1}{2\delta} \log \sum_{x,y} P(x,y) \left[\frac{P_X(x) P_Y(y)}{P(x,y)} \right]^{2\delta} \quad (49)$$

for $\delta \in (0, \frac{1}{4})$. In the limit $\delta \downarrow 0$, the RHS of (49) tends to $I(X;Y)$, so $\liminf_{\alpha \uparrow 1} J_\alpha(X;Y) \geq I(X;Y)$.

8) Observe that for all $\alpha > 0$ and $\alpha \neq 1$,

$$D_\alpha(P_{XY} || Q_X Q_Y)$$

$$= \frac{1}{\alpha - 1} \log \sum_{x,y} P(x,y)^\alpha [Q_X(x) Q_Y(y)]^{1-\alpha} \quad (50)$$

$$= \frac{1}{\alpha - 1} \log \sum_y \gamma^\alpha \left[\gamma^{-1} R(y) \right]^\alpha Q_Y(y)^{1-\alpha} \quad (51)$$

$$= \frac{\alpha}{\alpha - 1} \log \sum_y R(y) + D_\alpha(\gamma^{-1} R || Q_Y), \quad (52)$$

where (50) follows from the definition (4); (51) follows for any positive γ by identifying $R \colon \mathcal{Y} \to \mathbb{R}_{\geq 0}$ as

$$R(y) = \left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}}; \quad (53)$$

and (52) follows by choosing the normalization constant $\gamma = \sum_y R(y)$ so that $\gamma^{-1} R$ is a PMF. The claim now follows because

$$J_\alpha(X;Y)$$

$$= \min_{Q_X} \min_{Q_Y} D_\alpha(P_{XY} || Q_X Q_Y) \quad (54)$$

$$= \min_{Q_X} \frac{\alpha}{\alpha - 1} \log \sum_y R(y) \quad (55)$$

$$= \min_{Q_X} \frac{\alpha}{\alpha - 1} \log \sum_y \left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}}, \quad (56)$$

where (54) follows from the definition (11); (55) follows from (52) and from the nonnegativity of $D_\alpha(P||Q)$; and (56) follows from (53). We omit the proof that the RHS of (56) is a convex optimization problem if $\alpha \geq \frac{1}{2}$.

9) For $\alpha \in (0,1)$, we have [8, Theorem 30]

$$D_\alpha(P||Q) = \inf_R \left[D(R||Q) + \frac{\alpha}{1-\alpha} D(R||P) \right], \quad (57)$$

where the infimum is over all PMFs R. The claim follows by observing that[3]

$$J_\alpha(X;Y)$$

$$= \min_{Q_X, Q_Y} \inf_R \left[D(R||Q_X Q_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY}) \right] \quad (58)$$

$$= \inf_R \inf_{Q_X, Q_Y} \left[D(R||Q_X Q_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY}) \right] \quad (59)$$

$$= \inf_{R_{XY}} \left[I_{R_{XY}}(X;Y) + \frac{\alpha}{1-\alpha} D(R_{XY}||P_{XY}) \right] \quad (60)$$

$$= \min_{R_{XY}} \left[I_{R_{XY}}(X;Y) + \frac{\alpha}{1-\alpha} D(R_{XY}||P_{XY}) \right], \quad (61)$$

where (58) follows from (11) and (57); (59) follows by interchanging the order of the infima; (60) follows from (10); and (61) follows from a continuity argument.

10) For $\alpha > 1$, we have [8, Theorem 30]

$$D_\alpha(P||Q) = \sup_R \left[D(R||Q) + \frac{\alpha}{1-\alpha} D(R||P) \right], \quad (62)$$

where the supremum is over all PMFs R. A simple computation reveals that[3]

$$D(R||Q_X Q_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY})$$

$$= \frac{1}{\alpha - 1} H(R) + \sum_{x,y} R(x,y) \log \frac{P_{XY}(x,y)^{\frac{\alpha}{\alpha-1}}}{Q_X(x) Q_Y(y)} \quad (63)$$

is concave in R because $H(R)$ and linear functionals of R are concave in R; in addition, the LHS of (63) is convex in Q_Y and continuous in R and Q_Y.[4] Then,

$$\inf_{Q_Y} \sup_R \left[D(R||Q_X Q_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY}) \right]$$

$$= \sup_R \inf_{Q_Y} \left[D(R||Q_X Q_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY}) \right] \quad (64)$$

$$= \sup_R \left[D(R||Q_X R_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY}) \right], \quad (65)$$

where (64) can be justified by [16, Corollary 37.3.2] because the set of all PMFs is compact, convex, and nonempty and because the expression in brackets is

[3] For brevity, R is used to denote R_{XY}.

[4] Here, we ignore the issue that the Rényi divergence can be ∞. It is possible, but more involved, to justify the statements without this assumption.

continuous in R and Q_Y, convex in Q_Y, and concave in R; and (65) follows from a simple computation. Finally,

$$\begin{aligned} & J_\alpha(X;Y) \\ &= \min_{Q_X, Q_Y} \sup_R \left[D(R||Q_X Q_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY}) \right] && (66) \\ &= \inf_{Q_X} \sup_R \left[D(R||Q_X R_Y) + \frac{\alpha}{1-\alpha} D(R||P_{XY}) \right] && (67) \\ &= \sup_{R_{XY}} \left[I_{R_{XY}}(X;Y) + \frac{\alpha}{1-\alpha} D(R_{XY}||P_{XY}) \right] && (68) \\ &= \max_{R_{XY}} \left[I_{R_{XY}}(X;Y) + \frac{\alpha}{1-\alpha} D(R_{XY}||P_{XY}) \right], && (69) \end{aligned}$$

where (66) follows from (11) and (62); (67) follows from (65); (68) follows from similar steps as (63)–(65); and (69) follows from a continuity argument. The proofs of the other two claims are omitted.

11) The proofs of this and the next property are omitted. ■

V. Properties of $K_\alpha(X;Y)$

The relationship $K_\alpha(X;Y) = J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y})$ from Lemma 2 allows us to derive some properties of $K_\alpha(X;Y)$ from the properties of $J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y})$. But, unlike $J_\alpha(X;Y)$, $K_\alpha(X;Y)$ does not satisfy the data-processing inequality and is not monotonic in α.[5]

Theorem 2. *Let X, X_1, X_2, Y, Y_1, and Y_2 be random variables on finite sets. Then, $K_\alpha(X;Y)$ satisfies the following properties for all $\alpha > 0$:*

1) *$K_\alpha(X;Y) \geq 0$ with equality if and only if X and Y are independent (nonnegativity).*
2) *$K_\alpha(X;Y) = K_\alpha(Y;X)$ (symmetry).*
3) *$K_\alpha(X_1, X_2; Y_1, Y_2) = K_\alpha(X_1;Y_1) + K_\alpha(X_2;Y_2)$ if the pairs (X_1, Y_1) and (X_2, Y_2) are independent (additivity).*
4) *$K_\alpha(X;Y) \leq \log|\mathcal{X}|$ and $K_\alpha(X;Y) \leq \log|\mathcal{Y}|$.*

In addition,

5) *$K_1(X;Y) = I(X;Y)$.*
6) *$K_\alpha(X;Y)$ is continuous in α for all $\alpha > 0$.*
7) *$K_\alpha(X;X) = \begin{cases} 2H_{\frac{\alpha}{2-\alpha}}(X) - H_\alpha(X) & \text{if } \alpha \in (0,2), \\ \frac{\alpha}{\alpha-1} H_\infty(X) - H_\alpha(X) & \text{if } \alpha \geq 2. \end{cases}$*

VI. Operational Meaning of $K_\alpha(X;Y)$

The motivation to study $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ stems from [17], which extends the task-encoding problem studied in [12] to a distributed setting. It considers a discrete source $\{(X_i, Y_i)\}_{i=1}^\infty$ over a finite alphabet that emits pairs of random variables (X_i, Y_i). For any positive integer n, the sequences $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are encoded separately, and the decoder outputs the list of all pairs (x^n, y^n) that share the given description.[6] The goal is to minimize the ρ-th moment of the list size for some $\rho > 0$ as n goes to infinity. In the following theorem, necessary and sufficient conditions on the coding rates are given to drive the ρ-th moment of the list size asymptotically to one. (For the proof, see [17].)

Theorem 3. *Let $\{(X_i, Y_i)\}_{i=1}^\infty$ be a discrete source over a finite alphabet $\mathcal{X} \times \mathcal{Y}$. For a fixed $\rho > 0$, a rate pair (R_X, R_Y) is called achievable if there exists a sequence of encoders $\{(f_n, g_n)\}_{n=1}^\infty$,*

$$f_n \colon \mathcal{X}^n \to \{1, \ldots, \lfloor 2^{nR_X} \rfloor\}, \quad (70)$$

$$g_n \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor 2^{nR_Y} \rfloor\}, \quad (71)$$

such that

$$\lim_{n\to\infty} \mathbf{E}\Big[\big| \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : x^n = f_n(X^n) \wedge y^n = g_n(Y^n)\} \big|^\rho \Big] = 1. \quad (72)$$

For an i.i.d. source, the rate region is the set of pairs (R_X, R_Y) satisfying the following three conditions:

$$R_X \geq H_{\frac{1}{1+\rho}}(X), \quad (73)$$

$$R_Y \geq H_{\frac{1}{1+\rho}}(Y), \quad (74)$$

$$R_X + R_Y \geq H_{\frac{1}{1+\rho}}(X,Y) + K_{\frac{1}{1+\rho}}(X;Y). \quad (75)$$

Rate pairs (R_X, R_Y) outside this region are not achievable and rate pairs in the interior of this region are achievable.

[5] Although $K_\alpha(X;Y)$ is not monotonic in α, it is possible to show that the sum $K_\alpha(X;Y) + H_\alpha(X,Y)$ is nonincreasing in α.

[6] The list may also contain pairs with posterior probability zero; for a precise definition, see (72).

References

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

[2] A. Lapidoth and C. Pfister, "Two measures of dependence," preprint, arXiv:1607.02330.

[3] M. Tomamichel and M. Hayashi, "Operational interpretation of Rényi information measures via composite hypothesis testing against product and Markov distributions," preprint, arXiv:1511.04874.

[4] R. Sibson, "Information radius," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, vol. 14, no. 2, pp. 149–160, June 1969.

[5] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory*, I. Csiszár and P. Elias, Eds. Amsterdam: North-Holland Publishing, 1977, pp. 41–52.

[6] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 26–34, Jan. 1995.

[7] S. Verdú, "α-mutual information," in *Proc. 2015 Inf. Theory and Applicat. Workshop*, San Diego, CA, 2015, pp. 1–6.

[8] 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.

[9] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 269–287, Jan. 2007.

[10] J. L. Massey, "Guessing and entropy," in *Proc. 1994 IEEE Int. Symp. Inf. Theory*, Trondheim, 1994, p. 204.

[11] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, Jan. 1996.

[12] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5065–5076, Sept. 2014.

[13] M. A. Kumar and R. Sundaresan, "Minimization problems based on relative α-entropy I: forward projection," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 5063–5080, Sept. 2015.

[14] M. A. Kumar and R. Sundaresan, "Minimization problems based on relative α-entropy II: reverse projection," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 5081–5095, Sept. 2015.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006.

[16] R. T. Rockafellar, *Convex Analysis*, Princeton, NJ: Princeton Univ. Press, 1970.

[17] A. Bracher, A. Lapidoth, and C. Pfister, "Distributed guessing and task encoding," in preparation.