Article

Conditional Rényi Divergences and Horse Betting

Cédric Bleuler, Amos Lapidoth [ORCID] **and Christoph Pfister** *

Signal and Information Processing Laboratory, ETH Zurich, 8092 Zurich, Switzerland;
cedric.bleuler@gmail.com (C.B.); lapidoth@isi.ee.ethz.ch (A.L.)
* Correspondence: pfister@isi.ee.ethz.ch

check for updates

Abstract: Motivated by a horse betting problem, a new conditional Rényi divergence is introduced. It is compared with the conditional Rényi divergences that appear in the definitions of the dependence measures by Csiszár and Sibson, and the properties of all three are studied with emphasis on their behavior under data processing. In the same way that Csiszár's and Sibson's conditional divergence lead to the respective dependence measures, so does the new conditional divergence lead to the Lapidoth–Pfister mutual information. Moreover, the new conditional divergence is also related to the Arimoto–Rényi conditional entropy and to Arimoto's measure of dependence. In the second part of the paper, the horse betting problem is analyzed where, instead of Kelly's expected log-wealth criterion, a more general family of power-mean utility functions is considered. The key role in the analysis is played by the Rényi divergence, and in the setting where the gambler has access to side information, the new conditional Rényi divergence is key. The setting with side information also provides another operational meaning to the Lapidoth–Pfister mutual information. Finally, a universal strategy for independent and identically distributed races is presented that—without knowing the winning probabilities or the parameter of the utility function—asymptotically maximizes the gambler's utility function.

1. Introduction

As shown by Kelly [1,2], many of Shannon's information measures appear naturally in the context of horse gambling when the gambler's utility function is expected log-wealth. Here, we show that under a more general family of utility functions, gambling also provides a context for some of Rényi's information measures. Moreover, the setting where the gambler has side information motivates a new Rényi-like conditional divergence, which we study and compare to other conditional divergences. The proposed family of utility functions in the context of gambling with side information also provides another operational meaning to the Rényi-like mutual information that was recently proposed by Lapidoth and Pfister [3]: it measures the gambler's gain from the side information as measured by the increase in the minimax value of the two-player zero-sum game in which the bookmaker picks the odds and the gambler then places the bets based on these odds and her side information.

Deferring the gambling-based motivation to the second part of the paper, we first describe the different conditional divergences and study some of their properties with emphasis on their behavior under data processing. We also show that the new conditional Rényi divergence relates to the Lapidoth–Pfister mutual information in much the same way that Csiszár's and Sibson's conditional divergences relate to their corresponding mutual informations. Before discussing the conditional divergences, we first recall other information measures.

The Kullback–Leibler divergence (or relative entropy) is an important concept in information theory and statistics [2,4–6]. It is defined between two probability mass functions (PMFs) P and Q over a finite set $\mathcal{X}$ as

$$D(P\|Q) \triangleq \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}, \tag{1}$$

where $\log(\cdot)$ denotes the base-2 logarithm. Defining a conditional Kullback–Leibler divergence is straightforward because, as simple algebra shows, the two natural approaches lead to the same result:

$$D(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq \sum_{x\in\operatorname{supp}(P_X)} P(x)\,D(P_{Y|X=x}\|Q_{Y|X=x}) \tag{2}$$

$$= D(P_X P_{Y|X}\|P_X Q_{Y|X}), \tag{3}$$

where $\operatorname{supp}(P) \triangleq \{x\in\mathcal{X} : P(x) > 0\}$ denotes the support of P, and in (3) and throughout $P_X P_{Y|X}$ denotes the PMF on $\mathcal{X}\times\mathcal{Y}$ that assigns (x,y) the probability $P_X(x)P_{Y|X}(y|x)$.

The Rényi divergence of order α [7,8] between two PMFs P and Q is defined for all positive α's other than one as

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

A conditional Rényi divergence can be defined in more than one way. In this paper, we consider the following three definitions, two classic and one new:

$$D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq \sum_{x\in\operatorname{supp}(P_X)} P(x)\,D_\alpha(P_{Y|X=x}\|Q_{Y|X=x}), \tag{5}$$

$$D_\alpha^{\mathsf{s}}(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq D_\alpha(P_X P_{Y|X}\|P_X Q_{Y|X}), \tag{6}$$

$$D_\alpha^{\mathsf{1}}(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq \frac{\alpha}{\alpha - 1}\log\sum_{x\in\operatorname{supp}(P_X)} P(x)\,2^{\frac{\alpha-1}{\alpha}D_\alpha(P_{Y|X=x}\|Q_{Y|X=x})}, \tag{7}$$

where (5) is inspired by Csiszár [9]; (6) is inspired by Sibson [10]; and (7) is motivated by the horse betting problem discussed in Section 9. The first two conditional Rényi divergences were used to define the Rényi measures of dependence of Csiszár $I_\alpha^{\mathsf{c}}(X;Y)$ [9] and of Sibson $I_\alpha^{\mathsf{s}}(X;Y)$ [10]:

$$I_\alpha^{\mathsf{c}}(X;Y) \triangleq \min_{Q_Y} D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_Y|P_X), \tag{8}$$

$$I_\alpha^{\mathsf{s}}(X;Y) \triangleq \min_{Q_Y} D_\alpha^{\mathsf{s}}(P_{Y|X}\|Q_Y|P_X), \tag{9}$$

where the minimization is over all PMFs on the set $\mathcal{Y}$. (Gallager's E_0 function [11] and $I_\alpha^{\mathsf{s}}(X;Y)$ are in one-to-one correspondence; see (65) below.) The analogous minimization of $D_\alpha^{\mathsf{1}}(\cdot)$ leads to the Lapidoth–Pfister mutual information $J_\alpha(X;Y)$ [3]:

$$J_\alpha(X;Y) \triangleq \min_{Q_X,Q_Y} D_\alpha(P_{XY}\|Q_X Q_Y) \tag{10}$$

$$= \min_{Q_Y} D_\alpha^{\mathsf{1}}(P_{Y|X}\|Q_Y|P_X), \tag{11}$$

where (11) is proved in Proposition 5.

The first part of the paper is structured as follows: In Section 2, we discuss some preliminaries. In Sections 3–5, we study the properties of the three conditional Rényi divergences and their associated measure of dependence. In Section 6, we express the Arimoto–Rényi conditional entropy $H_\alpha(X|Y)$ and the Arimoto measure of dependence $I_\alpha^{\mathsf{a}}(X;Y)$ [12] in terms of $D_\alpha^{\mathsf{1}}(P_{X|Y}\|U_X|P_Y)$. In Section 7,

we relate the conditional Rényi divergences to each other and discuss the relations between the Rényi dependence measures.

The second part of the paper deals with horse gambling under our proposed family of power-mean utility functions. It is in this context that the Rényi divergence (Theorem 9) and the conditional Rényi divergence $D_\alpha^1(\cdot)$ (Theorem 10) appear naturally.

More specifically, consider a horse race with a finite nonempty set of horses $\mathcal{X}$, where a bookmaker offers odds $o(x)$-for-1 on each horse $x \in \mathcal{X}$, where $o\colon \mathcal{X} \to (0, \infty)$ [2] (Section 6.1). A gambler spends all her wealth placing bets on the horses. The fraction of her wealth that she bets on Horse $x \in \mathcal{X}$ is denoted $b(x) \geq 0$, which sums up to 1 over $x \in \mathcal{X}$, and the PMF b is her "betting strategy." The winning horse, which we denote X, is drawn according to the PMF p, where we assume $p(x) > 0$ for all $x \in \mathcal{X}$. The wealth relative (or end-to-beginning wealth ratio) is the random variable

$$S \triangleq b(X)o(X). \tag{12}$$

Hence, given an initial wealth γ, the gambler's wealth after the race is γS. We seek betting strategies that maximize the utility function

$$U_\beta \triangleq \begin{cases} \frac{1}{\beta} \log \mathrm{E}[S^\beta] & \text{if } \beta \neq 0, \\ \mathrm{E}[\log S] & \text{if } \beta = 0, \end{cases} \tag{13}$$

where $\beta \in \mathbb{R}$ is a parameter that accounts for the risk sensitivity. This optimization generalizes the following cases:

(a) In the limit as β tends to $-\infty$, we optimize the worst-case return. The optimal strategy is risk-free in the sense that S does not depend on the winning horse (see Proposition 8).

(b) If $\beta = 0$, then we optimize $\mathrm{E}[\log S]$, which is known as the doubling rate [2] (Section 6.1). The optimal strategy is proportional betting, i.e., to choose $b = p$ (see Remark 4).

(c) If $\beta = 1$, then we optimize $\mathrm{E}[S]$, the expected return. The optimal strategy is to put all the money on a horse that maximizes $p(x)o(x)$ (see Proposition 9).

(d) In general, if $\beta \geq 1$, then it is optimal to put all the money on one horse (see Proposition 9). This is risky: if that horse loses, the gambler will go broke.

(e) In the limit as β tends to $+\infty$, we optimize the best-case return. The optimal strategy is to put all the money on a horse that maximizes $o(x)$ (see Proposition 10).

Note that, for $\beta \neq 0$ and $\eta \triangleq 1 - \beta$, maximizing U_β is equivalent to maximizing

$$\mathrm{E}\left[\frac{S^{1-\eta}}{1-\eta}\right], \tag{14}$$

which is known in the finance literature as Constant Relative Risk Aversion (CRRA) [13,14].

We refer to our utility function as "power mean" because it can be written as the logarithm of a weighted power mean [15,16]:

$$U_\beta = \log\left[\sum_x p(x)\left(b(x)o(x)\right)^\beta\right]^{\frac{1}{\beta}}. \tag{15}$$

Because the power mean tends to the geometric mean as β tends to zero [15] (Problem 8.1), U_β is continuous at $\beta = 0$:

$$\lim_{\beta \to 0} U_\beta = \log \prod_x \left(b(x)o(x)\right)^{p(x)} \tag{16}$$

$$= \mathrm{E}[\log S]. \tag{17}$$

Campbell [17,18] used an exponential cost function with a similar structure to (15) to provide an operational meaning to the Rényi entropy in source coding. Other information-theoretic applications of exponential moments were studied in [19].

The second part of the paper is structured as follows: In Section 8, we relate the utility function U_β to the Rényi divergence (Theorem 9) and derive its optimal gambling strategy. In Section 9, we consider the situation where the gambler observes side information prior to betting, a situation that leads to the conditional Rényi divergence $D^1_\alpha(\cdot)$ (Theorem 10) and to a new operational meaning for the measure of dependence $J_\alpha(X;Y)$ (Theorem 11). In Section 10, we consider the situation where the gambler invests only part of her money. In Section 11, we present a universal strategy for independent and identically distributed (IID) races that requires neither knowledge of the winning probabilities nor of the parameter β of the utility function and yet asymptotically maximizes the utility function for all PMFs p and all $\beta \in \mathbb{R}$.

2. 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 $\mathcal{Y}$. An expression of the form $P_X P_{Y|X}$ denotes the PMF on $\mathcal{X} \times \mathcal{Y}$ that assigns (x, y) the probability $P_X(x) P_{Y|X}(y|x)$. 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 $\sum_x$ for $\sum_{x \in \mathcal{X}}$, $\min_{Q_X, Q_Y}$ for $\min_{(Q_X, Q_Y) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})}$, $P(x)$ for $P_X(x)$, and $P(y|x)$ for $P_{Y|X}(y|x)$. When $P(x)$ is 0, we define the conditional probability $P(y|x)$ as $1/|\mathcal{Y}|$. The conditional distribution of Y given $X = x$ is denoted by $P_{Y|X=x}$, thus

$$P_{Y|X=x}(y) = P(y|x). \tag{18}$$

We denote by $\mathbb{1}\{\text{condition}\}$ the indicator function that is one if the condition is satisfied and zero otherwise.

In the definition of the Kullback–Leibler divergence in (1), we use the conventions

$$0 \log \frac{0}{q} = 0 \quad \forall q \geq 0, \qquad p \log \frac{p}{0} = \infty \quad \forall p > 0. \tag{19}$$

In the definition of the Rényi divergence in (4), we read $P(x)^\alpha Q(x)^{1-\alpha}$ as $P(x)^\alpha / Q(x)^{\alpha-1}$ for $\alpha > 1$ and use the conventions

$$\frac{0}{0} = 0, \qquad \frac{p}{0} = \infty \quad \forall p > 0. \tag{20}$$

For α being zero, one, or infinity, we define by continuous extension of (4)

$$D_0(P\|Q) \triangleq -\log \sum_{x \in \mathrm{supp}(P)} Q(x), \tag{21}$$

$$D_1(P\|Q) \triangleq D(P\|Q), \tag{22}$$

$$D_\infty(P\|Q) \triangleq \log \max_x \frac{P(x)}{Q(x)}. \tag{23}$$

The Rényi divergence for negative α is defined as

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha - 1} \log \sum_x \frac{Q(x)^{1-\alpha}}{P(x)^{-\alpha}}. \tag{24}$$

(We use negative α in the proof of Proposition 1 (e) below and in Remark 6. More about negative orders can be found in [8] (Section V). For other applications of negative orders, see [20] (Proof of Theorem 1 and Example 1).)

The Rényi divergence satisfies the following basic properties:

Proposition 1. *Let P and Q be PMFs. Then, the Rényi divergence $D_\alpha(P\|Q)$ satisfies the following:*

(a)   *For all $\alpha \in [0, \infty]$, $D_\alpha(P\|Q) \geq 0$. If $\alpha \in (0, \infty]$, then $D_\alpha(P\|Q) = 0$ if and only if $P = Q$.*

(b)   *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 $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$.*

(c)   *The mapping $\alpha \mapsto D_\alpha(P\|Q)$ is continuous on $[0, \infty]$.*

(d)   *The mapping $\alpha \mapsto D_\alpha(P\|Q)$ is nondecreasing on $[0, \infty]$.*

(e)   *The mapping $\alpha \mapsto \frac{1-\alpha}{\alpha} D_\alpha(P\|Q)$ is nonincreasing on $(0, \infty)$.*

(f)   *The mapping $\alpha \mapsto (1-\alpha) D_\alpha(P\|Q)$ is concave on $[0, \infty)$.*

(g)   *The mapping $\alpha \mapsto (\alpha - 1) D_{1/\alpha}(P\|Q)$ is concave on $(0, \infty)$.*

(h)   *(Data-processing inequality.) Let $A_{X'|X}$ be a conditional PMF, and define the PMFs*

$$P'(x') \triangleq \sum_x P(x) A_{X'|X}(x'|x), \tag{25}$$

$$Q'(x') \triangleq \sum_x Q(x) A_{X'|X}(x'|x). \tag{26}$$

Then, for all $\alpha \in [0, \infty]$,

$$D_\alpha(P'\|Q') \leq D_\alpha(P\|Q). \tag{27}$$

Proof. See Appendix A. $\square$

All three conditional Rényi divergences reduce to the unconditional Rényi divergence when both $P_{Y|X}$ and $Q_{Y|X}$ are independent of X:

Remark 1. *Let P_Y, Q_Y, and P_X be PMFs. Then, for all $\alpha \in [0, \infty]$,*

$$D_\alpha^c(P_Y\|Q_Y|P_X) = D_\alpha^s(P_Y\|Q_Y|P_X) = D_\alpha^1(P_Y\|Q_Y|P_X) = D_\alpha(P_Y\|Q_Y). \tag{28}$$

Proof. This follows from the definitions of $D_\alpha^c(\cdot)$, $D_\alpha^s(\cdot)$, and $D_\alpha^1(\cdot)$ in (5)–(7). $\square$

3. Csiszár's Conditional Rényi Divergence

For a PMF P_X and conditional PMFs $P_{Y|X}$ and $Q_{Y|X}$, Csiszár's conditional Rényi divergence $D_\alpha^c(\cdot)$ is defined for every $\alpha \in [0, \infty]$ as

$$D_\alpha^c(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq \sum_{x \in \mathrm{supp}(P_X)} P(x) D_\alpha(P_{Y|X=x}\|Q_{Y|X=x}). \tag{29}$$

For $\alpha \in (0, 1) \cup (1, \infty)$,

$$D_\alpha^c(P_{Y|X}\|Q_{Y|X}|P_X) = \frac{1}{\alpha - 1} \sum_{x \in \mathrm{supp}(P_X)} P(x) \log \sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha}, \tag{30}$$

which follows from the definition of the Rényi divergence in (4). For α being zero, one, or infinity, we obtain from (21)–(23) and (2)

$$D_0^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X) = - \sum_{x\in\mathrm{supp}(P_X)} P(x)\log \sum_{y\in\mathrm{supp}(P_{Y|X=x})} Q(y|x), \tag{31}$$

$$D_1^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X) = D(P_{Y|X}\|Q_{Y|X}|P_X), \tag{32}$$

$$D_\infty^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X) = \sum_{x\in\mathrm{supp}(P_X)} P(x)\log\max_y \frac{P(y|x)}{Q(y|x)}. \tag{33}$$

Augustin [21] and later Csiszár [9] defined the measure of dependence

$$I_\alpha^{\mathsf{c}}(X;Y) \triangleq \min_{Q_Y} D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_Y|P_X). \tag{34}$$

Augustin used this measure to study the error exponents for channel coding with input constraints, while Csiszár used it to study generalized cutoff rates for channel coding with composition constraints. Nakiboğlu [22] studied more properties of $I_\alpha^{\mathsf{c}}(X;Y)$. Inter alia, he analyzed the minimax properties of the Augustin capacity

$$\sup_{P_X\in\mathcal{A}} I_\alpha^{\mathsf{c}}(P_X, P_{Y|X}) = \sup_{P_X\in\mathcal{A}} \min_{Q_Y} D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_Y|P_X), \tag{35}$$

where $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ is a constraint set. The Augustin capacity is used in [23] to establish the sphere packing bound for memoryless channels with cost constraints.

The rest of the section presents some properties of $D_\alpha^{\mathsf{c}}(\cdot)$. Being an average of Rényi divergences (see (29)), $D_\alpha^{\mathsf{c}}(\cdot)$ inherits many properties from the Rényi divergence:

Proposition 2. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. Then,*

(a) *For all $\alpha \in [0,\infty]$, $D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X) \geq 0$. If $\alpha \in (0,\infty]$, then $D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X) = 0$ if and only if $\big(P_{Y|X=x} = Q_{Y|X=x}$ for all $x \in \mathrm{supp}(P_X)\big)$.*

(b) *For all $\alpha \in [0,1)$, $D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X)$ is finite if and only if $\big(|\mathrm{supp}(P_{Y|X=x}) \cap \mathrm{supp}(Q_{Y|X=x})| > 0$ for all $x \in \mathrm{supp}(P_X)\big)$. For all $\alpha \in [1,\infty]$, $D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X)$ is finite if and only if $\big(\mathrm{supp}(P_{Y|X=x}) \subseteq \mathrm{supp}(Q_{Y|X=x})$ for all $x \in \mathrm{supp}(P_X)\big)$.*

(c) *The mapping $\alpha \mapsto D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X)$ is continuous on $[0,\infty]$.*

(d) *The mapping $\alpha \mapsto D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X)$ is nondecreasing on $[0,\infty]$.*

(e) *The mapping $\alpha \mapsto \frac{1-\alpha}{\alpha} D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X)$ is nonincreasing on $(0,\infty)$.*

(f) *The mapping $\alpha \mapsto (1-\alpha) D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X)$ is concave on $[0,\infty)$.*

(g) *The mapping $\alpha \mapsto (\alpha-1) D_{1/\alpha}^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X)$ is concave on $(0,\infty)$.*

Proof. These follow from (29) and the properties of the Rényi divergence (Proposition 1). For Parts (f) and (g), recall that a nonnegative weighted sum of concave functions is concave. □

We next consider data-processing inequalities for $D_\alpha^{\mathsf{c}}(\cdot)$. We distinguish between processing Y and processing X. The data-processing inequality for processing Y follows from the data-processing inequality for the (unconditional) Rényi divergence:

Theorem 1. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. For a conditional PMF $A_{Y'|XY}$, define*

$$P_{Y'|X}(y'|x) \triangleq \sum_y P_{Y|X}(y|x) \, A_{Y'|XY}(y'|x,y), \tag{36}$$

$$Q_{Y'|X}(y'|x) \triangleq \sum_y Q_{Y|X}(y|x) \, A_{Y'|XY}(y'|x,y). \tag{37}$$

Then, for all $\alpha \in [0,\infty]$,

$$D_\alpha^{\mathsf{c}}(P_{Y'|X}\|Q_{Y'|X}|P_X) \leq D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X). \tag{38}$$

Proof. See Appendix B. □

The following data-processing inequality for processing X holds for $\alpha \in [0,1]$ (as shown in Example 1 below, it does not extend to $\alpha \in (1,\infty]$):

Theorem 2. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. For a conditional PMF $B_{X'|X}$, define the PMFs*

$$P_{X'}(x') \triangleq \sum_x P_X(x) \, B_{X'|X}(x'|x), \tag{39}$$

$$B_{X|X'}(x|x') \triangleq \begin{cases} P_X(x) \, B_{X'|X}(x'|x)/P_{X'}(x') & \text{if } P_{X'}(x') > 0, \\ 1/|\mathcal{X}| & \text{otherwise,} \end{cases} \tag{40}$$

$$P_{Y|X'}(y|x') \triangleq \sum_x B_{X|X'}(x|x') \, P_{Y|X}(y|x), \tag{41}$$

$$Q_{Y|X'}(y|x') \triangleq \sum_x B_{X|X'}(x|x') \, Q_{Y|X}(y|x). \tag{42}$$

Then, for all $\alpha \in [0,1]$,

$$D_\alpha^{\mathsf{c}}(P_{Y|X'}\|Q_{Y|X'}|P_{X'}) \leq D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X). \tag{43}$$

Note that $P_{X'}$, $P_{Y|X'}$, and $Q_{Y|X'}$ in Theorem 2 can be obtained from the following marginalizations:

$$P_{X'}(x') \, P_{Y|X'}(y|x') = \sum_x P_X(x) \, B_{X'|X}(x'|x) \, P_{Y|X}(y|x), \tag{44}$$

$$P_{X'}(x') \, Q_{Y|X'}(y|x') = \sum_x P_X(x) \, B_{X'|X}(x'|x) \, Q_{Y|X}(y|x). \tag{45}$$

Proof of Theorem 2. See Appendix C. □

As a special case of Theorem 2, we obtain the following relation between the conditional and the unconditional Rényi divergence:

Corollary 1. *For a PMF P_X and conditional PMFs $P_{Y|X}$ and $Q_{Y|X}$, define the marginal PMFs*

$$P_Y(y) \triangleq \sum_x P_X(x) \, P_{Y|X}(y|x), \tag{46}$$

$$Q_Y(y) \triangleq \sum_x P_X(x) \, Q_{Y|X}(y|x). \tag{47}$$

Then, for all $\alpha \in [0,1]$,

$$D_\alpha(P_Y\|Q_Y) \leq D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X). \tag{48}$$

Proof. See Appendix D. □

Consider next $\alpha \in (1, \infty]$. It turns out that Corollary 1, and hence Theorem 2, cannot be extended to these values of α (not even if $Q_{Y|X}$ is restricted to be independent of X, i.e., if $Q_{Y|X} = Q_Y$):

Example 1. *Let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. For $\epsilon \in (0, 1)$, define the PMFs P_X, $Q_Y^{(\epsilon)}$, and $P_{Y|X}^{(\epsilon)}$ as*

$$P_X(0) = 0.5, \qquad P_X(1) = 0.5, \tag{49}$$

$$Q_Y^{(\epsilon)}(0) = 1 - \epsilon, \qquad Q_Y^{(\epsilon)}(1) = \epsilon, \tag{50}$$

$$P_{Y|X}^{(\epsilon)}(0|0) = 1 - \epsilon, \qquad P_{Y|X}^{(\epsilon)}(1|0) = \epsilon, \tag{51}$$

$$P_{Y|X}^{(\epsilon)}(0|1) = \epsilon, \qquad P_{Y|X}^{(\epsilon)}(1|1) = 1 - \epsilon. \tag{52}$$

Then, for every $\alpha \in (1, \infty]$, there exists an $\epsilon \in (0, 1)$ such that

$$D_\alpha\big(P_Y \| Q_Y^{(\epsilon)}\big) > D_\alpha^c\big(P_{Y|X}^{(\epsilon)} \| Q_Y^{(\epsilon)} | P_X\big), \tag{53}$$

where the PMF P_Y is defined by (46) and, irrespective of ϵ, satisfies $P_Y(0) = P_Y(1) = 0.5$.

Proof. See Appendix E. □

4. Sibson's Conditional Rényi Divergence

For a PMF P_X and conditional PMFs $P_{Y|X}$ and $Q_{Y|X}$, Sibson's conditional Rényi divergence $D_\alpha^s(\cdot)$ is defined for every $\alpha \in [0, \infty]$ as

$$D_\alpha^s(P_{Y|X} \| Q_{Y|X} | P_X) \triangleq D_\alpha(P_X P_{Y|X} \| P_X Q_{Y|X}). \tag{54}$$

For $\alpha \in (0, 1) \cup (1, \infty)$,

$$D_\alpha^s(P_{Y|X} \| Q_{Y|X} | P_X) = \frac{1}{\alpha - 1} \log \sum_{x \in \text{supp}(P_X)} P(x) \sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \tag{55}$$

$$= \frac{1}{\alpha - 1} \log \sum_{x \in \text{supp}(P_X)} P(x) 2^{(\alpha-1)D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})}, \tag{56}$$

where (55) and (56) follow from the definition of the Rényi divergence in (4). For α being zero, one, or infinity, we obtain from (21)–(23) and (3)

$$D_0^s(P_{Y|X} \| Q_{Y|X} | P_X) = -\log \sum_{x \in \text{supp}(P_X)} P(x) \sum_{y \in \text{supp}(P_{Y|X=x})} Q(y|x), \tag{57}$$

$$D_1^s(P_{Y|X} \| Q_{Y|X} | P_X) = D(P_{Y|X} \| Q_{Y|X} | P_X), \tag{58}$$

$$D_\infty^s(P_{Y|X} \| Q_{Y|X} | P_X) = \log \max_{x \in \text{supp}(P_X)} \max_y \frac{P(y|x)}{Q(y|x)}. \tag{59}$$

Sibson [10] defined the measure of dependence

$$I_\alpha^s(X; Y) \triangleq \min_{Q_Y} D_\alpha^s(P_{Y|X} \| Q_Y | P_X). \tag{60}$$

This minimum can be computed explicitly [10] (Corollary 2.3): For $\alpha \in (0, 1) \cup (1, \infty)$,

$$I_\alpha^s(X; Y) = \frac{\alpha}{\alpha - 1} \log \sum_y \left[\sum_x P(x) P(y|x)^\alpha \right]^{\frac{1}{\alpha}}, \tag{61}$$

and for α being one or infinity,

$$I_1^{\mathrm{s}}(X;Y) = I(X;Y), \tag{62}$$

$$I_\infty^{\mathrm{s}}(X;Y) = \log \sum_y \max_x P(y|x), \tag{63}$$

where $I(X;Y)$ denotes Shannon's mutual information.

The concavity and convexity properties of $D_\alpha^{\mathrm{s}}(\cdot)$ and $I_\alpha^{\mathrm{s}}(X;Y)$ were studied by Ho–Verdú [24]. More properties of $I_\alpha^{\mathrm{s}}(X;Y)$ were collected by Verdú [25]. The maximization of $I_\alpha^{\mathrm{s}}(X;Y)$ with respect to P_X and the minimax properties of $D_\alpha^{\mathrm{s}}(\cdot)$ were studied by Nakiboğlu [26] and Cai–Verdú [27].

The conditional Rényi divergence $D_\alpha^{\mathrm{s}}(\cdot)$ was used by Fong and Tan [28] to establish strong converse theorems for multicast networks. Yu and Tan [29] analyzed channel resolvability, among other measures, in terms of $D_\alpha^{\mathrm{s}}(\cdot)$.

From (61) we see that Gallager's E_0 function [11], which is defined as

$$E_0(\rho, P_X, P_{Y|X}) \triangleq -\log \sum_y \left[\sum_x P(x) P(y|x)^{\frac{1}{1+\rho}} \right]^{1+\rho}, \tag{64}$$

is in one-to-one correspondence to Sibson's measure of dependence:

$$I_\alpha^{\mathrm{s}}(X;Y) = \frac{\alpha}{1-\alpha} E_0\left(\frac{1-\alpha}{\alpha}, P_X, P_{Y|X} \right). \tag{65}$$

Gallager's E_0 function is important in channel coding: it appears in the random coding exponent [30] and in the sphere packing exponent [31,32] (see also Gallager [11]). The exponential strong converse theorem proved by Arimoto [33] also uses the E_0 function. Polyanskiy and Verdú [34] extended the exponential strong converse theorem to channels with feedback. Augustin [21] and Nakiboğlu [35,36] extended the sphere packing bound to channels with feedback.

The rest of the section presents some properties of $D_\alpha^{\mathrm{s}}(\cdot)$. Because $D_\alpha^{\mathrm{s}}(\cdot)$ can be written as an (unconditional) Rényi divergence (see (54)), it inherits many properties from the Rényi divergence:

Proposition 3. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. Then,*

(a) *For all $\alpha \in [0, \infty]$, $D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X) \geq 0$. If $\alpha \in (0, \infty]$, then $D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X) = 0$ if and only if $\big(P_{Y|X=x} = Q_{Y|X=x}$ for all $x \in \mathrm{supp}(P_X) \big)$.*

(b) *For all $\alpha \in [0, 1)$, $D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X)$ is finite if and only if $\big($ there exists an $x \in \mathrm{supp}(P_X)$ such that $|\mathrm{supp}(P_{Y|X=x}) \cap \mathrm{supp}(Q_{Y|X=x})| > 0 \big)$. For all $\alpha \in [1, \infty]$, $D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X)$ is finite if and only if $\big(\mathrm{supp}(P_{Y|X=x}) \subseteq \mathrm{supp}(Q_{Y|X=x})$ for all $x \in \mathrm{supp}(P_X) \big)$.*

(c) *The mapping $\alpha \mapsto D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X)$ is continuous on $[0, \infty]$.*

(d) *The mapping $\alpha \mapsto D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X)$ is nondecreasing on $[0, \infty]$.*

(e) *The mapping $\alpha \mapsto \frac{1-\alpha}{\alpha} D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X)$ is nonincreasing on $(0, \infty)$.*

(f) *The mapping $\alpha \mapsto (1-\alpha) D_\alpha^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X)$ is concave on $[0, \infty)$.*

(g) *The mapping $\alpha \mapsto (\alpha - 1) D_{1/\alpha}^{\mathrm{s}}(P_{Y|X} \| Q_{Y|X} | P_X)$ is concave on $(0, \infty)$.*

Proof. These follow from (54) and the properties of the Rényi divergence (Proposition 1). □

We next consider data-processing inequalities for $D_\alpha^{\mathrm{s}}(\cdot)$. We distinguish between processing Y and processing X. The data-processing inequality for processing Y follows from the data-processing inequality for the (unconditional) Rényi divergence:

Theorem 3. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. For a conditional PMF $A_{Y'|XY}$, define*

$$P_{Y'|X}(y'|x) \triangleq \sum_y P_{Y|X}(y|x) A_{Y'|XY}(y'|x,y), \tag{66}$$

$$Q_{Y'|X}(y'|x) \triangleq \sum_y Q_{Y|X}(y|x) A_{Y'|XY}(y'|x,y). \tag{67}$$

Then, for all $\alpha \in [0, \infty]$,

$$D_\alpha^{\mathsf{s}}(P_{Y'|X} \| Q_{Y'|X} | P_X) \leq D_\alpha^{\mathsf{s}}(P_{Y|X} \| Q_{Y|X} | P_X). \tag{68}$$

Proof. See Appendix F. □

The data-processing inequality for processing X similarly follows from the data-processing inequality for the (unconditional) Rényi divergence:

Theorem 4. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. For a conditional PMF $B_{X'|X}$, define the PMFs*

$$P_{X'}(x') \triangleq \sum_x P_X(x) B_{X'|X}(x'|x), \tag{69}$$

$$B_{X|X'}(x|x') \triangleq \begin{cases} P_X(x) B_{X'|X}(x'|x) / P_{X'}(x') & \text{if } P_{X'}(x') > 0, \\ 1/|\mathcal{X}| & \text{otherwise,} \end{cases} \tag{70}$$

$$P_{Y|X'}(y|x') \triangleq \sum_x B_{X|X'}(x|x') P_{Y|X}(y|x), \tag{71}$$

$$Q_{Y|X'}(y|x') \triangleq \sum_x B_{X|X'}(x|x') Q_{Y|X}(y|x). \tag{72}$$

Then, for all $\alpha \in [0, \infty]$,

$$D_\alpha^{\mathsf{s}}(P_{Y|X'} \| Q_{Y|X'} | P_{X'}) \leq D_\alpha^{\mathsf{s}}(P_{Y|X} \| Q_{Y|X} | P_X). \tag{73}$$

Proof. See Appendix G. □

As a special case of Theorem 4, we obtain the following relation between the conditional and the unconditional Rényi divergence:

Corollary 2. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. Define the marginal PMFs*

$$P_Y(y) \triangleq \sum_x P_X(x) P_{Y|X}(y|x), \tag{74}$$

$$Q_Y(y) \triangleq \sum_x P_X(x) Q_{Y|X}(y|x). \tag{75}$$

Then, for all $\alpha \in [0, \infty]$,

$$D_\alpha(P_Y \| Q_Y) \leq D_\alpha^{\mathsf{s}}(P_{Y|X} \| Q_{Y|X} | P_X). \tag{76}$$

Proof. This follows from Theorem 4 in the same way that Corollary 1 followed from Theorem 2. □

5. New Conditional Rényi Divergence

Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. For $\alpha \in (0,1) \cup (1,\infty)$, define

$$D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq \frac{\alpha}{\alpha-1} \log \sum_{x \in \text{supp}(P_X)} P(x) 2^{\frac{\alpha-1}{\alpha} D_\alpha(P_{Y|X=x}\|Q_{Y|X=x})} \tag{77}$$

$$= \frac{\alpha}{\alpha-1} \log \sum_{x \in \text{supp}(P_X)} P(x) \left[\sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \right]^{\frac{1}{\alpha}}, \tag{78}$$

where (78) follows from the definition of the Rényi divergence in (4). (Except for the sign, the exponential averaging in (77) is very similar to the one of the Arimoto–Rényi conditional entropy; compare with (147) below.) For α being zero, one, or infinity, we define by continuous extension of (77)

$$D_0^1(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq -\log \max_{x \in \text{supp}(P_X)} \sum_{y \in \text{supp}(P_{Y|X=x})} Q(y|x), \tag{79}$$

$$D_1^1(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq D(P_{Y|X}\|Q_{Y|X}|P_X), \tag{80}$$

$$D_\infty^1(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq \log \sum_{x \in \text{supp}(P_X)} P(x) \max_y \frac{P(y|x)}{Q(y|x)}. \tag{81}$$

This conditional Rényi divergence has an operational meaning in horse betting with side information (see Theorem 10 below). Before discussing the measure of dependence associated with $D_\alpha^1(\cdot)$, we establish the following alternative characterization of $D_\alpha^1(\cdot)$:

Proposition 4. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. Then, for all $\alpha \in [0,\infty]$,*

$$D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X) = \min_{Q_X} D_\alpha(P_X P_{Y|X}\|Q_X Q_{Y|X}). \tag{82}$$

Proof. We first treat the case $\alpha \in (0,1) \cup (1,\infty)$. Some algebra reveals that, for every PMF Q_X,

$$D_\alpha(P_X P_{Y|X}\|Q_X Q_{Y|X}) = D_\alpha\big(Q_X^{*(\alpha)}\|Q_X\big) + \frac{\alpha}{\alpha-1} \log \sum_{x \in \text{supp}(P_X)} P(x) \left[\sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \right]^{\frac{1}{\alpha}}, \tag{83}$$

where the PMF $Q_X^{*(\alpha)}$ is defined as

$$Q_X^{*(\alpha)}(x) \triangleq \frac{P(x) \left[\sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha}\right]^{1/\alpha}}{\sum_{x' \in \text{supp}(P_X)} P(x') \left[\sum_y P(y|x')^\alpha Q(y|x')^{1-\alpha}\right]^{1/\alpha}}. \tag{84}$$

The right-hand side (RHS) of (82) is thus equal to the minimum over Q_X of the RHS of (83). Since $D_\alpha\big(Q_X^{*(\alpha)}\|Q_X\big) \geq 0$ with equality if $Q_X = Q_X^{*(\alpha)}$ (Proposition 1 (a)), this minimum is equal to the second term on the RHS of (83), which, by (78), equals $D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X)$.

For $\alpha = 1$ and $\alpha = \infty$, (82) follows from the same argument using that, for every PMF Q_X,

$$D_1(P_X P_{Y|X}\|Q_X Q_{Y|X}) = D(P_X\|Q_X) + D(P_{Y|X}\|Q_{Y|X}|P_X), \tag{85}$$

$$D_\infty(P_X P_{Y|X}\|Q_X Q_{Y|X}) = D_\infty\big(Q_X^{*(\infty)}\|Q_X\big) + \log \sum_{x \in \text{supp}(P_X)} P(x) \max_y \frac{P(y|x)}{Q(y|x)}, \tag{86}$$

where the PMF $Q_X^{*(\infty)}$ is defined as

$$Q_X^{*(\infty)}(x) \triangleq \frac{P(x) \max_y \left[P(y|x)/Q(y|x) \right]}{\sum_{x' \in \text{supp}(P_X)} P(x') \max_y \left[P(y|x')/Q(y|x') \right]}. \tag{87}$$

For $\alpha = 0$, (82) holds because

$$\min_{Q_X} D_0(P_X P_{Y|X} \| Q_X Q_{Y|X}) = \min_{Q_X} -\log \sum_{x \in \text{supp}(P_X)} Q(x) \sum_{y \in \text{supp}(P_{Y|X=x})} Q(y|x) \tag{88}$$

$$= -\log \max_{Q_X} \sum_{x \in \text{supp}(P_X)} Q(x) \sum_{y \in \text{supp}(P_{Y|X=x})} Q(y|x) \tag{89}$$

$$= -\log \max_{x \in \text{supp}(P_X)} \sum_{y \in \text{supp}(P_{Y|X=x})} Q(y|x) \tag{90}$$

$$= D_0^1(P_{Y|X} \| Q_{Y|X} | P_X), \tag{91}$$

where (88) follows from the definition of $D_0(P\|Q)$ in (21), and (91) follows from (79). □

Tomamichel and Hayashi [37] and Lapidoth and Pfister [3] independently introduced and studied the dependence measure

$$J_\alpha(X;Y) \triangleq \min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X Q_Y). \tag{92}$$

(For some measure-theoretic properties of $J_\alpha(X;Y)$, see Aishwarya–Madiman [38].) The measure $J_\alpha(X;Y)$ can be related to the error exponents in a hypothesis testing problem where the samples are either from a known joint distribution or an unknown product distribution (see [37] (Equation (57)) and [39]). It also appears in horse betting with side information (see Theorem 11 below).

Similar to $I_\alpha^c(X;Y)$ in (34) and $I_\alpha^s(X;Y)$ in (60), the measure $J_\alpha(X;Y)$ can be expressed as a minimization involving the new conditional Rényi divergence:

Proposition 5. *Let P_{XY} be a joint PMF. Denote its marginal PMFs by P_X and P_Y and its conditional PMFs by $P_{Y|X}$ and $P_{X|Y}$, so $P_{XY} = P_X P_{Y|X} = P_Y P_{X|Y}$. Then, for all $\alpha \in [0, \infty]$,*

$$J_\alpha(X;Y) = \min_{Q_Y} D_\alpha^1(P_{Y|X} \| Q_Y | P_X) \tag{93}$$

$$= \min_{Q_X} D_\alpha^1(P_{X|Y} \| Q_X | P_Y). \tag{94}$$

Proof. Equation (93) holds because

$$\min_{Q_Y} D_\alpha^1(P_{Y|X} \| Q_Y | P_X) = \min_{Q_Y} \min_{Q_X} D_\alpha(P_X P_{Y|X} \| Q_X Q_Y) \tag{95}$$

$$= J_\alpha(X;Y), \tag{96}$$

where (95) follows from Proposition 4, and (96) follows from (92). Swapping the roles of X and Y establishes (94):

$$\min_{Q_X} D_\alpha^1(P_{X|Y} \| Q_X | P_Y) = \min_{Q_X} \min_{Q_Y} D_\alpha(P_Y P_{X|Y} \| Q_Y Q_X) \tag{97}$$

$$= J_\alpha(X;Y), \tag{98}$$

where (97) follows from Proposition 4, and (98) follows from (92). □

The rest of the section presents some properties of $D_\alpha^1(\cdot)$.

Proposition 6. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. Then,*

(a) *For all $\alpha \in [0, \infty]$, $D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) \geq 0$. If $\alpha \in (0, \infty]$, then $D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = 0$ if and only if $\left(P_{Y|X=x} = Q_{Y|X=x} \text{ for all } x \in \mathrm{supp}(P_X) \right)$.*

(b) *For all $\alpha \in [0, 1)$, $D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is finite if and only if (there exists an $x \in \mathrm{supp}(P_X)$ such that $|\mathrm{supp}(P_{Y|X=x}) \cap \mathrm{supp}(Q_{Y|X=x})| > 0$). For all $\alpha \in [1, \infty]$, $D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is finite if and only if $\left(\mathrm{supp}(P_{Y|X=x}) \subseteq \mathrm{supp}(Q_{Y|X=x}) \text{ for all } x \in \mathrm{supp}(P_X) \right)$.*

(c) *The mapping $\alpha \mapsto D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is continuous on $[0, \infty]$.*

(d) *The mapping $\alpha \mapsto D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is nondecreasing on $[0, \infty]$.*

(e) *The mapping $\alpha \mapsto \frac{1-\alpha}{\alpha} D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is nonincreasing on $(0, \infty)$.*

(f) *The mapping $\alpha \mapsto (1 - \alpha) D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is concave on $[0, 1]$.*

(g) *The mapping $\alpha \mapsto (\alpha - 1) D_{1/\alpha}^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is concave on $[1, \infty)$.*

Proof. We prove these properties as follows:

(a) For all $\alpha \in [0, \infty]$, Proposition 4 implies

$$D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \min_{Q_X} D_\alpha(P_X P_{Y|X} \| Q_X Q_{Y|X}). \tag{99}$$

The nonnegativity of $D_\alpha^1(\cdot)$ now follows from the nonnegativity of the Rényi divergence (Proposition 1 (a)). If $\left(P_{Y|X=x} = Q_{Y|X=x} \text{ for all } x \in \mathrm{supp}(P_X) \right)$, then $P_X P_{Y|X} = P_X Q_{Y|X}$. Hence, using $Q_X = P_X$ on the RHS of (99), $D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ equals zero. Conversely, if $\alpha \in (0, \infty]$ and $D_\alpha^1(\cdot) = 0$, then $P_X P_{Y|X} = Q_X Q_{Y|X}$ for some Q_X by Proposition 1 (a), which implies $\left(P_{Y|X=x} = Q_{Y|X=x} \text{ for all } x \in \mathrm{supp}(P_X) \right)$.

(b) This follows from the definitions in (77) and (79)–(81) and the conventions in (20).

(c) For $\alpha \in (0, 1) \cup (1, \infty)$, $D_\alpha^1(\cdot)$ is continuous because it is, by its definition in (77), a composition of continuous functions. The continuity at $\alpha = 1$ follows from a careful application of L'Hôpital's rule.

We next consider the continuity at $\alpha = 0$. Define $\tau \triangleq \min_{x \in \mathrm{supp}(P_X)} P(x)$. Then, for all $\alpha \in (0, 1)$,

$$(\alpha - 1) D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \alpha \log \sum_{x \in \mathrm{supp}(P_X)} P(x) 2^{\frac{\alpha-1}{\alpha} D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})} \tag{100}$$

$$\geq \alpha \log \sum_{x \in \mathrm{supp}(P_X)} \tau 2^{\frac{\alpha-1}{\alpha} D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})} \tag{101}$$

$$\geq \alpha \log \max_{x \in \mathrm{supp}(P_X)} \tau 2^{\frac{\alpha-1}{\alpha} D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})} \tag{102}$$

$$= \alpha \log \tau + \max_{x \in \mathrm{supp}(P_X)} (\alpha - 1) D_\alpha(P_{Y|X=x} \| Q_{Y|X=x}), \tag{103}$$

where (100) follows from the definition in (77). On the other hand, for all $\alpha \in (0, 1)$,

$$(\alpha - 1) D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \alpha \log \sum_{x \in \mathrm{supp}(P_X)} P(x) 2^{\frac{\alpha-1}{\alpha} D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})} \tag{104}$$

$$\leq \alpha \log \max_{x \in \mathrm{supp}(P_X)} 2^{\frac{\alpha-1}{\alpha} D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})} \tag{105}$$

$$= \max_{x \in \mathrm{supp}(P_X)} (\alpha - 1) D_\alpha(P_{Y|X=x} \| Q_{Y|X=x}). \tag{106}$$

Because $\lim_{\alpha \to 0} \alpha \log \tau = 0$, it follows from (103) and (106) and the sandwich theorem that

$$\lim_{\alpha \downarrow 0} D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha - 1} \max_{x \in \mathrm{supp}(P_X)} (\alpha - 1) D_\alpha(P_{Y|X=x} \| Q_{Y|X=x}) \tag{107}$$

$$= -\log \max_{x \in \mathrm{supp}(P_X)} \sum_{y \in \mathrm{supp}(P_{Y|X=x})} Q(y|x), \tag{108}$$

where (108) follows from the continuity of the Rényi divergence (Proposition 1 (c)) and the definition of $D_0(P \| Q)$ in (21).

We conclude with the continuity at $\alpha = \infty$. Observe that

$$\lim_{\alpha \to \infty} D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \lim_{\alpha \to \infty} \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathrm{supp}(P_X)} P(x) 2^{\frac{\alpha-1}{\alpha} D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})} \tag{109}$$

$$= \log \sum_{x \in \mathrm{supp}(P_X)} P(x) 2^{\lim_{\alpha \to \infty} D_\alpha(P_{Y|X=x} \| Q_{Y|X=x})} \tag{110}$$

$$= \log \sum_{x \in \mathrm{supp}(P_X)} P(x) \max_y \frac{P(y|x)}{Q(y|x)}, \tag{111}$$

where (109) follows from the definition in (77), and (111) follows from the continuity of the Rényi divergence (Proposition 1 (c)) and the definition of $D_\infty(P \| Q)$ in (23).

(d)   For all $\alpha \in [0, \infty]$, Proposition 4 implies

$$D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \min_{Q_X} D_\alpha(P_X P_{Y|X} \| Q_X Q_{Y|X}). \tag{112}$$

Because $\alpha \mapsto D_\alpha(P \| Q)$ is nonincreasing on $[0, \infty]$ (Proposition 1 (d)) and because the pointwise minimum preserves the monotonicity, the mapping $\alpha \mapsto D_\alpha^1(\cdot)$ is nonincreasing on $[0, \infty]$.

(e)   By Proposition 4,

$$\frac{1 - \alpha}{\alpha} D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \begin{cases} \min_{Q_X} \frac{1-\alpha}{\alpha} D_\alpha(P_X P_{Y|X} \| Q_X Q_{Y|X}) & \text{if } \alpha \in (0, 1], \\ \max_{Q_X} \frac{1-\alpha}{\alpha} D_\alpha(P_X P_{Y|X} \| Q_X Q_{Y|X}) & \text{if } \alpha \in (1, \infty). \end{cases} \tag{113}$$

By the nonnegativity of the Rényi divergence (Proposition 1 (a)), the RHS of (113) is nonnegative for $\alpha \in (0, 1]$ and nonpositive for $\alpha \in (1, \infty)$. Hence, it suffices to show separately that the mapping $\alpha \mapsto \frac{1-\alpha}{\alpha} D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is nonincreasing on $(0, 1]$ and on $(1, \infty)$. This is indeed the case: the mapping $\alpha \mapsto \frac{1-\alpha}{\alpha} D_\alpha(P_X P_{Y|X} \| Q_X Q_{Y|X})$ on the RHS of (113) is nonincreasing on $(0, \infty)$ (Proposition 1 (e)), and the monotonicity is preserved by the pointwise minimum and maximum, respectively.

(f)   For $\alpha \in [0, 1]$, Proposition 4 implies that

$$(1 - \alpha) D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X) = \min_{Q_X} \big[(1 - \alpha) D_\alpha(P_X P_{Y|X} \| Q_X Q_{Y|X}) \big]. \tag{114}$$

Because $\alpha \mapsto (1 - \alpha) D_\alpha(P_X P_{Y|X} \| Q_X Q_{Y|X})$ is concave on $[0, 1]$ (Proposition 1 (f)) and because the pointwise minimum preserves the concavity, the mapping $\alpha \mapsto (1 - \alpha) D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X)$ is concave on $[0, 1]$.

(g)   This follows from Proposition 1 (g) in the same way that Part (f) followed from Proposition 1 (f). □

We next consider data-processing inequalities for $D_\alpha^1(\cdot)$. We distinguish between processing Y and processing X. The data-processing inequality for processing Y follows from the data-processing inequality for the (unconditional) Rényi divergence:

Theorem 5. *Let P_X be a PMF, and let $P_{Y|X}$ and $Q_{Y|X}$ be conditional PMFs. For a conditional PMF $A_{Y'|XY}$, define*

$$P_{Y'|X}(y'|x) \triangleq \sum_{y} P_{Y|X}(y|x) A_{Y'|XY}(y'|x,y), \tag{115}$$

$$Q_{Y'|X}(y'|x) \triangleq \sum_{y} Q_{Y|X}(y|x) A_{Y'|XY}(y'|x,y). \tag{116}$$

Then, for all $\alpha \in [0, \infty]$,

$$D_\alpha^1(P_{Y'|X}\|Q_{Y'|X}|P_X) \leq D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X). \tag{117}$$

Proof. We prove (117) for $\alpha \in (0,1) \cup (1,\infty)$; the claim will then extend to $\alpha \in [0,\infty]$ by the continuity of $D_\alpha^1(\cdot)$ in α (Proposition 6 (c)). For every $x \in \mathrm{supp}(P_X)$, we can apply Proposition 1 (h) with the substitution of $A_{Y'|Y,X=x}$ for $A_{Y'|Y}$ to obtain

$$D_\alpha(P_{Y'|X=x}\|Q_{Y'|X=x}) \leq D_\alpha(P_{Y|X=x}\|Q_{Y|X=x}). \tag{118}$$

For $\alpha \in (0,1) \cup (1,\infty)$, (117) now follows from (77) and (118). □

Processing X is different. Consider first $Q_{Y|X}$ that does not depend on X. Then, writing $Q_{Y|X} = Q_Y$, we have the following result (which, as shown in Example 2 below, does not extend to general $Q_{Y|X}$):

Theorem 6. *Let P_X and Q_Y be PMFs, and let $P_{Y|X}$ be a conditional PMF. For a conditional PMF $B_{X'|X}$, define the PMFs*

$$P_{X'}(x') \triangleq \sum_{x} P_X(x) B_{X'|X}(x'|x), \tag{119}$$

$$B_{X|X'}(x|x') \triangleq \begin{cases} P_X(x) B_{X'|X}(x'|x)/P_{X'}(x') & \text{if } P_{X'}(x') > 0, \\ 1/|\mathcal{X}| & \text{otherwise,} \end{cases} \tag{120}$$

$$P_{Y|X'}(y|x') \triangleq \sum_{x} B_{X|X'}(x|x') P_{Y|X}(y|x). \tag{121}$$

Then, for all $\alpha \in [0, \infty]$,

$$D_\alpha^1(P_{Y|X'}\|Q_Y|P_{X'}) \leq D_\alpha^1(P_{Y|X}\|Q_Y|P_X). \tag{122}$$

Once we provide the operational meaning of $D_\alpha^1(\cdot)$ in horse betting with side information (Theorem 10 below), Theorem 6 will become very intuitive: it expresses the fact that preprocessing the side information cannot increase the gambler's utility; see Remark 8. Note that $P_{X'}$ and $P_{Y|X'}$ in Theorem 6 can be obtained from the following marginalization:

$$P_{X'}(x') P_{Y|X'}(y|x') = \sum_{x} P_X(x) B_{X'|X}(x'|x) P_{Y|X}(y|x). \tag{123}$$

Proof of Theorem 6. We show (122) for $\alpha \in (0,1) \cup (1,\infty)$; the claim will then extend to $\alpha \in [0,\infty]$ by the continuity of $D_\alpha^1(\cdot)$ in α (Proposition 6 (c)). Consider first $\alpha \in (1,\infty)$. Then, (122) holds because

$$\frac{\alpha - 1}{\alpha} D_\alpha^1(P_{Y|X'}\|Q_Y|P_{X'})$$

$$= \log \sum_{x' \in \mathrm{supp}(P_{X'})} P_{X'}(x') \left[\sum_{y} P_{Y|X'}(y|x')^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{124}$$

$$= \log \sum_{x' \in \text{supp}(P_{X'})} P_{X'}(x') \left[\sum_y \left[\sum_x B_{X|X'}(x|x') P_{Y|X}(y|x) Q_Y(y)^{\frac{1-\alpha}{\alpha}} \right]^\alpha \right]^{\frac{1}{\alpha}} \tag{125}$$

$$= \log \sum_{x' \in \text{supp}(P_{X'})} \left[\sum_y \left[\sum_{x \in \text{supp}(P_X)} P_X(x) B_{X'|X}(x'|x) P_{Y|X}(y|x) Q_Y(y)^{\frac{1-\alpha}{\alpha}} \right]^\alpha \right]^{\frac{1}{\alpha}} \tag{126}$$

$$\leq \log \sum_{x' \in \text{supp}(P_{X'})} \sum_{x \in \text{supp}(P_X)} \left[\sum_y \left[P_X(x) B_{X'|X}(x'|x) P_{Y|X}(y|x) Q_Y(y)^{\frac{1-\alpha}{\alpha}} \right]^\alpha \right]^{\frac{1}{\alpha}} \tag{127}$$

$$= \log \sum_{x \in \text{supp}(P_X)} P_X(x) \left[\sum_{x' \in \text{supp}(P_{X'})} B_{X'|X}(x'|x) \right] \left[\sum_y P_{Y|X}(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{128}$$

$$= \log \sum_{x \in \text{supp}(P_X)} P_X(x) \left[\sum_y P_{Y|X}(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{129}$$

$$= \frac{\alpha - 1}{\alpha} D_\alpha^1(P_{Y|X} \| Q_Y | P_X), \tag{130}$$

where (124) follows from (78); (125) follows from (121); (126) follows from (120); (127) follows from the Minkowski inequality [16] (III 2.4 Theorem 9); (129) holds because $P_X(x) > 0$ and $P_{X'}(x') = 0$ imply $B_{X'|X}(x'|x) = 0$, hence the first expression in square brackets on the left-hand side (LHS) of (129) equals one; and (130) follows from (78).

The proof for $\alpha \in (0,1)$ is very similar: (124)–(126) and (128)–(130) continue to hold, and (127) is reversed [16] (III 2.4 Theorem 9). Because now $\frac{\alpha-1}{\alpha} < 0$, (122) continues to hold for $\alpha \in (0,1)$. □

As a special case of Theorem 6, we obtain the following relation between the conditional and the unconditional Rényi divergence:

Corollary 3. *Let P_X and Q_Y be PMFs, and let $P_{Y|X}$ be a conditional PMF. Define the marginal PMF*

$$P_Y(y) \triangleq \sum_x P_X(x) P_{Y|X}(y|x). \tag{131}$$

Then, for all $\alpha \in [0, \infty]$,

$$D_\alpha(P_Y \| Q_Y) \leq D_\alpha^1(P_{Y|X} \| Q_Y | P_X). \tag{132}$$

Proof. This follows from Theorem 6 in the same way that Corollary 1 followed from Theorem 2. □

Consider next $Q_{Y|X}$ that does depend on X. It turns out that Corollary 3, and hence Theorem 6, cannot be extended to this setting:

Example 2. *Let $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1,2\}$. Define the PMFs P_X, $P_{Y|X}$, and $Q_{Y|X}$ as*

$$P_X(0) = 0.5, \qquad P_X(1) = 0.5, \tag{133}$$

$$P_{Y|X}(0|0) = 0.96, \qquad P_{Y|X}(1|0) = 0.02, \qquad P_{Y|X}(2|0) = 0.02, \tag{134}$$

$$P_{Y|X}(0|1) = 0.12, \qquad P_{Y|X}(1|1) = 0.02, \qquad P_{Y|X}(2|1) = 0.86, \tag{135}$$

$$Q_{Y|X}(0|0) = 0.06, \qquad Q_{Y|X}(1|0) = 0.92, \qquad Q_{Y|X}(2|0) = 0.02, \tag{136}$$

$$Q_{Y|X}(0|1) = 0.02, \qquad Q_{Y|X}(1|1) = 0.16, \qquad Q_{Y|X}(2|1) = 0.82. \tag{137}$$

Then, for $\alpha = 0.5$ and for $\alpha = 2$,

$$D_\alpha(P_Y \| Q_Y) > D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X), \tag{138}$$

where the PMFs P_Y and Q_Y are given by

$$P_Y(y) \triangleq \sum_x P_X(x) P_{Y|X}(y|x), \tag{139}$$

$$Q_Y(y) \triangleq \sum_x P_X(x) Q_{Y|X}(y|x). \tag{140}$$

Proof. Numerically, $D_{0.5}(P_Y \| Q_Y) \approx 1.11$ bits, which is larger than $D_{0.5}^1(P_{Y|X} \| Q_{Y|X} | P_X) \approx 0.93$ bits. Similarly, $D_2(P_Y \| Q_Y) \approx 2.95$ bits, which is larger than $D_2^1(P_{Y|X} \| Q_{Y|X} | P_X) \approx 2.75$ bits. $\square$

6. Relation to Arimoto's Measures

Before discussing Arimoto's measures, we first recall the definition of the Rényi entropy. The Rényi entropy of order α [7] is defined for all positive α's other than one as

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha} \log \sum_x P(x)^\alpha. \tag{141}$$

For α being zero, one, or infinity, we define by continuous extension of (141)

$$H_0(X) \triangleq \log |\mathrm{supp}(P_X)|, \tag{142}$$

$$H_1(X) \triangleq H(X), \tag{143}$$

$$H_\infty(X) \triangleq -\log \max_x P(x), \tag{144}$$

where $H(X)$ denotes Shannon's entropy. The Rényi entropy can be related to the Rényi divergence as follows:

$$H_\alpha(X) = \log |\mathcal{X}| - D_\alpha(P_X \| U_X), \tag{145}$$

where U_X denotes the uniform distribution over $\mathcal{X}$.

There are different ways to define a conditional Rényi entropy [40]; we use Arimoto's proposal. The Arimoto–Rényi conditional entropy of order α [12,38,40,41] is defined for positive α other than one as

$$H_\alpha(X|Y) \triangleq \frac{\alpha}{1-\alpha} \log \sum_{y \in \mathrm{supp}(P_Y)} P(y) \left[\sum_x P(x|y)^\alpha \right]^{\frac{1}{\alpha}} \tag{146}$$

$$= \frac{\alpha}{1-\alpha} \log \sum_{y \in \mathrm{supp}(P_Y)} P(y) 2^{\frac{1-\alpha}{\alpha} H_\alpha(P_{X|Y=y})}, \tag{147}$$

where (147) follows from the definition of the Rényi entropy in (141). The Arimoto–Rényi conditional entropy plays a key role in guessing with side information [20,42–44] and in task encoding with side information [45]; and it can be related to hypothesis testing [41]. For α being zero, one, or infinity, we define by continuous extension of (146)

$$H_0(X|Y) \triangleq \log \max_{y \in \mathrm{supp}(P_Y)} |\mathrm{supp}(P_{X|Y=y})|, \tag{148}$$

$$H_1(X|Y) \triangleq H(X|Y), \tag{149}$$

$$H_\infty(X|Y) \triangleq -\log \sum_{y \in \mathrm{supp}(P_Y)} P(y) \max_x P(x|y), \tag{150}$$

where $H(X|Y)$ denotes Shannon's conditional entropy. The analog of (145) for $H_\alpha(X|Y)$ is:

Remark 2. *For all* $\alpha \in [0, \infty]$,

$$H_\alpha(X|Y) = \log|\mathcal{X}| - D_\alpha^1(P_{X|Y}\|U_X|P_Y) \tag{151}$$

$$= \log|\mathcal{X}| - \min_{Q_Y} D_\alpha(P_Y P_{X|Y}\|Q_Y U_X). \tag{152}$$

Proof. Equation (151) follows, using some algebra, from the definition of $D_\alpha^1(\cdot)$ in (78)–(81); and (152) follows from Proposition 4. (The characterization in (152) previously appeared as [40] (Theorem 4).) □

Arimoto [12] also defined the following measure of dependence:

$$I_\alpha^{\mathsf{a}}(X;Y) \triangleq H_\alpha(X) - H_\alpha(X|Y) \tag{153}$$

$$= \frac{\alpha}{\alpha - 1} \log \sum_y \left[\sum_x \frac{P(x)^\alpha}{\sum_{x' \in \mathcal{X}} P(x')^\alpha} P(y|x)^\alpha \right]^{\frac{1}{\alpha}}, \tag{154}$$

where (154) follows from (141) and (146). Using Remark 2, we can express $I_\alpha^{\mathsf{a}}(X;Y)$ in terms of $D_\alpha^1(\cdot)$:

Remark 3. *For all* $\alpha \in [0, \infty]$,

$$I_\alpha^{\mathsf{a}}(X;Y) = D_\alpha^1(P_{X|Y}\|U_X|P_Y) - D_\alpha(P_X\|U_X). \tag{155}$$

Proof. This follows from (145), (151), and (153). □

7. Relations Between the Conditional Rényi Divergences and the Rényi Dependence Measures

In this section, we first establish the greater-or-equal-than order between the conditional Rényi divergences, where the order depends on whether $\alpha \in [0, 1]$ or $\alpha \in [1, \infty]$. We then show that this implies the same order between the dependence measures derived from the conditional Rényi divergences. Finally, we remark that many of the dependence measures coincide when they are maximized over all PMFs P_X.

Proposition 7. *For all* $\alpha \in [0, \infty]$,

$$D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X) \leq D_\alpha^{\mathsf{s}}(P_{Y|X}\|Q_{Y|X}|P_X). \tag{156}$$

Proof. This holds because

$$D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X) = \min_{Q_X} D_\alpha(P_X P_{Y|X}\|Q_X Q_{Y|X}) \tag{157}$$

$$\leq D_\alpha(P_X P_{Y|X}\|P_X Q_{Y|X}) \tag{158}$$

$$= D_\alpha^{\mathsf{s}}(P_{Y|X}\|Q_{Y|X}|P_X), \tag{159}$$

where (157) follows from Proposition 4, and (159) follows from the definition of $D_\alpha^{\mathsf{s}}(\cdot)$ in (54). □

Theorem 7. *For all* $\alpha \in [0, 1]$,

$$D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X) \leq D_\alpha^{\mathsf{s}}(P_{Y|X}\|Q_{Y|X}|P_X) \leq D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X). \tag{160}$$

For all $\alpha \in [1, \infty]$,

$$D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X) \leq D_\alpha^1(P_{Y|X}\|Q_{Y|X}|P_X) \leq D_\alpha^{\mathsf{s}}(P_{Y|X}\|Q_{Y|X}|P_X). \tag{161}$$

Proof. For both $\alpha \in [0, 1]$ and $\alpha \in [1, \infty]$, the relation $D_\alpha^1(\cdot) \leq D_\alpha^{\mathsf{s}}(\cdot)$ follows from Proposition 7.

We next show that $D_\alpha^s(\cdot) \leq D_\alpha^c(\cdot)$ for $\alpha \in [0,1]$. We show this for $\alpha \in (0,1)$; the claim will then extend to $\alpha \in [0,1]$ by the continuity in α of $D_\alpha^s(\cdot)$ and $D_\alpha^c(\cdot)$ (Proposition 3 (c) and Proposition 2 (c)). For $\alpha \in (0,1)$,

$$(\alpha - 1) D_\alpha^s(P_{Y|X} \| Q_{Y|X} | P_X) = \log \sum_{x \in \mathrm{supp}(P_X)} P(x) \sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \tag{162}$$

$$\geq \sum_{x \in \mathrm{supp}(P_X)} P(x) \log \sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \tag{163}$$

$$= (\alpha - 1) D_\alpha^c(P_{Y|X} \| Q_{Y|X} | P_X), \tag{164}$$

where (162) follows from (55); (163) follows from Jensen's inequality because $\log(\cdot)$ is a concave function; and (164) follows from (30). The proof of the claim for $\alpha \in (0,1)$ is finished by dividing (162)–(164) by $\alpha - 1$, which reverses the inequality because $\alpha - 1 < 0$.

We conclude by showing that $D_\alpha^c(\cdot) \leq D_\alpha^1(\cdot)$ for $\alpha \in [1, \infty]$. We show this for $\alpha \in (1, \infty)$; the claim will then extend to $\alpha \in [1, \infty]$ by the continuity of $D_\alpha^c(\cdot)$ and $D_\alpha^1(\cdot)$ in α (Proposition 2 (c) and Proposition 6 (c)). For $\alpha \in (1, \infty)$,

$$D_\alpha^c(P_{Y|X} \| Q_{Y|X} | P_X) = \sum_{x \in \mathrm{supp}(P_X)} P(x) \frac{1}{\alpha - 1} \log \sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \tag{165}$$

$$= \frac{\alpha}{\alpha - 1} \sum_{x \in \mathrm{supp}(P_X)} P(x) \log \left[\sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{166}$$

$$\leq \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathrm{supp}(P_X)} P(x) \left[\sum_y P(y|x)^\alpha Q(y|x)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{167}$$

$$= D_\alpha^1(P_{Y|X} \| Q_{Y|X} | P_X), \tag{168}$$

where (165) follows from (30); (167) follows from Jensen's inequality because $\log(\cdot)$ is a concave function; and (168) follows from (78). $\square$

Corollary 4. *For all $\alpha \in [0,1]$,*

$$J_\alpha(X;Y) \leq I_\alpha^s(X;Y) \leq I_\alpha^c(X;Y). \tag{169}$$

For all $\alpha \in [1, \infty]$,

$$I_\alpha^c(X;Y) \leq J_\alpha(X;Y) \leq I_\alpha^s(X;Y). \tag{170}$$

Proof. By (34) and (60) and Proposition 5, respectively,

$$I_\alpha^c(X;Y) = \min_{Q_Y} D_\alpha^c(P_{Y|X} \| Q_Y | P_X), \tag{171}$$

$$I_\alpha^s(X;Y) = \min_{Q_Y} D_\alpha^s(P_{Y|X} \| Q_Y | P_X), \tag{172}$$

$$J_\alpha(X;Y) = \min_{Q_Y} D_\alpha^1(P_{Y|X} \| Q_Y | P_X). \tag{173}$$

The corollary now follows from (171)–(173) and Theorem 7. $\square$

Despite $I_\alpha^c(X;Y)$, $I_\alpha^s(X;Y)$, $I_\alpha^a(X;Y)$, and $J_\alpha(X;Y)$ being different measures, they often coincide when maximized over all PMFs P_X:

Theorem 8. *For every conditional PMF $P_{Y|X}$ and every $\alpha \in (0,1) \cup (1,\infty)$,*

$$\max_{P_X} I_\alpha^c(P_X, P_{Y|X}) = \max_{P_X} I_\alpha^s(P_X, P_{Y|X}) \tag{174}$$

$$= \max_{P_X} I_\alpha^a(P_X, P_{Y|X}). \tag{175}$$

In addition, for every conditional PMF $P_{Y|X}$ and every $\alpha \in [\frac{1}{2}, 1) \cup (1, \infty)$,

$$\max_{P_X} J_\alpha(P_X, P_{Y|X}) = \max_{P_X} I_\alpha^s(P_X, P_{Y|X}). \tag{176}$$

For $\alpha \in (0, \frac{1}{2})$, the situation is different: there exists a conditional PMF $P_{Y|X}$ such that, for every $\alpha \in (0, \frac{1}{2})$,

$$\max_{P_X} J_\alpha(P_X, P_{Y|X}) < \max_{P_X} I_\alpha^s(P_X, P_{Y|X}). \tag{177}$$

Proof. Equation (174) follows from [9] (Proposition 1); (175) follows from [12] (Lemma 1); and (176) follows from [38] (Theorem V.1) for $\alpha \in (1, \infty)$.

We next establish (176) for $\alpha \in [\frac{1}{2}, 1)$. Observe that, for $\alpha \in [\frac{1}{2}, 1)$, (176) is equivalent to

$$\max_{P_X} -2^{\frac{\alpha-1}{\alpha} J_\alpha(P_X, P_{Y|X})} = \max_{P_X} -2^{\frac{\alpha-1}{\alpha} I_\alpha^s(P_X, P_{Y|X})}. \tag{178}$$

For $\alpha \in [\frac{1}{2}, 1)$, (178) holds because

$$\max_{P_X} -2^{\frac{\alpha-1}{\alpha} J_\alpha(P_X, P_{Y|X})} = \max_{P_X} \min_{Q_Y} -2^{\frac{\alpha-1}{\alpha} D_\alpha^1(P_{Y|X} \| Q_Y | P_X)} \tag{179}$$

$$= -\min_{P_X} \max_{Q_Y} \sum_x P_X(x) \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{180}$$

$$= -\max_{Q_Y} \min_{P_X} \sum_x P_X(x) \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{181}$$

$$= -\max_{Q_Y} \min_x \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{182}$$

$$= -\left[\max_{Q_Y} \min_x \sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{183}$$

$$= -\left[\max_{Q_Y} \min_{P_X} \sum_x P_X(x) \sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{184}$$

$$= -\left[\min_{P_X} \max_{Q_Y} \sum_x P_X(x) \sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{185}$$

$$= -\min_{P_X} \max_{Q_Y} \left[\sum_x P_X(x) \sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{186}$$

$$= \max_{P_X} \min_{Q_Y} -2^{\frac{\alpha-1}{\alpha} D_\alpha^s(P_{Y|X} \| Q_Y | P_X)} \tag{187}$$

$$= \max_{P_X} -2^{\frac{\alpha-1}{\alpha} I_\alpha^s(P_X, P_{Y|X})}, \tag{188}$$

where (179) follows from Proposition 5; (180) follows from (78); (181) and (185) follow from a minimax theorem and are justified below; (187) follows from (55); and (188) follows from (60).

To justify (181), we apply the minimax theorem [46] (Corollary 37.3.2) to the function $f \colon \mathcal{P}(\mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}$,

$$f(Q_Y, P_X) = \sum_x P_X(x) \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}}. \tag{189}$$

The sets of all PMFs over $\mathcal{X}$ and over $\mathcal{Y}$ are convex and compact; the function f is jointly continuous in the pair (Q_Y, P_X) because it is a composition of continuous functions; for every $Q_Y \in \mathcal{P}(\mathcal{Y})$, the function f is linear and hence convex in P_X; and it only remains to show that the function f is concave in Q_Y for every $P_X \in \mathcal{P}(\mathcal{X})$. Indeed, for every $\lambda, \lambda' \in [0,1]$ with $\lambda + \lambda' = 1$, every $Q_Y, Q_Y' \in \mathcal{P}(\mathcal{Y})$, and every $P_X \in \mathcal{P}(\mathcal{X})$,

$$f(\lambda Q_Y + \lambda' Q_Y', P_X) \tag{190}$$

$$= \sum_x P_X(x) \left[\sum_y P(y|x)^\alpha \left[\lambda Q_Y(y) + \lambda' Q_Y'(y) \right]^{1-\alpha} \right]^{\frac{1}{\alpha}} \tag{191}$$

$$= \sum_x P_X(x) \left[\sum_y \left[\lambda P(y|x)^{\frac{\alpha}{1-\alpha}} Q_Y(y) + \lambda' P(y|x)^{\frac{\alpha}{1-\alpha}} Q_Y'(y) \right]^{1-\alpha} \right]^{\frac{1}{1-\alpha} \cdot \frac{1-\alpha}{\alpha}} \tag{192}$$

$$\geq \sum_x P_X(x) \left\{ \left[\sum_y \left[\lambda P(y|x)^{\frac{\alpha}{1-\alpha}} Q_Y(y) \right]^{1-\alpha} \right]^{\frac{1}{1-\alpha}} + \left[\sum_y \left[\lambda' P(y|x)^{\frac{\alpha}{1-\alpha}} Q_Y'(y) \right]^{1-\alpha} \right]^{\frac{1}{1-\alpha}} \right\}^{\frac{1-\alpha}{\alpha}} \tag{193}$$

$$= \sum_x P_X(x) \left\{ \lambda \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{1-\alpha}} + \lambda' \left[\sum_y P(y|x)^\alpha Q_Y'(y)^{1-\alpha} \right]^{\frac{1}{1-\alpha}} \right\}^{\frac{1-\alpha}{\alpha}} \tag{194}$$

$$\geq \sum_x P_X(x) \left\{ \lambda \left[\sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} + \lambda' \left[\sum_y P(y|x)^\alpha Q_Y'(y)^{1-\alpha} \right]^{\frac{1}{\alpha}} \right\} \tag{195}$$

$$= \lambda f(Q_Y, P_X) + \lambda' f(Q_Y', P_X), \tag{196}$$

where (193) follows from the reverse Minkowski inequality [16] (III 2.4 Theorem 9) because $\alpha \in [\frac{1}{2}, 1)$; and (195) holds because the function $z \mapsto z^{(1-\alpha)/\alpha}$ is concave for $\alpha \in [\frac{1}{2}, 1)$.

The justification of (185) is very similar to that of (181); here, we apply the minimax theorem to the function $g \colon \mathcal{P}(\mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}$,

$$g(Q_Y, P_X) = \sum_x P_X(x) \sum_y P(y|x)^\alpha Q_Y(y)^{1-\alpha}. \tag{197}$$

Compared to the justification of (181), the only essential difference lies in showing that the function g is concave in Q_Y for every $P_X \in \mathcal{P}(\mathcal{X})$: here, this follows easily from the concavity of the function $z \mapsto z^{1-\alpha}$ for $\alpha \in [\frac{1}{2}, 1)$.

We conclude the proof by establishing (177). Let $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and let the conditional PMF $P_{Y|X}$ be given by $P_{Y|X}(y|x) = \mathbb{1}\{y = x\}$. (This corresponds to a binary noiseless channel.) Then, denoting by U_X the uniform distribution over $\mathcal{X}$,

$$\max_{P_X} I_\alpha^{\mathsf{s}}(P_X, P_{Y|X}) \geq I_\alpha^{\mathsf{s}}(U_X, P_{Y|X}) \tag{198}$$

$$= \log 2, \tag{199}$$

where (199) follows from (61). On the other hand, for every $\alpha \in (0, \frac{1}{2})$ and every PMF P_X,

$$J_\alpha(P_X, P_{Y|X}) = \frac{\alpha}{1-\alpha} H_\infty(P_X) \tag{200}$$

$$\leq \frac{\alpha}{1-\alpha} \log 2 \tag{201}$$

$$< \log 2, \tag{202}$$

where (200) follows from [3] (Lemma 11); (201) follows from (144); and (202) holds because $\alpha \in (0, \frac{1}{2})$. Inequality (177) now follows from (199) and (202). □

8. Horse Betting

In this section, we analyze horse betting with a gambler investing all her money. Recall from the introduction that the winning horse X is distributed according to the PMF p, where we assume $p(x) > 0$ for all $x \in \mathcal{X}$; that the odds offered by the bookmaker are denoted by $o \colon \mathcal{X} \to (0, \infty)$; that the fraction of her wealth that the gambler bets on Horse $x \in \mathcal{X}$ is denoted $b(x) \geq 0$; that the wealth relative is the random variable $S \triangleq b(X) o(X)$; and that we seek betting strategies that maximize the utility function

$$U_\beta \triangleq \begin{cases} \frac{1}{\beta} \log \mathrm{E}[S^\beta] & \text{if } \beta \neq 0, \\ \mathrm{E}[\log S] & \text{if } \beta = 0. \end{cases} \tag{203}$$

Because the gambler invests all her money, b is a PMF. As in [47] (Section 10.3), define the constant

$$c \triangleq \left[\sum_x \frac{1}{o(x)} \right]^{-1} \tag{204}$$

and the PMF

$$r(x) \triangleq \frac{c}{o(x)}. \tag{205}$$

Using these definitions, the utility function U_β can be decomposed as follows:

Theorem 9. *Let $\beta \in (-\infty, 1)$, and let b be a PMF. Then,*

$$U_\beta = \log c + D_{\frac{1}{1-\beta}}(p\|r) - D_{1-\beta}(g^{(\beta)}\|b), \tag{206}$$

where the PMF $g^{(\beta)}$ is given by

$$g^{(\beta)}(x) \triangleq \frac{p(x)^{\frac{1}{1-\beta}} o(x)^{\frac{\beta}{1-\beta}}}{\sum_{x' \in \mathcal{X}} p(x')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}}}. \tag{207}$$

Thus, choosing $b = g^{(\beta)}$ uniquely maximizes U_β among all PMFs b.

The three terms in (206) can be interpreted as follows:

1.　The first term, $\log c$, depends only on the odds and is related to the fairness of the odds. The odds are called subfair if $c < 1$, fair if $c = 1$, and superfair if $c > 1$.
2.　The second term, $D_{1/(1-\beta)}(p\|r)$, is related to the bookmaker's estimate of the winning probabilities. It is zero if and only if the odds are inversely proportional to the winning probabilities.

3.　The third term, $-D_{1-\beta}(g^{(\beta)}\|b)$, is related to the gambler's estimate of the winning probabilities. It is zero if and only if b is equal to $g^{(\beta)}$.

Remark 4. *For $\beta = 0$, (206) reduces to the following decomposition of the doubling rate $\mathrm{E}[\log S]$:*

$$\mathrm{E}[\log S] = \log c + D(p\|r) - D(p\|b). \tag{208}$$

(This decomposition appeared previously in [47] (Section 10.3).) Equation (208) implies that the doubling rate is maximized by proportional gambling, i.e., that $\mathrm{E}[\log S]$ is maximized if and only if b is equal to p.

Remark 5. *Considering the limits $\beta \to -\infty$ and $\beta \uparrow 1$, the PMF $g^{(\beta)}$ satisfies, for every $x \in \mathcal{X}$,*

$$\lim_{\beta \to -\infty} g^{(\beta)}(x) = \frac{c}{o(x)}, \tag{209}$$

$$\lim_{\beta \uparrow 1} g^{(\beta)}(x) = \frac{p(x)\,\mathbb{1}\{x \in \mathcal{S}\}}{\sum_{x' \in \mathcal{X}} p(x')\,\mathbb{1}\{x' \in \mathcal{S}\}}, \tag{210}$$

where the set $\mathcal{S}$ is defined as $\mathcal{S} \triangleq \{x' \in \mathcal{X} : p(x')o(x') = \max_x[p(x)o(x)]\}$. It follows from Proposition 8 below that the RHS of (209) is the unique maximizer of $\lim_{\beta \to -\infty} U_\beta$; and it follows from the proof of Proposition 9 below that the RHS of (210) is a maximizer (not necessarily unique) of U_1.

Proof of Remark 5. Recall that we assume $p(x) > 0$ for every $x \in \mathcal{X}$. Then, (209) follows from (207) and the definition of c in (204). To establish (210), define $\tau \triangleq \max_x[p(x)o(x)]$ and observe that, for every $x \in \mathcal{X}$,

$$\lim_{\beta \uparrow 1} g^{(\beta)}(x) = \lim_{\beta \uparrow 1} \frac{p(x)\left[p(x)o(x)/\tau\right]^{\frac{\beta}{1-\beta}}}{\sum_{x' \in \mathcal{X}} p(x')\left[p(x')o(x')/\tau\right]^{\frac{\beta}{1-\beta}}} \tag{211}$$

$$= \frac{p(x)\,\mathbb{1}\{x \in \mathcal{S}\}}{\sum_{x' \in \mathcal{X}} p(x')\,\mathbb{1}\{x' \in \mathcal{S}\}}, \tag{212}$$

where (211) follows from (207) and some algebra; and (212) is justified as follows: if $x \in \mathcal{S}$, then $\left[p(x)o(x)/\tau\right]^{\beta/(1-\beta)}$ equals one; and if $x \notin \mathcal{S}$, then $\left[p(x)o(x)/\tau\right]^{\beta/(1-\beta)}$ tends to zero as $\beta \uparrow 1$ because $p(x)o(x)/\tau < 1$ and because $\lim_{\beta \uparrow 1} \frac{\beta}{1-\beta} = +\infty$.　□

Remark 6. *Using the definition in (24) for the Rényi divergence of negative orders, it is not difficult to see from the proof of Theorem 9 below that (206) also holds for $\beta > 1$. However, because the Rényi divergence of negative orders is nonpositive instead of nonnegative, the above interpretation is not valid anymore; in particular, for $\beta > 1$, choosing $b = g^{(\beta)}$ is in general not optimal.*

Proof of Theorem 9. We first show the maximization claim. The only term on the RHS of (206) that depends on b is $-D_{1-\beta}(g^{(\beta)}\|b)$. Because $1 - \beta > 0$, this term is maximized if and only if $b = g^{(\beta)}$ (Proposition 1 (a)).

We now establish (206) for $\beta \in (-\infty, 0) \cup (0, 1)$; we omit the proof for $\beta = 0$, which can be found in [47] (Section 10.3). For $\beta \in (-\infty, 0) \cup (0, 1)$,

$$U_\beta = \frac{1}{\beta} \log \sum_x p(x)\,b(x)^\beta o(x)^\beta. \tag{213}$$

For every $x \in \mathcal{X}$,

$$p(x) b(x)^\beta o(x)^\beta = \left[\sum_{x' \in \mathcal{X}} p(x')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}} \right]^{1-\beta} \cdot g^{(\beta)}(x)^{1-\beta} b(x)^\beta, \tag{214}$$

which follows from (207). Now, (206) holds because

$$U_\beta = \frac{1-\beta}{\beta} \log \sum_{x' \in \mathcal{X}} p(x')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}} + \frac{1}{\beta} \log \sum_x g^{(\beta)}(x)^{1-\beta} b(x)^\beta \tag{215}$$

$$= \frac{1-\beta}{\beta} \log \sum_{x' \in \mathcal{X}} p(x')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}} - D_{1-\beta}(g^{(\beta)} \| b) \tag{216}$$

$$= \log c + \frac{1-\beta}{\beta} \log \sum_{x' \in \mathcal{X}} p(x')^{\frac{1}{1-\beta}} r(x')^{\frac{-\beta}{1-\beta}} - D_{1-\beta}(g^{(\beta)} \| b) \tag{217}$$

$$= \log c + D_{\frac{1}{1-\beta}}(p \| r) - D_{1-\beta}(g^{(\beta)} \| b), \tag{218}$$

where (215) follows from (213) and (214); (216) follows from identifying the Rényi divergence (recall that $g^{(\beta)}$ and b are PMFs); (217) follows from (205); and (218) follows from identifying the Rényi divergence (recall that r is a PMF).　□

The rest of the section presents the cases $\beta \to -\infty$, $\beta \geq 1$, and $\beta \to +\infty$.

Proposition 8. *Let b be a PMF. Then,*

$$\lim_{\beta \to -\infty} U_\beta = \log \min_x [b(x) o(x)] \tag{219}$$

$$\leq \log c. \tag{220}$$

Inequality (220) holds with equality if and only if $b(x) = c/o(x)$ for all $x \in \mathcal{X}$.

Observe that if $b(x) = c/o(x)$ for all $x \in \mathcal{X}$, then $S = c$ with probability one, i.e., S does not depend on the winning horse.

Proof of Proposition 8. Equation (219) holds because

$$\lim_{\beta \to -\infty} U_\beta = \lim_{\beta \to -\infty} \log \left[\sum_x p(x) \left(b(x) o(x) \right)^\beta \right]^{\frac{1}{\beta}} \tag{221}$$

$$= \log \min_x [b(x) o(x)], \tag{222}$$

where (222) holds because, in the limit as β tends to $-\infty$, the power mean tends to the minimum (since p is a PMF with $p(x) > 0$ for all $x \in \mathcal{X}$ [15] (Chapter 8)).

We show (220) by contradiction. Assume that there exists a PMF b that does not satisfy (220), thus

$$b(x) o(x) > c \tag{223}$$

for all $x \in \mathcal{X}$. Then,

$$1 = \sum_x b(x) \tag{224}$$

$$> \sum_x \frac{c}{o(x)} \tag{225}$$

$$= 1, \tag{226}$$

where (224) holds because b is a PMF; (225) follows from (223); and (226) follows from the definition of c in (204). Because $1 > 1$ is impossible, such a b cannot exist, which establishes (220).

It is not difficult to see that (220) holds with equality if $b(x) = c/o(x)$ for all $x \in \mathcal{X}$. We therefore focus on establishing that if (220) holds with equality, then $b(x) = c/o(x)$ for all $x \in \mathcal{X}$. Observe first that, if (220) holds with equality, then, for all $x \in \mathcal{X}$,

$$b(x)o(x) \geq c. \tag{227}$$

We now claim that (227) holds with equality for all $x \in \mathcal{X}$. Indeed, if this were not the case, then there would exist an $x' \in \mathcal{X}$ for which $b(x')o(x') > c$, thus (224)–(226) would hold, which would lead to a contradiction. Hence, if (220) holds with equality, then $b(x) = c/o(x)$ for all $x \in \mathcal{X}$. $\square$

Proposition 9. *Let $\beta \geq 1$, and let b be a PMF. Then,*

$$U_\beta \leq \log \max_x \left[p(x)^{1/\beta} o(x) \right]. \tag{228}$$

Equality in (228) can be achieved by choosing $b(x) = \mathbb{1}\{x = x'\}$ for some $x' \in \mathcal{X}$ satisfying

$$p(x')^{1/\beta} o(x') = \max_x \left[p(x)^{1/\beta} o(x) \right]. \tag{229}$$

Remark 7. *Proposition 9 implies that if $\beta \geq 1$, then it is optimal to bet on a single horse. Unless $|\mathcal{X}| = 1$, this is not the case when $\beta < 1$: When $\beta < 1$, an optimal betting strategy requires placing a bet on every horse. This follows from Theorem 9 and our assumption that $p(x)$ and $o(x)$ are all positive.*

Proof of Proposition 9. Inequality (228) holds because

$$U_\beta = \frac{1}{\beta} \log \sum_x p(x) b(x)^\beta o(x)^\beta \tag{230}$$

$$\leq \frac{1}{\beta} \log \sum_x p(x) b(x) o(x)^\beta \tag{231}$$

$$\leq \frac{1}{\beta} \log \sum_x b(x) \cdot \max_{x' \in \mathcal{X}} \left[p(x') o(x')^\beta \right] \tag{232}$$

$$= \frac{1}{\beta} \log \max_{x' \in \mathcal{X}} \left[p(x') o(x')^\beta \right] \tag{233}$$

$$= \log \max_{x' \in \mathcal{X}} \left[p(x')^{1/\beta} o(x') \right], \tag{234}$$

where (231) holds because $b(x) \in [0,1]$ and $\beta \geq 1$, and (233) holds because b is a PMF. It is not difficult to see that (228) holds with equality if $b(x) = \mathbb{1}\{x = x'\}$ for some $x' \in \mathcal{X}$ satisfying (229). $\square$

Proposition 10. *Let b be a PMF. Then,*

$$\lim_{\beta \to +\infty} U_\beta = \log \max_x \left[b(x) o(x) \right] \tag{235}$$

$$\leq \log \max_x o(x). \tag{236}$$

Equality in (236) can be achieved by choosing $b(x) = \mathbb{1}\{x = x'\}$ for some $x' \in \mathcal{X}$ satisfying

$$o(x') = \max_x o(x). \tag{237}$$

Proof. Equation (235) holds because

$$\lim_{\beta \to +\infty} U_\beta = \lim_{\beta \to +\infty} \log \left[\sum_x p(x) \left(b(x) o(x) \right)^\beta \right]^{\frac{1}{\beta}} \tag{238}$$

$$= \log \max_x \left[b(x) o(x) \right], \tag{239}$$

where (239) holds because in the limit as β tends to $+\infty$, the power mean tends to the maximum (since p is a PMF with $p(x) > 0$ for all $x \in \mathcal{X}$ [15] (Chapter 8)). Inequality (236) holds because $b(x) \leq 1$ for all $x \in \mathcal{X}$. It is not difficult to see that (236) holds with equality if $b(x) = \mathbb{1}\{x = x'\}$ for some $x' \in \mathcal{X}$ satisfying (237). $\quad\square$

9. Horse Betting with Side Information

In this section, we study the horse betting problem where the gambler observes some side information Y before placing her bets. This setting leads to the conditional Rényi divergence $D_\alpha^1(\cdot)$ discussed in Section 5 (see Theorem 10). In addition, it provides a new operational meaning to the dependence measure $J_\alpha(X;Y)$ (see Theorem 11).

We adapt our notation as follows: The joint PMF of X and Y is denoted p_{XY}. (Recall that X denotes the winning horse.) We drop the assumption that the winning probabilities $p(x)$ are positive, but we assume that $p(y) > 0$ for all $y \in \mathcal{Y}$. We continue to assume that the gambler invests all her wealth, so a betting strategy is now a conditional PMF $b_{X|Y}$, and the wealth relative S is

$$S \triangleq b(X|Y) o(X). \tag{240}$$

As in Section 8, define the constant

$$c \triangleq \left[\sum_x \frac{1}{o(x)} \right]^{-1} \tag{241}$$

and the PMF

$$r_X(x) \triangleq \frac{c}{o(x)}. \tag{242}$$

The following decomposition of the utility function U_β parallels that of Theorem 9:

Theorem 10. *Let $\beta \in (-\infty, 1)$. Then,*

$$U_\beta = \log c + D_{\frac{1}{1-\beta}}^1 (p_{X|Y} \| r_X | p_Y) - D_{1-\beta} (g_{X|Y}^{(\beta)} g_Y^{(\beta)} \| b_{X|Y} g_Y^{(\beta)}), \tag{243}$$

where the conditional PMF $g_{X|Y}^{(\beta)}$ and the PMF $g_Y^{(\beta)}$ are given by

$$g_{X|Y}^{(\beta)}(x|y) \triangleq \frac{p(x|y)^{\frac{1}{1-\beta}} o(x)^{\frac{\beta}{1-\beta}}}{\sum_{x'} p(x'|y)^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}}}, \tag{244}$$

$$g_Y^{(\beta)}(y) \triangleq \frac{p(y) \left[\sum_{x'} p(x'|y)^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}} \right]^{1-\beta}}{\sum_{y'} p(y') \left[\sum_{x'} p(x'|y')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}} \right]^{1-\beta}}. \tag{245}$$

Thus, choosing $b_{X|Y} = g_{X|Y}^{(\beta)}$ uniquely maximizes U_β among all conditional PMFs $b_{X|Y}$.

Proof. We first show that U_β is uniquely maximized by $g_{X|Y}^{(\beta)}$. The only term on the RHS of (243) that depends on $b_{X|Y}$ is $-D_{1-\beta}\big(g_{X|Y}^{(\beta)}g_Y^{(\beta)}\|b_{X|Y}g_Y^{(\beta)}\big)$. Because $1-\beta > 0$, this term is maximized if and only if $b_{X|Y}g_Y^{(\beta)} = g_{X|Y}^{(\beta)}g_Y^{(\beta)}$ (Proposition 1 (a)). By our assumptions that $p(y) > 0$ for all $y \in \mathcal{Y}$ and $o(x) > 0$ for all $x \in \mathcal{X}$, we have $g_Y^{(\beta)}(y) > 0$ for all $y \in \mathcal{Y}$. Consequently, $b_{X|Y}g_Y^{(\beta)} = g_{X|Y}^{(\beta)}g_Y^{(\beta)}$ if and only if $b_{X|Y} = g_{X|Y}^{(\beta)}$.

Consider now (243) for $\beta = 0$. For $\beta = 0$, (243) reduces to

$$\mathrm{E}[\log S] = \log c + D(p_{X|Y}p_Y\|r_X p_Y) - D(p_{X|Y}p_Y\|b_{X|Y}p_Y), \tag{246}$$

and some algebra reveals that (246) holds.

We conclude with establishing (243) for $\beta \in (-\infty, 0) \cup (0, 1)$. For $\beta \in (-\infty, 0) \cup (0, 1)$,

$$U_\beta = \frac{1}{\beta}\log \sum_{x,y} p(x,y)\,b(x|y)^\beta o(x)^\beta. \tag{247}$$

For every $x \in \mathcal{X}$ and every $y \in \mathcal{Y}$,

$$p(x,y)\,b(x|y)^\beta o(x)^\beta = \sum_{y'\in\mathcal{Y}} p(y')\left[\sum_{x'\in\mathcal{X}} p(x'|y')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}}\right]^{1-\beta} \cdot g_Y^{(\beta)}(y)\,g_{X|Y}^{(\beta)}(x|y)^{1-\beta}b(x|y)^\beta, \tag{248}$$

which follows from (244) and (245). Now, (243) holds because

$$
\begin{aligned}
U_\beta &= \frac{1}{\beta}\log\sum_{y'\in\mathcal{Y}} p(y')\left[\sum_{x'\in\mathcal{X}} p(x'|y')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}}\right]^{1-\beta}\\
&\quad + \frac{1}{\beta}\log\sum_{x,y}\big[g_{X|Y}^{(\beta)}(x|y)\,g_Y^{(\beta)}(y)\big]^{1-\beta}\big[b(x|y)\,g_Y^{(\beta)}(y)\big]^\beta \tag{249}\\[4pt]
&= \frac{1}{\beta}\log\sum_{y'\in\mathcal{Y}} p(y')\left[\sum_{x'\in\mathcal{X}} p(x'|y')^{\frac{1}{1-\beta}} o(x')^{\frac{\beta}{1-\beta}}\right]^{1-\beta} - D_{1-\beta}\big(g_{X|Y}^{(\beta)}g_Y^{(\beta)}\|b_{X|Y}g_Y^{(\beta)}\big) \tag{250}\\[4pt]
&= \log c + \frac{1}{\beta}\log\sum_{y'\in\mathcal{Y}} p(y')\left[\sum_{x'\in\mathcal{X}} p(x'|y')^{\frac{1}{1-\beta}} r_X(x')^{\frac{-\beta}{1-\beta}}\right]^{1-\beta} - D_{1-\beta}\big(g_{X|Y}^{(\beta)}g_Y^{(\beta)}\|b_{X|Y}g_Y^{(\beta)}\big) \tag{251}\\[4pt]
&= \log c + D_{\frac{1}{1-\beta}}^{1}(p_{X|Y}\|r_X|p_Y) - D_{1-\beta}\big(g_{X|Y}^{(\beta)}g_Y^{(\beta)}\|b_{X|Y}g_Y^{(\beta)}\big), \tag{252}
\end{aligned}
$$

where (249) follows from (247) and (248) and the fact that $g_Y^{(\beta)}(y) = g_Y^{(\beta)}(y)^{1-\beta}g_Y^{(\beta)}(y)^\beta$; (250) follows by identifying the Rényi divergence; (251) follows from (242); and (252) follows by identifying the conditional Rényi divergence using (78). □

Remark 8. *It follows from Theorem 10 that, if the gambler gambles optimally, then, for $\beta \in (-\infty, 1)$,*

$$U_\beta = \log c + D_{\frac{1}{1-\beta}}^{1}(p_{X|Y}\|r_X|p_Y). \tag{253}$$

Operationally, it is clear that preprocessing the side information cannot increase the gambler's utility, i.e., that, for every conditional PMF $p_{Y'|Y}$,

$$D_{\frac{1}{1-\beta}}^{1}(p_{X|Y'}\|r_X|p_{Y'}) \le D_{\frac{1}{1-\beta}}^{1}(p_{X|Y}\|r_X|p_Y), \tag{254}$$

where $p_{X|Y'}$ and $p_{Y'}$ are derived from the joint PMF $p_{XYY'}$ given by

$$p_{XYY'}(x, y, y') = p_Y(y)\, p_{X|Y}(x|y)\, p_{Y'|Y}(y'|y). \tag{255}$$

This provides the intuition for Theorem 6, where (254) is shown directly.

The extreme case is when the preprocessing maps the side information to a constant and hence leads to the case where the side information is absent. In this case, Y' is deterministic and $p_{X|Y'}$ equals p_X. Theorem 9 and Theorem 10 then lead to the following relation between the conditional and unconditional Rényi divergence:

$$D_{\frac{1}{1-\beta}}(p_X \| r_X) \leq D^1_{\frac{1}{1-\beta}}(p_{X|Y} \| r_X | p_Y), \tag{256}$$

where the marginal PMF p_X is given by

$$p_X(x) = \sum_y p_{XY}(x, y). \tag{257}$$

This motivates Corollary 3, where (256) is derived from (254).

The last result of this section provides a new operational meaning to the Lapidoth–Pfister mutual information $J_\alpha(X;Y)$: assuming that $\beta \in (-\infty, 1)$ and that the gambler knows the winning probabilities, $J_{1/(1-\beta)}(X;Y)$ measures how much the side information that is available to the gambler but not the bookmaker increases the gambler's smallest guaranteed utility for a fixed level of fairness c. To see this, consider first the setting without side information. By Theorem 9, the gambler chooses $b = g^{(\beta)}$ to maximize her utility, where $g^{(\beta)}$ is defined in (207). Then, using the nonnegativity of the Rényi divergence (Proposition 1 (a)), the following lower bound on the gambler's utility follows from (206):

$$U_\beta \geq \log c. \tag{258}$$

We call the RHS of (258) the smallest guaranteed utility for a fixed level of fairness c because (258) holds with equality if the bookmaker chooses the odds inversely proportional to the winning probabilities. Comparing (258) with (259) below, we see that the difference due to the side information is $J_{1/(1-\beta)}(X;Y)$. Note that $J_{1/(1-\beta)}(X;Y)$ is typically not the difference between the utility with and without side information; this is because the odds for which (258) and (259) hold with equality are typically not the same.

Theorem 11. *Let $\beta \in (-\infty, 1)$. If $b_{X|Y}$ is equal to $g^{(\beta)}_{X|Y}$ from Theorem 10, then*

$$U_\beta \geq \log c + J_{\frac{1}{1-\beta}}(X;Y). \tag{259}$$

Moreover, for every $c > 0$, there exist odds $o \colon \mathcal{X} \to (0, \infty)$ such that (259) holds with equality.

Proof. For this choice of $b_{X|Y}$, (259) holds because

$$U_\beta = \log c + D^1_{\frac{1}{1-\beta}}(p_{X|Y} \| r_X | p_Y) \tag{260}$$

$$\geq \log c + \min_{\tilde{r}_X \in \mathcal{P}(\mathcal{X})} D^1_{\frac{1}{1-\beta}}(p_{X|Y} \| \tilde{r}_X | p_Y) \tag{261}$$

$$= \log c + J_{\frac{1}{1-\beta}}(X;Y), \tag{262}$$

where (260) follows from Theorem 10, and (262) follows from Proposition 5.

Fix now $c > 0$, let $\tilde{r}_X^*$ achieve the minimum on the RHS of (261), and choose the odds

$$o(x) = \frac{c}{\tilde{r}_X^*(x)}. \tag{263}$$

Then, (261) holds with equality because $r_X = \tilde{r}_X^*$ by (241) and (242). $\quad\square$

10. Horse Betting with Part of the Money

In this section, we treat the possibility that the gambler does not invest all her wealth. We restrict ourselves to the setting without side information and to $\beta \in (-\infty, 0) \cup (0, 1)$. (For the case $\beta = 0$, see [47] (Section 10.5).) We assume that $p(x) > 0$ and $o(x) > 0$ for all $x \in \mathcal{X}$. Denote by $b(0)$ the fraction of her wealth that the gambler does not use for betting. (We assume $0 \notin \mathcal{X}$.) Then, $b \colon \mathcal{X} \cup \{0\} \to [0, 1]$ is a PMF, and the wealth relative S is the random variable

$$S \triangleq b(0) + b(X)o(X). \tag{264}$$

As in Section 8, define the constant

$$c \triangleq \left[\sum_x \frac{1}{o(x)} \right]^{-1}. \tag{265}$$

We treat the cases $c < 1$ and $c \geq 1$ separately, starting with the latter. If $c \geq 1$, then it is optimal to invest all the money:

Proposition 11. *Assume $c \geq 1$, let $\beta \in \mathbb{R}$, and let b be a PMF on $\mathcal{X} \cup \{0\}$ with utility U_β. Then, there exists a PMF b' on $\mathcal{X} \cup \{0\}$ with $b'(0) = 0$ and utility $U'_\beta \geq U_\beta$.*

Proof. Choose the PMF b' as follows:

$$b'(x) = \begin{cases} \frac{c}{o(x)} \cdot b(0) + b(x) & \text{if } x \in \mathcal{X}, \\ 0 & \text{if } x = 0. \end{cases} \tag{266}$$

Then, for every $x \in \mathcal{X}$,

$$b'(0) + b'(x)o(x) = c \cdot b(0) + b(x)o(x) \tag{267}$$
$$\geq b(0) + b(x)o(x), \tag{268}$$

where (268) holds because $c \geq 1$ by assumption. For $\beta > 0$, $U'_\beta \geq U_\beta$ holds because (268) implies $\mathrm{E}[S'^\beta] \geq \mathrm{E}[S^\beta]$. For $\beta < 0$ and $\beta = 0$, $U'_\beta \geq U_\beta$ follows similarly from (268). $\quad\square$

On the other hand, if $\beta < 1$ and the odds are subfair, i.e., if $c < 1$, then Claim (c) of the following theorem shows that investing all the money is not optimal:

Theorem 12. *Assume $c < 1$, let $\beta \in (-\infty, 0) \cup (0, 1)$, and let b^* be a PMF on $\mathcal{X} \cup \{0\}$ that maximizes U_β among all PMFs b. Defining*

$$\mathcal{S} \triangleq \{ x \in \mathcal{X} : b^*(x) > 0 \}, \tag{269}$$

$$\Gamma \triangleq \frac{1 - \sum_{x \in \mathcal{S}} p(x)}{1 - \sum_{x \in \mathcal{S}} \frac{1}{o(x)}}, \tag{270}$$

$$\gamma(x) \triangleq \max\left\{ 0, \Gamma^{\frac{1}{\beta-1}} p(x)^{\frac{1}{1-\beta}} o(x)^{\frac{\beta}{1-\beta}} - \frac{1}{o(x)} \right\} \quad \forall x \in \mathcal{X}, \tag{271}$$

the following claims hold:

(a) *Both the numerator and denominator on the RHS of* (270) *are positive, so* Γ *is well-defined and positive.*

(b) *For every* $x \in \mathcal{X}$,

$$b^*(x) = \gamma(x) b^*(0). \tag{272}$$

(c) *The quantity* $b^*(0)$ *satisfies*

$$b^*(0) = \frac{1}{1 + \sum_{x \in \mathcal{X}} \gamma(x)}. \tag{273}$$

In particular, $b^*(0) > 0$.

Claim (b) implies that for every $x \in \mathcal{X}$, $b^*(x) > 0$ if and only if $p(x)o(x) > \Gamma$. Ordering the elements $x_1, x_2, \ldots$ of $\mathcal{X}$ such that $p(x_1)o(x_1) \geq p(x_2)o(x_2) \geq \ldots$, the set $\mathcal{S}$ thus has a special structure: it is either empty or equal to $\{x_1, x_2, \ldots, x_k\}$ for some integer k. To maximize U_β, the following procedure can be used: for every $\mathcal{S}$ with the above structure, compute the corresponding b according to (270)–(273); and from these b's, take one that maximizes U_β. This procedure leads to an optimal solution: an optimal solution b^* exists because we are optimizing a continuous function over a compact set, and b^* corresponds to a set $\mathcal{S}$ that will be considered by the procedure.

Proof of Theorem 12. The proof is based on the Karush–Kuhn–Tucker conditions. By separately considering the cases $\beta \in (0, 1)$ and $\beta < 0$, we first show that, for $\beta \in (-\infty, 0) \cup (0, 1)$, a strategy $b(\cdot)$ is optimal if and only if the following conditions are satisfied for some $\mu \in \mathbb{R}$:

$$\sum_{x \in \mathcal{X}} p(x) \left(b(0) + b(x) o(x) \right)^{\beta-1} \begin{cases} = \mu & \text{if } b(0) > 0, \\ \leq \mu & \text{if } b(0) = 0, \end{cases} \tag{274}$$

and, for every $x \in \mathcal{X}$,

$$p(x)o(x) \left(b(0) + b(x) o(x) \right)^{\beta-1} \begin{cases} = \mu & \text{if } b(x) > 0, \\ \leq \mu & \text{if } b(x) = 0. \end{cases} \tag{275}$$

Consider first $\beta \in (0, 1)$, and define the function $\tau \colon \mathcal{P}(\mathcal{X} \cup \{0\}) \to \mathbb{R}$,

$$\tau(b) \triangleq \sum_{x \in \mathcal{X}} p(x) \left(b(0) + b(x) o(x) \right)^{\beta}. \tag{276}$$

Since $\beta > 0$ and since the logarithm is an increasing function, maximizing $U_\beta = \frac{1}{\beta} \log \mathrm{E}[S^\beta]$ over b is equivalent to maximizing $\tau(b)$. Observe that τ is concave, thus, by the Karush–Kuhn–Tucker conditions [11] (Theorem 4.4.1), it is maximized by a PMF b if and only if there exists a $\lambda \in \mathbb{R}$ such that (i) for all $x \in \mathcal{X} \cup \{0\}$ with $b(x) > 0$,

$$\frac{\partial \tau}{\partial b(x)}(b) = \lambda, \tag{277}$$

and (ii) for all $x \in \mathcal{X} \cup \{0\}$ with $b(x) = 0$,

$$\frac{\partial \tau}{\partial b(x)}(b) \leq \lambda. \tag{278}$$

Henceforth, we use the following notation: to designate that (i) and (ii) both hold, we write

$$\frac{\partial \tau}{\partial b(x)}(b) \begin{cases} = \lambda & \text{if } b(x) > 0, \\ \leq \lambda & \text{if } b(x) = 0. \end{cases} \tag{279}$$

Dividing both sides of (279) by $\beta > 0$ and defining $\mu \triangleq \frac{\lambda}{\beta}$, we obtain that (279) is equivalent to

$$\frac{1}{\beta} \cdot \frac{\partial \tau}{\partial b(x)}(b) \begin{cases} = \mu & \text{if } b(x) > 0, \\ \leq \mu & \text{if } b(x) = 0. \end{cases} \tag{280}$$

Now, (280) translates to (274) for $x = 0$ and to (275) for $x \in \mathcal{X}$.

Consider now $\beta < 0$, and define τ as in (276). Then, because $\beta < 0$, maximizing $U_\beta = \frac{1}{\beta} \log \mathrm{E}[S^\beta]$ is equivalent to minimizing τ. The function τ is convex, thus Inequality (278) is reversed. Dividing by $\beta < 0$ again reverses the inequalities, thus (280), (274), and (275) continue to hold for $\beta < 0$.

Having established that, for all $\beta \in (-\infty, 0) \cup (0, 1)$, a strategy b is optimal if and only if (274) and (275) hold, we next continue with the proof. Let $\beta \in (-\infty, 0) \cup (0, 1)$, and let b^* be a PMF on $\mathcal{X} \cup \{0\}$ that maximizes U_β. By the above discussion, (274) and (275) are satisfied by b^* for some $\mu \in \mathbb{R}$. The LHS of (274) is positive, so $\mu > 0$. We now show that for all $x \in \mathcal{X}$,

$$b^*(x) = \max\left\{0, \left[\frac{p(x)o(x)^\beta}{\mu}\right]^{\frac{1}{1-\beta}} - \frac{b^*(0)}{o(x)}\right\}. \tag{281}$$

To this end, fix $x \in \mathcal{X}$. If $b^*(x) > 0$, then (275) implies

$$b^*(x) = \left[\frac{p(x)o(x)^\beta}{\mu}\right]^{\frac{1}{1-\beta}} - \frac{b^*(0)}{o(x)}, \tag{282}$$

and the RHS of (282) is equal to the RHS of (281) because, being equal to $b^*(x)$, it is positive. If $b^*(x) = 0$, then (275) implies

$$\left[\frac{p(x)o(x)^\beta}{\mu}\right]^{\frac{1}{1-\beta}} - \frac{b^*(0)}{o(x)} \leq 0, \tag{283}$$

so the RHS of (281) is zero and (281) hence holds.

Having established (281), we next show that $b^*(\hat{x}) = 0$ for some $\hat{x} \in \mathcal{X}$. For a contradiction, assume that $b^*(x) > 0$ for all $x \in \mathcal{X}$. Then,

$$\sum_{x \in \mathcal{X}} p(x) \left(b^*(0) + b^*(x)o(x)\right)^{\beta-1} = \mu \sum_{x \in \mathcal{X}} \frac{1}{o(x)} \tag{284}$$

$$> \mu, \tag{285}$$

where (284) follows from (275), and (285) holds because $c < 1$ by assumption. However, this is impossible: (285) contradicts (274).

Let now $\hat{x} \in \mathcal{X}$ be such that $b^*(\hat{x}) = 0$. Then, by (281),

$$\left[\frac{p(\hat{x})o(\hat{x})^\beta}{\mu}\right]^{\frac{1}{1-\beta}} - \frac{b^*(0)}{o(\hat{x})} \leq 0. \tag{286}$$

Because $p(\hat{x})$ and $o(\hat{x})$ are positive, this implies $b^*(0) > 0$. Thus, by (274),

$$\sum_{x \in \mathcal{X}} p(x) \left(b^*(0) + b^*(x)o(x)\right)^{\beta-1} = \mu. \tag{287}$$

Splitting the sum on the LHS of (287) depending on whether $b^*(x) > 0$ or $b^*(x) = 0$, we obtain

$$\mu = \sum_{x \in \mathcal{S}} p(x) \left(b^*(0) + b^*(x) o(x)\right)^{\beta-1} + \sum_{x \notin \mathcal{S}} p(x) \left(b^*(0) + b^*(x) o(x)\right)^{\beta-1} \tag{288}$$

$$= \sum_{x \in \mathcal{S}} \frac{\mu}{o(x)} + \sum_{x \notin \mathcal{S}} p(x) b^*(0)^{\beta-1} \tag{289}$$

$$= \mu \sum_{x \in \mathcal{S}} \frac{1}{o(x)} + b^*(0)^{\beta-1} \left[1 - \sum_{x \in \mathcal{S}} p(x)\right], \tag{290}$$

where (289) follows from (275). Rearranging (290), we obtain

$$\mu \left[1 - \sum_{x \in \mathcal{S}} \frac{1}{o(x)}\right] = b^*(0)^{\beta-1} \left[1 - \sum_{x \in \mathcal{S}} p(x)\right]. \tag{291}$$

Recall that $\mu > 0$ and $b^*(0) > 0$. In addition, $1 - \sum_{x \in \mathcal{S}} p(x) > 0$ because $b^*(\hat{x}) = 0$ and hence $\hat{x} \notin \mathcal{S}$. Thus, $1 - \sum_{x \in \mathcal{S}} \frac{1}{o(x)} > 0$, so both the numerator and denominator in the definition of Γ in (270) are positive, which establishes Claim (a), namely that Γ is well-defined and positive.

To establish Claim (b), note that (291) and (270) imply that μ is given by

$$\mu = b^*(0)^{\beta-1} \Gamma, \tag{292}$$

which, when substituted into (281), yields (272).

We conclude by proving Claim (c). Because b^* is a PMF on $\mathcal{X} \cup \{0\}$,

$$1 = b^*(0) + \sum_{x \in \mathcal{X}} b^*(x) \tag{293}$$

$$= b^*(0) \left[1 + \sum_{x \in \mathcal{X}} \gamma(x)\right], \tag{294}$$

where (294) follows from (272). Rearranging (294) yields (273). $\square$

11. Universal Betting for IID Races

In this section, we present a universal gambling strategy for IID races that requires neither knowledge of the winning probabilities nor of the parameter β of the utility function and yet asymptotically maximizes the utility function for all PMFs p and all $\beta \in \mathbb{R}$. Consider n consecutive horse races, where the winning horse in the ith race is denoted X_i for $i \in \{1, \dots, n\}$. We assume that $X_1, \dots, X_n$ are IID according to the PMF p, where $p(x) > 0$ for all $x \in \mathcal{X}$. In every race, the bookmaker offers the same odds $o \colon \mathcal{X} \to (0, \infty)$, and the gambler spends all her wealth placing bets on the horses. The gambler plays race-after-race, i.e., before placing bets for a race, she is revealed the winning horse of the previous race and receives the money from the bookmaker. Her betting strategy is hence a sequence of conditional PMFs $\left(b_{X_1}, b_{X_2|X_1}, b_{X_3|X_1 X_2}, \dots, b_{X_n|X_1 X_2 \cdots X_{n-1}}\right)$. The wealth relative is the random variable

$$S_n \triangleq \prod_{i=1}^{n} b(X_i|X_1, \dots, X_{i-1}) o(X_i). \tag{295}$$

We seek betting strategies that maximize the utility function

$$U_{\beta,n} \triangleq \begin{cases} \frac{1}{\beta} \log \mathrm{E}[S_n^{\beta}] & \text{if } \beta \neq 0, \\ \mathrm{E}[\log S_n] & \text{if } \beta = 0. \end{cases} \tag{296}$$

We first establish that to maximize $U_{\beta,n}$ for a fixed $\beta \in \mathbb{R}$, it suffices to use the same betting strategy in every race; see Theorem 13. We then show that the individual-sequence-universal strategy by Cover–Ordentlich [48] allows to asymptotically achieve the same normalized utility without knowing p or β (see Theorem 14).

For a fixed $\beta \in \mathbb{R}$, let the PMF b^* be a betting strategy that maximizes the single-race utility U_β discussed in Section 8, and denote by U_β^* the utility associated with b^*. Using the same betting strategy b^* over n races leads to the utility $U_{\beta,n}$, and it follows from (295) and (296) that

$$U_{\beta,n} = n U_\beta^*. \tag{297}$$

As we show next, $n U_\beta^*$ is the maximum utility that can be achieved among all betting strategies:

Theorem 13. *Let $\beta \in \mathbb{R}$, and let $\left(b_{X_1}, b_{X_2|X_1}, b_{X_3|X_1X_2}, \ldots, b_{X_n|X_1X_2\cdots X_{n-1}}\right)$ be a sequence of conditional PMFs. Then,*

$$U_{\beta,n} \leq n U_\beta^*. \tag{298}$$

Proof. We show (298) for $\beta > 0$; analogous arguments establish (298) for $\beta < 0$ and $\beta = 0$. We prove (298) by induction on n. For $n = 1$, (298) holds because U_β^* is the maximum single-race utility. Assume now $n \geq 2$ and that (298) is valid for $n - 1$. For $\beta > 0$, (298) holds because

$$U_{\beta,n} = \frac{1}{\beta} \log \mathrm{E}[S_n^\beta] \tag{299}$$

$$= \frac{1}{\beta} \log \sum_{x_1,\ldots,x_n} P(x_1) \cdots P(x_n) \prod_{i=1}^n b(x_i|x^{i-1})^\beta o(x_i)^\beta \tag{300}$$

$$= \frac{1}{\beta} \log \sum_{x_1,\ldots,x_{n-1}} P(x_1) \cdots P(x_{n-1}) \left[\prod_{i=1}^{n-1} b(x_i|x^{i-1})^\beta o(x_i)^\beta \right] \sum_{x_n} P(x_n) b(x_n|x^{n-1})^\beta o(x_n)^\beta \tag{301}$$

$$\leq \frac{1}{\beta} \log \sum_{x_1,\ldots,x_{n-1}} P(x_1) \cdots P(x_{n-1}) \left[\prod_{i=1}^{n-1} b(x_i|x^{i-1})^\beta o(x_i)^\beta \right] \max_{b \in \mathcal{P}(\mathcal{X})} \sum_{x_n} P(x_n) b(x_n)^\beta o(x_n)^\beta \tag{302}$$

$$= \frac{1}{\beta} \log \sum_{x_1,\ldots,x_{n-1}} P(x_1) \cdots P(x_{n-1}) \left[\prod_{i=1}^{n-1} b(x_i|x^{i-1})^\beta o(x_i)^\beta \right] \sum_{x_n} P(x_n) b^*(x_n)^\beta o(x_n)^\beta \tag{303}$$

$$= U_{\beta,n-1} + U_\beta^* \tag{304}$$

$$\leq (n-1) U_\beta^* + U_\beta^* \tag{305}$$

$$= n U_\beta^*, \tag{306}$$

where (303) holds because b^* maximizes the single-race utility U_β, and (305) holds because (298) is valid for $n - 1$. □

In portfolio theory, Cover–Ordentlich [48] (Definition 1) proposed a universal strategy. Adapted to our setting, it leads to the following sequence of conditional PMFs:

$$\hat{b}(x_i|x^{i-1}) = \frac{\int_{b \in \mathcal{P}(\mathcal{X})} b(x_i) S_{i-1}(b, x^{i-1}) d\mu(b)}{\int_{b \in \mathcal{P}(\mathcal{X})} S_{i-1}(b, x^{i-1}) d\mu(b)}, \tag{307}$$

where $i \in \{1, 2, \ldots\}$; μ is the Dirichlet$(1/2, \ldots, 1/2)$ distribution on $\mathcal{P}(\mathcal{X})$; $S_0(b, x^0) \triangleq 1$; and

$$S_i(b, x^i) \triangleq \prod_{j=1}^i b(x_j) o(x_j). \tag{308}$$

This strategy depends neither on the winning probabilities p nor on the parameter β. Denoting the utility (296) associated with the strategy $\hat{b}(x_i|x^{i-1})$ by $\hat{U}_{\beta,n}$, we have the following result:

Theorem 14. *For every $\beta \in \mathbb{R}$,*

$$n U_\beta^* - \log 2 - \frac{|\mathcal{X}| - 1}{2} \log(n+1) \leq \hat{U}_{\beta,n} \tag{309}$$

$$\leq n U_\beta^*. \tag{310}$$

Hence,

$$\lim_{n \to \infty} \frac{1}{n} \hat{U}_{\beta,n} = U_\beta^*. \tag{311}$$

Proof. Inequality (310) follows from Theorem 13; and (311) follows from (309) and (310) and the sandwich theorem. It thus remains to establish (309): We do so for $\beta > 0$; analogous arguments establish (309) for $\beta < 0$ and $\beta = 0$. For a fixed sequence $x^n \in \mathcal{X}^n$, let $\tilde{b}$ be a PMF on $\mathcal{X}$ that maximizes $S_n(b, x^n)$, and denote the wealth relative in (295) associated with using $\tilde{b}$ in every race by $\tilde{S}_n(x^n)$, thus

$$\tilde{S}_n(x^n) = \max_{b \in \mathcal{P}(\mathcal{X})} \prod_{i=1}^{n} b(x_i) o(x_i). \tag{312}$$

Let $\hat{S}_n(x^n)$ denote the wealth relative in (295) associated with the strategy $\hat{b}(x_i|x^{i-1})$ and the sequence x^n. Using [48] (Theorem 2) it follows that, for every $x^n \in \mathcal{X}^n$,

$$\hat{S}_n(x^n) \geq \frac{1}{2(n+1)^{(|\mathcal{X}|-1)/2}} \tilde{S}_n(x^n). \tag{313}$$

This implies that (309) holds for $\beta > 0$ because

$$\hat{U}_{\beta,n} = \frac{1}{\beta} \log \mathrm{E}\big[\hat{S}_n(X^n)^\beta\big] \tag{314}$$

$$\geq \frac{1}{\beta} \log \mathrm{E}\big[\tilde{S}_n(X^n)^\beta\big] - \log 2 - \frac{|\mathcal{X}|-1}{2} \log(n+1) \tag{315}$$

$$\geq \frac{1}{\beta} \log \sum_{x_1,\ldots,x_n} P(x_1) \cdots P(x_n) \prod_{i=1}^{n} b^*(x_i)^\beta o(x_i)^\beta - \log 2 - \frac{|\mathcal{X}|-1}{2} \log(n+1) \tag{316}$$

$$= n U_\beta^* - \log 2 - \frac{|\mathcal{X}|-1}{2} \log(n+1), \tag{317}$$

where (315) follows from (313), and (316) follows from (312). $\square$

Remark 9. *As discussed in Section 8, the optimal single-race betting strategy varies significantly with different values of β, thus it might be a bit surprising that the Cover–Ordentlich strategy is not only universal with respect to the winning probabilities, but also with respect to β. This is due to the following two reasons: First, for fixed winning probabilities and a fixed β, it is optimal to use the same betting strategy in every race (see Theorem 13). Second, for every $x^n \in \mathcal{X}^n$, the wealth relative of the Cover–Ordentlich strategy is not much worse than that of using the same strategy $b(\cdot)$ in every race, irrespective of $b(\cdot)$ (see (313)). Hence, irrespective of the optimal single-race betting strategy, the Cover–Ordentlich strategy is able to asymptotically achieve the same normalized utility.*

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

Appendix A. Proof of Proposition 1

These properties mostly follow from van Erven–Harremoës [8]:

(a) See [8] (Theorem 8).

(b) This follows from the definitions in (4) and (21)–(23) and the conventions in (20).

(c) This follows from [8] (Theorem 7) and the fact that $\lim_{\alpha \to 1} D_\alpha(P\|Q) = D(P\|Q)$ by L'Hôpital's rule. (Note that $\alpha \mapsto D_\alpha(P\|Q)$ does not need to be continuous at $\alpha = 1$ when the alphabets are not finite; see the discussion after [8] (Equation (18)).)

(d) See [8] (Theorem 3).

(e) Let $\alpha, \alpha' \in (0, \infty)$ satisfy $\alpha \le \alpha'$. Then,

$$\frac{1-\alpha}{\alpha} D_\alpha(P\|Q) = D_{1-\alpha}(Q\|P) \tag{A1}$$

$$\ge D_{1-\alpha'}(Q\|P) \tag{A2}$$

$$= \frac{1-\alpha'}{\alpha'} D_{\alpha'}(P\|Q), \tag{A3}$$

where (A1) and (A3) follow from [8] (Lemma 10), and (A2) holds because the Rényi divergence, extended to negative orders, is nondecreasing ([8] (Theorem 39)).

(f) See [8] (Corollary 2).

(g) For $\alpha \in (0, \infty)$,

$$(\alpha - 1) D_{1/\alpha}(P\|Q) = \alpha \left(1 - \frac{1}{\alpha}\right) D_{1/\alpha}(P\|Q) \tag{A4}$$

$$= \alpha \inf_R \left[\frac{1}{\alpha} D(R\|P) + \left(1 - \frac{1}{\alpha}\right) D(R\|Q)\right] \tag{A5}$$

$$= \inf_R \left[D(R\|P) + (\alpha - 1) D(R\|Q)\right], \tag{A6}$$

where (A5) follows from [8] (Theorem 30). Hence, $(\alpha - 1) D_{1/\alpha}(P\|Q)$ is concave in α because the expression in square brackets on the RHS of (A6) is concave in α for every R and because the pointwise infimum preserves the concavity.

(h) See [8] (Theorem 9).

Appendix B. Proof of Theorem 1

Beginning with (29),

$$D_\alpha^{\mathsf{c}}(P_{Y'|X}\|Q_{Y'|X}|P_X) = \sum_{x \in \operatorname{supp}(P_X)} P(x) D_\alpha(P_{Y'|X=x}\|Q_{Y'|X=x}) \tag{A7}$$

$$\le \sum_{x \in \operatorname{supp}(P_X)} P(x) D_\alpha(P_{Y|X=x}\|Q_{Y|X=x}) \tag{A8}$$

$$= D_\alpha^{\mathsf{c}}(P_{Y|X}\|Q_{Y|X}|P_X), \tag{A9}$$

where (A8) follows by applying, separately for every $x \in \operatorname{supp}(P_X)$, Proposition 1 (h) with the conditional PMF $A_{Y'|Y,X=x}$.

Appendix C. Proof of Theorem 2

We show (43) for $\alpha \in (0,1)$; the claim then extends to $\alpha \in [0,1]$ by the continuity of $D_\alpha^c(\cdot)$ in α (Proposition 2 (c)). Let $\alpha \in (0,1)$. Keeping in mind that $\alpha - 1 < 0$, (43) holds because

$$
(\alpha - 1)\, D_\alpha^c(P_{Y|X'} \| Q_{Y|X'} | P_{X'})
$$

$$
= \sum_{x' \in \mathrm{supp}(P_{X'})} P_{X'}(x') \log \sum_y P_{Y|X'}(y|x')^\alpha \, Q_{Y|X'}(y|x')^{1-\alpha} \tag{A10}
$$

$$
= \sum_{x' \in \mathrm{supp}(P_{X'})} P_{X'}(x') \log \sum_y \left[\sum_x B_{X|X'}(x|x') P_{Y|X}(y|x) \right]^\alpha \left[\sum_x B_{X|X'}(x|x') Q_{Y|X}(y|x) \right]^{1-\alpha} \tag{A11}
$$

$$
\geq \sum_{x' \in \mathrm{supp}(P_{X'})} P_{X'}(x') \log \sum_y \sum_x B_{X|X'}(x|x') P_{Y|X}(y|x)^\alpha \, Q_{Y|X}(y|x)^{1-\alpha} \tag{A12}
$$

$$
= \sum_{x' \in \mathrm{supp}(P_{X'})} P_{X'}(x') \log \sum_{x \in \mathrm{supp}(P_X)} B_{X|X'}(x|x') \sum_y P_{Y|X}(y|x)^\alpha \, Q_{Y|X}(y|x)^{1-\alpha} \tag{A13}
$$

$$
\geq \sum_{x' \in \mathrm{supp}(P_{X'})} P_{X'}(x') \sum_{x \in \mathrm{supp}(P_X)} B_{X|X'}(x|x') \log \sum_y P_{Y|X}(y|x)^\alpha \, Q_{Y|X}(y|x)^{1-\alpha} \tag{A14}
$$

$$
= \sum_{x \in \mathrm{supp}(P_X)} P_X(x) \left[\sum_{x' \in \mathrm{supp}(P_{X'})} B_{X'|X}(x'|x) \right] \log \sum_y P_{Y|X}(y|x)^\alpha \, Q_{Y|X}(y|x)^{1-\alpha} \tag{A15}
$$

$$
= \sum_{x \in \mathrm{supp}(P_X)} P_X(x) \log \sum_y P_{Y|X}(y|x)^\alpha \, Q_{Y|X}(y|x)^{1-\alpha} \tag{A16}
$$

$$
= (\alpha - 1)\, D_\alpha^c(P_{Y|X} \| Q_{Y|X} | P_X), \tag{A17}
$$

where (A10) follows from (30); (A11) follows from (41) and (42); (A12) follows from Hölder's inequality; (A13) holds because $B_{X|X'}(x|x') = 0$ if $P_{X'}(x') > 0$ and $P_X(x) = 0$; (A14) follows from Jensen's inequality because $\log(\cdot)$ is concave; (A15) follows from (40); (A16) holds because $P_X(x) > 0$ and $P_{X'}(x') = 0$ imply $B_{X'|X}(x'|x) = 0$, hence the expression in square brackets on the LHS of (A16) equals one; and (A17) follows from (30).

Appendix D. Proof of Corollary 1

Applying Theorem 2 with $\mathcal{X}' \triangleq \{1\}$ and the conditional PMF $B_{X'|X}(x'|x) \triangleq 1$, we obtain

$$
D_\alpha^c(P_{Y|X'} \| Q_{Y|X'} | P_{X'}) \leq D_\alpha^c(P_{Y|X} \| Q_{Y|X} | P_X). \tag{A18}
$$

To complete the proof of (48), observe that

$$
D_\alpha^c(P_{Y|X'} \| Q_{Y|X'} | P_{X'}) = D_\alpha^c(P_Y \| Q_Y | P_{X'}) \tag{A19}
$$

$$
= D_\alpha(P_Y \| Q_Y), \tag{A20}
$$

where (A19) holds because (41) and (46) imply $P_{Y|X'}(y|x') = P_Y(y)$ and because (42) and (47) imply $Q_{Y|X'}(y|x') = Q_Y(y)$; and (A20) follows from Remark 1.

Appendix E. Proof of Example 1

If $\alpha = \infty$, then it can be verified numerically that (53) holds for $\epsilon = 0.1$. Fix now $\alpha \in (1, \infty)$. Then, for all $\epsilon \in (0,1)$,

$$
D_\alpha\left(P_Y \| Q_Y^{(\epsilon)}\right) = \frac{1}{\alpha - 1} \log \left[0.5^\alpha (1 - \epsilon)^{1-\alpha} + 0.5^\alpha \epsilon^{1-\alpha} \right] \tag{A21}
$$

$$
\geq \frac{1}{\alpha - 1} \log \left[0.5^\alpha \epsilon^{1-\alpha} \right] \tag{A22}
$$

$$= \frac{\alpha}{\alpha - 1} \log 0.5 + \log \frac{1}{\epsilon}. \tag{A23}$$

The RHS of (53) satisfies, for sufficiently small ϵ,

$$D_\alpha^c\big(P_{Y|X}^{(\epsilon)}\|Q_Y^{(\epsilon)}|P_X\big) = 0.5 \cdot 0 + 0.5 \cdot D_\alpha\big(P_{Y|X=1}^{(\epsilon)}\|Q_Y^{(\epsilon)}\big) \tag{A24}$$

$$= \frac{0.5}{\alpha - 1} \log\Big[\epsilon^\alpha (1 - \epsilon)^{1-\alpha} + (1 - \epsilon)^\alpha \epsilon^{1-\alpha}\Big] \tag{A25}$$

$$= \frac{0.5}{\alpha - 1} \log\Big[\epsilon^{1-\alpha} \big((1 - \epsilon)^\alpha + \epsilon^{2\alpha-1} (1 - \epsilon)^{1-\alpha}\big)\Big] \tag{A26}$$

$$\leq \frac{0.5}{\alpha - 1} \log\Big[2\epsilon^{1-\alpha}\Big] \tag{A27}$$

$$= \frac{0.5}{\alpha - 1} \log 2 + 0.5 \log \frac{1}{\epsilon}, \tag{A28}$$

where (A27) holds for sufficiently small ϵ because $\lim_{\epsilon\downarrow0}\big((1 - \epsilon)^\alpha + \epsilon^{2\alpha-1}(1 - \epsilon)^{1-\alpha}\big) = 1$. Because $\lim_{\epsilon\downarrow0} \log \frac{1}{\epsilon} = \infty$, (53) follows from (A23) and (A28) for sufficiently small ϵ.

Appendix F. Proof of Theorem 3

Observe that, for all $x' \in \mathcal{X}$ and all $y' \in \mathcal{Y}'$,

$$P_X(x')P_{Y'|X}(y'|x') = \sum_{x,y} P_X(x)P_{Y|X}(y|x)\, \mathbb{1}\{x' = x\}\, A_{Y'|XY}(y'|x,y), \tag{A29}$$

$$P_X(x')Q_{Y'|X}(y'|x') = \sum_{x,y} P_X(x)Q_{Y|X}(y|x)\, \mathbb{1}\{x' = x\}\, A_{Y'|XY}(y'|x,y). \tag{A30}$$

Hence, (68) follows from (54) and

$$D_\alpha(P_X P_{Y'|X}\|P_X Q_{Y'|X}) \leq D_\alpha(P_X P_{Y|X}\|P_X Q_{Y|X}), \tag{A31}$$

which follows from the data-processing inequality for the Rényi divergence by substituting $\mathbb{1}_{X'=X}A_{Y'|XY}$ for $A_{X'Y'|XY}$ in Proposition 1 (h).

Appendix G. Proof of Theorem 4

Observe that, for all $x' \in \mathcal{X}'$ and all $y' \in \mathcal{Y}$,

$$P_{X'}(x')P_{Y|X'}(y'|x') = \sum_{x,y} P_X(x)P_{Y|X}(y|x)\, B_{X'|X}(x'|x)\, \mathbb{1}\{y' = y\}, \tag{A32}$$

$$P_{X'}(x')Q_{Y|X'}(y'|x') = \sum_{x,y} P_X(x)Q_{Y|X}(y|x)\, B_{X'|X}(x'|x)\, \mathbb{1}\{y' = y\}. \tag{A33}$$

Hence, (73) follows from (54) and

$$D_\alpha(P_{X'} P_{Y|X'}\|P_{X'} Q_{Y|X'}) \leq D_\alpha(P_X P_{Y|X}\|P_X Q_{Y|X}), \tag{A34}$$

which follows from the data-processing inequality for the Rényi divergence by substituting $B_{X'|X}\mathbb{1}_{Y'=Y}$ for $A_{X'Y'|XY}$ in Proposition 1 (h).

References

1. Kelly, J.L., Jr. A new interpretation of information rate. *Bell Syst. Tech. J.* **1956**, *35*, 917–926. [CrossRef]
2. 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.
3. Lapidoth, A.; Pfister, C. Two measures of dependence. *Entropy* **2019**, *21*, 778. [CrossRef]

4. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

5. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011; ISBN 978-0-521-19681-9.

6. Csiszár, I.; Shields, P.C. *Information Theory and Statistics: A Tutorial*; now Publishers: Hanover, MA, USA, 2004; ISBN 978-1-933019-05-5.

7. 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.

8. van Erven, T.; Harremoës, P. Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [CrossRef]

9. Csiszár, I. Generalized cutoff rates and Rényi's information measures. *IEEE Trans. Inf. Theory* **1995**, *41*, 26–34. [CrossRef]

10. Sibson, R. Information radius. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **1969**, *14*, 149–160. [CrossRef]

11. Gallager, R.G. *Information Theory and Reliable Communication*; John Wiley & Sons: Hoboken, NJ, USA, 1968; ISBN 978-0-471-29048-3.

12. 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.

13. Eeckhoudt, L.; Gollier, C.; Schlesinger, H. *Economic and Financial Decisions under Risk*; Princeton University Press: Princeton, NJ, USA, 2005; ISBN 978-0-691-12215-1.

14. Soklakov, A.N. Economics of disagreement – financial intuition for the Rényi divergence. *arXiv* **2018**, arXiv:1811.08308.

15. Steele, J.M. *The Cauchy–Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*; Cambridge University Press: Cambridge, UK, 2004; ISBN 978-0-521-54677-5.

16. Bullen, P.S. *Handbook of Means and Their Inequalities*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003; ISBN 978-1-4020-1522-9.

17. Campbell, L.L. A coding theorem and Rényi's entropy. *Inf. Control* **1965**, *8*, 423–429. [CrossRef]

18. Campbell, L.L. Definition of entropy by means of a coding problem. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **1966**, *6*, 113–118. [CrossRef]

19. Merhav, N. On optimum strategies for minimizing the exponential moments of a loss function. *Commun. Inf. Syst.* **2011**, *11*, 343–368. [CrossRef]

20. 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]

21. Augustin, U. Noisy Channels. Habilitation Thesis, Universität Erlangen–Nürnberg, Erlangen, Germany, 1978.

22. Nakiboğlu, B. The Augustin capacity and center. *Probl. Inf. Transm.* **2019**, *55*, 299–342. [CrossRef]

23. Nakiboğlu, B. The sphere packing bound for memoryless channels. *arXiv* **2018**, arXiv:1804.06372.

24. Ho, S.-W.; Verdú, S. Convexity/concavity of Rényi entropy and α-mutual information. In Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015; pp. 745–749. [CrossRef]

25. 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]

26. Nakiboğlu, B. The Rényi capacity and center. *IEEE Trans. Inf. Theory* **2019**, *65*, 841–860. [CrossRef]

27. Cai, C.; Verdú, S. Conditional Rényi divergence saddlepoint and the maximization of α-mutual information. *Entropy* **2019**, *21*, 969. [CrossRef]

28. Fong, S.L.; Tan, V.Y.F. Strong converse theorems for classes of multimessage multicast networks: A Rényi divergence approach. *IEEE Trans. Inf. Theory* **2016**, *62*, 4953–4967. [CrossRef]

29. Yu, L.; Tan, V.Y.F. Rényi resolvability and its applications to the wiretap channel. *IEEE Trans. Inf. Theory* **2019**, *65*, 1862–1897. [CrossRef]

30. Gallager, R.G. A simple derivation of the coding theorem and some applications. *IEEE Trans. Inf. Theory* **1965**, *11*, 3–18. [CrossRef]

31. Shannon, C.E.; Gallager, R.G.; Berlekamp, E.R. Lower bounds to error probability for coding on discrete memoryless channels. I. *Inf. Control* **1967**, *10*, 65–103. [CrossRef]

32. Shannon, C.E.; Gallager, R.G.; Berlekamp, E.R. Lower bounds to error probability for coding on discrete memoryless channels. II. *Inf. Control* **1967**, *10*, 522–552. [CrossRef]

33. Arimoto, S. On the converse to the coding theorem for discrete memoryless channels. *IEEE Trans. Inf. Theory* **1973**, *19*, 357–359. [CrossRef]

34. Polyanskiy, Y.; Verdú, S. Arimoto channel coding converse and Rényi divergence. In Proceedings of the 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Allerton, IL, USA, 29 September–1 October 2010; pp. 1327–1333. [CrossRef]

35. Nakiboğlu, B. The sphere packing bound via Augustin's method. *IEEE Trans. Inf. Theory* **2019**, *65*, 816–840. [CrossRef]

36. Nakiboğlu, B. The sphere packing bound for DSPCs with feedback à la Augustin. *IEEE Trans. Commun.* **2019**, *67*, 7456–7467. [CrossRef]

37. 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]

38. 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. [CrossRef]

39. 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]

40. Fehr, S.; Berens, S. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 6801–6810. [CrossRef]

41. Sason, I.; Verdú, S. Arimoto–Rényi conditional entropy and Bayesian *M*-ary hypothesis testing. *IEEE Trans. Inf. Theory* **2018**, *64*, 4–25. [CrossRef]

42. Arıkan, E. An inequality on guessing and its application to sequential decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105. [CrossRef]

43. Sundaresan, R. Guessing under source uncertainty. *IEEE Trans. Inf. Theory* **2007**, *53*, 269–287. [CrossRef]

44. Bracher, A.; Lapidoth, A.; Pfister, C. Guessing with distributed encoders. *Entropy* **2019**, *21*, 298. [CrossRef]

45. Bunte, C.; Lapidoth, A. Encoding tasks and Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 5065–5076. [CrossRef]

46. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970; ISBN 978-0-691-01586-6.

47. Moser, S.M. Information Theory (Lecture Notes), version 6.6. 2018. Available online: http://moser-isi.ethz.ch/scripts.html (accessed on 8 March 2020).

48. Cover, T.M.; Ordentlich, E. Universal portfolios with side information. *IEEE Trans. Inf. Theory* **1996**, *42*, 348–363. [CrossRef]