entropy

MDPI

Article

Two Measures of Dependence

Amos Lapidoth and Christoph Pfister *

Signal and Information Processing Laboratory, ETH Zurich, 8092 Zurich, Switzerland
* Correspondence: pfister@isi.ee.ethz.ch

Received: 5 July 2019; Accepted: 5 August 2019; Published: 8 August 2019

check for updates

Abstract: Two families of dependence measures between random variables are introduced. They are based on the Rényi divergence of order α and the relative α-entropy, respectively, and both dependence measures reduce to Shannon's mutual information when their order α is one. The first measure shares many properties with the mutual information, including the data-processing inequality, and can be related to the optimal error exponents in composite hypothesis testing. The second measure does not satisfy the data-processing inequality, but appears naturally in the context of distributed task encoding.

Keywords: data processing; dependence measure; relative α-entropy; Rényi divergence; Rényi entropy

---

1. Introduction

The solutions to many information-theoretic problems can be expressed using Shannon's information measures such as entropy, relative entropy, and mutual information. Other problems require Rényi's information measures, which generalize Shannon's. In this paper, we analyze two Rényi measures of dependence, $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$, between random variables X and Y taking values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$, with $\alpha \in [0, \infty]$ being a parameter. (Our notation is similar to the one used for the mutual information: technically, $J_\alpha(\cdot)$ and $K_\alpha(\cdot)$ are functions not of X and Y, but of their joint probability mass function (PMF) P_{XY}.) For $\alpha \in [0, \infty]$, we define $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ as

$$J_\alpha(X;Y) \triangleq \min_{(Q_X,Q_Y)\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})} D_\alpha(P_{XY}\|Q_X Q_Y), \tag{1}$$

$$K_\alpha(X;Y) \triangleq \min_{(Q_X,Q_Y)\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})} \Delta_\alpha(P_{XY}\|Q_X Q_Y), \tag{2}$$

where $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ denote the set of all PMFs over $\mathcal{X}$ and $\mathcal{Y}$, respectively; $D_\alpha(P\|Q)$ denotes the Rényi divergence of order α (see (50) ahead); and $\Delta_\alpha(P\|Q)$ denotes the relative α-entropy (see (55) ahead). As shown in Proposition 7, $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ are in fact closely related.

The measures $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ have the following operational meanings (see Section 3): $J_\alpha(X;Y)$ is related to the optimal error exponents in testing whether the observed independent and identically distributed (IID) samples were generated according to the joint PMF P_{XY} or an unknown product PMF; and $K_\alpha(X;Y)$ appears as a penalty term in the sum-rate constraint of distributed task encoding.

The measures $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ share many properties with Shannon's mutual information [1], and both are equal to the mutual information when α is one. Except for some special cases, we have no closed-form expressions for $J_\alpha(X;Y)$ or $K_\alpha(X;Y)$. As illustrated in Figure 1, unless α is one, the minimum in the definitions of $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ is typically not achieved by $Q_X = P_X$ and $Q_Y = P_Y$. (When α is one, then the minimum is always achieved by $Q_X = P_X$ and $Q_Y = P_Y$; this follows from Proposition 8 and the fact that $D_1(P_{XY}\|Q_X Q_Y) = \Delta_1(P_{XY}\|Q_X Q_Y) = D(P_{XY}\|Q_X Q_Y)$.)

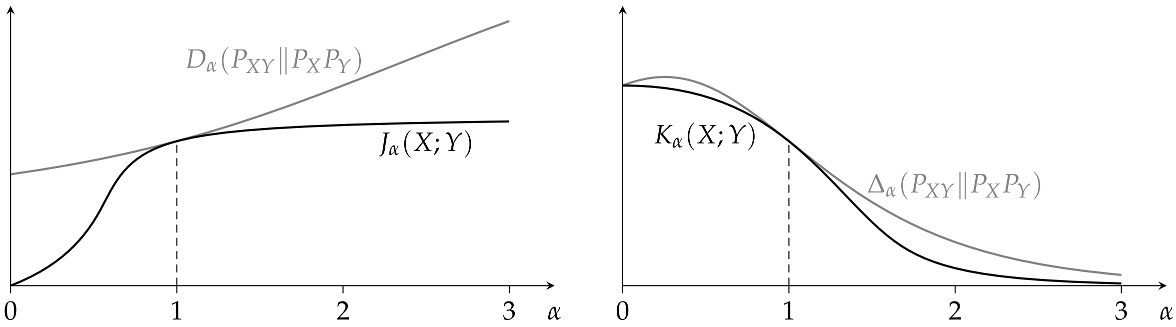

Figure 1. (**Left**) $J_\alpha(X;Y)$ and $D_\alpha(P_{XY}\|P_XP_Y)$ versus α. (**Right**) $K_\alpha(X;Y)$ and $\Delta_\alpha(P_{XY}\|P_XP_Y)$ versus α. In both plots, X is Bernoulli with $\Pr(X=1)=0.2$, and Y is equal to X.

The rest of this paper is organized as follows. In Section 2, we review other generalizations of the mutual information. In Section 3, we discuss the operational meanings of $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$. In Section 4, we recall the required Rényi information measures and prove some preparatory results. In Section 5, we state the properties of $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$. In Section 6, we prove these properties.

2. Related Work

The measure $J_\alpha(X;Y)$ was discovered independently from the authors of the present paper by Tomamichel and Hayashi [2] (Equation (58)), who, for the case when $\alpha > \frac{1}{2}$, derived some of its properties in [2] (Appendix A-C).

Other Rényi-based measures of dependence appeared in the past. Notable are those by Sibson [3], Arimoto [4], and Csiszár [5], respectively denoted by $I_\alpha^{\mathsf{s}}(\cdot)$, $I_\alpha^{\mathsf{a}}(\cdot)$, and $I_\alpha^{\mathsf{c}}(\cdot)$:

$$I_\alpha^{\mathsf{s}}(X;Y) \triangleq \frac{\alpha}{\alpha-1}\log\sum_y\left[\sum_x P(x)P(y|x)^\alpha\right]^{\frac{1}{\alpha}} \tag{3}$$

$$= \min_{Q_Y} D_\alpha(P_{XY}\|P_XQ_Y), \tag{4}$$

$$I_\alpha^{\mathsf{a}}(X;Y) \triangleq H_\alpha(X) - H_\alpha(X|Y) \tag{5}$$

$$= \frac{\alpha}{\alpha-1}\log\sum_y\left[\sum_x \frac{P(x)^\alpha}{\sum_{x'\in\mathcal{X}}P(x')^\alpha}P(y|x)^\alpha\right]^{\frac{1}{\alpha}}, \tag{6}$$

$$I_\alpha^{\mathsf{c}}(X;Y) \triangleq \min_{Q_Y}\sum_x P(x)D_\alpha(P_{Y|X=x}\|Q_Y), \tag{7}$$

where, throughout the paper, $\log(\cdot)$ denotes the base-2 logarithm; $D_\alpha(P\|Q)$ denotes the Rényi divergence of order α (see (50) ahead); $H_\alpha(X)$ denotes the Rényi entropy of order α (see (45) ahead); and $H_\alpha(X|Y)$ denotes the Arimoto–Rényi conditional entropy [4,6,7], which is defined for positive α other than one as

$$H_\alpha(X|Y) \triangleq \frac{\alpha}{1-\alpha}\log\sum_y\left[\sum_x P(x,y)^\alpha\right]^{\frac{1}{\alpha}}. \tag{8}$$

(Equation (4) follows from Proposition 9 ahead, and (6) follows from (45) and (8).) An overview of $I_\alpha^{\mathsf{s}}(\cdot)$, $I_\alpha^{\mathsf{a}}(\cdot)$, and $I_\alpha^{\mathsf{c}}(\cdot)$ is provided in [8]. Another Rényi-based measure of dependence can be found in [9] (Equation (19)):

$$I_\alpha^{\mathsf{t}}(X;Y) \triangleq D_\alpha(P_{XY}\|P_XP_Y). \tag{9}$$

The relation between $I_\alpha^{\mathsf{c}}(X;Y)$, $J_\alpha(X;Y)$, and $I_\alpha^{\mathsf{s}}(X;Y)$ for $\alpha > 1$ was established recently:

Proposition 1 ([10] (Theorem IV.1))**.** *For every PMF P_{XY} and every $\alpha > 1$,*

$$I_\alpha^{\mathsf{c}}(X;Y) \le J_\alpha(X;Y) \tag{10}$$

$$\le I_\alpha^{\mathsf{s}}(X;Y). \tag{11}$$

Proof. This is proved in [10] for a measure-theoretic setting. Here, we specialize the proof to finite alphabets. We first prove (10):

$$J_\alpha(X;Y) = \min_{Q_Y} \min_{Q_X} D_\alpha(P_{XY} \| Q_X Q_Y) \tag{12}$$

$$= \min_{Q_Y} \frac{\alpha}{\alpha-1} \log \sum_x \left[\sum_y P(x,y)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{13}$$

$$= \min_{Q_Y} \frac{\alpha}{\alpha-1} \log \sum_x P(x) \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{14}$$

$$\ge \min_{Q_Y} \frac{\alpha}{\alpha-1} \sum_x P(x) \log \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{15}$$

$$= \min_{Q_Y} \sum_x P(x) \frac{1}{\alpha-1} \log \sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \tag{16}$$

$$= I_\alpha^{\mathsf{c}}(X;Y), \tag{17}$$

where (12) follows from the definition of $J_\alpha(X;Y)$ in (1); (13) follows from Proposition 9 ahead with the roles of Q_X and Q_Y swapped; (15) follows from Jensen's inequality because $\log(\cdot)$ is concave and because $\frac{\alpha}{\alpha-1} > 0$; and (17) follows from the definition of $I_\alpha^{\mathsf{c}}(X;Y)$ in (7).

We next prove (11):

$$J_\alpha(X;Y) = \min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X Q_Y) \tag{18}$$

$$\le \min_{Q_Y} D_\alpha(P_{XY} \| P_X Q_Y) \tag{19}$$

$$= I_\alpha^{\mathsf{s}}(X;Y), \tag{20}$$

where (18) follows from the definition of $J_\alpha(X;Y)$ in (1), and (20) follows from (4). □

Many of the above Rényi information measures coincide when they are maximized over P_X with $P_{Y|X}$ held fixed: for every conditional PMF $P_{Y|X}$ and every positive α other than one,

$$\max_{P_X} I_\alpha^{\mathsf{a}}(P_X P_{Y|X}) = \max_{P_X} I_\alpha^{\mathsf{s}}(P_X P_{Y|X}) \tag{21}$$

$$= \max_{P_X} I_\alpha^{\mathsf{c}}(P_X P_{Y|X}), \tag{22}$$

where $P_X P_{Y|X}$ denotes the joint PMF of X and Y; (21) follows from [4] (Lemma 1); and (22) follows from [5] (Proposition 1). It was recently established that, for $\alpha > 1$, this is also true for $J_\alpha(X;Y)$:

Proposition 2 ([10] (Theorem V.1))**.** *For every conditional PMF $P_{Y|X}$ and every $\alpha > 1$,*

$$\max_{P_X} J_\alpha(P_X P_{Y|X}) = \max_{P_X} I_\alpha^{\mathsf{s}}(P_X P_{Y|X}). \tag{23}$$

Proof. By Proposition 1, we have for all $\alpha > 1$

$$\max_{P_X} I_\alpha^{\mathsf{c}}(P_X P_{Y|X}) \le \max_{P_X} J_\alpha(P_X P_{Y|X}) \tag{24}$$
$$\le \max_{P_X} I_\alpha^{\mathsf{s}}(P_X P_{Y|X}). \tag{25}$$

By (22), the left-hand side (LHS) of (24) is equal to the right-hand side (RHS) of (25), so (24) and (25) both hold with equality. □

Dependence measures can also be based on the f-divergence $D_f(P\|Q)$ [11–13]. Every convex function $f\colon (0,\infty) \to \mathbb{R}$ satisfying $f(1) = 0$ induces a dependence measure, namely

$$I_f(X;Y) \triangleq D_f(P_{XY}\|P_X P_Y) \tag{26}$$
$$= \sum_{x,y} P(x)P(y) f\left(\frac{P(x,y)}{P(x)P(y)}\right), \tag{27}$$

where (27) follows from the definition of the f-divergence. (For $f(t) = t\log t$, $I_f(X;Y)$ is the mutual information.) Such dependence measures are used for example in [14], and a construction equivalent to (27) is studied in [15].

3. Operational Meanings

In this section, we discuss the operational meaning of $J_\alpha(X;Y)$ in hypothesis testing (Section 3.1) and of $K_\alpha(X;Y)$ in distributed task encoding (Section 3.2).

3.1. Testing Against Independence and $J_\alpha(X;Y)$

Consider the hypothesis testing problem of guessing whether an observed sequence of pairs was drawn IID from some given joint PMF P_{XY} or IID from some unknown product distribution. Thus, based on a sequence of pairs of random variables $\{(X_i, Y_i)\}_{i=1}^n$, two hypotheses have to be distinguished:

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 and Q_Y are arbitrary PMFs over $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Associated with every deterministic test $T_n\colon \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ and pair (Q_X, Q_Y) are the type-I error probability $P_{XY}^{\times n}[T_n(X^n, Y^n) = 1]$ and the type-II error probability $(Q_X Q_Y)^{\times n}[T_n(X^n, Y^n) = 0]$, where $R_{XY}^{\times n}[\mathcal{A}]$ denotes the probability of an event $\mathcal{A}$ when $\{(X_i, Y_i)\}_{i=1}^n$ are IID according to R_{XY}. We seek sequences of tests whose worst-case type-II error probability decays exponentially faster than $2^{-n\mathsf{E}_\mathsf{Q}}$. To be more specific, for a fixed $\mathsf{E}_\mathsf{Q} \in \mathbb{R}$, denote by $\mathcal{T}(\mathsf{E}_\mathsf{Q})$ the set of all sequences of deterministic tests $\{T_n\}_{n=1}^\infty$ for which

$$\liminf_{n\to\infty} \min_{Q_X, Q_Y} -\frac{1}{n}\log\big((Q_X Q_Y)^{\times n}[T_n(X^n, Y^n) = 0]\big) > \mathsf{E}_\mathsf{Q}, \tag{28}$$

where $\log(\cdot)$ denotes the base-2 logarithm. Note that (28) implies—but is not equivalent to—that for n sufficiently large, $(Q_X Q_Y)^{\times n}[T_n(X^n, Y^n) = 0] \le 2^{-n\mathsf{E}_\mathsf{Q}}$ for all $(Q_X, Q_Y) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$. For a fixed $\mathsf{E}_\mathsf{Q} \in \mathbb{R}$, the optimal type-I error exponent that can be asymptotically achieved under the constraint (28) is given by

$$\mathsf{E}_\mathsf{P}(\mathsf{E}_\mathsf{Q}) \triangleq \sup_{\{T_n\}_{n=1}^\infty \in \mathcal{T}(\mathsf{E}_\mathsf{Q})} \liminf_{n\to\infty} -\frac{1}{n}\log\big(P_{XY}^{\times n}[T_n(X^n, Y^n) = 1]\big). \tag{29}$$

The measure $J_\alpha(X;Y)$ appears as follows: In [2] (first part of (57)), it is shown that for $\mathsf{E_Q}$ sufficiently close to $I(X;Y)$,

$$\mathsf{E_P}(\mathsf{E_Q}) = \sup_{\alpha \in (\frac{1}{2},1]} \frac{1-\alpha}{\alpha}\big(J_\alpha(X;Y) - \mathsf{E_Q}\big), \tag{30}$$

and in [16] (Theorem 3), it is shown that for all $\mathsf{E_Q} \in \mathbb{R}$,

$$\mathsf{E_P^{**}}(\mathsf{E_Q}) = \sup_{\alpha \in (0,1]} \frac{1-\alpha}{\alpha}\big(J_\alpha(X;Y) - \mathsf{E_Q}\big), \tag{31}$$

where $\mathsf{E_P^{**}}(\cdot)$ denotes the Fenchel biconjugate of $\mathsf{E_P}(\cdot)$. In general, the Fenchel biconjugation cannot be omitted because sometimes [16] (Equation (11) and Example 14)

$$\mathsf{E_P}(\mathsf{E_Q}) \neq \mathsf{E_P^{**}}(\mathsf{E_Q}). \tag{32}$$

For large values of $\mathsf{E_Q}$, the optimal type-I error tends to one as n tends to infinity. In this case, the type-I strong-converse exponent [17,18], which is defined for a sequence of tests $\{T_n\}_{n=1}^\infty$ as

$$\mathsf{SC_P} \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log\big(1 - P_{XY}^{\times n}[T_n(X^n,Y^n)=1]\big), \tag{33}$$

measures how fast the type-I error tends to one as n tends to infinity (smaller values correspond to lower error probabilities). For a fixed $\mathsf{E_Q} \in \mathbb{R}$, the optimal type-I strong-converse exponent that can be asymptotically achieved under the constraint (28) is given by

$$\mathsf{SC_P}(\mathsf{E_Q}) \triangleq \inf_{\{T_n\}_{n=1}^\infty \in \mathcal{T}(\mathsf{E_Q})} \limsup_{n\to\infty} -\frac{1}{n}\log\big(1 - P_{XY}^{\times n}[T_n(X^n,Y^n)=1]\big). \tag{34}$$

In [2] (second part of (57)), it is shown that for $\mathsf{E_Q}$ sufficiently close to $I(X;Y)$,

$$\mathsf{SC_P}(\mathsf{E_Q}) = \sup_{\alpha>1} \frac{1-\alpha}{\alpha}\big(J_\alpha(X;Y) - \mathsf{E_Q}\big). \tag{35}$$

Here, the same $\frac{1-\alpha}{\alpha}\big(J_\alpha(X;Y) - \mathsf{E_Q}\big)$ expression appears as in (30) and (31), but with a different set of α's to optimize over.

3.2. Distributed Task Encoding and $K_\alpha(X;Y)$

The task-encoding problem studied in [19] can be extended to a distributed setting as follows [20]: A source $\{(X_i,Y_i)\}_{i=1}^\infty$ emits pairs of random variables (X_i,Y_i) taking values in a finite alphabet $\mathcal{X}\times\mathcal{Y}$. For a fixed rate pair $(\mathsf{R_X},\mathsf{R_Y}) \in \mathbb{R}^2_{\geq 0}$ and a positive integer n, the sequences $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are described separately using $\lfloor 2^{n\mathsf{R_X}}\rfloor$ and $\lfloor 2^{n\mathsf{R_Y}}\rfloor$ labels, respectively. The decoder produces a list comprising all the pairs (x^n,y^n) whose description matches the given labels, and the goal is to minimize the ρ-th moment of the list size as n tends to infinity (for some $\rho>0$).

For a fixed $\rho>0$, a rate pair $(\mathsf{R_X},\mathsf{R_Y}) \in \mathbb{R}^2_{\geq 0}$ is called achievable if there exists a sequence of encoders $\{(f_n,g_n)\}_{n=1}^\infty$,

$$f_n\colon \mathcal{X}^n \to \{1,\ldots,\lfloor 2^{n\mathsf{R_X}}\rfloor\}, \tag{36}$$

$$g_n\colon \mathcal{Y}^n \to \{1,\ldots,\lfloor 2^{n\mathsf{R_Y}}\rfloor\}, \tag{37}$$

such that the ρ-th moment of the list size tends to one as n tends to infinity, i.e.,

$$\lim_{n\to\infty} \mathrm{E}\big[|\mathcal{L}(X^n,Y^n)|^\rho\big] = 1, \tag{38}$$

where

$$\mathcal{L}(x^n, y^n) \triangleq \{(\tilde{x}^n, \tilde{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n : f_n(\tilde{x}^n) = f_n(x^n) \wedge g_n(\tilde{y}^n) = g_n(y^n)\}. \tag{39}$$

For a memoryless source and a fixed $\rho > 0$, rate pairs in the interior of the region $\mathcal{R}(\rho)$ defined next are achievable, while those outside $\mathcal{R}(\rho)$ are not achievable [20] (Theorem 1). The region $\mathcal{R}(\rho)$ is defined as the set of all rate pairs $(\mathsf{R_X}, \mathsf{R_Y})$ satisfying the following inequalities simultaneously:

$$\mathsf{R_X} \geq H_{\frac{1}{1+\rho}}(X), \tag{40}$$

$$\mathsf{R_Y} \geq H_{\frac{1}{1+\rho}}(Y), \tag{41}$$

$$\mathsf{R_X} + \mathsf{R_Y} \geq H_{\frac{1}{1+\rho}}(X, Y) + K_{\frac{1}{1+\rho}}(X; Y), \tag{42}$$

where $H_\alpha(X)$ denotes the Rényi entropy of order α (see (45) ahead).

To better understand the role of $K_\alpha(X; Y)$, suppose that the sequences $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ were allowed to be described jointly using $\lfloor 2^{n\mathsf{R_X}} \rfloor \cdot \lfloor 2^{n\mathsf{R_Y}} \rfloor \approx 2^{n(\mathsf{R_X}+\mathsf{R_Y})}$ labels. Then, by [19] (Theorem I.2), all rate pairs $(\mathsf{R_X}, \mathsf{R_Y}) \in \mathbb{R}^2_{\geq 0}$ satisfying the following inequality with strict inequality would be achievable, while those not satisfying the inequality would not:

$$\mathsf{R_X} + \mathsf{R_Y} \geq H_{\frac{1}{1+\rho}}(X, Y). \tag{43}$$

Comparing (42) and (43), we see that the measure $K_\alpha(X; Y)$ appears as a penalty term on the sum-rate constraint incurred by requiring that the sequences be described separately as opposed to jointly.

4. Preliminaries

Throughout the paper, $\log(\cdot)$ denotes the base-2 logarithm, $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, P_{XY} denotes a joint PMF over $\mathcal{X} \times \mathcal{Y}$, Q_X denotes a PMF over $\mathcal{X}$, and Q_Y denotes a PMF over $\mathcal{Y}$. We use P and Q as generic PMFs over a finite set $\mathcal{X}$. We denote by $\mathrm{supp}(P) \triangleq \{x \in \mathcal{X} : P(x) > 0\}$ the support of P, and by $\mathcal{P}(\mathcal{X})$ the set of all PMFs over $\mathcal{X}$. When clear from the context, we often omit sets and subscripts: for example, we write $\min_{Q_X, Q_Y}$ for $\min_{(Q_X, Q_Y) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})}$, $\sum_x$ for $\sum_{x \in \mathcal{X}}$, $P(x)$ for $P_X(x)$, and $P(y|x)$ for $P_{Y|X}(y|x)$. Whenever a conditional probability $P(y|x)$ is undefined because $P(x) = 0$, we define $P(y|x) \triangleq 1/|\mathcal{Y}|$. We denote by $\mathbb{1}\{\mathsf{condition}\}$ the indicator function that is one if the condition is satisfied and zero otherwise. In the definitions below, we use the following conventions:

$$\frac{0}{0} = 0, \qquad \frac{p}{0} = \infty \quad \forall p > 0, \qquad 0 \log 0 = 0, \qquad \beta \log 0 = -\infty \quad \forall \beta > 0. \tag{44}$$

The Rényi entropy of order α [21] is defined for positive α other than one as

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha} \log \sum_x P(x)^\alpha. \tag{45}$$

For α being zero, one, or infinity, we define by continuous extension of (45)

$$H_0(X) \triangleq \log |\mathrm{supp}(P)|, \tag{46}$$

$$H_1(X) \triangleq H(X), \tag{47}$$

$$H_\infty(X) \triangleq -\log \max_x P(x), \tag{48}$$

where $H(X)$ is the Shannon entropy. With this extension to $\alpha \in \{0, 1, \infty\}$, the Rényi entropy satisfies the following basic properties:

Proposition 3 ([5]). *Let P be a PMF. Then,*

(i) For all $\alpha \in [0,\infty]$, $H_\alpha(X) \le \log|\mathcal{X}|$. If $\alpha \in (0,\infty]$, then $H_\alpha(X) = \log|\mathcal{X}|$ if and only if X is distributed uniformly over $\mathcal{X}$.
(ii) The mapping $\alpha \mapsto H_\alpha(X)$ is nonincreasing on $[0,\infty]$.
(iii) The mapping $\alpha \mapsto H_\alpha(X)$ is continuous on $[0,\infty]$.

The relative entropy (or Kullback–Leibler divergence) is defined as

$$D(P\|Q) \triangleq \sum_x P(x) \log \frac{P(x)}{Q(x)}. \tag{49}$$

The Rényi divergence of order α [21,22] is defined for positive α other than one as

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha - 1} \log \sum_x P(x)^\alpha Q(x)^{1-\alpha}, \tag{50}$$

where we read $P(x)^\alpha Q(x)^{1-\alpha}$ as $P(x)^\alpha / Q(x)^{\alpha-1}$ if $\alpha > 1$. For α being zero, one, or infinity, we define by continuous extension of (50)

$$D_0(P\|Q) \triangleq -\log \sum_{x \in \mathrm{supp}(P)} Q(x), \tag{51}$$

$$D_1(P\|Q) \triangleq D(P\|Q), \tag{52}$$

$$D_\infty(P\|Q) \triangleq \log \max_x \frac{P(x)}{Q(x)}. \tag{53}$$

With this extension to $\alpha \in \{0, 1, \infty\}$, the Rényi divergence satisfies the following basic properties:

Proposition 4. *Let P and Q be PMFs. Then,*

(i) For all $\alpha \in [0,1)$, $D_\alpha(P\|Q)$ is finite if and only if $|\mathrm{supp}(P) \cap \mathrm{supp}(Q)| > 0$. For all $\alpha \in [1,\infty]$, $D_\alpha(P\|Q)$ is finite if and only if $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$.
(ii) For all $\alpha \in [0,\infty]$, $D_\alpha(P\|Q) \ge 0$. If $\alpha \in (0,\infty]$, then $D_\alpha(P\|Q) = 0$ if and only if $P = Q$.
(iii) For every $\alpha \in [0,\infty]$, the mapping $Q \mapsto D_\alpha(P\|Q)$ is continuous.
(iv) The mapping $\alpha \mapsto D_\alpha(P\|Q)$ is nondecreasing on $[0,\infty]$.
(v) The mapping $\alpha \mapsto D_\alpha(P\|Q)$ is continuous on $[0,\infty]$.

Proof. Part (i) follows from the definition of $D_\alpha(P\|Q)$ and the conventions (44), and Parts (ii)–(v) are shown in [22]. □

The Rényi divergence for negative α is defined as

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha - 1} \log \sum_x \frac{Q(x)^{1-\alpha}}{P(x)^{-\alpha}}. \tag{54}$$

(We use negative α only in Lemma 19. More about negative orders can be found in [22] (Section V). For other applications of negative orders, see [23] (Proof of Theorem 1 and Example 1).)

The relative α-entropy [24,25] is defined for positive α other than one as

$$\Delta_\alpha(P\|Q) \triangleq \frac{\alpha}{1-\alpha} \log \sum_x P(x) Q(x)^{\alpha-1} + \log \sum_x Q(x)^\alpha - \frac{1}{1-\alpha} \log \sum_x P(x)^\alpha, \tag{55}$$

where we read $P(x)Q(x)^{\alpha-1}$ as $P(x)/Q(x)^{1-\alpha}$ if $\alpha < 1$. The relative α-entropy appears in mismatched guessing [26], mismatched source coding [26] (Theorem 8), and mismatched task encoding [19]

(Section IV). It also arises in robust parameter estimation and constrained compression settings [25] (Section II). For α being zero, one, or infinity, we define by continuous extension of (55)

$$\Delta_0(P\|Q) \triangleq \begin{cases} \log \frac{|\mathrm{supp}(Q)|}{|\mathrm{supp}(P)|} & \text{if } \mathrm{supp}(P) \subseteq \mathrm{supp}(Q), \\ \infty & \text{otherwise}, \end{cases} \tag{56}$$

$$\Delta_1(P\|Q) \triangleq D(P\|Q), \tag{57}$$

$$\Delta_\infty(P\|Q) \triangleq \log \frac{\max_x P(x)}{|\mathrm{argmax}(Q)|^{-1} \sum_{x \in \mathrm{argmax}(Q)} P(x)}, \tag{58}$$

where $\mathrm{argmax}(Q) \triangleq \{x \in \mathcal{X} : Q(x) = \max_{x' \in \mathcal{X}} Q(x')\}$ and $|\mathrm{argmax}(Q)|$ is the cardinality of this set. With this extension to $\alpha \in \{0, 1, \infty\}$, the relative α-entropy satisfies the following basic properties:

Proposition 5. *Let P and Q be PMFs. Then,*

- *(i) For all $\alpha \in [0,1]$, $\Delta_\alpha(P\|Q)$ is finite if and only if* $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$. *For all $\alpha \in (1,\infty)$, $\Delta_\alpha(P\|Q)$ is finite if and only if* $|\mathrm{supp}(P) \cap \mathrm{supp}(Q)| > 0$.
- *(ii) For all $\alpha \in [0,\infty]$, $\Delta_\alpha(P\|Q) \geq 0$. If $\alpha \in (0,\infty)$, then $\Delta_\alpha(P\|Q) = 0$ if and only if $P = Q$.*
- *(iii) For every $\alpha \in (0,\infty)$, the mapping $Q \mapsto \Delta_\alpha(P\|Q)$ is continuous.*
- *(iv) The mapping $\alpha \mapsto \Delta_\alpha(P\|Q)$ is continuous on $[0,\infty]$.*

(Part (i) differs from [19] (Proposition IV.1), where the conventions for $\alpha > 1$ differ from ours. Our conventions are compatible with [24,25], and, as stated in Part (iii), they result in the continuity of the mapping $Q \mapsto \Delta_\alpha(P\|Q)$.)

Proof of Proposition 5. Part (i) follows from the definition of $\Delta_\alpha(P\|Q)$ in (55) and the conventions (44). For $\alpha \in (0,1) \cup (1,\infty)$, Part (ii) follows from [19] (Proposition IV.1); for $\alpha = 1$, Part (ii) holds because $\Delta_1(P\|Q) = D(P\|Q)$; and for $\alpha \in \{0,\infty\}$, Part (ii) follows from the definition of $\Delta_\alpha(P\|Q)$. Part (iii) follows from the definition of $\Delta_\alpha(P\|Q)$, and Part (iv) follows from [19] (Proposition IV.1). □

In the rest of this section, we prove some auxiliary results that we need later (Propositions 6–9). We first establish the relation between $D_\alpha(P\|Q)$ and $\Delta_\alpha(P\|Q)$.

Proposition 6 ([26] (Section V, Property 4))**.** *Let P and Q be PMFs, and let $\alpha > 0$. Then,*

$$\Delta_\alpha(P\|Q) = D_{\frac{1}{\alpha}}(\widetilde{P}\|\widetilde{Q}), \tag{59}$$

where the PMFs $\widetilde{P}$ and $\widetilde{Q}$ are given by

$$\widetilde{P}(x) \triangleq \frac{P(x)^\alpha}{\sum_{x' \in \mathcal{X}} P(x')^\alpha}, \tag{60}$$

$$\widetilde{Q}(x) \triangleq \frac{Q(x)^\alpha}{\sum_{x' \in \mathcal{X}} Q(x')^\alpha}. \tag{61}$$

Proof. If $\alpha = 1$, then (59) holds because $\widetilde{P} = P$, $\widetilde{Q} = Q$, and $\Delta_1(P\|Q) = D_1(P\|Q) = D(P\|Q)$. Now let $\alpha \neq 1$. Because $\widetilde{P}(x)$ and $\widetilde{Q}(x)$ are zero if and only if $P(x)$ and $Q(x)$ are zero, respectively, the LHS of (59) is finite if and only if its RHS is finite. If $D_{1/\alpha}(\widetilde{P}\|\widetilde{Q})$ is finite, then (59) follows from a simple computation. □

In light of Proposition 6, $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ are related as follows:

Proposition 7. *Let P_{XY} be a joint PMF, and let $\alpha > 0$. Then,*

$$K_\alpha(X;Y) = J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}), \tag{62}$$

where the joint PMF of $\widetilde{X}$ and $\widetilde{Y}$ is given by

$$\widetilde{P}_{XY}(x,y) \triangleq \frac{P_{XY}(x,y)^\alpha}{\sum_{(x',y')\in\mathcal{X}\times\mathcal{Y}} P_{XY}(x',y')^\alpha}. \tag{63}$$

Proof. Let $\alpha > 0$. For fixed PMFs Q_X and Q_Y, define the transformed PMFs $\widetilde{Q_X Q_Y}$, $\widetilde{Q}_X$, and $\widetilde{Q}_Y$ as

$$\widetilde{Q_X Q_Y}(x,y) \triangleq \frac{[Q_X(x)\,Q_Y(y)]^\alpha}{\sum_{(x',y')\in\mathcal{X}\times\mathcal{Y}}[Q_X(x')\,Q_Y(y')]^\alpha}, \tag{64}$$

$$\widetilde{Q}_X(x) \triangleq \frac{Q_X(x)^\alpha}{\sum_{x'\in\mathcal{X}} Q_X(x')^\alpha}, \tag{65}$$

$$\widetilde{Q}_Y(y) \triangleq \frac{Q_Y(y)^\alpha}{\sum_{y'\in\mathcal{Y}} Q_Y(y')^\alpha}. \tag{66}$$

Then,

$$K_\alpha(X;Y) = \min_{Q_X,Q_Y} \Delta_\alpha(P_{XY}\|Q_X Q_Y) \tag{67}$$

$$= \min_{Q_X,Q_Y} D_{\frac{1}{\alpha}}(\widetilde{P}_{XY}\|\widetilde{Q_X Q_Y}) \tag{68}$$

$$= \min_{Q_X,Q_Y} D_{\frac{1}{\alpha}}(\widetilde{P}_{XY}\|\widetilde{Q}_X\widetilde{Q}_Y) \tag{69}$$

$$= \min_{Q_X,Q_Y} D_{\frac{1}{\alpha}}(\widetilde{P}_{XY}\|Q_X Q_Y) \tag{70}$$

$$= J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}), \tag{71}$$

where (67) holds by the definition of $K_\alpha(X;Y)$; (68) follows from Proposition 6; (69) holds because $\widetilde{Q_X Q_Y} = \widetilde{Q}_X\widetilde{Q}_Y$; (70) holds because the transformations (65) and (66) are bijective on the set of PMFs over $\mathcal{X}$ and $\mathcal{Y}$, respectively; and (71) holds by the definition of $J_\alpha(X;Y)$. □

The next proposition provides a characterization of the mutual information that parallels the definitions of $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$. Because $D_1(P\|Q) = \Delta_1(P\|Q) = D(P\|Q)$, this also shows that $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ reduce to the mutual information when α is one.

Proposition 8 ([27] (Theorem 3.4))**.** *Let P_{XY} be a joint PMF. Then, for all PMFs Q_X and Q_Y,*

$$D(P_{XY}\|Q_X Q_Y) \ge D(P_{XY}\|P_X P_Y), \tag{72}$$

with equality if and only if $Q_X = P_X$ and $Q_Y = P_Y$. Thus,

$$I(X;Y) = \min_{Q_X,Q_Y} D(P_{XY}\|Q_X Q_Y). \tag{73}$$

Proof. A simple computation reveals that

$$D(P_{XY}\|Q_X Q_Y) = D(P_{XY}\|P_X P_Y) + D(P_X\|Q_X) + D(P_Y\|Q_Y), \tag{74}$$

which implies (72) because $D(P\|Q) \geq 0$ with equality if and only if $P = Q$. Thus, (73) holds because $I(X;Y) = D(P_{XY}\|P_X P_Y)$. □

The last proposition of this section is about a precursor to $J_\alpha(X;Y)$, namely, the minimization of $D_\alpha(P_{XY}\|Q_X Q_Y)$ with respect to Q_Y only, which can be carried out explicitly. (This proposition extends [5] (Equation (13)) and [2] (Lemma 29).)

Proposition 9. *Let P_{XY} be a joint PMF and Q_X a PMF. Then, for every $\alpha \in (0,1) \cup (1,\infty)$,*

$$\min_{Q_Y} D_\alpha(P_{XY}\|Q_X Q_Y) = \frac{\alpha}{\alpha-1} \log \sum_y \left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}}, \tag{75}$$

with the conventions of (44). If the RHS of (75) is finite, then the minimum is achieved uniquely by

$$Q_Y^*(y) = \frac{\left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha}\right]^{\frac{1}{\alpha}}}{\sum_{y'\in\mathcal{Y}} \left[\sum_x P(x,y')^\alpha Q_X(x)^{1-\alpha}\right]^{\frac{1}{\alpha}}}. \tag{76}$$

For $\alpha = \infty$,

$$\min_{Q_Y} D_\infty(P_{XY}\|Q_X Q_Y) = \log \sum_y \max_x \frac{P(x,y)}{Q_X(x)}, \tag{77}$$

with the conventions of (44). If the RHS of (77) is finite, then the minimum is achieved uniquely by

$$Q_Y^*(y) = \frac{\max_x [P(x,y)/Q_X(x)]}{\sum_{y'\in\mathcal{Y}} \max_x [P(x,y')/Q_X(x)]}. \tag{78}$$

Proof. We first treat the case $\alpha \in (0,1) \cup (1,\infty)$. If the RHS of (75) is infinite, then the conventions imply that $D_\alpha(P_{XY}\|Q_X Q_Y)$ is infinite for every $Q_Y \in \mathcal{P}(\mathcal{Y})$, so (75) holds. Otherwise, if the RHS of (75) is finite, then the PMF Q_Y^* given by (76) is well-defined, and a simple computation shows that for every $Q_Y \in \mathcal{P}(\mathcal{Y})$,

$$D_\alpha(P_{XY}\|Q_X Q_Y) = \frac{\alpha}{\alpha-1} \log \sum_y \left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}} + D_\alpha(Q_Y^*\|Q_Y). \tag{79}$$

The only term on the RHS of (79) that depends on Q_Y is $D_\alpha(Q_Y^*\|Q_Y)$. Because $D_\alpha(Q_Y^*\|Q_Y) \geq 0$ with equality if and only if $Q_Y = Q_Y^*$ (Proposition 4), (79) implies (75) and (76).

The case $\alpha = \infty$ is analogous: if the RHS of (77) is infinite, then the LHS of (77) is infinite, too; and if the RHS of (77) is finite, then the PMF Q_Y^* given by (78) is well-defined, and a simple computation shows that for every $Q_Y \in \mathcal{P}(\mathcal{Y})$,

$$D_\infty(P_{XY}\|Q_X Q_Y) = \log \sum_y \max_x \frac{P(x,y)}{Q_X(x)} + D_\infty(Q_Y^*\|Q_Y). \tag{80}$$

The only term on the RHS of (80) that depends on Q_Y is $D_\infty(Q_Y^*\|Q_Y)$. Because $D_\infty(Q_Y^*\|Q_Y) \geq 0$ with equality if and only if $Q_Y = Q_Y^*$ (Proposition 4), (80) implies (77) and (78). □

5. Two Measures of Dependence

We state the properties of $J_\alpha(X;Y)$ in Theorem 1 and those of $K_\alpha(X;Y)$ in Theorem 2. The enumeration labels in the theorems refer to the lemmas in Section 6 where the properties are proved. (The enumeration labels are not consecutive because, in order to avoid forward references in the proofs, the order of the results in Section 6 is not the same as here.)

Theorem 1. *Let X, X_1, X_2, Y, Y_1, Y_2, and Z be random variables taking values in finite sets. Then:*

(Lemma 1) *For every $\alpha \in [0, \infty]$, the minimum in the definition of $J_\alpha(X;Y)$ exists and is finite.*

The following properties of the mutual information $I(X;Y)$ [28] (Chapter 2) are also satisfied by $J_\alpha(X;Y)$:

(Lemma 2) *For all $\alpha \in [0, \infty]$, $J_\alpha(X;Y) \geq 0$. If $\alpha \in (0, \infty]$, then $J_\alpha(X;Y) = 0$ if and only if X and Y are independent (nonnegativity).*
(Lemma 3) *For all $\alpha \in [0, \infty]$, $J_\alpha(X;Y) = J_\alpha(Y;X)$ (symmetry).*
(Lemma 4) *If $X \multimap Y \multimap Z$ form a Markov chain, then $J_\alpha(X;Z) \leq J_\alpha(X;Y)$ for all $\alpha \in [0, \infty]$ (data-processing inequality).*
(Lemma 12) *If the pairs (X_1, Y_1) and (X_2, Y_2) are independent, then $J_\alpha(X_1, X_2; Y_1, Y_2) = J_\alpha(X_1;Y_1) + J_\alpha(X_2;Y_2)$ for all $\alpha \in [0, \infty]$ (additivity).*
(Lemma 13) *For all $\alpha \in [0, \infty]$, $J_\alpha(X;Y) \leq \log|\mathcal{X}|$ with equality if and only if $\big(\alpha \in [\frac{1}{2}, \infty]$, X is distributed uniformly over $\mathcal{X}$, and $H(X|Y) = 0\big)$.*
(Lemma 14) *For every $\alpha \in [1, \infty]$, $J_\alpha(X;Y)$ is concave in P_X for fixed $P_{Y|X}$.*

Moreover:

(Lemma 5) $J_0(X;Y) = 0$.
(Lemma 6) *Let $f\colon \{1, \ldots, |\mathcal{X}|\} \to \mathcal{X}$ and $g\colon \{1, \ldots, |\mathcal{Y}|\} \to \mathcal{Y}$ be bijective functions, and let A be the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix whose Row-i Column-j entry $\mathsf{A}_{i,j}$ equals $\sqrt{P_{XY}(f(i), g(j))}$. Then,*

$$J_{\frac{1}{2}}(X;Y) = -2 \log \sigma_1(\mathsf{A}), \tag{81}$$

where $\sigma_1(\mathsf{A})$ denotes the largest singular value of A. (Because the singular values of a matrix are invariant under row and column permutations, the result does not depend on f or g.)
(Lemma 7) $J_1(X;Y) = I(X;Y)$.
(Lemma 8) *For all $\alpha > 0$,*

$$(1-\alpha)\, J_\alpha(X;Y) = \min_{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \big[(1-\alpha)\, D(R_{XY} \| R_X R_Y) + \alpha\, D(R_{XY} \| P_{XY})\big]. \tag{82}$$

Thus, being the minimum of concave functions in α, the mapping $\alpha \mapsto (1-\alpha)\, J_\alpha(X;Y)$ is concave on $(0, \infty)$.
(Lemma 9) *The mapping $\alpha \mapsto J_\alpha(X;Y)$ is nondecreasing on $[0, \infty]$.*
(Lemma 10) *The mapping $\alpha \mapsto J_\alpha(X;Y)$ is continuous on $[0, \infty]$.*
(Lemma 11) *If $X = Y$ with probability one, then*

$$J_\alpha(X;Y) = \begin{cases} \frac{\alpha}{1-\alpha} H_\infty(X) & \text{if } \alpha \in [0, \frac{1}{2}], \\ H_{\frac{\alpha}{2\alpha-1}}(X) & \text{if } \alpha > \frac{1}{2}, \\ H_{\frac{1}{2}}(X) & \text{if } \alpha = \infty. \end{cases} \tag{83}$$

The minimization problem in the definition of $J_\alpha(X;Y)$ has the following characteristics:

(Lemma 15) *For every $\alpha \in [\frac{1}{2}, \infty]$, the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$ is convex, i.e., for all $\lambda, \lambda' \in [0,1]$ with $\lambda + \lambda' = 1$, all $Q_X, Q'_X \in \mathcal{P}(\mathcal{X})$, and all $Q_Y, Q'_Y \in \mathcal{P}(\mathcal{Y})$,*

$$D_\alpha\big(P_{XY} \| (\lambda Q_X + \lambda' Q'_X)(\lambda Q_Y + \lambda' Q'_Y)\big) \leq \lambda\, D_\alpha(P_{XY} \| Q_X Q_Y) + \lambda'\, D_\alpha(P_{XY} \| Q'_X Q'_Y). \tag{84}$$

For $\alpha \in [0, \frac{1}{2})$, the mapping need not be convex.

(Lemma 16) *Let $\alpha \in (0,1) \cup (1,\infty)$. If (Q_X^*, Q_Y^*) achieves the minimum in the definition of $J_\alpha(X;Y)$, then there exist positive normalization constants c and d such that*

$$Q_X^*(x) = c\left[\sum_y P(x,y)^\alpha Q_Y^*(y)^{1-\alpha}\right]^{\frac{1}{\alpha}} \quad \forall x \in \mathcal{X}, \tag{85}$$

$$Q_Y^*(y) = d\left[\sum_x P(x,y)^\alpha Q_X^*(x)^{1-\alpha}\right]^{\frac{1}{\alpha}} \quad \forall y \in \mathcal{Y}, \tag{86}$$

with the conventions of (44). The case $\alpha = \infty$ is similar: if (Q_X^*, Q_Y^*) achieves the minimum in the definition of $J_\infty(X;Y)$, then there exist positive normalization constants c and d such that

$$Q_X^*(x) = c \max_y \frac{P(x,y)}{Q_Y^*(y)} \quad \forall x \in \mathcal{X}, \tag{87}$$

$$Q_Y^*(y) = d \max_x \frac{P(x,y)}{Q_X^*(x)} \quad \forall y \in \mathcal{Y}, \tag{88}$$

with the conventions of (44). (If $\alpha = 1$, then $Q_X^* = P_X$ and $Q_Y^* = P_Y$ by Proposition 8.) Thus, for all $\alpha \in (0,\infty]$, both inclusions $\operatorname{supp}(Q_X^*) \subseteq \operatorname{supp}(P_X)$ and $\operatorname{supp}(Q_Y^*) \subseteq \operatorname{supp}(P_Y)$ hold.

(Lemma 20) *For every $\alpha \in (\frac{1}{2}, \infty]$, the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$ has a unique minimizer. This need not be the case when $\alpha \in [0, \frac{1}{2}]$.*

The measure $J_\alpha(X;Y)$ can also be expressed as follows:

(Lemma 17) *For all $\alpha \in (0,\infty]$,*

$$J_\alpha(X;Y) = \min_{Q_X} \phi_\alpha(Q_X), \tag{89}$$

where $\phi_\alpha(Q_X)$ is defined as

$$\phi_\alpha(Q_X) \triangleq \min_{Q_Y} D_\alpha(P_{XY} \| Q_X Q_Y) \tag{90}$$

and is given explicitly as follows: for $\alpha \in (0,1) \cup (1,\infty)$,

$$\phi_\alpha(Q_X) = \frac{\alpha}{\alpha - 1} \log \sum_y \left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha}\right]^{\frac{1}{\alpha}}, \tag{91}$$

with the conventions of (44); and for $\alpha \in \{1, \infty\}$,

$$\phi_1(Q_X) = D(P_{XY} \| Q_X P_Y), \tag{92}$$

$$\phi_\infty(Q_X) = \log \sum_y \max_x \frac{P(x,y)}{Q_X(x)}, \tag{93}$$

with the conventions of (44). For every $\alpha \in [\frac{1}{2}, \infty]$, the mapping $Q_X \mapsto \phi_\alpha(Q_X)$ is convex. For $\alpha \in (0, \frac{1}{2})$, the mapping need not be convex.

(Lemma 18) *For all $\alpha \in (0,1) \cup (1,\infty]$,*

$$J_\alpha(X;Y) = \begin{cases} \min\limits_{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \psi_\alpha(R_{XY}) & \text{if } \alpha \in (0,1), \\ \max\limits_{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \psi_\alpha(R_{XY}) & \text{if } \alpha \in (1,\infty], \end{cases} \tag{94}$$

where

$$\psi_\alpha(R_{XY}) \triangleq \begin{cases} D(R_{XY}\|R_X R_Y) + \dfrac{\alpha}{1-\alpha} D(R_{XY}\|P_{XY}) & \text{if } \alpha \in (0,1)\cup(1,\infty), \\ D(R_{XY}\|R_X R_Y) - D(R_{XY}\|P_{XY}) & \text{if } \alpha = \infty. \end{cases} \tag{95}$$

For every $\alpha \in (1,\infty]$, the mapping $R_{XY} \mapsto \psi_\alpha(R_{XY})$ is concave. For all $\alpha \in (1,\infty]$ and all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, the statement $J_\alpha(X;Y) = \psi_\alpha(R_{XY})$ is equivalent to $\psi_\alpha(R_{XY}) = D_\alpha(P_{XY}\|R_X R_Y)$.

(Lemma 19) *For all $\alpha \in (0,1)\cup(1,\infty)$,*

$$J_\alpha(X;Y) = \min_{R_X \ll P_X} \frac{1}{\alpha-1}\Big[D_{\frac{\alpha}{\alpha-1}}(P_X\|R_X) - \alpha E_0\big(\tfrac{1-\alpha}{\alpha}, R_X\big)\Big], \tag{96}$$

where the minimization is over all PMFs R_X satisfying $R_X \ll P_X$ (i.e., $\operatorname{supp}(R_X) \subseteq \operatorname{supp}(P_X)$); $D_\alpha(P\|Q)$ for negative α is given by (54); and Gallager's E_0 function [29] is defined as

$$E_0(\rho, R_X) \triangleq -\log \sum_y \Big[\sum_x R_X(x)\, P(y|x)^{\frac{1}{1+\rho}}\Big]^{1+\rho}. \tag{97}$$

We now move on to the properties of $K_\alpha(X;Y)$. Some of these properties are derived from their counterparts of $J_\alpha(X;Y)$ using the relation $K_\alpha(X;Y) = J_{1/\alpha}(\widetilde{X};\widetilde{Y})$ described in Proposition 7.

Theorem 2. *Let X, X_1, X_2, Y, Y_1, Y_2, and Z be random variables taking values in finite sets. Then:*

(Lemma 21) *For every $\alpha \in [0,\infty]$, the minimum in the definition of $K_\alpha(X;Y)$ in (2) exists and is finite.*

The following properties of the mutual information $I(X;Y)$ are also satisfied by $K_\alpha(X;Y)$:

(Lemma 22) *For all $\alpha \in [0,\infty]$, $K_\alpha(X;Y) \ge 0$. If $\alpha \in (0,\infty)$, then $K_\alpha(X;Y) = 0$ if and only if X and Y are independent (nonnegativity).*

(Lemma 23) *For all $\alpha \in [0,\infty]$, $K_\alpha(X;Y) = K_\alpha(Y;X)$ (symmetry).*

(Lemma 34) *If the pairs (X_1,Y_1) and (X_2,Y_2) are independent, then $K_\alpha(X_1,X_2;Y_1,Y_2) = K_\alpha(X_1;Y_1) + K_\alpha(X_2;Y_2)$ for all $\alpha \in [0,\infty]$ (additivity).*

(Lemma 35) *For all $\alpha \in [0,\infty]$, $K_\alpha(X;Y) \le \log|\mathcal{X}|$.*

Unlike the mutual information, $K_\alpha(X;Y)$ does not satisfy the data-processing inequality:

(Lemma 36) *There exists a Markov chain $X \multimap\!\!- Y \multimap\!\!- Z$ for which $K_2(X;Z) > K_2(X;Y)$.*

Moreover:

(Lemma 24) *For all $\alpha \in (0,\infty)$,*

$$K_\alpha(X;Y) + H_\alpha(X,Y) = \min_{Q_X,Q_Y} -\log M_{\frac{\alpha-1}{\alpha}}(Q_X,Q_Y), \tag{98}$$

where $M_\beta(Q_X,Q_Y)$ is the following weighted power mean [30] (Chapter III): For $\beta \in \mathbb{R}\setminus\{0\}$,

$$M_\beta(Q_X,Q_Y) \triangleq \Big[\sum_{x,y} P(x,y)[Q_X(x)\,Q_Y(y)]^\beta\Big]^{\frac{1}{\beta}}, \tag{99}$$

where for $\beta < 0$, we read $P(x,y)[Q_X(x)Q_Y(y)]^\beta$ as $P(x,y)/[Q_X(x)Q_Y(y)]^{-\beta}$ and use the conventions (44); and for $\beta = 0$, using the convention $0^0 = 1$,

$$M_0(Q_X,Q_Y) \triangleq \prod_{x,y}[Q_X(x)\,Q_Y(y)]^{P(x,y)}. \tag{100}$$

(Lemma 25) *For* $\alpha = 0$,

$$K_0(X;Y) = \log \frac{|\mathrm{supp}(P_X P_Y)|}{|\mathrm{supp}(P_{XY})|} \tag{101}$$

$$\geq \min_{Q_X, Q_Y} \log \max_{(x,y) \in \mathrm{supp}(P_{XY})} \frac{1}{Q_X(x) Q_Y(y)} - \log |\mathrm{supp}(P_{XY})| \tag{102}$$

$$= \lim_{\alpha \downarrow 0} K_\alpha(X;Y), \tag{103}$$

where in the RHS of (102), *we use the conventions* (44). *The inequality can be strict, so* $\alpha \mapsto K_\alpha(X;Y)$ *need not be continuous at* $\alpha = 0$.

(Lemma 26) $K_1(X;Y) = I(X;Y)$.

(Lemma 27) *Let* $f\colon \{1, \ldots, |\mathcal{X}|\} \to \mathcal{X}$ *and* $g\colon \{1, \ldots, |\mathcal{Y}|\} \to \mathcal{Y}$ *be bijective functions, and let* B *be the* $|\mathcal{X}| \times |\mathcal{Y}|$ *matrix whose Row-i Column-j entry* $\mathsf{B}_{i,j}$ *equals* $P_{XY}(f(i), g(j))$. *Then,*

$$K_2(X;Y) = -2 \log \sigma_1(\mathsf{B}) - H_2(X,Y), \tag{104}$$

where $\sigma_1(\mathsf{B})$ *denotes the largest singular value of* B. *(Because the singular values of a matrix are invariant under row and column permutations, the result does not depend on* f *or* g*.)*

(Lemma 28) $K_\infty(X;Y) = 0$.

(Lemma 29) *The mapping* $\alpha \mapsto K_\alpha(X;Y)$ *need not be monotonic on* $[0, \infty]$.

(Lemma 30) *The mapping* $\alpha \mapsto K_\alpha(X;Y) + H_\alpha(X,Y)$ *is nonincreasing on* $[0, \infty]$.

(Lemma 31) *The mapping* $\alpha \mapsto K_\alpha(X;Y)$ *is continuous on* $(0, \infty]$. *(See Lemma 25 for the behavior at* $\alpha = 0$*.)*

(Lemma 32) *If* $X = Y$ *with probability one, then*

$$K_\alpha(X;Y) = \begin{cases} 2H_{\frac{\alpha}{2-\alpha}}(X) - H_\alpha(X) & \text{if } \alpha \in [0,2), \\ \frac{\alpha}{\alpha-1} H_\infty(X) - H_\alpha(X) & \text{if } \alpha \geq 2, \\ 0 & \text{if } \alpha = \infty. \end{cases} \tag{105}$$

(Lemma 33) *For every* $\alpha \in (0,2)$, *the mapping* $(Q_X, Q_Y) \mapsto \Delta_\alpha(P_{XY} \| Q_X Q_Y)$ *in the definition of* $K_\alpha(X;Y)$ *in* (2) *has a unique minimizer. This need not be the case when* $\alpha \in \{0\} \cup [2, \infty]$.

6. Proofs

In this section, we prove the properties of $J_\alpha(X;Y)$ and $K_\alpha(X;Y)$ stated in Section 5.

Lemma 1. *For every* $\alpha \in [0, \infty]$, *the minimum in the definition of* $J_\alpha(X;Y)$ *exists and is finite.*

Proof. Let $\alpha \in [0, \infty]$. Then $\inf_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X Q_Y)$ is finite because $D_\alpha(P_{XY} \| P_X P_Y)$ is finite and because the Rényi divergence is nonnegative. The minimum exists because the set $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ is compact and the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$ is continuous. □

Lemma 2. *For all* $\alpha \in [0, \infty]$, $J_\alpha(X;Y) \geq 0$. *If* $\alpha \in (0, \infty]$, *then* $J_\alpha(X;Y) = 0$ *if and only if* X *and* Y *are independent (nonnegativity).*

Proof. The nonnegativity follows from the definition of $J_\alpha(X;Y)$ because the Rényi divergence is nonnegative for $\alpha \in [0, \infty]$. If X and Y are independent, then $P_{XY} = P_X P_Y$, and the choice $Q_X = P_X$ and $Q_Y = P_Y$ in the definition of $J_\alpha(X;Y)$ achieves $J_\alpha(X;Y) = 0$. Conversely, if $J_\alpha(X;Y) = 0$, then there exist PMFs Q_X^* and Q_Y^* satisfying $D_\alpha(P_{XY} \| Q_X^* Q_Y^*) = 0$. If, in addition, $\alpha \in (0, \infty]$, then $P_{XY} = Q_X^* Q_Y^*$ by Proposition 4, and hence X and Y are independent. □

Lemma 3. *For all* $\alpha \in [0, \infty]$, $J_\alpha(X;Y) = J_\alpha(Y;X)$ *(symmetry).*

Proof. The definition of $J_\alpha(X;Y)$ is symmetric in X and Y. □

Lemma 4. *If $X \multimap Y \multimap Z$ form a Markov chain, then $J_\alpha(X;Z) \le J_\alpha(X;Y)$ for all $\alpha \in [0,\infty]$ (data-processing inequality).*

Proof. Let $X \multimap Y \multimap Z$ form a Markov chain, and let $\alpha \in [0,\infty]$. Let $\hat{Q}_X$ and $\hat{Q}_Y$ be PMFs that achieve the minimum in the definition of $J_\alpha(X;Y)$, so

$$J_\alpha(X;Y) = D_\alpha(P_{XY} \| \hat{Q}_X \hat{Q}_Y). \tag{106}$$

Define the PMF $\hat{Q}_Z$ as

$$\hat{Q}_Z(z) \triangleq \sum_y \hat{Q}_Y(y)\, P_{Z|Y}(z|y). \tag{107}$$

(As noted in the preliminaries, we define $P_{Z|Y}(z|y) \triangleq 1/|\mathcal{Z}|$ when $P_Y(y) = 0$.) We show below that

$$D_\alpha(P_{XZ} \| \hat{Q}_X \hat{Q}_Z) \le D_\alpha(P_{XY} \| \hat{Q}_X \hat{Q}_Y), \tag{108}$$

which implies the data-processing inequality because

$$\begin{aligned} J_\alpha(X;Z) &\le D_\alpha(P_{XZ} \| \hat{Q}_X \hat{Q}_Z) && (109)\\ &\le D_\alpha(P_{XY} \| \hat{Q}_X \hat{Q}_Y) && (110)\\ &= J_\alpha(X;Y), && (111) \end{aligned}$$

where (109) holds by the definition of $J_\alpha(X;Z)$; (110) follows from (108); and (111) follows from (106).

The proof of (108) is based on the data-processing inequality for the Rényi divergence. Define the conditional PMF $A_{X'Z'|XY}$ as

$$A_{X'Z'|XY}(x',z'|x,y) \triangleq \mathbb{1}\{x' = x\}\, P_{Z|Y}(z'|y). \tag{112}$$

If $(X,Y) \sim P_{XY}$, then the marginal distribution of X' and Z' is

$$\begin{aligned} (P_{XY} A_{X'Z'|XY})(x',z') &= \sum_{x,y} P_{XY}(x,y)\, A_{X'Z'|XY}(x',z'|x,y) && (113)\\ &= \sum_y P_{XY}(x',y)\, P_{Z|Y}(z'|y) && (114)\\ &= \sum_y P_{XY}(x',y)\, P_{Z|XY}(z'|x',y) && (115)\\ &= P_{XZ}(x',z'), && (116) \end{aligned}$$

where (114) follows from (112); and (115) holds because X, Y, and Z form a Markov chain. If $(X,Y) \sim \hat{Q}_X \hat{Q}_Y$, then the marginal distribution of X' and Z' is

$$\begin{aligned} (\hat{Q}_X \hat{Q}_Y A_{X'Z'|XY})(x',z') &= \sum_{x,y} \hat{Q}_X(x)\, \hat{Q}_Y(y)\, A_{X'Z'|XY}(x',z'|x,y) && (117)\\ &= \sum_y \hat{Q}_X(x')\, \hat{Q}_Y(y)\, P_{Z|Y}(z'|y) && (118)\\ &= \hat{Q}_X(x')\, \hat{Q}_Z(z'), && (119) \end{aligned}$$

where (118) follows from (112), and (119) follows from (107). Finally, we are ready to prove (108):

$$\begin{aligned} D_\alpha(P_{XZ} \| \hat{Q}_X \hat{Q}_Z) &= D_\alpha\big((P_{XY} A_{X'Z'|XY}) \| (\hat{Q}_X \hat{Q}_Y A_{X'Z'|XY})\big) && (120)\\ &\le D_\alpha(P_{XY} \| \hat{Q}_X \hat{Q}_Y), && (121) \end{aligned}$$

where (120) follows from (116) and (119), and where (121) follows from the data-processing inequality for the Rényi divergence [22] (Theorem 9). □

Lemma 5. $J_0(X;Y) = 0$.

Proof. By Lemma 2, $J_0(X;Y) \geq 0$, so it suffices to show that $J_0(X;Y) \leq 0$. Let $(\hat{x},\hat{y}) \in \mathcal{X} \times \mathcal{Y}$ satisfy $P_{XY}(\hat{x},\hat{y}) > 0$. Define the PMF $\hat{Q}_X$ as $\hat{Q}_X(x) \triangleq \mathbb{1}\{x = \hat{x}\}$ and the PMF $\hat{Q}_Y$ as $\hat{Q}_Y(y) \triangleq \mathbb{1}\{y = \hat{y}\}$. Then, $D_0(P_{XY} \| \hat{Q}_X \hat{Q}_Y) = 0$, so $J_0(X;Y) \leq 0$ by the definition of $J_0(X;Y)$. □

Lemma 6. *Let $f\colon \{1,\ldots,|\mathcal{X}|\} \to \mathcal{X}$ and $g\colon \{1,\ldots,|\mathcal{Y}|\} \to \mathcal{Y}$ be bijective functions, and let A be the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix whose Row-i Column-j entry $\mathsf{A}_{i,j}$ equals $\sqrt{P_{XY}(f(i),g(j))}$. Then,*

$$J_{\frac{1}{2}}(X;Y) = -2\log \sigma_1(\mathsf{A}), \tag{122}$$

where $\sigma_1(\mathsf{A})$ denotes the largest singular value of A. (Because the singular values of a matrix are invariant under row and column permutations, the result does not depend on f or g.)

Proof. By the definitions of $J_\alpha(X;Y)$ and the Rényi divergence,

$$J_{\frac{1}{2}}(X;Y) = -2\log \max_{Q_X,Q_Y} \sum_{x,y} \sqrt{Q_X(x)}\,\sqrt{P(x,y)}\,\sqrt{Q_Y(y)}. \tag{123}$$

The claim follows from (123) because

$$\max_{Q_X,Q_Y} \sum_{x,y} \sqrt{Q_X(x)}\,\sqrt{P(x,y)}\,\sqrt{Q_Y(y)} = \max_{\|\mathsf{u}\|_2=\|\mathsf{v}\|_2=1} \mathsf{u}^\mathsf{T}\mathsf{A}\mathsf{v} \tag{124}$$

$$= \max_{\|\mathsf{v}\|_2=1} \|\mathsf{A}\mathsf{v}\|_2 \tag{125}$$

$$= \sigma_1(\mathsf{A}), \tag{126}$$

where u and v are column vectors with $|\mathcal{X}|$ and $|\mathcal{Y}|$ elements, respectively; (124) is shown below; (125) follows from the Cauchy–Schwarz inequality $|\mathsf{u}^\mathsf{T}\mathsf{A}\mathsf{v}| \leq \|\mathsf{u}\|_2\,\|\mathsf{A}\mathsf{v}\|_2$, which holds with equality if u and $\mathsf{A}\mathsf{v}$ are linearly dependent; and (126) holds because the spectral norm of a matrix is equal to its largest singular value [31] (Example 5.6.6).

We now prove (124). Let u and v be vectors that satisfy $\|\mathsf{u}\|_2 = \|\mathsf{v}\|_2 = 1$, and define the PMFs $\hat{Q}_X$ and $\hat{Q}_Y$ as $\hat{Q}_X(x) \triangleq \mathsf{u}^2_{f^{-1}(x)}$ and $\hat{Q}_Y(y) \triangleq \mathsf{v}^2_{g^{-1}(y)}$, where f^{-1} and g^{-1} denote the inverse functions of f and g, respectively. Then,

$$\mathsf{u}^\mathsf{T}\mathsf{A}\mathsf{v} = \sum_{i,j} \mathsf{u}_i \mathsf{A}_{i,j} \mathsf{v}_j \tag{127}$$

$$\leq \sum_{i,j} |\mathsf{u}_i|\, \mathsf{A}_{i,j}\, |\mathsf{v}_j| \tag{128}$$

$$= \sum_{x,y} \sqrt{\hat{Q}_X(x)}\,\sqrt{P(x,y)}\,\sqrt{\hat{Q}_Y(y)} \tag{129}$$

$$\leq \max_{Q_X,Q_Y} \sum_{x,y} \sqrt{Q_X(x)}\,\sqrt{P(x,y)}\,\sqrt{Q_Y(y)}, \tag{130}$$

where (128) holds because all the entries of A are nonnegative, and in (129), we changed the summation variables to $x \triangleq f(i)$ and $y \triangleq g(j)$. It remains to show that equality can be achieved in (128) and (130). To that end, let Q^*_X and Q^*_Y be PMFs that achieve the maximum on the RHS of (130), and define the vectors u and v as $\mathsf{u}_i \triangleq Q^*_X(f(i))^{1/2}$ and $\mathsf{v}_j \triangleq Q^*_Y(g(j))^{1/2}$. Then, $\|\mathsf{u}\|_2 = \|\mathsf{v}\|_2 = 1$, and (128) and (130) hold with equality, which proves (124). □

Lemma 7. $J_1(X;Y) = I(X;Y)$.

Proof. This follows from Proposition 8 because $D_1(P_{XY} \| Q_X Q_Y)$ in the definition of $J_1(X;Y)$ is equal to $D(P_{XY} \| Q_X Q_Y)$. □

Lemma 8. *For all* $\alpha > 0$,

$$(1-\alpha) J_\alpha(X;Y) = \min_{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \big[(1-\alpha) D(R_{XY} \| R_X R_Y) + \alpha D(R_{XY} \| P_{XY}) \big]. \tag{131}$$

Thus, being the minimum of concave functions in α*, the mapping* $\alpha \mapsto (1-\alpha) J_\alpha(X;Y)$ *is concave on* $(0,\infty)$.

Proof. For $\alpha = 1$, (131) holds because $D(R_{XY} \| P_{XY}) \geq 0$ with equality if $R_{XY} = P_{XY}$. For $\alpha \in (0,1)$,

$$(1-\alpha) J_\alpha(X;Y) = \min_{Q_X, Q_Y} (1-\alpha) D_\alpha(P_{XY} \| Q_X Q_Y) \tag{132}$$

$$= \min_{Q_X, Q_Y} \min_{R_{XY}} \big[(1-\alpha) D(R_{XY} \| Q_X Q_Y) + \alpha D(R_{XY} \| P_{XY}) \big] \tag{133}$$

$$= \min_{R_{XY}} \big[(1-\alpha) D(R_{XY} \| R_X R_Y) + \alpha D(R_{XY} \| P_{XY}) \big], \tag{134}$$

where (132) holds by the definition of $J_\alpha(X;Y)$; (133) follows from [22] (Theorem 30); and (134) follows from Proposition 8 after swapping the minima.

For $\alpha > 1$, define the sets

$$\mathcal{Q} \triangleq \{ (Q_X, Q_Y) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) : \operatorname{supp}(Q_X Q_Y) = \mathcal{X} \times \mathcal{Y} \}, \tag{135}$$

$$\mathcal{R} \triangleq \{ R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \operatorname{supp}(R_{XY}) \subseteq \operatorname{supp}(P_{XY}) \}. \tag{136}$$

Then,

$$(1-\alpha) J_\alpha(X;Y) = \sup_{(Q_X,Q_Y) \in \mathcal{Q}} (1-\alpha) D_\alpha(P_{XY} \| Q_X Q_Y) \tag{137}$$

$$= \sup_{(Q_X,Q_Y) \in \mathcal{Q}} \min_{R_{XY} \in \mathcal{R}} \big[(1-\alpha) D(R_{XY} \| Q_X Q_Y) + \alpha D(R_{XY} \| P_{XY}) \big] \tag{138}$$

$$= \min_{R_{XY} \in \mathcal{R}} \sup_{(Q_X,Q_Y) \in \mathcal{Q}} \big[(1-\alpha) D(R_{XY} \| Q_X Q_Y) + \alpha D(R_{XY} \| P_{XY}) \big] \tag{139}$$

$$= \min_{R_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \big[(1-\alpha) D(R_{XY} \| R_X R_Y) + \alpha D(R_{XY} \| P_{XY}) \big], \tag{140}$$

where (137) follows from the definition of $J_\alpha(X;Y)$ because $1-\alpha < 0$ and because the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$ is continuous; (138) follows from [22] (Theorem 30); (139) follows from a minimax theorem and is justified below; and (140) follows from Proposition 8, a continuity argument, and the observation that $D(R_{XY} \| P_{XY})$ is infinite if $R_{XY} \notin \mathcal{R}$.

We now verify the conditions of Ky Fan's minimax theorem [32] (Theorem 2), which will establish (139). (We use Ky Fan's minimax theorem because it does not require that the set $\mathcal{Q}$ be compact, and having a noncompact set $\mathcal{Q}$ helps to guarantee that the function f defined next takes on finite values only. A brief proof of Ky Fan's minimax theorem appears in [33].) Let the function $f \colon \mathcal{R} \times \mathcal{Q} \to \mathbb{R}$ be defined by the expression in square brackets in (139), i.e.,

$$f(R_{XY}, Q_X, Q_Y) \triangleq (1-\alpha) D(R_{XY} \| Q_X Q_Y) + \alpha D(R_{XY} \| P_{XY}). \tag{141}$$

We check that

(i) the sets $\mathcal{Q}$ and $\mathcal{R}$ are convex;
(ii) the set $\mathcal{R}$ is compact;
(iii) the function f is real-valued;
(iv) for every $(Q_X, Q_Y) \in \mathcal{Q}$, the function f is continuous in R_{XY};
(v) for every $(Q_X, Q_Y) \in \mathcal{Q}$, the function f is convex in R_{XY}; and
(vi) for every $R_{XY} \in \mathcal{R}$, the function f is concave in the pair (Q_X, Q_Y).

Indeed, Parts (i) and (ii) are easy to see; Part (iii) holds because both relative entropies on the RHS of (141) are finite by our definitions of $\mathcal{Q}$ and $\mathcal{R}$; and to show Parts (iv)–(vi), we rewrite f as:

$$f(R_{XY}, Q_X, Q_Y) = -H(R_{XY}) - \alpha \sum_{x,y} R_{XY}(x,y) \log P(x,y) \\ + (\alpha - 1) \sum_x R_X(x) \log Q_X(x) + (\alpha - 1) \sum_y R_Y(y) \log Q_Y(y). \tag{142}$$

From (142), we see that Part (iv) holds by our definitions of $\mathcal{Q}$ and $\mathcal{R}$; Part (v) holds because the entropy is a concave function (so $-H(R_{XY})$ is convex), because linear functionals of R_{XY} are convex, and because the sum of convex functions is convex; and Part (vi) holds because the logarithm is a concave function and because a nonnegative weighted sum of concave functions is concave. (In Ky Fan's theorem, weaker conditions than Parts (i)–(vi) are required, but it is not difficult to see that Parts (i)–(vi) are sufficient.)

The last claim, namely, that the mapping $\alpha \mapsto (1-\alpha) J_\alpha(X;Y)$ is concave on $(0,\infty)$, is true because the expression in square brackets on the RHS of (131) is concave in α for every R_{XY} and because the pointwise minimum preserves the concavity. □

Lemma 9. *The mapping $\alpha \mapsto J_\alpha(X;Y)$ is nondecreasing on $[0,\infty]$.*

Proof. This is true because for every $\alpha, \alpha' \in [0,\infty]$ with $\alpha \le \alpha'$,

$$\min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X Q_Y) \le \min_{Q_X, Q_Y} D_{\alpha'}(P_{XY} \| Q_X Q_Y), \tag{143}$$

which holds because the Rényi divergence is nondecreasing in α (Proposition 4). □

Lemma 10. *The mapping $\alpha \mapsto J_\alpha(X;Y)$ is continuous on $[0,\infty]$.*

Proof. By Lemma 8, the mapping $\alpha \mapsto (1-\alpha) J_\alpha(X;Y)$ is concave on $(0,\infty)$, thus it is continuous on $(0,\infty)$, which implies that $\alpha \mapsto J_\alpha(X;Y)$ is continuous on $(0,1) \cup (1,\infty)$.

We next prove the continuity at $\alpha = 0$. Let Q_X^* and Q_Y^* be PMFs that achieve the minimum in the definition of $J_0(X;Y)$. Then, for all $\alpha \ge 0$,

$$D_0(P_{XY} \| Q_X^* Q_Y^*) = J_0(X;Y) \tag{144}$$

$$\le J_\alpha(X;Y) \tag{145}$$

$$\le D_\alpha(P_{XY} \| Q_X^* Q_Y^*), \tag{146}$$

where (145) holds because $\alpha \mapsto J_\alpha(X;Y)$ is nondecreasing (Lemma 9), and (146) holds by the definition of $J_\alpha(X;Y)$. The Rényi divergence is continuous in α (Proposition 4), so (144)–(146) and the sandwich theorem imply that $J_\alpha(X;Y)$ is continuous at $\alpha = 0$.

We continue with the continuity at $\alpha = \infty$. Define

$$\tau \triangleq \min_{(x,y) \in \operatorname{supp}(P_{XY})} P(x,y). \tag{147}$$

Then, for all $\alpha > 1$,

$$J_\infty(X;Y) \geq J_\alpha(X;Y) \tag{148}$$
$$= \min_{Q_X,Q_Y} \frac{1}{\alpha-1} \log \sum_{x,y} P(x,y) \frac{P(x,y)^{\alpha-1}}{[Q_X(x)\,Q_Y(y)]^{\alpha-1}} \tag{149}$$
$$\geq \min_{Q_X,Q_Y} \frac{1}{\alpha-1} \log \max_{x,y} \frac{\tau\, P(x,y)^{\alpha-1}}{[Q_X(x)\,Q_Y(y)]^{\alpha-1}} \tag{150}$$
$$= \frac{1}{\alpha-1} \log \tau + \min_{Q_X,Q_Y} \log \max_{x,y} \frac{P(x,y)}{Q_X(x)\,Q_Y(y)} \tag{151}$$
$$= \frac{1}{\alpha-1} \log \tau + J_\infty(X;Y), \tag{152}$$

where (148) holds because $\alpha \mapsto J_\alpha(X;Y)$ is nondecreasing (Lemma 9), and (149) and (152) hold by the definitions of $J_\alpha(X;Y)$ and the Rényi divergence. The RHS of (152) tends to $J_\infty(X;Y)$ as α tends to infinity, so $J_\alpha(X;Y)$ is continuous at $\alpha = \infty$ by the sandwich theorem.

It remains to show the continuity at $\alpha = 1$. Let $\alpha \in (\frac{3}{4}, 1) \cup (1, \frac{5}{4})$, and let $\delta \triangleq |1-\alpha| \in (0, \frac{1}{4})$. Then, for all PMFs Q_X and Q_Y,

$$2^{-\delta D_\alpha(P_{XY}\|Q_X Q_Y)} \leq 2^{-\delta D_{1-\delta}(P_{XY}\|Q_X Q_Y)} \tag{153}$$
$$= \sum_{x,y} P(x,y) \left[\frac{Q_X(x)\,Q_Y(y)}{P(x,y)} \right]^\delta \tag{154}$$
$$= \sum_{x,y} P(x,y) \left[\frac{P_X(x)\,P_Y(y)}{P(x,y)} \right]^\delta \left[\frac{Q_X(x)\,Q_Y(y)}{P_X(x)\,P_Y(y)} \right]^\delta \tag{155}$$
$$\leq \left\{ \sum_{x,y} P(x,y) \left[\frac{P_X(x)\,P_Y(y)}{P(x,y)} \right]^{2\delta} \right\}^{\frac{1}{2}} \cdot \left\{ \sum_{x,y} P(x,y) \left[\frac{Q_X(x)\,Q_Y(y)}{P_X(x)\,P_Y(y)} \right]^{2\delta} \right\}^{\frac{1}{2}} \tag{156}$$
$$\leq \left\{ \sum_{x,y} P(x,y) \left[\frac{P_X(x)\,P_Y(y)}{P(x,y)} \right]^{2\delta} \right\}^{\frac{1}{2}} \tag{157}$$
$$= 2^{-\delta D_{1-2\delta}(P_{XY}\|P_X P_Y)}, \tag{158}$$

where (153) holds because $1-\delta \leq \alpha$ and because the Rényi divergence is nondecreasing in α (Proposition 4); (156) follows from the Cauchy–Schwarz inequality; and (157) holds because

$$\left\{ \sum_{x,y} P(x,y) \left[\frac{Q_X(x)\,Q_Y(y)}{P_X(x)\,P_Y(y)} \right]^{2\delta} \right\}^{\frac{1}{2}} \leq \left\{ \sum_{x} P_X(x) \left[\frac{Q_X(x)}{P_X(x)} \right]^{4\delta} \right\}^{\frac{1}{4}} \cdot \left\{ \sum_{y} P_Y(y) \left[\frac{Q_Y(y)}{P_Y(y)} \right]^{4\delta} \right\}^{\frac{1}{4}} \tag{159}$$
$$= 2^{-\delta D_{1-4\delta}(P_X\|Q_X)} \cdot 2^{-\delta D_{1-4\delta}(P_Y\|Q_Y)} \tag{160}$$
$$\leq 1, \tag{161}$$

where (159) follows from the Cauchy–Schwarz inequality, and (161) holds because $1-4\delta > 0$ and because the Rényi divergence is nonnegative for positive orders (Proposition 4). Thus, for all $\alpha \in (\frac{3}{4}, \frac{5}{4})$,

$$D_{1-2|1-\alpha|}(P_{XY}\|P_X P_Y) \leq \min_{Q_X,Q_Y} D_\alpha(P_{XY}\|Q_X Q_Y) \tag{162}$$
$$= J_\alpha(X;Y) \tag{163}$$
$$\leq D_\alpha(P_{XY}\|P_X P_Y), \tag{164}$$

where (162) follows from (158) if $\alpha \neq 1$ and from Proposition 8 if $\alpha = 1$; and (164) holds by the definition of $J_\alpha(X;Y)$. The Rényi divergence is continuous in α (Proposition 4), thus (162)–(164) and the sandwich theorem imply that $J_\alpha(X;Y)$ is continuous at $\alpha = 1$. □

Lemma 11. *If $X = Y$ with probability one, then*

$$J_\alpha(X;Y) = \begin{cases} \frac{\alpha}{1-\alpha} H_\infty(X) & \text{if } \alpha \in [0, \frac{1}{2}], \\ H_{\frac{\alpha}{2\alpha-1}}(X) & \text{if } \alpha > \frac{1}{2}, \\ H_{\frac{1}{2}}(X) & \text{if } \alpha = \infty. \end{cases} \tag{165}$$

Proof. We show below that (165) holds for $\alpha \in (0,1) \cup (1,\infty)$. Thus, (165) holds also for $\alpha \in \{0,1,\infty\}$ because both its sides are continuous in α: its LHS by Lemma 10, and its RHS by the continuity of the Rényi entropy (Proposition 3).

Fix $\alpha \in (0,1) \cup (1,\infty)$. Then,

$$J_\alpha(X;Y) = \min_{Q_X} \min_{Q_Y} D_\alpha(P_{XY} \| Q_X Q_Y) \tag{166}$$

$$= \min_{Q_X} \frac{\alpha}{\alpha-1} \log \sum_y \left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{167}$$

$$= \min_{Q_X} \frac{\alpha}{\alpha-1} \log \sum_x P_X(x) Q_X(x)^{\frac{1-\alpha}{\alpha}}, \tag{168}$$

where (167) follows from Proposition 9, and (168) holds because

$$P_{XY}(x,y) = \begin{cases} P_X(x) & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \tag{169}$$

First consider the case $\alpha > \frac{1}{2}$. Define $\gamma \triangleq \sum_x P_X(x)^{\frac{\alpha}{2\alpha-1}}$. Then, for all $Q_X \in \mathcal{P}(\mathcal{X})$,

$$\frac{\alpha}{\alpha-1} \log \sum_x P_X(x) Q_X(x)^{\frac{1-\alpha}{\alpha}} = \frac{\alpha}{\alpha-1} \log \sum_x \left[\gamma \gamma^{-1} P_X(x)^{\frac{\alpha}{2\alpha-1}}\right]^{\frac{2\alpha-1}{\alpha}} Q_X(x)^{\frac{1-\alpha}{\alpha}} \tag{170}$$

$$= \frac{2\alpha-1}{\alpha-1} \log \gamma + D_{\frac{2\alpha-1}{\alpha}}(\gamma^{-1} P_X^{\frac{\alpha}{2\alpha-1}} \| Q_X) \tag{171}$$

$$= H_{\frac{\alpha}{2\alpha-1}}(X) + D_{\frac{2\alpha-1}{\alpha}}(\gamma^{-1} P_X^{\frac{\alpha}{2\alpha-1}} \| Q_X), \tag{172}$$

where (171) holds because $x \mapsto \gamma^{-1} P_X(x)^{\frac{\alpha}{2\alpha-1}}$ is a PMF. Because $\frac{2\alpha-1}{\alpha} > 0$, Proposition 4 implies that $D_{(2\alpha-1)/\alpha}(P\|Q) \geq 0$ with equality if $Q = P$. This together with (168) and (172) establishes (165).

Now consider the case $\alpha \in (0, \frac{1}{2}]$. For all $Q_X \in \mathcal{P}(\mathcal{X})$,

$$\sum_x P_X(x) Q_X(x)^{\frac{1-\alpha}{\alpha}} \leq \sum_x P_X(x) Q_X(x) \tag{173}$$

$$\leq \sum_x \left[\max_{x'} P_X(x')\right] Q_X(x) \tag{174}$$

$$= \max_x P_X(x), \tag{175}$$

where (173) holds because $Q_X(x) \in [0,1]$ for all $x \in \mathcal{X}$ and because $\frac{1-\alpha}{\alpha} \geq 1$. The inequalities (173) and (174) both hold with equality when $Q_X(x) = \mathbb{1}\{x = x^*\}$, where $x^* \in \mathcal{X}$ is such that $P_X(x^*) = \max_x P_X(x)$. Thus,

$$\max_{Q_X} \sum_x P_X(x) Q_X(x)^{\frac{1-\alpha}{\alpha}} = \max_x P_X(x). \tag{176}$$

Now (165) follows:

$$\begin{aligned} J_\alpha(X;Y) &= \min_{Q_X} \frac{\alpha}{\alpha-1} \log \sum_x P_X(x) Q_X(x)^{\frac{1-\alpha}{\alpha}} && (177) \\ &= \frac{\alpha}{\alpha-1} \log \max_{Q_X} \sum_x P_X(x) Q_X(x)^{\frac{1-\alpha}{\alpha}} && (178) \\ &= \frac{\alpha}{\alpha-1} \log \max_x P_X(x) && (179) \\ &= \frac{\alpha}{1-\alpha} H_\infty(X), && (180) \end{aligned}$$

where (177) follows from (168); (178) holds because $\frac{\alpha}{\alpha-1} < 0$; (179) follows from (176); and (180) follows from the definition of $H_\infty(X)$. □

Lemma 12. *If the pairs (X_1, Y_1) and (X_2, Y_2) are independent, then $J_\alpha(X_1, X_2; Y_1, Y_2) = J_\alpha(X_1; Y_1) + J_\alpha(X_2; Y_2)$ for all $\alpha \in [0, \infty]$ (additivity).*

Proof. Let the pairs (X_1, Y_1) and (X_2, Y_2) be independent. For $\alpha \in (0,1) \cup (1, \infty)$, we establish the lemma by showing the following two inequalities:

$$J_\alpha(X_1, X_2; Y_1, Y_2) \leq J_\alpha(X_1; Y_1) + J_\alpha(X_2; Y_2), \tag{181}$$

$$J_\alpha(X_1, X_2; Y_1, Y_2) \geq J_\alpha(X_1; Y_1) + J_\alpha(X_2; Y_2). \tag{182}$$

Because $J_\alpha(X;Y)$ is continuous in α (Lemma 10), this will also establish the lemma for $\alpha \in \{0, 1, \infty\}$.

To show (181), let $Q^*_{X_1}$ and $Q^*_{Y_1}$ be PMFs that achieve the minimum in the definition of $J_\alpha(X_1; Y_1)$, and let $Q^*_{X_2}$ and $Q^*_{Y_2}$ be PMFs that achieve the minimum in the definition of $J_\alpha(X_2; Y_2)$, so

$$J_\alpha(X_1; Y_1) = D_\alpha(P_{X_1 Y_1} \| Q^*_{X_1} Q^*_{Y_1}), \tag{183}$$

$$J_\alpha(X_2; Y_2) = D_\alpha(P_{X_2 Y_2} \| Q^*_{X_2} Q^*_{Y_2}). \tag{184}$$

Then, (181) holds because

$$\begin{aligned} J_\alpha(X_1, X_2; Y_1, Y_2) &\leq D_\alpha(P_{X_1 X_2 Y_1 Y_2} \| Q^*_{X_1} Q^*_{X_2} Q^*_{Y_1} Q^*_{Y_2}) && (185) \\ &= D_\alpha(P_{X_1 Y_1} \| Q^*_{X_1} Q^*_{Y_1}) + D_\alpha(P_{X_2 Y_2} \| Q^*_{X_2} Q^*_{Y_2}) && (186) \\ &= J_\alpha(X_1; Y_1) + J_\alpha(X_2; Y_2), && (187) \end{aligned}$$

where (185) holds by the definition of $J_\alpha(X_1, X_2; Y_1, Y_2)$ as a minimum; (186) follows from a simple computation using the independence hypothesis $P_{X_1 X_2 Y_1 Y_2} = P_{X_1 Y_1} P_{X_2 Y_2}$; and (187) follows from (183) and (184).

To establish (182), we consider the cases $\alpha > 1$ and $\alpha < 1$ separately, starting with $\alpha > 1$. Let $\hat{Q}_{X_1 X_2}$ and $\hat{Q}_{Y_1 Y_2}$ be PMFs that achieve the minimum in the definition of $J_\alpha(X_1, X_2; Y_1, Y_2)$, so

$$J_\alpha(X_1, X_2; Y_1, Y_2) = D_\alpha(P_{X_1 X_2 Y_1 Y_2} \| \hat{Q}_{X_1 X_2} \hat{Q}_{Y_1 Y_2}). \tag{188}$$

Define the function $f\colon \mathcal{X}_1 \times \mathcal{Y}_1 \to \mathbb{R} \cup \{\infty\}$ as

$$f(x_1, y_1) \triangleq \sum_{x_2, y_2} P_{X_2Y_2}(x_2, y_2)^{\alpha} \left[\hat{Q}_{X_2|X_1}(x_2|x_1)\, \hat{Q}_{Y_2|Y_1}(y_2|y_1) \right]^{1-\alpha}, \tag{189}$$

and let $(x_1', y_1') \in \mathcal{X}_1 \times \mathcal{Y}_1$ be such that

$$f(x_1', y_1') = \min_{x_1, y_1} f(x_1, y_1). \tag{190}$$

Define the PMFs Q'_{X_2} and Q'_{Y_2} as

$$Q'_{X_2}(x_2) \triangleq \hat{Q}_{X_2|X_1}(x_2|x_1'), \tag{191}$$
$$Q'_{Y_2}(y_2) \triangleq \hat{Q}_{Y_2|Y_1}(y_2|y_1'). \tag{192}$$

Then,

$$2^{(\alpha-1) J_\alpha(X_1, X_2; Y_1, Y_2)} = 2^{(\alpha-1) D_\alpha(P_{X_1X_2Y_1Y_2} \| \hat{Q}_{X_1X_2} \hat{Q}_{Y_1Y_2})} \tag{193}$$
$$= \sum_{x_1, x_2, y_1, y_2} \left[P_{X_1Y_1}(x_1, y_1)\, P_{X_2Y_2}(x_2, y_2) \right]^{\alpha} \left[\hat{Q}_{X_1X_2}(x_1, x_2)\, \hat{Q}_{Y_1Y_2}(y_1, y_2) \right]^{1-\alpha} \tag{194}$$
$$= \sum_{x_1, y_1} P_{X_1Y_1}(x_1, y_1)^{\alpha} \left[\hat{Q}_{X_1}(x_1)\, \hat{Q}_{Y_1}(y_1) \right]^{1-\alpha} f(x_1, y_1) \tag{195}$$
$$\geq \sum_{x_1, y_1} P_{X_1Y_1}(x_1, y_1)^{\alpha} \left[\hat{Q}_{X_1}(x_1)\, \hat{Q}_{Y_1}(y_1) \right]^{1-\alpha} f(x_1', y_1') \tag{196}$$
$$= 2^{(\alpha-1) D_\alpha(P_{X_1Y_1} \| \hat{Q}_{X_1} \hat{Q}_{Y_1}) + (\alpha-1) D_\alpha(P_{X_2Y_2} \| Q'_{X_2} Q'_{Y_2})}, \tag{197}$$

where (193) follows from (188); (194) holds by the independence hypothesis $P_{X_1X_2Y_1Y_2} = P_{X_1Y_1} P_{X_2Y_2}$; (195) follows from (189); (196) follows from (190); and (197) follows from (191) and (192). Taking the logarithm and multiplying by $\frac{1}{\alpha-1} > 0$ establishes (182):

$$J_\alpha(X_1, X_2; Y_1, Y_2) \geq D_\alpha(P_{X_1Y_1} \| \hat{Q}_{X_1} \hat{Q}_{Y_1}) + D_\alpha(P_{X_2Y_2} \| Q'_{X_2} Q'_{Y_2}) \tag{198}$$
$$\geq J_\alpha(X_1; Y_1) + J_\alpha(X_2; Y_2), \tag{199}$$

where (199) holds by the definition of $J_\alpha(X_1; Y_1)$ and $J_\alpha(X_2; Y_2)$.

The proof of (182) for $\alpha \in (0,1)$ is essentially the same as for $\alpha > 1$: Replace the minimum in (190) by a maximum. Inequality (196) is then reversed, but (198) continues to hold because $\frac{1}{\alpha-1} < 0$. Inequality (199) also continues to hold, and (198) and (199) together imply (182). □

Lemma 13. *For all $\alpha \in [0, \infty]$, $J_\alpha(X; Y) \leq \log |\mathcal{X}|$ with equality if and only if $\big(\alpha \in [\frac{1}{2}, \infty]$, X is distributed uniformly over $\mathcal{X}$, and $H(X|Y) = 0\big)$.*

Proof. Throughout the proof, define $X' \triangleq X$. We first show that $J_\alpha(X; Y) \leq \log |\mathcal{X}|$ for all $\alpha \in [0, \infty]$:

$$J_\alpha(X; Y) \leq J_\alpha(X; X') \tag{200}$$
$$\leq J_\infty(X; X') \tag{201}$$
$$= H_{\frac{1}{2}}(X) \tag{202}$$
$$\leq \log |\mathcal{X}|, \tag{203}$$

where (200) follows from the data-processing inequality (Lemma 4) because $X \multimap X' \multimap Y$ form a Markov chain; (201) holds because $J_\alpha(X; X')$ is nondecreasing in α (Lemma 9); (202) follows from Lemma 11; and (203) follows from Proposition 3.

We now show that (200)–(203) can hold with equality only if the following conditions all hold:

(i) $\alpha \in [\frac{1}{2}, \infty]$;
(ii) X is distributed uniformly over $\mathcal{X}$; and
(iii) $H(X|Y) = 0$, i.e., for every $y \in \operatorname{supp}(P_Y)$, there exists an $x \in \mathcal{X}$ for which $P(x|y) = 1$.

Indeed, if $\alpha < \frac{1}{2}$, then Lemma 11 implies that

$$J_\alpha(X; X') = \frac{\alpha}{1-\alpha} H_\infty(X). \tag{204}$$

Because $\frac{\alpha}{1-\alpha} < 1$ for such α's and because $H_\infty(X) \leq \log|\mathcal{X}|$ (Proposition 3), the RHS of (204) is strictly smaller than $\log|\mathcal{X}|$. This, together with (200), shows that Part (i) is a necessary condition. The necessity of Part (ii) follows from (203): if X is not distributed uniformly over $\mathcal{X}$, then (203) holds with strict inequality (Proposition 3). As to the necessity of Part (iii),

$$\begin{aligned}
J_\alpha(X;Y) &\leq J_\infty(X;Y) && (205)\\
&= \min_{Q_X} \min_{Q_Y} D_\infty(P_{XY} \| Q_X Q_Y) && (206)\\
&= \min_{Q_X} \log \sum_y \max_x \frac{P(x,y)}{Q_X(x)} && (207)\\
&\leq \log \sum_y \max_x \frac{P(y)\,P(x|y)}{1/|\mathcal{X}|} && (208)\\
&= \log|\mathcal{X}| + \log \sum_y P(y) \max_x P(x|y) && (209)\\
&\leq \log|\mathcal{X}|, && (210)
\end{aligned}$$

where (205) holds because $J_\alpha(X;Y)$ is nondecreasing in α (Lemma 9); (207) follows from Proposition 9; and (208) follows from choosing Q_X to be the uniform distribution. The inequality (210) is strict when Part (iii) does not hold, so Part (iii) is a necessary condition.

It remains to show that when Parts (i)–(iii) all hold, $J_\alpha(X;Y) = \log|\mathcal{X}|$. By (203), $J_\alpha(X;Y) \leq \log|\mathcal{X}|$ always holds, so it suffices to show that Parts (i)–(iii) together imply $J_\alpha(X;Y) \geq \log|\mathcal{X}|$. Indeed,

$$\begin{aligned}
J_\alpha(X;Y) &\geq J_{\frac{1}{2}}(X;Y) && (211)\\
&\geq J_{\frac{1}{2}}(X;X') && (212)\\
&= H_\infty(X) && (213)\\
&= \log|\mathcal{X}|, && (214)
\end{aligned}$$

where (211) holds because Part (i) implies that $\alpha \geq \frac{1}{2}$ and because $J_\alpha(X;Y)$ is nondecreasing in α (Lemma 9); (212) follows from the data-processing inequality (Lemma 4) because Part (iii) implies that $X \multimap Y \multimap X'$ form a Markov chain; (213) follows from Lemma 11; and (214) follows from Part (ii). □

Lemma 14. *For every $\alpha \in [1, \infty]$, $J_\alpha(X;Y)$ is concave in P_X for fixed $P_{Y|X}$.*

Proof. We prove the claim for $\alpha \in (1, \infty)$; for $\alpha \in \{1, \infty\}$ the claim will then hold because $J_\alpha(X;Y)$ is continuous in α (Lemma 10).

Fix $\alpha \in (1, \infty)$. Let $\lambda, \lambda' \in [0, 1]$ with $\lambda + \lambda' = 1$, let P_X and P'_X be PMFs, let $P_{Y|X}$ be a conditional PMF, and define $f \colon \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R} \cup \{\infty\}$ as

$$f(x, Q_Y) \triangleq \left[\sum_y P_{Y|X}(y|x)^\alpha \, Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}}. \tag{215}$$

Denoting $J_\alpha(X;Y)$ by $J_\alpha(P_X P_{Y|X})$,

$$\begin{aligned}
&J_\alpha\big((\lambda P_X + \lambda' P'_X) P_{Y|X}\big) \\
&\quad = \min_{Q_Y} \min_{Q_X} D_\alpha\big((\lambda P_X + \lambda' P'_X) P_{Y|X} \| Q_X Q_Y\big) &(216)\\
&\quad = \min_{Q_Y} \frac{\alpha}{\alpha-1} \log \sum_x \left[\sum_y [\lambda P_X(x) + \lambda' P'_X(x)]^\alpha P_{Y|X}(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} &(217)\\
&\quad = \min_{Q_Y} \frac{\alpha}{\alpha-1} \log \sum_x [\lambda P_X(x) + \lambda' P'_X(x)] \left[\sum_y P_{Y|X}(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} &(218)\\
&\quad = \min_{Q_Y} \frac{\alpha}{\alpha-1} \log \left[\lambda \sum_x P_X(x) f(x, Q_Y) + \lambda' \sum_x P'_X(x) f(x, Q_Y) \right] &(219)\\
&\quad \geq \min_{Q_Y} \frac{\alpha}{\alpha-1} \left[\lambda \log \sum_x P_X(x) f(x, Q_Y) + \lambda' \log \sum_x P'_X(x) f(x, Q_Y) \right] &(220)\\
&\quad \geq \lambda \min_{Q_Y} \frac{\alpha}{\alpha-1} \log \sum_x P_X(x) f(x, Q_Y) + \lambda' \min_{Q_Y} \frac{\alpha}{\alpha-1} \log \sum_x P'_X(x) f(x, Q_Y) &(221)\\
&\quad = \lambda J_\alpha(P_X P_{Y|X}) + \lambda' J_\alpha(P'_X P_{Y|X}), &(222)
\end{aligned}$$

where (217) follows from Proposition 9 with the roles of Q_X and Q_Y swapped; (220) holds because $\log(\cdot)$ is concave; (221) holds because optimizing Q_Y separately cannot be worse than optimizing a common Q_Y; and (222) can be established using steps similar to (216)–(218). □

Lemma 15. *For every $\alpha \in [\frac{1}{2}, \infty]$, the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$ is convex, i.e., for all $\lambda, \lambda' \in [0, 1]$ with $\lambda + \lambda' = 1$, all $Q_X, Q'_X \in \mathcal{P}(\mathcal{X})$, and all $Q_Y, Q'_Y \in \mathcal{P}(\mathcal{Y})$,*

$$D_\alpha\big(P_{XY} \| (\lambda Q_X + \lambda' Q'_X)(\lambda Q_Y + \lambda' Q'_Y)\big) \leq \lambda D_\alpha(P_{XY} \| Q_X Q_Y) + \lambda' D_\alpha(P_{XY} \| Q'_X Q'_Y). \tag{223}$$

For $\alpha \in [0, \frac{1}{2})$, the mapping need not be convex.

Proof. We establish (223) for $\alpha \in [\frac{1}{2}, 1)$ and for $\alpha \in (1, \infty)$, which also establishes (223) for $\alpha \in \{1, \infty\}$ because the Rényi divergence is continuous in α (Proposition 4). Afterwards, we provide an example where (223) is violated for all $\alpha \in [0, \frac{1}{2})$.

We begin with the case where $\alpha \in [\frac{1}{2}, 1)$:

$$\begin{aligned}
&2^{(\alpha-1)\lambda D_\alpha(P_{XY} \| Q_X Q_Y) + (\alpha-1)\lambda' D_\alpha(P_{XY} \| Q'_X Q'_Y)} \\
&\quad = \left[\sum_{x,y} P(x,y)^\alpha [Q_X(x) Q_Y(y)]^{1-\alpha} \right]^\lambda \cdot \left[\sum_{x,y} P(x,y)^\alpha [Q'_X(x) Q'_Y(y)]^{1-\alpha} \right]^{\lambda'} &(224)\\
&\quad \leq \lambda \sum_{x,y} P(x,y)^\alpha [Q_X(x) Q_Y(y)]^{1-\alpha} + \lambda' \sum_{x,y} P(x,y)^\alpha [Q'_X(x) Q'_Y(y)]^{1-\alpha} &(225)\\
&\quad = \sum_{x,y} P(x,y)^\alpha \left[\sqrt{\lambda} Q_X(x)^{1-\alpha} \sqrt{\lambda} Q_Y(y)^{1-\alpha} + \sqrt{\lambda'} Q'_X(x)^{1-\alpha} \sqrt{\lambda'} Q'_Y(y)^{1-\alpha} \right] &(226)\\
&\quad \leq \sum_{x,y} P(x,y)^\alpha \sqrt{\lambda Q_X(x)^{2(1-\alpha)} + \lambda' Q'_X(x)^{2(1-\alpha)}} \sqrt{\lambda Q_Y(y)^{2(1-\alpha)} + \lambda' Q'_Y(y)^{2(1-\alpha)}} &(227)
\end{aligned}$$

$$\leq \sum_{x,y} P(x,y)^\alpha \left[\lambda Q_X(x) + \lambda' Q'_X(x)\right]^{1-\alpha} \sqrt{\lambda Q_Y(y)^{2(1-\alpha)} + \lambda' Q'_Y(y)^{2(1-\alpha)}} \tag{228}$$

$$\leq \sum_{x,y} P(x,y)^\alpha \left[\lambda Q_X(x) + \lambda' Q'_X(x)\right]^{1-\alpha} \left[\lambda Q_Y(y) + \lambda' Q'_Y(y)\right]^{1-\alpha} \tag{229}$$

$$= 2^{(\alpha-1)D_\alpha(P_{XY}\|(\lambda Q_X+\lambda' Q'_X)(\lambda Q_Y+\lambda' Q'_Y))}, \tag{230}$$

where (225) follows from the arithmetic mean-geometric mean inequality; (227) follows from the Cauchy–Schwarz inequality; and (228) and (229) hold because the mapping $z \mapsto z^{2(1-\alpha)}$ is concave on $\mathbb{R}_{\geq 0}$ for $\alpha \in [\frac{1}{2}, 1)$. Taking the logarithm and multiplying by $\frac{1}{\alpha-1} < 0$ establishes (223).

Now, consider $\alpha \in (1, \infty)$. Then,

$$2^{(\alpha-1)D_\alpha(P_{XY}\|(\lambda Q_X+\lambda' Q'_X)(\lambda Q_Y+\lambda' Q'_Y))}$$

$$= \sum_{x,y} P(x,y)^\alpha \left[\lambda Q_X(x) + \lambda' Q'_X(x)\right]^{1-\alpha} \left[\lambda Q_Y(y) + \lambda' Q'_Y(y)\right]^{1-\alpha} \tag{231}$$

$$\leq \sum_{x,y} P(x,y)^\alpha \left[Q_X(x)^\lambda Q'_X(x)^{\lambda'}\right]^{1-\alpha} \left[Q_Y(y)^\lambda Q'_Y(y)^{\lambda'}\right]^{1-\alpha} \tag{232}$$

$$= \sum_{x,y} P(x,y)^\alpha \left[Q_X(x) Q_Y(y)\right]^{(1-\alpha)\lambda} \left[Q'_X(x) Q'_Y(y)\right]^{(1-\alpha)\lambda'} \tag{233}$$

$$\leq \left[\sum_{x,y} P(x,y)^\alpha \left[Q_X(x) Q_Y(y)\right]^{1-\alpha}\right]^\lambda \cdot \left[\sum_{x,y} P(x,y)^\alpha \left[Q'_X(x) Q'_Y(y)\right]^{1-\alpha}\right]^{\lambda'} \tag{234}$$

$$= 2^{(\alpha-1)\lambda D_\alpha(P_{XY}\|Q_X Q_Y) + (\alpha-1)\lambda' D_\alpha(P_{XY}\|Q'_X Q'_Y)}, \tag{235}$$

where (232) follows from the arithmetic mean-geometric mean inequality and the fact that the mapping $z \mapsto z^{1-\alpha}$ is decreasing on $\mathbb{R}_{>0}$ for $\alpha > 1$, and (234) follows from Hölder's inequality. Taking the logarithm and multiplying by $\frac{1}{\alpha-1} > 0$ establishes (223).

Finally, we show that the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY}\|Q_X Q_Y)$ does not need to be convex for $\alpha \in [0, \frac{1}{2})$. Let X be uniformly distributed over $\{0, 1\}$, and let $Y = X$. Then, for all $\alpha \in [0, \frac{1}{2})$,

$$D_\alpha\big(P_{XY}\|(0.5, 0.5)(0.5, 0.5)\big) > 0.5 D_\alpha\big(P_{XY}\|(1, 0)(1, 0)\big) + 0.5 D_\alpha\big(P_{XY}\|(0, 1)(0, 1)\big), \tag{236}$$

because the LHS of (236) is equal to $\log 2$, and the RHS of (236) is equal to $\frac{\alpha}{1-\alpha} \log 2$. □

Lemma 16. *Let $\alpha \in (0, 1) \cup (1, \infty)$. If (Q^*_X, Q^*_Y) achieves the minimum in the definition of $J_\alpha(X; Y)$, then there exist positive normalization constants c and d such that*

$$Q^*_X(x) = c \left[\sum_y P(x,y)^\alpha Q^*_Y(y)^{1-\alpha}\right]^{\frac{1}{\alpha}} \quad \forall x \in \mathcal{X}, \tag{237}$$

$$Q^*_Y(y) = d \left[\sum_x P(x,y)^\alpha Q^*_X(x)^{1-\alpha}\right]^{\frac{1}{\alpha}} \quad \forall y \in \mathcal{Y}, \tag{238}$$

with the conventions of (44). The case $\alpha = \infty$ is similar: if (Q^*_X, Q^*_Y) achieves the minimum in the definition of $J_\infty(X; Y)$, then there exist positive normalization constants c and d such that

$$Q^*_X(x) = c \max_y \frac{P(x,y)}{Q^*_Y(y)} \quad \forall x \in \mathcal{X}, \tag{239}$$

$$Q^*_Y(y) = d \max_x \frac{P(x,y)}{Q^*_X(x)} \quad \forall y \in \mathcal{Y}, \tag{240}$$

with the conventions of (44). *(If* $\alpha = 1$*, then* $Q_X^* = P_X$ *and* $Q_Y^* = P_Y$ *by Proposition* 8.*) Thus, for all* $\alpha \in (0, \infty]$*, both inclusions* $\operatorname{supp}(Q_X^*) \subseteq \operatorname{supp}(P_X)$ *and* $\operatorname{supp}(Q_Y^*) \subseteq \operatorname{supp}(P_Y)$ *hold.*

Proof. If (Q_X^*, Q_Y^*) achieves the minimum in the definition of $J_\alpha(X;Y)$, then

$$\min_{Q_Y} D_\alpha(P_{XY} \| Q_X^* Q_Y) = D_\alpha(P_{XY} \| Q_X^* Q_Y^*). \tag{241}$$

Hence, (238) and (240) follow from (76) and (78) of Proposition 9 because $D_\alpha(P_{XY} \| Q_X^* Q_Y^*) = J_\alpha(X;Y)$ is finite. Swapping the roles of Q_X and Q_Y establishes (237) and (239). For $\alpha \in (0,1) \cup (1,\infty)$ the claimed inclusions follow from (237) and (238); for $\alpha = \infty$ from (239) and (240); and for $\alpha = 1$ from Proposition 8. □

Lemma 17. *For all* $\alpha \in (0, \infty]$*,*

$$J_\alpha(X;Y) = \min_{Q_X} \phi_\alpha(Q_X), \tag{242}$$

where $\phi_\alpha(Q_X)$ *is defined as*

$$\phi_\alpha(Q_X) \triangleq \min_{Q_Y} D_\alpha(P_{XY} \| Q_X Q_Y) \tag{243}$$

and is given explicitly as follows: for $\alpha \in (0,1) \cup (1,\infty)$*,*

$$\phi_\alpha(Q_X) = \frac{\alpha}{\alpha - 1} \log \sum_y \left[\sum_x P(x,y)^\alpha \, Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}}, \tag{244}$$

with the conventions of (44)*; and for* $\alpha \in \{1, \infty\}$*,*

$$\phi_1(Q_X) = D(P_{XY} \| Q_X P_Y), \tag{245}$$

$$\phi_\infty(Q_X) = \log \sum_y \max_x \frac{P(x,y)}{Q_X(x)}, \tag{246}$$

with the conventions of (44)*. For every* $\alpha \in [\frac{1}{2}, \infty]$*, the mapping* $Q_X \mapsto \phi_\alpha(Q_X)$ *is convex. For* $\alpha \in (0, \frac{1}{2})$*, the mapping need not be convex.*

Proof. We first establish (242) and (244)–(246): (242) follows from the definition of $J_\alpha(X;Y)$; (244) and (246) follow from Proposition 9; and (245) holds because

$$\min_{Q_Y} D(P_{XY} \| Q_X Q_Y) = \min_{Q_Y} [D(P_{XY} \| Q_X P_Y) + D(P_Y \| Q_Y)] \tag{247}$$

$$= D(P_{XY} \| Q_X P_Y), \tag{248}$$

where (247) follows from a simple computation, and (248) holds because $D(P_Y \| Q_Y) \geq 0$ with equality if $Q_Y = P_Y$.

We now show that the mapping $Q_X \mapsto \phi_\alpha(Q_X)$ is convex for every $\alpha \in [\frac{1}{2}, \infty]$. To that end, let $\alpha \in [\frac{1}{2}, \infty]$, let $\lambda, \lambda' \in [0,1]$ with $\lambda + \lambda' = 1$, and let $Q_X, Q_X' \in \mathcal{P}(\mathcal{X})$. Let $\hat{Q}_Y$ and $\hat{Q}_Y'$ be PMFs that achieve the minimum in the definitions of $\phi_\alpha(Q_X)$ and $\phi_\alpha(Q_X')$, respectively. Then,

$$\phi_\alpha(\lambda Q_X + \lambda' Q_X') \leq D_\alpha\big(P_{XY} \| (\lambda Q_X + \lambda' Q_X')(\lambda \hat{Q}_Y + \lambda' \hat{Q}_Y')\big) \tag{249}$$

$$\leq \lambda D_\alpha(P_{XY} \| Q_X \hat{Q}_Y) + \lambda' D_\alpha(P_{XY} \| Q_X' \hat{Q}_Y') \tag{250}$$

$$= \lambda \phi_\alpha(Q_X) + \lambda' \phi_\alpha(Q_X'), \tag{251}$$

where (249) holds by the definition of $\phi_\alpha(\cdot)$; (250) holds because $D_\alpha(P_{XY}\|Q_XQ_Y)$ is convex in the pair (Q_X, Q_Y) for $\alpha \in [\frac{1}{2}, \infty]$ (Lemma 15); and (251) follows from our choice of $\hat{Q}_Y$ and $\hat{Q}'_Y$.

Finally, we show that the mapping $Q_X \mapsto \phi_\alpha(Q_X)$ need not be convex for $\alpha \in (0, \frac{1}{2})$. Let X be uniformly distributed over $\{0,1\}$, and let $Y = X$. Then, for all $\alpha \in (0, \frac{1}{2})$,

$$\phi_\alpha\big((0.5, 0.5)\big) > 0.5\phi_\alpha\big((1,0)\big) + 0.5\phi_\alpha\big((0,1)\big), \tag{252}$$

because the LHS of (252) is equal to $\log 2$, and the RHS of (252) is equal to $\frac{\alpha}{1-\alpha}\log 2$. □

Lemma 18. *For all* $\alpha \in (0,1) \cup (1,\infty]$,

$$J_\alpha(X;Y) = \begin{cases} \min\limits_{R_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})} \psi_\alpha(R_{XY}) & \text{if } \alpha \in (0,1), \\ \max\limits_{R_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})} \psi_\alpha(R_{XY}) & \text{if } \alpha \in (1,\infty], \end{cases} \tag{253}$$

where

$$\psi_\alpha(R_{XY}) \triangleq \begin{cases} D(R_{XY}\|R_XR_Y) + \dfrac{\alpha}{1-\alpha}D(R_{XY}\|P_{XY}) & \text{if } \alpha \in (0,1)\cup(1,\infty), \\ D(R_{XY}\|R_XR_Y) - D(R_{XY}\|P_{XY}) & \text{if } \alpha = \infty. \end{cases} \tag{254}$$

For every $\alpha \in (1,\infty]$, *the mapping* $R_{XY} \mapsto \psi_\alpha(R_{XY})$ *is concave. For all* $\alpha \in (1,\infty]$ *and all* $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, *the statement* $J_\alpha(X;Y) = \psi_\alpha(R_{XY})$ *is equivalent to* $\psi_\alpha(R_{XY}) = D_\alpha(P_{XY}\|R_XR_Y)$.

Proof. For $\alpha \in (0,1)\cup(1,\infty)$, (253) follows from Lemma 8 by dividing by $1-\alpha$, which is positive or negative depending on whether α is smaller than or greater than one. For $\alpha = \infty$, we establish (253) as follows: By Lemma 10, its LHS is continuous at $\alpha = \infty$. We argue below that its RHS is continuous at $\alpha = \infty$, i.e., that

$$\lim_{\alpha\to\infty} \max_{R_{XY}} \psi_\alpha(R_{XY}) = \max_{R_{XY}} \psi_\infty(R_{XY}). \tag{255}$$

Because (253) holds for $\alpha \in (1,\infty)$ and because both its sides are continuous at $\alpha = \infty$, it must also hold for $\alpha = \infty$.

We now establish (255). Let R^*_{XY} be a PMF that achieves the maximum on the RHS of (255). Then, for all $\alpha > 1$,

$$\begin{aligned} \psi_\infty(R^*_{XY}) &= \max_{R_{XY}} \psi_\infty(R_{XY}) && (256)\\ &\geq \max_{R_{XY}} \psi_\alpha(R_{XY}) && (257)\\ &\geq \psi_\alpha(R^*_{XY}), && (258) \end{aligned}$$

where (257) holds because, by (254), $\psi_\infty(R_{XY}) = \psi_\alpha(R_{XY}) + \frac{1}{\alpha-1}D(R_{XY}\|P_{XY}) \geq \psi_\alpha(R_{XY})$ for all $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$. By (254), $\alpha \mapsto \psi_\alpha(R^*_{XY})$ is continuous at $\alpha = \infty$, so the RHS of (258) approaches $\psi_\infty(R^*_{XY})$ as α tends to infinity, and (255) follows from the sandwich theorem.

We now show that $R_{XY} \mapsto \psi_\alpha(R_{XY})$ is concave for $\alpha \in (1,\infty]$. A simple computation reveals that for all $\alpha \in (1,\infty)$,

$$\psi_\alpha(R_{XY}) = H(R_X) + H(R_Y) + \frac{1}{\alpha-1}H(R_{XY}) + \frac{\alpha}{\alpha-1}\sum_{x,y} R_{XY}(x,y)\log P(x,y). \tag{259}$$

Because the entropy is a concave function and because a nonnegative weighted sum of concave functions is concave, this implies that $\psi_\alpha(R_{XY})$ is concave in R_{XY} for $\alpha \in (1, \infty)$. By (254), $\alpha \mapsto \psi_\alpha(R_{XY})$ is continuous at $\alpha = \infty$, so $\psi_\alpha(R_{XY})$ is concave in R_{XY} also for $\alpha = \infty$.

We next show that if $\alpha \in (1, \infty]$ and $\psi_\alpha(R_{XY}) = D_\alpha(P_{XY} \| R_X R_Y)$, then $J_\alpha(X;Y) = \psi_\alpha(R_{XY})$. Let $\alpha \in (1, \infty]$, and let R_{XY} be a PMF that satisfies $\psi_\alpha(R_{XY}) = D_\alpha(P_{XY} \| R_X R_Y)$. Then,

$$\psi_\alpha(R_{XY}) \leq J_\alpha(X;Y) \tag{260}$$

$$\leq D_\alpha(P_{XY} \| R_X R_Y), \tag{261}$$

where (260) follows from (253), and (261) holds by the definition of $J_\alpha(X;Y)$. Because $\psi_\alpha(R_{XY})$ is equal to $D_\alpha(P_{XY} \| R_X R_Y)$, both inequalities hold with equality, which implies the claim.

Finally, we show that if $\alpha \in (1, \infty]$ and $J_\alpha(X;Y) = \psi_\alpha(R_{XY})$, then $\psi_\alpha(R_{XY}) = D_\alpha(P_{XY} \| R_X R_Y)$. We first consider $\alpha \in (1, \infty)$. Let R_{XY} be a PMF that satisfies $J_\alpha(X;Y) = \psi_\alpha(R_{XY})$, and let Q_X^* and Q_Y^* be PMFs that achieve the minimum in the definition of $J_\alpha(X;Y)$. Then,

$$J_\alpha(X;Y) = \psi_\alpha(R_{XY}) \tag{262}$$

$$= D(R_{XY} \| R_X R_Y) + \frac{\alpha}{1-\alpha} D(R_{XY} \| P_{XY}) \tag{263}$$

$$\leq D(R_{XY} \| Q_X^* Q_Y^*) + \frac{\alpha}{1-\alpha} D(R_{XY} \| P_{XY}) \tag{264}$$

$$\leq D_\alpha(P_{XY} \| Q_X^* Q_Y^*) \tag{265}$$

$$= J_\alpha(X;Y), \tag{266}$$

where (264) follows from Proposition 8, and (265) follows from [22] (Theorem 30). Thus, all inequalities hold with equality. Because (264) holds with equality, $Q_X^* = R_X$ and $Q_Y^* = R_Y$ by Proposition 8. Hence, $\psi_\alpha(R_{XY}) = D_\alpha(P_{XY} \| Q_X^* Q_Y^*) = D_\alpha(P_{XY} \| R_X R_Y)$ as desired. We now consider $\alpha = \infty$. Here, (262)–(266) remain valid after replacing $\frac{\alpha}{1-\alpha}$ by -1. (Now, (265) follows from a short computation.) Consequently, $\psi_\alpha(R_{XY}) = D_\alpha(P_{XY} \| R_X R_Y)$ holds also for $\alpha = \infty$. □

Lemma 19. *For all* $\alpha \in (0,1) \cup (1, \infty)$,

$$J_\alpha(X;Y) = \min_{R_X \ll P_X} \frac{1}{\alpha - 1} \Big[D_{\frac{\alpha}{\alpha-1}}(P_X \| R_X) - \alpha\, E_0\big(\tfrac{1-\alpha}{\alpha}, R_X\big) \Big], \tag{267}$$

where the minimization is over all PMFs R_X *satisfying* $R_X \ll P_X$ $\big($*i.e.*, $\operatorname{supp}(R_X) \subseteq \operatorname{supp}(P_X)\big)$; $D_\alpha(P \| Q)$ *for negative* α *is given by* (54)*; and Gallager's* E_0 *function* [29] *is defined as*

$$E_0(\rho, R_X) \triangleq -\log \sum_y \Bigg[\sum_x R_X(x)\, P(y|x)^{\frac{1}{1+\rho}} \Bigg]^{1+\rho}. \tag{268}$$

Proof. Let $\alpha \in (0,1) \cup (1, \infty)$, and define the set $\mathcal{R} \triangleq \{R_X \in \mathcal{P}(\mathcal{X}) : \operatorname{supp}(R_X) \subseteq \operatorname{supp}(P_X)\}$. We establish (267) by showing that for all $R_X \in \mathcal{R}$,

$$\frac{1}{\alpha - 1} \Big[D_{\frac{\alpha}{\alpha-1}}(P_X \| R_X) - \alpha\, E_0\big(\tfrac{1-\alpha}{\alpha}, R_X\big) \Big] \geq J_\alpha(X;Y), \tag{269}$$

with equality for some $R_X \in \mathcal{R}$.

Fix $R_X \in \mathcal{R}$. If the LHS of (269) is infinite, then (269) holds trivially. Otherwise, define the PMF $\hat{Q}_X$ as

$$\hat{Q}_X(x) \triangleq \frac{P_X(x)^{\frac{\alpha}{\alpha-1}} R_X(x)^{\frac{1}{1-\alpha}}}{\sum_{x' \in \mathcal{X}} P_X(x')^{\frac{\alpha}{\alpha-1}} R_X(x')^{\frac{1}{1-\alpha}}}, \tag{270}$$

where we use the convention that $0^{\frac{\alpha}{\alpha-1}} \cdot 0^{\frac{1}{1-\alpha}} = 0$. (The RHS of (270) is finite whenever the LHS of (269) is finite.) Then, (269) holds because

$$J_\alpha(X;Y) = \min_{Q_X} \frac{\alpha}{\alpha-1} \log \sum_y \left[\sum_x P(x,y)^\alpha Q_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{271}$$

$$\leq \frac{\alpha}{\alpha-1} \log \sum_y \left[\sum_x P(x,y)^\alpha \hat{Q}_X(x)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{272}$$

$$= \log \sum_x P_X(x)^{\frac{\alpha}{\alpha-1}} R_X(x)^{\frac{1}{1-\alpha}} + \frac{\alpha}{\alpha-1} \log \sum_y \left[\sum_x R_X(x) P(y|x)^\alpha \right]^{\frac{1}{\alpha}} \tag{273}$$

$$= \frac{1}{\alpha-1} \Big[D_{\frac{\alpha}{\alpha-1}}(P_X \| R_X) - \alpha E_0\big(\tfrac{1-\alpha}{\alpha}, R_X\big) \Big], \tag{274}$$

where (271) follows from Lemma 17, and (273) follows from (270) using some algebra. It remains to show that there exists an $R_X \in \mathcal{R}$ for which (272) holds with equality. To that end, let Q_X^* be a PMF that achieves the minimum on the RHS of (271), and define the PMF R_X as

$$R_X(x) \triangleq \frac{P_X(x)^\alpha Q_X^*(x)^{1-\alpha}}{\sum_{x' \in \mathcal{X}} P_X(x')^\alpha Q_X^*(x')^{1-\alpha}}, \tag{275}$$

where we use the convention that $0^\alpha \cdot 0^{1-\alpha} = 0$. Because $\operatorname{supp}(Q_X^*) \subseteq \operatorname{supp}(P_X)$ (Lemma 16), the definitions (275) and (270) imply that $\hat{Q}_X = Q_X^*$. Hence, (272) holds with equality for this $R_X \in \mathcal{R}$. □

Lemma 20. *For every $\alpha \in (\frac{1}{2}, \infty]$, the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$ has a unique minimizer. This need not be the case when $\alpha \in [0, \frac{1}{2}]$.*

Proof. First consider $\alpha \in (\frac{1}{2}, 1)$. Let (Q_X^*, Q_Y^*) and $(\hat{Q}_X, \hat{Q}_Y)$ be pairs of PMFs that both minimize $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$. We establish uniqueness by arguing that (Q_X^*, Q_Y^*) and $(\hat{Q}_X, \hat{Q}_Y)$ must be identical. Observe that

$$J_\alpha(X;Y) \leq D_\alpha\big(P_{XY} \| (0.5 Q_X^* + 0.5 \hat{Q}_X)(0.5 Q_Y^* + 0.5 \hat{Q}_Y)\big) \tag{276}$$

$$\leq 0.5 D_\alpha(P_{XY} \| Q_X^* Q_Y^*) + 0.5 D_\alpha(P_{XY} \| \hat{Q}_X \hat{Q}_Y) \tag{277}$$

$$= J_\alpha(X;Y), \tag{278}$$

where (276) holds by the definition of $J_\alpha(X;Y)$, and (277) follows from Lemma 15. Hence, (277) holds with equality, which implies that (228) in the proof of Lemma 15 holds with equality, i.e.,

$$\sum_{x,y} P(x,y)^\alpha \sqrt{0.5 Q_X^*(x)^{2(1-\alpha)} + 0.5 \hat{Q}_X(x)^{2(1-\alpha)}} \sqrt{0.5 Q_Y^*(y)^{2(1-\alpha)} + 0.5 \hat{Q}_Y(y)^{2(1-\alpha)}}$$
$$= \sum_{x,y} P(x,y)^\alpha \big[0.5 Q_X^*(x) + 0.5 \hat{Q}_X(x)\big]^{1-\alpha} \sqrt{0.5 Q_Y^*(y)^{2(1-\alpha)} + 0.5 \hat{Q}_Y(y)^{2(1-\alpha)}}. \tag{279}$$

We first argue that $Q_X^* = \hat{Q}_X$. Since Q_X^* and $\hat{Q}_X$ are PMFs, it suffices to show that $Q_X^*(x) = \hat{Q}_X(x)$ for every $x \in \operatorname{supp}(\hat{Q}_X)$. Let $\hat{x} \in \operatorname{supp}(\hat{Q}_X)$. Because $\operatorname{supp}(\hat{Q}_X) \subseteq \operatorname{supp}(P_X)$ (Lemma 16), there exists a $\hat{y} \in \mathcal{Y}$ such that $P(\hat{x}, \hat{y}) > 0$. Again by Lemma 16, this implies that $\hat{Q}_Y(\hat{y}) > 0$. Because the mapping $z \mapsto z^{2(1-\alpha)}$ is strictly concave on $\mathbb{R}_{\geq 0}$ for $\alpha \in (\frac{1}{2}, 1)$, it follows from (279) that $Q_X^*(\hat{x}) = \hat{Q}_X(\hat{x})$. Swapping the roles of Q_X and Q_Y, we obtain that $Q_Y^* = \hat{Q}_Y$.

For $\alpha = 1$, the minimizer is unique by Proposition 8 because $D_1(P_{XY} \| Q_X Q_Y) = D(P_{XY} \| Q_X Q_Y)$.

Now consider $\alpha \in (1, \infty]$. Here, we establish uniqueness via the characterization of $J_\alpha(X;Y)$ provided by Lemma 18. Let $\psi_\alpha(R_{XY})$ be defined as in Lemma 18. Let R_{XY} be a PMF that satisfies

$J_\alpha(X;Y) = \psi_\alpha(R_{XY})$, and let (Q_X^*, Q_Y^*) be a pair of PMFs that minimizes $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$. If $\alpha \in (1, \infty)$, then (264) in the proof of Lemma 18 holds with equality, i.e.,

$$D(R_{XY} \| R_X R_Y) + \frac{\alpha}{1-\alpha} D(R_{XY} \| P_{XY}) = D(R_{XY} \| Q_X^* Q_Y^*) + \frac{\alpha}{1-\alpha} D(R_{XY} \| P_{XY}). \tag{280}$$

Because the LHS of (280) is finite, Proposition 8 implies that $Q_X^* = R_X$ and $Q_Y^* = R_Y$, thus the minimizer is unique. As shown in the proof of Lemma 18, (280) remains valid for $\alpha = \infty$ after replacing $\frac{\alpha}{1-\alpha}$ by -1, thus the same argument establishes the uniqueness for $\alpha = \infty$.

Finally, we show that, for $\alpha \in [0, \frac{1}{2}]$, the mapping $(Q_X, Q_Y) \mapsto D_\alpha(P_{XY} \| Q_X Q_Y)$ can have more than one minimizer. Let X be uniformly distributed over $\{0, 1\}$, and let $Y = X$. Then, for all $\alpha \in [0, \frac{1}{2}]$,

$$J_\alpha(X;Y) = \frac{\alpha}{1-\alpha} \log 2 \tag{281}$$

$$= D_\alpha\big(P_{XY} \| (1,0)(1,0)\big) \tag{282}$$

$$= D_\alpha\big(P_{XY} \| (0,1)(0,1)\big), \tag{283}$$

where (281) follows from Lemma 11. □

Lemma 21. *For every $\alpha \in [0, \infty]$, the minimum in the definition of $K_\alpha(X;Y)$ in (2) exists and is finite.*

Proof. Let $\alpha \in [0, \infty]$, and denote by U_X and U_Y the uniform distribution over $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then $\inf_{Q_X, Q_Y} \Delta_\alpha(P_{XY} \| Q_X Q_Y)$ is finite because $\Delta_\alpha(P_{XY} \| U_X U_Y)$ is finite and because the relative α-entropy is nonnegative (Proposition 5). For $\alpha \in (0, \infty)$, the minimum exists because the set $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ is compact and the mapping $(Q_X, Q_Y) \mapsto \Delta_\alpha(P_{XY} \| Q_X Q_Y)$ is continuous. For $\alpha \in \{0, \infty\}$, the minimum exists because $(Q_X, Q_Y) \mapsto \Delta_\alpha(P_{XY} \| Q_X Q_Y)$ takes on only a finite number of values: if $\alpha = 0$, then $\Delta_\alpha(P_{XY} \| Q_X Q_Y)$ depends on $Q_X Q_Y$ only via $\operatorname{supp}(Q_X Q_Y) \subseteq \mathcal{X} \times \mathcal{Y}$; and if $\alpha = \infty$, then $\Delta_\alpha(P_{XY} \| Q_X Q_Y)$ depends on $Q_X Q_Y$ only via $\operatorname{argmax}(Q_X Q_Y) \subseteq \mathcal{X} \times \mathcal{Y}$. □

Lemma 22. *For all $\alpha \in [0, \infty]$, $K_\alpha(X;Y) \geq 0$. If $\alpha \in (0, \infty)$, then $K_\alpha(X;Y) = 0$ if and only if X and Y are independent (nonnegativity).*

Proof. The nonnegativity follows from the definition of $K_\alpha(X;Y)$ because the relative α-entropy is nonnegative for $\alpha \in [0, \infty]$ (Proposition 5). If X and Y are independent, then $P_{XY} = P_X P_Y$, and the choice $Q_X = P_X$ and $Q_Y = P_Y$ in the definition of $K_\alpha(X;Y)$ achieves $K_\alpha(X;Y) = 0$. Conversely, if $K_\alpha(X;Y) = 0$, then there exist PMFs Q_X^* and Q_Y^* satisfying $\Delta_\alpha(P_{XY} \| Q_X^* Q_Y^*) = 0$. If, in addition, $\alpha \in (0, \infty)$, then $P_{XY} = Q_X^* Q_Y^*$ by Proposition 5, and hence X and Y are independent. □

Lemma 23. *For all $\alpha \in [0, \infty]$, $K_\alpha(X;Y) = K_\alpha(Y;X)$ (symmetry).*

Proof. The definition of $K_\alpha(X;Y)$ is symmetric in X and Y. □

Lemma 24. *For all $\alpha \in (0, \infty)$,*

$$K_\alpha(X;Y) + H_\alpha(X,Y) = \min_{Q_X, Q_Y} -\log M_{\frac{\alpha-1}{\alpha}}(Q_X, Q_Y), \tag{284}$$

where $M_\beta(Q_X, Q_Y)$ is the following weighted power mean [30] (Chapter III): For $\beta \in \mathbb{R} \setminus \{0\}$,

$$M_\beta(Q_X, Q_Y) \triangleq \left[\sum_{x,y} P(x,y) [Q_X(x) Q_Y(y)]^\beta \right]^{\frac{1}{\beta}}, \tag{285}$$

where for $\beta < 0$, we read $P(x,y)[Q_X(x)\,Q_Y(y)]^\beta$ as $P(x,y)/[Q_X(x)\,Q_Y(y)]^{-\beta}$ and use the conventions (44); and for $\beta = 0$, using the convention $0^0 = 1$,

$$M_0(Q_X, Q_Y) \triangleq \prod_{x,y} [Q_X(x)\,Q_Y(y)]^{P(x,y)}. \tag{286}$$

Proof. Let $\alpha \in (0, \infty)$, and define the PMF $\widetilde{P}_{XY}$ as

$$\widetilde{P}_{XY}(x,y) \triangleq \frac{P_{XY}(x,y)^\alpha}{\sum_{(x',y') \in \mathcal{X} \times \mathcal{Y}} P_{XY}(x',y')^\alpha}. \tag{287}$$

Then,

$$K_\alpha(X;Y) = J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}) \tag{288}$$

$$= \min_{Q_X, Q_Y} D_{\frac{1}{\alpha}}(\widetilde{P}_{XY} \| Q_X Q_Y), \tag{289}$$

where (288) follows from Proposition 7, and (289) follows from the definition of $J_{1/\alpha}(\widetilde{X};\widetilde{Y})$. A simple computation reveals that for all PMFs Q_X and Q_Y,

$$D_{\frac{1}{\alpha}}(\widetilde{P}_{XY} \| Q_X Q_Y) = -\log M_{\frac{\alpha-1}{\alpha}}(Q_X, Q_Y) - H_\alpha(X,Y). \tag{290}$$

Hence, (284) follows from (289) and (290). □

Lemma 25. *For $\alpha = 0$,*

$$K_0(X;Y) = \log \frac{|\mathrm{supp}(P_X P_Y)|}{|\mathrm{supp}(P_{XY})|} \tag{291}$$

$$\geq \min_{Q_X, Q_Y} \log \max_{(x,y) \in \mathrm{supp}(P_{XY})} \frac{1}{Q_X(x)\,Q_Y(y)} - \log |\mathrm{supp}(P_{XY})| \tag{292}$$

$$= \lim_{\alpha \downarrow 0} K_\alpha(X;Y), \tag{293}$$

where in the RHS of (292), we use the conventions (44). The inequality can be strict, so $\alpha \mapsto K_\alpha(X;Y)$ need not be continuous at $\alpha = 0$.

Proof. We first prove (291). Recall that

$$\Delta_0(P_{XY} \| Q_X Q_Y) = \begin{cases} \log \frac{|\mathrm{supp}(Q_X Q_Y)|}{|\mathrm{supp}(P_{XY})|} & \text{if } \mathrm{supp}(P_{XY}) \subseteq \mathrm{supp}(Q_X Q_Y), \\ \infty & \text{otherwise.} \end{cases} \tag{294}$$

Observe that $\Delta_0(P_{XY} \| Q_X Q_Y)$ is finite only if $\mathrm{supp}(P_X) \subseteq \mathrm{supp}(Q_X)$ and $\mathrm{supp}(P_Y) \subseteq \mathrm{supp}(Q_Y)$. For such PMFs Q_X and Q_Y, we have $|\mathrm{supp}(Q_X Q_Y)| \geq |\mathrm{supp}(P_X P_Y)|$. Thus, for all PMFs Q_X and Q_Y,

$$\Delta_0(P_{XY} \| Q_X Q_Y) \geq \log \frac{|\mathrm{supp}(P_X P_Y)|}{|\mathrm{supp}(P_{XY})|}. \tag{295}$$

Choosing $Q_X = P_X$ and $Q_Y = P_Y$ achieves equality in (295), which establishes (291).

We now show (292). Let Q_X and Q_Y be the uniform distributions over $\mathrm{supp}(P_X)$ and $\mathrm{supp}(P_Y)$, respectively. Then,

$$\log \max_{(x,y) \in \mathrm{supp}(P_{XY})} \frac{1}{Q_X(x)\,Q_Y(y)} - \log |\mathrm{supp}(P_{XY})| = \log \frac{|\mathrm{supp}(P_X P_Y)|}{|\mathrm{supp}(P_{XY})|}, \tag{296}$$

and hence (292) holds.

We next establish (293). To that end, define

$$\tau \triangleq \min_{(x,y)\in\mathrm{supp}(P_{XY})} P(x,y). \tag{297}$$

We bound $K_\alpha(X;Y) + H_\alpha(X,Y)$ as follows: For all $\alpha \in (0,1)$,

$$K_\alpha(X;Y) + H_\alpha(X,Y) = \min_{Q_X,Q_Y} \frac{\alpha}{1-\alpha} \log \sum_{x,y} P(x,y)[Q_X(x)\,Q_Y(y)]^{\frac{\alpha-1}{\alpha}} \tag{298}$$

$$\geq \min_{Q_X,Q_Y} \frac{\alpha}{1-\alpha} \log \sum_{(x,y)\in\mathrm{supp}(P_{XY})} \tau\,[Q_X(x)\,Q_Y(y)]^{\frac{\alpha-1}{\alpha}} \tag{299}$$

$$\geq \min_{Q_X,Q_Y} \frac{\alpha}{1-\alpha} \log \max_{(x,y)\in\mathrm{supp}(P_{XY})} \tau\,[Q_X(x)\,Q_Y(y)]^{\frac{\alpha-1}{\alpha}} \tag{300}$$

$$= \min_{Q_X,Q_Y} \log \max_{(x,y)\in\mathrm{supp}(P_{XY})} \frac{1}{Q_X(x)\,Q_Y(y)} - \frac{\alpha}{1-\alpha} \log \frac{1}{\tau}, \tag{301}$$

where (298) follows from Lemma 24. Similarly, for all $\alpha \in (0,1)$,

$$K_\alpha(X;Y) + H_\alpha(X,Y) = \min_{Q_X,Q_Y} \frac{\alpha}{1-\alpha} \log \sum_{x,y} P(x,y)[Q_X(x)\,Q_Y(y)]^{\frac{\alpha-1}{\alpha}} \tag{302}$$

$$\leq \min_{Q_X,Q_Y} \frac{\alpha}{1-\alpha} \log \max_{(x,y)\in\mathrm{supp}(P_{XY})} [Q_X(x)\,Q_Y(y)]^{\frac{\alpha-1}{\alpha}} \tag{303}$$

$$= \min_{Q_X,Q_Y} \log \max_{(x,y)\in\mathrm{supp}(P_{XY})} \frac{1}{Q_X(x)\,Q_Y(y)}, \tag{304}$$

where (302) is the same as (298). Now (293) follows from (301), (304), and the sandwich theorem because $\lim_{\alpha\downarrow 0} \frac{\alpha}{1-\alpha} \log \frac{1}{\tau} = 0$ and because $\lim_{\alpha\downarrow 0} H_\alpha(X,Y) = \log|\mathrm{supp}(P_{XY})|$ (Proposition 3).

Finally, we provide an example for which (292) holds with strict inequality. Let $\mathcal{X} = \{1,2,3\}$, let $\mathcal{Y} = \{1,2\}$, and let (X,Y) be uniformly distributed over $\{(1,1),(2,2),(3,1)\}$. The LHS of (292) then equals $\log 2$. Using

$$Q_X(x) \triangleq \begin{cases} 0.28 & \text{if } x \in \{1,3\}, \\ 0.44 & \text{if } x = 2, \end{cases} \tag{305}$$

$$Q_Y(y) \triangleq \begin{cases} 0.60 & \text{if } y = 1, \\ 0.40 & \text{if } y = 2, \end{cases} \tag{306}$$

we see that the RHS of (292) is upper bounded by $\log \frac{5.952\ldots}{3}$, which is smaller than $\log 2$. □

Lemma 26. $K_1(X;Y) = I(X;Y)$.

Proof. The claim follows from Proposition 8 because $\Delta_1(P_{XY}\|Q_X Q_Y)$ in the definition of $K_1(X;Y)$ is equal to $D(P_{XY}\|Q_X Q_Y)$. □

Lemma 27. *Let $f\colon \{1,\ldots,|\mathcal{X}|\} \to \mathcal{X}$ and $g\colon \{1,\ldots,|\mathcal{Y}|\} \to \mathcal{Y}$ be bijective functions, and let B be the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix whose Row-i Column-j entry $\mathsf{B}_{i,j}$ equals $P_{XY}(f(i),g(j))$. Then,*

$$K_2(X;Y) = -2\log\sigma_1(\mathsf{B}) - H_2(X,Y), \tag{307}$$

where $\sigma_1(\mathsf{B})$ denotes the largest singular value of B. (Because the singular values of a matrix are invariant under row and column permutations, the result does not depend on f or g.)

Proof. Let $(\widetilde{X}, \widetilde{Y})$ be distributed according to the joint PMF

$$\widetilde{P}_{XY}(x,y) \triangleq [\beta P_{XY}(x,y)]^2, \tag{308}$$

where

$$\beta \triangleq \left[\sum_{x,y} P_{XY}(x,y)^2\right]^{-\frac{1}{2}}. \tag{309}$$

Then,

$$K_2(X;Y) = J_{\frac{1}{2}}(\widetilde{X};\widetilde{Y}) \tag{310}$$
$$= -2\log\sigma_1(\beta \mathsf{B}) \tag{311}$$
$$= -2\log[\beta\sigma_1(\mathsf{B})] \tag{312}$$
$$= -2\log\sigma_1(\mathsf{B}) - H_2(X,Y), \tag{313}$$

where (310) follows from Proposition 7; (311) follows from Lemma 6 and (308); (312) holds because $\beta > 0$; and (313) follows from the definition of $H_2(X,Y)$. □

Lemma 28. $K_\infty(X;Y) = 0$.

Proof. Let the pair $(\hat{x}, \hat{y})$ be such that $P(\hat{x},\hat{y}) = \max_{x,y} P(x,y)$, and define the PMFs $\hat{Q}_X$ and $\hat{Q}_Y$ as $\hat{Q}_X(x) = \mathbb{1}\{x = \hat{x}\}$ and $\hat{Q}_Y(y) = \mathbb{1}\{y = \hat{y}\}$. Then, $\Delta_\infty(P_{XY}\|\hat{Q}_X\hat{Q}_Y) = 0$, so $K_\infty(X;Y) \le 0$. Because $K_\infty(X;Y) \ge 0$ (Lemma 22), this implies $K_\infty(X;Y) = 0$. □

Lemma 29. *The mapping* $\alpha \mapsto K_\alpha(X;Y)$ *need not be monotonic on* $[0,\infty]$.

Proof. Let P_{XY} be such that $\operatorname{supp}(P_{XY}) = \mathcal{X} \times \mathcal{Y}$ and $I(X;Y) > 0$. Then,

$$K_0(X;Y) = 0, \tag{314}$$
$$K_1(X;Y) > 0, \tag{315}$$
$$K_\infty(X;Y) = 0, \tag{316}$$

which follow from Lemmas 25, 26, and 28, respectively. Thus, $\alpha \mapsto K_\alpha(X;Y)$ is not monotonic on $[0,\infty]$. □

Lemma 30. *The mapping* $\alpha \mapsto K_\alpha(X;Y) + H_\alpha(X,Y)$ *is nonincreasing on* $[0,\infty]$.

Proof. We first show the monotonicity for $\alpha \in (0,\infty)$. To that end, let $\alpha, \alpha' \in (0,\infty)$ with $\alpha \le \alpha'$, and let $M_\beta(Q_X, Q_Y)$ be defined as in (285) and (286). Then, for all PMFs Q_X and Q_Y,

$$M_{\frac{\alpha-1}{\alpha}}(Q_X,Q_Y) \le M_{\frac{\alpha'-1}{\alpha'}}(Q_X,Q_Y), \tag{317}$$

which follows from the power mean inequality [30] (III 3.1.1 Theorem 1) because $\frac{\alpha-1}{\alpha} \le \frac{\alpha'-1}{\alpha'}$. Hence,

$$K_\alpha(X;Y) + H_\alpha(X,Y) = \min_{Q_X,Q_Y} -\log M_{\frac{\alpha-1}{\alpha}}(Q_X,Q_Y) \tag{318}$$
$$\ge \min_{Q_X,Q_Y} -\log M_{\frac{\alpha'-1}{\alpha'}}(Q_X,Q_Y) \tag{319}$$
$$= K_{\alpha'}(X;Y) + H_{\alpha'}(X,Y), \tag{320}$$

where (318) and (320) follow from Lemma 24, and (319) follows from (317).

The monotonicity extends to $\alpha = 0$ because

$$K_0(X;Y) + H_0(X,Y) \geq \lim_{\alpha \downarrow 0} K_\alpha(X;Y) + H_0(X,Y) \tag{321}$$
$$= \lim_{\alpha \downarrow 0} [K_\alpha(X;Y) + H_\alpha(X,Y)], \tag{322}$$

where (321) follows from Lemma 25, and (322) holds because $\alpha \mapsto H_\alpha(X,Y)$ is continuous at $\alpha = 0$ (Proposition 3).

The monotonicity extends to $\alpha = \infty$ because for all $\alpha \in (0,\infty)$,

$$K_\alpha(X;Y) + H_\alpha(X,Y) \geq H_\alpha(X,Y) \tag{323}$$
$$\geq H_\infty(X,Y) \tag{324}$$
$$= K_\infty(X;Y) + H_\infty(X,Y), \tag{325}$$

where (323) holds because $K_\alpha(X;Y) \geq 0$ (Lemma 22); (324) holds because $H_\alpha(X,Y)$ is nonincreasing in α (Proposition 3); and (325) holds because $K_\infty(X;Y) = 0$ (Lemma 28). □

Lemma 31. *The mapping $\alpha \mapsto K_\alpha(X;Y)$ is continuous on $(0,\infty]$. (See Lemma 25 for the behavior at $\alpha = 0$.)*

Proof. Because $\alpha \mapsto H_\alpha(X,Y)$ is continuous on $[0,\infty]$ (Proposition 3), it suffices to show that the mapping $\alpha \mapsto K_\alpha(X;Y) + H_\alpha(X,Y)$ is continuous on $(0,\infty]$. We first show that it is continuous on $(0,1) \cup (1,\infty)$ by showing that $\alpha \mapsto (1-\frac{1}{\alpha})\,[K_\alpha(X;Y) + H_\alpha(X,Y)]$ is concave and hence continuous on $(0,\infty)$. For a fixed $\alpha \in (0,\infty)$, let $(\widetilde{X},\widetilde{Y})$ be distributed according to the joint PMF

$$\widetilde{P}_{XY}(x,y) \triangleq \frac{P_{XY}(x,y)^\alpha}{\sum_{(x',y')\in\mathcal{X}\times\mathcal{Y}} P_{XY}(x',y')^\alpha}. \tag{326}$$

Then, for all $\alpha \in (0,\infty)$,

$$(1-\tfrac{1}{\alpha})\,[K_\alpha(X;Y) + H_\alpha(X,Y)]$$
$$= (1-\tfrac{1}{\alpha}) J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}) + (1-\tfrac{1}{\alpha}) H_\alpha(X,Y) \tag{327}$$
$$= \min_{R_{XY}} \Big[(1-\tfrac{1}{\alpha}) D(R_{XY} \| R_X R_Y) + \tfrac{1}{\alpha} D(R_{XY} \| \widetilde{P}_{XY}) + (1-\tfrac{1}{\alpha}) H_\alpha(X,Y) \Big] \tag{328}$$
$$= \min_{R_{XY}} \Big[(1-\tfrac{1}{\alpha}) D(R_{XY} \| R_X R_Y) + (1-\tfrac{1}{\alpha}) H(R_{XY}) + D(R_{XY} \| P_{XY}) \Big], \tag{329}$$

where (327) follows from Proposition 7; (328) follows from Lemma 8; and (329) follows from a short computation. For every $R_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, the expression in square brackets on the RHS of (329) is concave in α because the mapping $\alpha \mapsto 1 - \frac{1}{\alpha}$ is concave on $(0,\infty)$ and because $D(R_{XY}\|R_X R_Y)$ and $H(R_{XY})$ are nonnegative. The pointwise minimum preserves the concavity, thus the LHS of (327) is concave in α and hence continuous in $\alpha \in (0,\infty)$. This implies that $\alpha \mapsto K_\alpha(X;Y) + H_\alpha(X,Y)$ and hence $\alpha \mapsto K_\alpha(X;Y)$ is continuous on $(0,1) \cup (1,\infty)$.

We now establish continuity at $\alpha = \infty$. Let $(\hat{x},\hat{y})$ be such that $P(\hat{x},\hat{y}) = \max_{x,y} P(x,y)$; define the PMFs $\hat{Q}_X$ and $\hat{Q}_Y$ as $\hat{Q}_X(x) \triangleq \mathbb{1}\{x = \hat{x}\}$ and $\hat{Q}_Y(y) \triangleq \mathbb{1}\{y = \hat{y}\}$; and let $M_\beta(Q_X, Q_Y)$ be defined as in (285). Then, for all $\alpha \in (1,\infty)$,

$$K_\infty(X;Y) + H_\infty(X,Y) \leq K_\alpha(X;Y) + H_\alpha(X,Y) \tag{330}$$
$$\leq -\log M_{\frac{\alpha-1}{\alpha}}(\hat{Q}_X, \hat{Q}_Y) \tag{331}$$
$$= \frac{\alpha}{\alpha-1} H_\infty(X,Y) \tag{332}$$

$$= K_\infty(X;Y) + \frac{\alpha}{\alpha-1} H_\infty(X,Y), \tag{333}$$

where (330) holds because $K_\alpha(X;Y) + H_\alpha(X,Y)$ is nonincreasing in α (Lemma 30); (331) follows from Lemma 24; (332) follows from the definitions of $M_\beta(Q_X, Q_Y)$ in (285) and $H_\infty(X,Y)$ in (48); and (333) holds because $K_\infty(X;Y) = 0$ (Lemma 28). Because $\lim_{\alpha\to\infty} \frac{\alpha}{\alpha-1} = 1$, (330)–(333) and the sandwich theorem imply that $\alpha \mapsto K_\alpha(X;Y) + H_\alpha(X,Y)$ is continuous at $\alpha = \infty$. This and the continuity of $\alpha \mapsto H_\alpha(X,Y)$ at $\alpha = \infty$ (Proposition 3) establish the continuity of $\alpha \mapsto K_\alpha(X;Y)$ at $\alpha = \infty$.

It remains to show the continuity at $\alpha = 1$. Let $\alpha \in (\frac{4}{5}, 1) \cup (1, \frac{4}{3})$, and define $\delta \triangleq \frac{|\alpha-1|}{\alpha} \in (0, \frac{1}{4})$. (These definitions ensure that on the RHS of (340) ahead, $1 - 4\delta$ will be positive.) Let $M_\beta(Q_X, Q_Y)$ be defined as in (285) and (286). Then, for all PMFs Q_X and Q_Y,

$$M_{\frac{\alpha-1}{\alpha}}(Q_X, Q_Y) \le M_\delta(Q_X, Q_Y) \tag{334}$$

$$= \left[\sum_{x,y} P(x,y) \, [P_X(x) P_Y(y)]^\delta \left[\frac{Q_X(x) Q_Y(y)}{P_X(x) P_Y(y)} \right]^\delta \right]^{\frac{1}{\delta}} \tag{335}$$

$$\le \left[\sum_{x,y} P(x,y) \, [P_X(x) P_Y(y)]^{2\delta} \right]^{\frac{1}{2\delta}} \cdot \left[\sum_{x,y} P(x,y) \left[\frac{Q_X(x) Q_Y(y)}{P_X(x) P_Y(y)} \right]^{2\delta} \right]^{\frac{1}{2\delta}} \tag{336}$$

$$\le \left[\sum_{x,y} P(x,y) \, [P_X(x) P_Y(y)]^{2\delta} \right]^{\frac{1}{2\delta}} \tag{337}$$

$$= M_{2\delta}(P_X, P_Y), \tag{338}$$

where (334) follows from the power mean inequality [30] (III 3.1.1 Theorem 1) because $\frac{\alpha-1}{\alpha} \le \delta$; (336) follows from the Cauchy–Schwarz inequality; and (337) holds because

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

$$\le \left[\sum_{x} P_X(x) \left[\frac{Q_X(x)}{P_X(x)} \right]^{4\delta} \right]^{\frac{1}{4\delta}} \cdot \left[\sum_{y} P_Y(y) \left[\frac{Q_Y(y)}{P_Y(y)} \right]^{4\delta} \right]^{\frac{1}{4\delta}} \tag{339}$$

$$= 2^{-D_{1-4\delta}(P_X \| Q_X)} \cdot 2^{-D_{1-4\delta}(P_Y \| Q_Y)} \tag{340}$$

$$\le 1, \tag{341}$$

where (339) follows from the Cauchy–Schwarz inequality, and (341) holds because $1 - 4\delta > 0$ and because the Rényi divergence is nonnegative for positive orders (Proposition 4). Thus, for all $\alpha \in (\frac{4}{5}, \frac{4}{3})$,

$$-\log M_{\frac{2|\alpha-1|}{\alpha}}(P_X, P_Y) \le \min_{Q_X, Q_Y} -\log M_{\frac{\alpha-1}{\alpha}}(Q_X, Q_Y) \tag{342}$$

$$\le -\log M_{\frac{\alpha-1}{\alpha}}(P_X, P_Y), \tag{343}$$

where (342) follows from (338) if $\alpha \ne 1$ and from Proposition 8 and a simple computation if $\alpha = 1$. By Lemma 24, this implies that for all $\alpha \in (\frac{4}{5}, \frac{4}{3})$,

$$-\log M_{\frac{2|\alpha-1|}{\alpha}}(P_X, P_Y) \le K_\alpha(X;Y) + H_\alpha(X,Y) \tag{344}$$

$$\le -\log M_{\frac{\alpha-1}{\alpha}}(P_X, P_Y). \tag{345}$$

Because $\beta \mapsto M_\beta(P_X, P_Y)$ is continuous at $\beta = 0$ [30] (III 1 Theorem 2(b)), (344)–(345) and the sandwich theorem imply that $\alpha \mapsto K_\alpha(X;Y) + H_\alpha(X,Y)$ is continuous at $\alpha = 1$. This and the continuity of $\alpha \mapsto H_\alpha(X,Y)$ at $\alpha = 1$ (Proposition 3) establish the continuity of $\alpha \mapsto K_\alpha(X;Y)$ at $\alpha = 1$. □

Lemma 32. *If $X = Y$ with probability one, then*

$$K_\alpha(X;Y) = \begin{cases} 2H_{\frac{\alpha}{2-\alpha}}(X) - H_\alpha(X) & \text{if } \alpha \in [0,2), \\ \frac{\alpha}{\alpha-1} H_\infty(X) - H_\alpha(X) & \text{if } \alpha \geq 2, \\ 0 & \text{if } \alpha = \infty. \end{cases} \tag{346}$$

Proof. We first treat the cases $\alpha = 0$, $\alpha = 1$, and $\alpha = \infty$. For $\alpha = 0$, (346) holds because

$$K_0(X;Y) = \log \frac{|\text{supp}(P_X P_Y)|}{|\text{supp}(P_{XY})|} \tag{347}$$

$$= \log|\text{supp}(P_X)| \tag{348}$$

$$= H_0(X), \tag{349}$$

where (347) follows from Lemma 25, and (348) holds because the hypothesis $\Pr[X = Y] = 1$ implies that $|\text{supp}(P_X P_Y)| = |\text{supp}(P_X)|^2$ and $|\text{supp}(P_{XY})| = |\text{supp}(P_X)|$. For $\alpha = 1$, (346) holds because $K_1(X;Y) = I(X;Y)$ (Lemma 26) and because $\Pr[X = Y] = 1$ implies that $I(X;Y) = H(X) = H_1(X)$. For $\alpha = \infty$, (346) holds because $K_\infty(X;Y) = 0$ (Lemma 28).

Now let $\alpha \in (0,1) \cup (1,\infty)$, and let $(\widetilde{X}, \widetilde{Y})$ be distributed according to the joint PMF

$$\widetilde{P}_{XY}(x,y) \triangleq \frac{P_{XY}(x,y)^\alpha}{\sum_{(x',y') \in \mathcal{X} \times \mathcal{Y}} P_{XY}(x',y')^\alpha} \tag{350}$$

$$= \frac{P_X(x)^\alpha}{\sum_{x' \in \mathcal{X}} P_X(x')^\alpha} \mathbb{1}\{x = y\}, \tag{351}$$

where (351) holds because $P_{XY}(x,y) = P_X(x)\mathbb{1}\{x = y\}$ for all $x \in \mathcal{X}$ and all $y \in \mathcal{Y}$. If $\alpha < 2$, then (346) holds because

$$K_\alpha(X;Y) = J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}) \tag{352}$$

$$= H_{\frac{1}{2-\alpha}}(\widetilde{X}) \tag{353}$$

$$= \frac{2-\alpha}{1-\alpha} \log \sum_x \left[\frac{P_X(x)^\alpha}{\sum_{x' \in \mathcal{X}} P_X(x')^\alpha} \right]^{\frac{1}{2-\alpha}} \tag{354}$$

$$= 2H_{\frac{\alpha}{2-\alpha}}(X) - H_\alpha(X), \tag{355}$$

where (352) follows from Proposition 7; (353) follows from Lemma 11 because $\Pr[\widetilde{X} = \widetilde{Y}] = 1$ and because $\frac{1}{\alpha} > \frac{1}{2}$; and (355) follows from a simple computation. If $\alpha \geq 2$, then (346) holds because

$$K_\alpha(X;Y) = J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}) \tag{356}$$

$$= \frac{1}{\alpha - 1} H_\infty(\widetilde{X}) \tag{357}$$

$$= \frac{-1}{\alpha - 1} \log \max_x \frac{P_X(x)^\alpha}{\sum_{x' \in \mathcal{X}} P_X(x')^\alpha} \tag{358}$$

$$= \frac{\alpha}{\alpha - 1} H_\infty(X) - H_\alpha(X), \tag{359}$$

where (356) follows from Proposition 7; (357) follows from Lemma 11 because $\Pr[\widetilde{X} = \widetilde{Y}] = 1$ and because $\frac{1}{\alpha} \leq \frac{1}{2}$; and (359) follows from a simple computation. □

Lemma 33. *For every $\alpha \in (0,2)$, the mapping $(Q_X, Q_Y) \mapsto \Delta_\alpha(P_{XY} \| Q_X Q_Y)$ in the definition of $K_\alpha(X;Y)$ in (2) has a unique minimizer. This need not be the case when $\alpha \in \{0\} \cup [2,\infty]$.*

Proof. Let $\alpha \in (0,2)$. By Proposition 7, $K_\alpha(X;Y) = J_{1/\alpha}(\widetilde{X};\widetilde{Y})$, where the pair $(\widetilde{X},\widetilde{Y})$ is distributed according to the joint PMF $\widetilde{P}_{XY}$ defined in Proposition 7. The mapping $(Q_X, Q_Y) \mapsto D_{1/\alpha}(\widetilde{P}_{XY} \| Q_X Q_Y)$ in the definition of $J_{1/\alpha}(\widetilde{X};\widetilde{Y})$ has a unique minimizer by Lemma 20 because $\frac{1}{\alpha} > \frac{1}{2}$. By Proposition 6, there is a bijection between the minimizers of $D_{1/\alpha}(\widetilde{P}_{XY} \| Q_X Q_Y)$ and $\Delta_\alpha(P_{XY} \| Q_X Q_Y)$, so the mapping $(Q_X, Q_Y) \mapsto \Delta_\alpha(P_{XY} \| Q_X Q_Y)$ also has a unique minimizer.

We next show that for $\alpha \in \{0\} \cup [2,\infty]$, the mapping $(Q_X, Q_Y) \mapsto \Delta_\alpha(P_{XY} \| Q_X Q_Y)$ can have more than one minimizer. Let X be uniformly distributed over $\{0,1\}$, and let $Y = X$. Then, by Lemma 32,

$$K_\alpha(X;Y) = \begin{cases} \log 2 & \text{if } \alpha = 0, \\ \frac{1}{\alpha-1} \log 2 & \text{if } \alpha \geq 2, \\ 0 & \text{if } \alpha = \infty. \end{cases} \tag{360}$$

If $\alpha = 0$, then it follows from the definition of $\Delta_0(P\|Q)$ in (56) that $\Delta_0(P_{XY} \| Q_X Q_Y) = \log 2$ whenever $\operatorname{supp}(Q_X) = \operatorname{supp}(Q_Y) = \{0,1\}$, so the minimizer is not unique. Otherwise, if $\alpha \in [2,\infty]$, it can be verified that

$$\Delta_\alpha\big(P_{XY} \| (1,0)(1,0)\big) = \Delta_\alpha\big(P_{XY} \| (0,1)(0,1)\big) \tag{361}$$

$$= \begin{cases} \frac{1}{\alpha-1} \log 2 & \text{if } \alpha \geq 2, \\ 0 & \text{if } \alpha = \infty, \end{cases} \tag{362}$$

so the minimizer is not unique in this case either. □

Lemma 34. *If the pairs (X_1, Y_1) and (X_2, Y_2) are independent, then $K_\alpha(X_1, X_2; Y_1, Y_2) = K_\alpha(X_1;Y_1) + K_\alpha(X_2;Y_2)$ for all $\alpha \in [0,\infty]$ (additivity).*

Proof. We first treat the cases $\alpha = 0$ and $\alpha = \infty$. For $\alpha = 0$, the claim is true because

$$K_0(X_1, X_2; Y_1, Y_2) = \log \frac{|\operatorname{supp}(P_{X_1X_2} P_{Y_1Y_2})|}{|\operatorname{supp}(P_{X_1X_2Y_1Y_2})|} \tag{363}$$

$$= \log \frac{|\operatorname{supp}(P_{X_1} P_{Y_1})| \cdot |\operatorname{supp}(P_{X_2} P_{Y_2})|}{|\operatorname{supp}(P_{X_1Y_1})| \cdot |\operatorname{supp}(P_{X_2Y_2})|} \tag{364}$$

$$= K_0(X_1;Y_1) + K_0(X_2;Y_2), \tag{365}$$

where (363) and (365) follow from Lemma 25, and (364) follows from the independence hypothesis $P_{X_1X_2Y_1Y_2} = P_{X_1Y_1} P_{X_2Y_2}$. For $\alpha = \infty$, the claim is true because $K_\infty(X;Y) = 0$ (Lemma 28).

Now let $\alpha \in (0,\infty)$, and let $(\widetilde{X}_1, \widetilde{X}_2, \widetilde{Y}_1, \widetilde{Y}_2)$ be distributed according to the joint PMF

$$\widetilde{P}_{X_1X_2Y_1Y_2}(x_1, x_2, y_1, y_2) \triangleq \frac{P_{X_1X_2Y_1Y_2}(x_1, x_2, y_1, y_2)^\alpha}{\sum_{x_1', x_2', y_1', y_2'} P_{X_1X_2Y_1Y_2}(x_1', x_2', y_1', y_2')^\alpha} \tag{366}$$

$$= \frac{P_{X_1Y_1}(x_1, y_1)^\alpha}{\sum_{x_1', y_1'} P_{X_1Y_1}(x_1', y_1')^\alpha} \cdot \frac{P_{X_2Y_2}(x_2, y_2)^\alpha}{\sum_{x_2', y_2'} P_{X_2Y_2}(x_2', y_2')^\alpha}, \tag{367}$$

where (367) follows from the independence hypothesis $P_{X_1X_2Y_1Y_2} = P_{X_1Y_1} P_{X_2Y_2}$. Then,

$$K_\alpha(X_1, X_2; Y_1, Y_2) = J_{\frac{1}{\alpha}}(\widetilde{X}_1, \widetilde{X}_2; \widetilde{Y}_1, \widetilde{Y}_2) \tag{368}$$

$$= J_{\frac{1}{\alpha}}(\widetilde{X}_1; \widetilde{Y}_1) + J_{\frac{1}{\alpha}}(\widetilde{X}_2; \widetilde{Y}_2) \tag{369}$$

$$= K_\alpha(X_1;Y_1) + K_\alpha(X_2;Y_2), \tag{370}$$

where (368) and (370) follow from Proposition 7, and (369) follows from Lemma 12 because the pairs $(\widetilde{X}_1, \widetilde{Y}_1)$ and $(\widetilde{X}_2, \widetilde{Y}_2)$ are independent by (367). □

Lemma 35. *For all $\alpha \in [0, \infty]$, $K_\alpha(X;Y) \le \log|\mathcal{X}|$.*

Proof. For $\alpha = 0$, this is true because

$$K_0(X;Y) = \log \frac{|\mathrm{supp}(P_X P_Y)|}{|\mathrm{supp}(P_{XY})|} \tag{371}$$

$$\le \log \frac{|\mathcal{X}| \cdot |\mathrm{supp}(P_Y)|}{|\mathrm{supp}(P_{XY})|} \tag{372}$$

$$\le \log|\mathcal{X}|, \tag{373}$$

where (371) follows from Lemma 25. For $\alpha \in (0, \infty)$, the claim is true because

$$K_\alpha(X;Y) = J_{\frac{1}{\alpha}}(\widetilde{X};\widetilde{Y}) \tag{374}$$

$$\le \log|\mathcal{X}|, \tag{375}$$

where (374) follows from Proposition 7, and (375) follows from Lemma 13. For $\alpha = \infty$, the claim is true because $K_\infty(X;Y) = 0$ (Lemma 28). □

Lemma 36. *There exists a Markov chain $X \multimap Y \multimap Z$ for which $K_2(X;Z) > K_2(X;Y)$.*

Proof. Let the Markov chain $X \multimap Y \multimap Z$ be given by

$P_{XY}(x,y)$	$y = 0$	$y = 1$
$x = 0$	0.6	0
$x = 1$	0	0.4

$P_{Z\|Y}(z\|y)$	$z = 0$	$z = 1$
$y = 0$	0.9	0.1
$y = 1$	0	1

Using Lemma 27, we see that $K_2(X;Z) \approx 0.605$ bits, which is larger than $K_2(X;Y) \approx 0.531$ bits. □

Author Contributions: Writing–original draft preparation, A.L. and C.P.; writing–review and editing, A.L. and C.P.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
2. Tomamichel, M.; Hayashi, M. Operational interpretation of Rényi information measures via composite hypothesis testing against product and Markov distributions. *IEEE Trans. Inf. Theory* **2018**, *64*, 1064–1082. [CrossRef]
3. Sibson, R. Information radius. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **1969**, *14*, 149–160. [CrossRef]
4. Arimoto, S. Information measures and capacity of order α for discrete memoryless channels. In *Topics in Information Theory*; Csiszár, I., Elias, P., Eds.; North-Holland Publishing Company: Amsterdam, The Netherlands, 1977; pp. 41–52, ISBN 0-7204-0699-4.
5. Csiszár, I. Generalized cutoff rates and Rényi's information measures. *IEEE Trans. Inf. Theory* **1995**, *41*, 26–34. [CrossRef]
6. Fehr, S.; Berens, S. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 6801–6810. [CrossRef]
7. Sason, I.; Verdú, S. Arimoto–Rényi conditional entropy and Bayesian M-ary hypothesis testing. *IEEE Trans. Inf. Theory* **2018**, *64*, 4–25. [CrossRef]

8. Verdú, S. α-mutual information. In Proceedings of the 2015 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 1–6 February 2015; pp. 1–6. [CrossRef]
9. Tridenski, S.; Zamir, R.; Ingber, A. The Ziv–Zakai–Rényi bound for joint source-channel coding. *IEEE Trans. Inf. Theory* **2015**, *61*, 4293–4315. [CrossRef]
10. Aishwarya, G.; Madiman, M. Remarks on Rényi versions of conditional entropy and mutual information. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 1117–1121.
11. Csiszár, I.; Shields, P.C. *Information Theory and Statistics: A Tutorial*; now Publishers: Hanover, MA, USA, 2004; ISBN 978-1-933019-05-5.
12. Liese, F.; Vajda, I. On divergences and informations in statistics and information theory. *IEEE Trans. Inf. Theory* **2006**, *52*, 4394–4412. [CrossRef]
13. Sason, I.; Verdú, S. f-divergence inequalities. *IEEE Trans. Inf. Theory* **2016**, *62*, 5973–6006. [CrossRef]
14. Jiao, J.; Han, Y.; Weissman, T. Dependence measures bounding the exploration bias for general measurements. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 1475–1479. [CrossRef]
15. Ziv, J.; Zakai, M. On functionals satisfying a data-processing theorem. *IEEE Trans. Inf. Theory* **1973**, *19*, 275–283. [CrossRef]
16. Lapidoth, A.; Pfister, C. Testing against independence and a Rényi information measure. In Proceedings of the 2018 IEEE Information Theory Workshop (ITW), Guangzhou, China, 25–29 November 2018; pp. 1–5. [CrossRef]
17. Han, T.S.; Kobayashi, K. The strong converse theorem for hypothesis testing. *IEEE Trans. Inf. Theory* **1989**, *35*, 178–180. [CrossRef]
18. Nakagawa, K.; Kanaya, F. On the converse theorem in statistical hypothesis testing. *IEEE Trans. Inf. Theory* **1993**, *39*, 623–628. [CrossRef]
19. Bunte, C.; Lapidoth, A. Encoding tasks and Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 5065–5076. [CrossRef]
20. Bracher, A.; Lapidoth, A.; Pfister, C. Distributed task encoding. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 1993–1997. [CrossRef]
21. Rényi, A. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; Volume 1, pp. 547–561.
22. van Erven, T.; Harremoës, P. Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [CrossRef]
23. Sason, I.; Verdú, S. Improved bounds on lossless source coding and guessing moments via Rényi measures. *IEEE Trans. Inf. Theory* **2018**, *64*, 4323–4346. [CrossRef]
24. Ashok Kumar, M.; Sundaresan, R. Minimization problems based on relative α-entropy I: Forward projection. *IEEE Trans. Inf. Theory* **2015**, *61*, 5063–5080. [CrossRef]
25. Ashok Kumar, M.; Sundaresan, R. Minimization problems based on relative α-entropy II: Reverse projection. *IEEE Trans. Inf. Theory* **2015**, *61*, 5081–5095. [CrossRef]
26. Sundaresan, R. Guessing under source uncertainty. *IEEE Trans. Inf. Theory* **2007**, *53*, 269–287. [CrossRef]
27. Polyanskiy, Y.; Wu, Y. Lecture Notes on Information Theory. 2017. Available online: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf (accessed on 18 August 2017).
28. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006; ISBN 978-0-471-24195-9.
29. Gallager, R.G. *Information Theory and Reliable Communication*; John Wiley & Sons: Hoboken, NJ, USA, 1968; ISBN 978-0-471-29048-3.
30. Bullen, P.S. *Handbook of Means and Their Inequalities*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003; ISBN 978-1-4020-1522-9.
31. Horn, R.A.; Johnson, C.R. *Matrix Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2013; ISBN 978-0-521-83940-2.

32. Fan, K. Minimax theorems. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 42–47. [CrossRef] [PubMed]
33. Borwein, J.M.; Zhuang, D. On Fan's minimax theorem. *Math. Program.* **1986**, *34*, 232–234. [CrossRef]