The Rényi Capacity and Center

Barış Nakiboğlu

Canım halam Fatma Nakiboğlu Aydiç'in anısına adanmıştır.
Dedicated to the memory of my dear aunt Fatma Nakiboğlu Aydiç.

Abstract—Rényi's information measures —the Rényi information, mean, capacity, radius, and center— are analyzed relying on the elementary properties of the Rényi divergence and the power means. The van Erven-Harremoës conjecture is proved for any positive order and for any set of probability measures on a given measurable space and a generalization of it is established for the constrained variant of the problem. The finiteness of the order α Rényi capacity is shown to imply the continuity of the Rényi capacity on $(0, \alpha]$ and the uniform equicontinuity of the Rényi information, both as a family of functions of the order indexed by the priors and as a family of functions of the prior indexed by the orders. The Rényi capacities and centers of various families of Poisson processes are derived as examples.

CONTENTS

e-mail:bnakib@metu.edu.tr

I. INTRODUCTION

Information transmission problems are often posed on models with finite sample spaces or on models with specific noise structures, such as Gaussian or Poisson models. As a result, certain fundamental observations such as the minimax theorem for the Shannon capacity in terms of the Kullback-Leibler divergence or the existence of a unique "capacity achieving output distribution", i.e. the existence of a unique Shannon center, are established either for models with finite sample spaces or for specific noise structures. In [56], Kemperman proved these assertions far more generally by interpreting the channel as a set of probability measures on a given measurable space.

In a sense, Kemperman tacitly suggests a purely measure theoretic understanding of the Shannon capacity and center that is separated from their significance in the information transmission problems. Even without the generality afforded by the measure theoretic framework, such an understanding is appealing because Shannon capacity and center come up in various information transmission problems, with very different operational meanings. Consider for example a finite set $\mathcal{W}$ of probability mass functions on a finite output set $\mathcal{Y}$.

- If we interpret $\mathcal{W}$ as a discrete channel that is to be used multiple times, then the Shannon capacity of $\mathcal{W}$ is the largest rate at which one can communicate reliably via the channel $\mathcal{W}$, [87].
- If we interpret $\mathcal{W}$ as a collection of sources that is to be encoded by a lossless variable length source code, then the Shannon capacity is a lower bound on the worst redundancy among the members of $\mathcal{W}$, which is off at most by one for some lossless variable length source code, [29], [36], [81].

In this paper we propose an analogous measure theoretic understanding for the Rényi capacity and center. Our interest in these concepts stems from their operational significance in the channel coding problem; we elucidate that operational significance in our concurrent paper [73]. Because of the generality of the measure theoretic model we adopt in this paper, we can discuss in [73] the operational significance of these concepts for a diverse family of channels in a unified framework. In the current paper our main aim is to present an analysis starting from the measure theoretic first principles and the elementary properties of the Rényi divergence. We will first present a brief overview of the Rényi information, divergence, and mean. Then we proceed with the analysis of the Rényi capacity and center.

Deriving the technical results employed in [73] is one of the main aims of the current paper; however, the scope of our analysis is not restricted to the needs of the particular

analysis we present in [73]. We aim to build a more complete understanding of Rényi's information measures that might lead us to new analysis techniques for the problems we investigate in [73] or for other information transmission problems involving Rényi's information measures. Our abstract and general framework is conducive to this purpose; in addition it allows us to observe certain phenomena that cannot be observed in simpler models. For example,[1] the Rényi capacity is either a continuous function of the order on $(0, \infty)$ or a finite and continuous function of the order on $(0, \phi]$ that is infinite on (ϕ, ∞) for some $\phi \in [1, \infty)$. This dichotomy, however, cannot be observed with models with finite $\mathcal{W}$ or finite $\mathcal{Y}$ because the Rényi capacity is bounded if either $\mathcal{W}$ or $\mathcal{Y}$ is finite.

In [79], Rényi provided an axiomatic characterization of a family of divergences for pairs of probability mass functions on a given finite sample space; the resulting family of divergences, parametrized by positive real numbers, are named after him. The definition of the Rényi divergence has been extended to pairs of probability measures. Recently, van Erven and Harremoës provided a comprehensive investigation of various properties of the Rényi divergence in [31]. For any α in $[0, \infty]$, the order α Rényi divergence between probability measures w and q, denoted by $D_\alpha(w \| q)$, is zero when w is equal to q and non-negative when w is not equal to q. Hence, given a measurable space $(\mathcal{Y}, \mathcal{Y})$ we can use the order α Rényi divergence to measure the spread of any set of probability measures $\mathcal{W}$ relative to any probability measure q on $(\mathcal{Y}, \mathcal{Y})$ as follows:

$$S_{\alpha, \mathcal{W}}(q) \triangleq \sup_{w \in \mathcal{W}} D_\alpha(w \| q). \tag{1}$$

$S_{\alpha, \mathcal{W}}(q)$ is called *the order α Rényi radius of $\mathcal{W}$ relative to q*. By taking the infimum of $S_{\alpha, \mathcal{W}}(q)$ over all probability measures q on $(\mathcal{Y}, \mathcal{Y})$, we get an absolute measure of the spread of $\mathcal{W}$, called *the order α Rényi radius of $\mathcal{W}$*,

$$S_{\alpha, \mathcal{W}} \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{w \in \mathcal{W}} D_\alpha(w \| q). \tag{2}$$

Any probability measure q on the measurable space $(\mathcal{Y}, \mathcal{Y})$ satisfying $S_{\alpha, \mathcal{W}}(q) = S_{\alpha, \mathcal{W}}$, is called *an order α Rényi center of $\mathcal{W}$*. The order one Rényi divergence is the Kullback-Leibler divergence; hence the order one Rényi radius and center are the Shannon radius and center referred to in [56].

The Shannon capacity, defined as the supremum of the mutual information, is another measure of the spread of a set of probability measures on a given measurable space. In order to have a parametric generalization of the Shannon capacity, similar to the one provided by the Rényi radius to the Shannon radius, we need a parametric generalization of the mutual information. Sibson [94] proposed one such parametric generalization using the Rényi divergence, called the Rényi information, see Definition 4. For any set of probability measures $\mathcal{W}$ on a given measurable space $(\mathcal{Y}, \mathcal{Y})$, probability mass function p on $\mathcal{W}$, and positive real number α, $I_\alpha(p; \mathcal{W})$ is *the order α Rényi information*[2] *for prior p*. The order one Rényi information equals to the mutual information. For other

positive real orders, the order α Rényi information can be described in terms of Gallager's function introduced in [35]:

$$I_\alpha(p; \mathcal{W}) = \frac{E_0(\rho, p)}{\rho} \Big|_{\rho = \frac{1-\alpha}{\alpha}} \qquad \forall \alpha \in \mathbb{R}_+ \setminus \{1\} \tag{3}$$

where Gallager's function $E_0(\rho, p)$ is defined for $\rho > -1$ as

$$E_0(\rho, p) \triangleq -\ln \int \left(\sum_w p(w) (\tfrac{\mathrm{d}w}{\mathrm{d}\nu})^{\frac{1}{1+\rho}} \right)^{1+\rho} \nu(\mathrm{d}y). \tag{4}$$

The order α Rényi capacity $C_{\alpha, \mathcal{W}}$ is defined as the supremum of the order α Rényi information $I_\alpha(p; \mathcal{W})$ over all priors p.

There are at least two other ways to define the Rényi information for which the order one Rényi information is equal to the mutual information: one by Arimoto [4] and another one by Augustin [6] and Csiszár [25]. A review of these three definitions of the Rényi information has recently been provided by Verdú [105]. Assuming $\mathcal{W}$ and $\mathcal{Y}$ to be finite sets, Csiszár showed that the order α Rényi capacity for all three definitions of the Rényi information are equal to one another and to the order α Rényi radius, [25, Prop. 1].

The extension of Kemperman's result [56, Thm. 1] about the Shannon capacity and center given in Theorem 1, presented in the following, is among the most important observations about the Rényi capacity and center. Theorem 1 establishes the equality of $C_{\alpha, \mathcal{W}}$ and $S_{\alpha, \mathcal{W}}$ for any positive order α and set of probability measures $\mathcal{W}$. Furthermore, it asserts the existence of a unique order α Rényi center $q_{\alpha, \mathcal{W}}$ whenever $C_{\alpha, \mathcal{W}}$ is finite and characterizes the unique order α Rényi center in terms of the order α Rényi means. These observations, however, have been reported in various forms before, at least partially. In [5], Augustin considered the orders in $(0, 1)$, proved a result equivalent to Theorem 1 for finite $\mathcal{W}$'s and described how this result can be extended to arbitrary $\mathcal{W}$'s. Later, Augustin established a result, [6, Thm. 26.6′], that implies Theorem 1 for all orders in α in $(0, 2)$. Csiszár [25, Prop. 1] proved the equality $C_{\alpha, \mathcal{W}} = S_{\alpha, \mathcal{W}}$ for arbitrary positive order α assuming $\mathcal{W}$ and $\mathcal{Y}$ are finite sets.

The equality of capacity to radius and the existence of a unique center, are phenomena that have been observed repeatedly in various contexts. In order to clarify the standing of Theorem 1 among these results, we provide a more comprehensive discussion of the previous work on these fundamental observations in §I-A.

The current paper and the concurrent paper [73] grew out of a desire to understand Augustin's proofs of the sphere packing bound given in [5] and [6] more intuitively. Augustin's proofs are important because, among other things, they are the only proofs of the sphere packing bound for non-stationary product channels, even for the case of discrete channels. Concepts of Rényi capacity, radius, and center provide a way to express the principal novelty of Augustin's method in a succinct and intuitive way. We discuss the novel observation underlying Augustin's method and its promise briefly in §I-B.

Similar to Theorem 1, some of the observations that we discuss in the paper have been reported before either in terms of Rényi's information measures [25], [94] or in terms of other related quantities, such as Gallager's function, [5], [6], [35], [37]. But we also have a number of new observations that

---

[1]This dichotomy is an immediate consequence of Lemma 15, see page 11.

[2]Sibson defines "the information radius of order α" through an infimum and then derives a closed form expression for it in [94, Thm. 2.2]. We take that closed form expression as the definition of the order α Rényi information.

have not been reported before. We provide a tally of our most important contributions in §I-C.

We conclude the current section with a summary of our notational conventions presented in §I-D. It is worth mentioning that only §I-D is necessary to understand the rest of the paper; readers may bypass other parts of the current section depending on their interest and background.

The Rényi entropy [79] is another information measure, that is intimately related to the information measures discussed in this paper. The Rényi entropy [7], [77] and its variants [4], [33], [84], [98] are of interest by themselves [14], [45], [52], [78]; in addition they have been used to pose projection problems [61]–[63] related to guessing [3], [85], [96] and various questions about the information transmission problems [10], [12], [97]. Recently, there has been a revived interest in Rényi's information measures and their operational significance [11], [15]–[17], [27], [34], [91], [100], [104], in general.

A. Radius, Center, and Capacity

The concepts of radius and center, as we use them, are analogous to their counter parts in Euclidean geometry. Let $\mathcal{W}$ be a set of points in the n dimensional Euclidean space $\mathbb{R}^n$ and q be a point in the same space. Then one measure of the spread of $\mathcal{W}$ relative to q is the infimum of the radii of the q-centered spheres including all points of $\mathcal{W}$, called *the Chebyshev radius of $\mathcal{W}$ relative to q*:

$$\mathcal{S}_{\mathcal{W}}(q) \triangleq \sup_{w \in \mathcal{W}} \|w - q\|_2 \qquad \forall \mathcal{W} \subset \mathbb{R}^n, q \in \mathbb{R}^n.$$

If we do not require the centers of the spheres to be at a given point q, then we get an absolute measure of the spread of $\mathcal{W}$, called *the Chebyshev radius of $\mathcal{W}$*:

$$\mathcal{S}_{\mathcal{W}} \triangleq \inf_{q \in \mathbb{R}^n} \sup_{w \in \mathcal{W}} \|w - q\|_2 \qquad \forall \mathcal{W} \subset \mathbb{R}^n.$$

If $\mathcal{S}_{\mathcal{W}}$ is finite, then there exists[3] a unique Chebyshev center $q_{\mathcal{W}}$ satisfying $\mathcal{S}_{\mathcal{W}}(q_{\mathcal{W}}) = \mathcal{S}_{\mathcal{W}}$.

For any set of points in a metric space $(\mathcal{X}, d)$, one can define the Chebyshev radius by replacing $\mathbb{R}^n$ with $\mathcal{X}$ and $\|w - q\|_2$ with $d(w, q)$ in the definition. However, neither the existence nor the uniqueness of the Chebyshev center is a foregone conclusion for such generalizations. Garkavi [39, Thm. 1] provides a three point set in a Banach space that does not have a Chebyshev center. In the Hamming space of length two binary strings, both $(0, 0)$ and $(1, 1)$ are Chebyshev centers of the set $\mathcal{W} = \{(0, 1), (1, 0)\}$. See [2, Ch. 15], for a discussion of these concepts on the inner product spaces.

The Chebyshev radius is, in a sense, special because it is defined via the distance measure —the metric corresponding to the norm of the space for normed spaces and the metric of the space for metric spaces— that is a part of the description of the space. In principle, one can measure the relative and the absolute spread of the subsets of $\mathcal{X}$ using any non-negative function g on $\mathcal{X} \times \mathcal{X}$ satisfying $g(x, x) = 0$ for all $x \in \mathcal{X}$ and define a center accordingly. However, neither the existence nor the uniqueness of such a center is guaranteed.

When $\mathcal{X}$ in the above formulation is the space of all probability measures $\mathcal{P}(\mathcal{Y})$ on a measurable space $(\mathcal{Y}, \mathscr{Y})$, one can measure the spread of a subset $\mathcal{W}$ of $\mathcal{P}(\mathcal{Y})$ using the Kullback-Leibler divergence. The resulting radius is nothing but the Shannon radius of $\mathcal{W}$ and whenever the Shannon radius is finite the existence of a unique Shannon center follows from Kemperman's result [56, Thm. 1]. The other assertion of Kemperman's result [56, Thm. 1] is the equality of the Shannon radius of $\mathcal{W}$ and the Shannon capacity of $\mathcal{W}$, defined as the supremum of the mutual information $I(p; \mathcal{W})$ over all probability mass functions p on $\mathcal{W}$. For the case where both $\mathcal{W}$ and $\mathcal{Y}$ are finite sets, Kemperman's result was already known at the time [37, Thm. 4.5.1]; in [56] Kemperman attributes this special case to Shannon [89]. For the case when $\mathcal{Y}$ is a finite set, first Gallager [36, Thm. A] and then Davisson and Leon-Garcia [29, Thm. 3] proved results equivalent to Kemperman's. Later, Haussler [49] proved Kemperman's result assuming $\mathcal{Y}$ to be a complete separable metric space, i.e. Polish space, and $\mathscr{Y}$ to be the associated Borel σ-algebra.

Theorem 1, which we prove in the following, extends Kemperman's result to the Rényi capacity and center of other orders. The existence of a unique center under the finite capacity hypothesis and the equality of the capacity and the radius have been confirmed in other contexts, as well.

1) Radius for f-Divergence: Csiszár [19], [21], Morimoto [65], and Ali and Silvey [1] defined the f-divergence using convex functions, satisfying $f(1) = 0$. The Kullback-Leibler divergence[4] is the f-divergence corresponding to the function $f(x) = x \ln x$. For any convex function f satisfying $f(1) = 0$, the absolute and relative f-*radius* are defined in terms of the corresponding f-divergence as follows:

$$\mathcal{S}_{f, \mathcal{W}}(q) \triangleq \sup_{w \in \mathcal{W}} \mathcal{D}_f(w \| q),$$
$$\mathcal{S}_{f, \mathcal{W}} \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{w \in \mathcal{W}} \mathcal{D}_f(w \| q).$$

The f-information and *the f-capacity* are defined in terms of corresponding f-divergence as follows

$$\mathcal{I}_f(p; \mathcal{W}) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} \mathcal{D}_f(p \circledast \mathcal{W} \| p \otimes q),$$
$$\mathcal{C}_{f, \mathcal{W}} \triangleq \sup_{p \in \mathcal{P}(\mathcal{W})} \mathcal{I}_f(p; \mathcal{W})$$

where $p \circledast \mathcal{W}$ is the probability measure whose marginal distribution on the support of p is p and whose conditional distribution is w and $p \otimes q$ is the product measure.

The mutual information[5] is the f-information corresponding to $f(x) = x \ln x$. For $\mathcal{W}$'s that are finite, Csiszár proved the following two assertions, see [23, Thm. 3.2]:

- $\mathcal{C}_{f, \mathcal{W}} = \mathcal{S}_{f, \mathcal{W}}$ for any f that is strictly convex at 1.
- There exists a unique f-center for any f that is strictly convex, provided that $\mathcal{S}_{f, \mathcal{W}}$ is finite.

For f's that are strictly convex, it seems both assertions of Csiszár [23, Thm. 3.2] can be extended to arbitrary $\mathcal{W}$'s

---

[3]The existence follows from the extreme value theorem for lower semi-continuous functions. The uniqueness is a result of the uniform convexity of finite dimensional Euclidean spaces.

[4]For positive finite orders other than one the Rényi divergence is not an f-divergence itself; but it can be written in terms of an f-divergence: $D_\alpha(w \| q) = \frac{1}{\alpha - 1} \ln(1 + (\alpha - 1)\mathcal{D}_f(w \| q))$ for $f(x) = \frac{x^\alpha - 1}{\alpha - 1}$, as previously pointed out in [20, (14)], [21, (1.10)], [22, (6)], [82, (1)], [83, (80)].

[5]For positive finite orders other than one the Rényi information can be written in terms of an f-information, using the analogous relation for divergences: $I_\alpha(p; \mathcal{W}) = \frac{1}{\alpha - 1} \ln(1 + (\alpha - 1)\mathcal{I}_f(p; \mathcal{W}))$ for $f(x) = \frac{x^\alpha - 1}{\alpha - 1}$.

using the technique employed by Kemperman, as Kemperman himself suggested in [56]. Gushchin and Zhdanov [44] proved that $\mathcal{C}_{f,\mathcal{W}}$ equals to $\mathcal{S}_{f,\mathcal{W}}$ for any convex function f and any set of probability measures $\mathcal{W}$ provided that $\mathcal{Y}$ is a complete separable metric space, i.e. Polish space, and $\mathcal{Y}$ is the associated Borel σ-algebra.

2) Radius in Quantum Information Theory: In this paper, we assume $\mathcal{W}$ to be a set of probability measures on a given measurable space. This is a generalization of the case when $\mathcal{W}$ is a set of probability mass functions on a given finite set $\mathcal{Y}$, i.e. the finite sample space case. Another generalization of the finite sample space case is obtained by assuming $\mathcal{W}$ to be a set of $|\mathcal{Y}|$-by-$|\mathcal{Y}|$ positive semidefinite, trace one, Hermitian matrices. In quantum information theory such matrices are called the density matrices; they represent the states of a $|\mathcal{Y}|$ dimensional Hilbert space $\mathcal{H}$, [51, §1.2]. The set of all such states is denoted by $\mathcal{S}(\mathcal{H})$. There is a one-to-one correspondence between the diagonal members of $\mathcal{S}(\mathcal{H})$ and the probability mass functions on $\mathcal{Y}$. As a result, statements about subsets of $\mathcal{S}(\mathcal{H})$ can be interpreted as generalizations of the corresponding statements about sets of probability mass functions on $\mathcal{Y}$.

The definition of the Kullback-Leibler divergence has been extended to the members of $\mathcal{S}(\mathcal{H})$; it is, however, customarily called *the quantum relative entropy* [51, §3.1.1]:

$$D(w\|q) \triangleq \operatorname{Tr} w(\ln w - \ln q) \qquad \forall w, q \in \mathcal{S}(\mathcal{H}). \tag{5}$$

This definition can be interpreted as an extension because for the diagonal members of $\mathcal{S}(\mathcal{H})$, the quantum relative entropy as defined in (5) is equal to the Kullback-Leibler divergence between the corresponding probability mass functions. For any subset $\mathcal{W}$ of $\mathcal{S}(\mathcal{H})$, the quantum Shannon radius is defined as $\inf_{q \in \mathcal{S}(\mathcal{H})} \sup_{w \in \mathcal{W}} D(w\|q)$.

The definition of mutual information has been extended as well, but it is called *the transmission information* [51, §4.1.1]:

$$I(p;\mathcal{W}) \triangleq \sum_{w \in \mathcal{W}} p(w) D(w\|q_p) \qquad \forall p \in \mathcal{P}(\mathcal{W}) \tag{6}$$

where $q_p = \sum_{w \in \mathcal{W}} p(w) w$. Note that when $\mathcal{W}$ includes only diagonal members of $\mathcal{S}(\mathcal{H})$, the above quantity equals to the mutual information for the prior p on the corresponding set of probability mass functions. The quantum Shannon capacity is defined as the supremum of $I(p;\mathcal{W})$ over all probability mass functions p on $\mathcal{W}$ with finite support.

The quantum Shannon capacity and radius are equal to one another for arbitrary $\mathcal{W} \subset \mathcal{S}(\mathcal{H})$ provided that $\mathcal{H}$ is a finite dimensional Hilbert space,[6] [51, Thm. 4.1], [74, Thm. 3.5], [86, (19)]. This implies the equality of Shannon capacity and radius in the classical case provided that $\mathcal{Y}$ is a finite set. However, neither Kemperman's result in [56] nor the weaker result by Haussler in [49] require $\mathcal{Y}$ to be finite. Thus those results are not subsumed by the quantum Information theoretic versions of Kemperman's result presented in [51], [74], [86].

The situation is similar for the quantum Rényi capacity, radius, and center. All the results on the equality of the quantum Rényi capacity and radius that we are aware of [27, Thm. 6], [51, (4.74)], [58, Lemma I.3], [66, Thm. IV.8], [67, Prop. 4.2], [106, Lemma 14] assume $\mathcal{W}$ to be a subset of $\mathcal{S}(\mathcal{H})$ for a finite dimensional Hilbert space $\mathcal{H}$. Hence, to the best of our knowledge, Theorem 1 is not subsumed by any of the known results in quantum information theory.

B. Augustin's Method and the Rényi Center

Augustin's proof of the sphere packing bound in [5] is one of the first few complete proofs of the sphere packing bound. Unlike its contemporaries by Shannon, Gallager and Berlekamp in [88] and by Haroutunian in [46], Augustin's proof does not assume either the stationarity of the channel or the finiteness of the input set because it does not rely on a type based expurgation (i.e. a fixed composition argument). After decades, Augustin's proofs in [5] and [6] are still the only proofs of the sphere packing bound for non-stationary product channels, even in the finite input alphabet case. Augustin's method has been applied to problems with feedback, as well. Using a variant of his method, Augustin provides a proof sketch for the derivation of the sphere packing bound for codes on discrete stationary product channels with feedback in [6]; see [72] for a complete proof following this proof sketch. What we call the discrete stationary product channels with feedback are customarily called DMCs with feedback.

Despite their strength and generality, Augustin's derivations of the sphere packing bound is scarcely known to date, even among the specialists working on related problems. In [73, §IV], we derive sphere packing bounds using Augustin's method in a way that makes the roles of the Rényi capacity and center more salient and precise. Our bound for the product channels is sharper than the corresponding bounds in [5] and [6]. In [73, §V], we present a new proof of the sphere packing bound for the discrete product channels with feedback that facilitates the ideas of Haroutunian [47] and Sheverdyaev [92], as well as Augustin [5], [6]. Our new proof for the case with feedback holds for non-stationary channels satisfying certain stationarity hypothesis. In [73, Appendix B], we discuss other aspects of the operational significance of Rényi capacity and information for the channel coding problem.

The generality and strength of Augustin's results compel one to ask: What is the principle behind Augustin's proofs of the sphere packing bound? A succinct answer exists for those who are already familiar with the concepts of Rényi capacity, radius and center.[7] In our judgment, the novel observation behind Augustin's proofs is the following:

$$\lim_{\phi \to \alpha} S_{\alpha,\mathcal{W}}(q_{\phi,\mathcal{W}}) = C_{\alpha,\mathcal{W}}.$$

In words, by choosing ϕ close enough to α, the order α Rényi radius relative to the order ϕ Rényi center can be made arbitrarily close to the order α Rényi capacity, which equals

---

[6]Results in [74] and [86] were proved with additional assumptions. In [74], Ohya, Petz, and Watanabe assumed $\mathcal{W}$ to be the image of an arbitrary Hilbert space under the channeling transformation. In [86], Shumacher and Westmoreland assumed $\mathcal{W}$ to be a closed convex set. The existence of a unique quantum Shannon center is implicit in both [74] and [86].

[7]To be precise, Augustin does not work with Rényi's information measures either in [5] or in [6]. It is, however, possible to restate his observations in terms of Rényi's information measures. His approach is eloquent and insightful, irrespective of the terms he chose to employ.

to the order α Rényi radius. This observation seems benign enough to hold for other parametric families of divergences and corresponding capacities, radii, and centers. Thus we believe that Augustin's method can probably be used to derive tight outer bounds in other information transmission problems.

C. Main Contributions

(1) If $\mathcal{W}$ and $\mathcal{Y}$ are finite sets, the continuity of the Rényi information is evident, both as a function of the order and as a function of the prior. In their proof of the sphere packing bound [88, p. 101], while proving the continuity of the Rényi capacity in the order on $(0,1)$ —for the finite $\mathcal{W}$ and $\mathcal{Y}$ case— Shannon, Gallager, and Berlekamp asserted that the Rényi information is in fact equicontinuous as a family of functions of the order on $(0,1)$ indexed by the priors. We strengthen their assertion by replacing the finiteness hypothesis on the sets $\mathcal{W}$ and $\mathcal{Y}$ with a finiteness hypothesis for the Rényi capacity, including orders greater than one, and establishing uniformity of the equicontinuity, see Lemma 16-(f). Furthermore, we show that the Rényi information is, also, uniformly equicontinuous when considered as a family of functions of the prior indexed by the orders, see Lemma 16-(e).

(2) Reflecting on [31, Thm. 37] for countable $\mathcal{Y}$'s at $\alpha = \infty$, van Erven and Harremoës conjectured the following:

Conjecture ([31, Conjecture 1]). *If $S_{\alpha,\mathcal{W}} < \infty$ for an α in $(0,\infty]$ and a $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$ then there exists a unique $q_{\alpha,\mathcal{W}} \in \mathcal{P}(\mathcal{Y})$ satisfying $S_{\alpha,\mathcal{W}} = \sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha,\mathcal{W}})$. Furthermore, for all $q \in \mathcal{P}(\mathcal{Y})$ we have*

$$\sup_{w \in \mathcal{W}} D_\alpha(w \| q) \geq S_{\alpha,\mathcal{W}} + D_\alpha(q_{\alpha,\mathcal{W}} \| q).$$

This conjecture is confirmed in Lemma 19 for the first time.[8] Lemma 19 implicitly asserts the existence of a unique $q_{\alpha,\mathcal{W}}$, which is proved in Theorem 1. This assertion, however, is not entirely new; Augustin proved an equivalent assertion for orders in $(0,2)$ in [6, Thm. 26.6'] and gave a proof sketch for an equivalent assertion for orders in $(0,1)$ in [5].

In Appendix A, we define $C_{\alpha,\mathcal{W},\mathcal{A}}$ as the supremum of $I_\alpha(p;\mathcal{W})$ over all priors p in $\mathcal{A}$ and generalize the van Erven-Harremoës bound to the convex $\mathcal{A}$ case, see Definition 10 and Lemma 25.

(3) Our framework allows us to pose and answer certain questions that are non-trivial only for infinite $\mathcal{W}$'s, i.e. infinite subsets of $\mathcal{P}(\mathcal{Y})$.

   (a) There exists a countable subset $\mathcal{W}'$ of $\mathcal{W}$ such that $C_{\alpha,\mathcal{W}'} = C_{\alpha,\mathcal{W}}$ for all α in $[0,\infty]$, Lemma 15-(b).

   (b) If $C_\eta,\mathcal{W}$ is finite, then for all $\epsilon > 0$ there exists a finite subset $\mathcal{W}'$ of $\mathcal{W}$ such that $C_{\alpha,\mathcal{W}'} > C_{\alpha,\mathcal{W}} - \epsilon$ for all α in $[\epsilon,\eta]$, Lemma 15-(g).

   (c) $C_{\alpha,\mathsf{cl}\mathcal{W}} = C_{\alpha,\mathcal{W}}$ for all α in $(0,\infty]$ where $\mathsf{cl}\mathcal{W}$ is the closure of $\mathcal{W}$ in the topology of setwise convergence, Lemma 24-(b). This has been pointed out by

---

[8] We were notified in [48] that van Erven and Harremoës had a proof establishing their conjecture in [31] under some regularity conditions, at the time.

Csiszár and Körner for α equals one case for finite $\mathcal{Y}$ in [26, Problem 8.10(b)].

D. Notational Conventions

For any set $\mathcal{Y}$, we denote the set of all subsets of $\mathcal{Y}$ by $2^\mathcal{Y}$ and the set of all probability measures on finite subsets of $\mathcal{Y}$ by $\mathcal{P}(\mathcal{Y})$. For each $p \in \mathcal{P}(\mathcal{Y})$, i.e. for each probability mass function (p.m.f.), we denote the set of all y's in $\mathcal{Y}$ for which $p(y) > 0$, by $\mathsf{supp}(p)$ and call it the support of p.

We call the pair $(\mathcal{Y}, \mathcal{Y})$ a measurable space iff $\mathcal{Y}$ is a σ-algebra of the subsets of $\mathcal{Y}$. On a measurable space $(\mathcal{Y}, \mathcal{Y})$, we denote the set of all finite signed measures by $\mathcal{M}(\mathcal{Y})$, the set of all finite measures by $\mathcal{M}_0^+(\mathcal{Y})$, the set of all non-zero finite measures by $\mathcal{M}^+(\mathcal{Y})$, and the set of all probability measures by $\mathcal{P}(\mathcal{Y})$. A countable collection $\mathcal{E}$ of the subsets of $\mathcal{Y}$ is called a $\mathcal{Y}$-measurable partition of $\mathcal{Y}$ iff $\cup_{\mathcal{E} \in \mathcal{E}} \mathcal{E} = \mathcal{Y}$, $\emptyset \notin \mathcal{E}$, $\mathcal{E} \cap \tilde{\mathcal{E}} = \emptyset$ for all $\mathcal{E}, \tilde{\mathcal{E}} \in \mathcal{E}$, and $\mathcal{E} \subset \mathcal{Y}$, [8, Def. 10.8.1].

A measure μ on the measurable space $(\mathcal{Y}, \mathcal{Y})$ is absolutely continuous with respect to another measure ν on $(\mathcal{Y}, \mathcal{Y})$, i.e. $\mu \prec \nu$, iff $\mu(\mathcal{E}) = 0$ for any $\mathcal{E} \in \mathcal{Y}$ such that $\nu(\mathcal{E}) = 0$. Measures μ and ν are equivalent, i.e. $\mu \sim \nu$, iff $\mu \prec \nu$ and $\nu \prec \mu$. Measures μ and ν are singular, i.e. $\mu \perp \nu$, iff there exists an $\mathcal{E} \in \mathcal{Y}$ such that $\mu(\mathcal{E}) = \nu(\mathcal{Y} \setminus \mathcal{E}) = 0$.

A subset $\mathcal{W}$ of $\mathcal{M}^+(\mathcal{Y})$ is absolutely continuous with respect to a measure ν, i.e. $\mathcal{W} \prec \nu$, iff $w \prec \nu$ for all $w \in \mathcal{W}$. A σ-finite measure ν is a reference measure for $\mathcal{W}$ iff $\mathcal{W} \prec \nu$. A subset $\mathcal{W}$ of $\mathcal{M}^+(\mathcal{Y})$ is uniformly absolutely continuous with respect to ν, i.e. $\mathcal{W} \prec^{uni} \nu$, iff for every $\epsilon > 0$ there exists a $\delta > 0$ such that $w(\mathcal{E}) < \epsilon$ for all $w \in \mathcal{W}$ provided that $\nu(\mathcal{E}) < \delta$. By [93, p. 366 & Thm. 2], $\mu \prec \nu$ iff $\{\mu\} \prec^{uni} \nu$. Two subsets $\mathcal{W}$ and $\mathcal{U}$ of $\mathcal{P}(\mathcal{Y})$ are singular, i.e. $\mathcal{W} \perp \mathcal{U}$, iff there exists an $\mathcal{E} \in \mathcal{Y}$ such that $w(\mathcal{E}) = 0$ for all $w \in \mathcal{W}$ and $u(\mathcal{Y} \setminus \mathcal{E}) = 0$ for all $u \in \mathcal{U}$.

We denote the Borel σ-algebra for the usual topology of the real numbers by $\mathcal{B}(\mathbb{R})$. We denote the essential supremum of a $\mathcal{Y}$-measurable, i.e. $(\mathcal{Y}, \mathcal{B}(\mathbb{R}))$-measurable, function f for the measure ν on $(\mathcal{Y}, \mathcal{Y})$ by $\mathrm{ess\,sup}_\nu f(y)$, i.e.

$$\mathrm{ess\,sup}_\nu f \triangleq \inf\{\gamma : \nu(\{y : f(y) > \gamma\}) = 0\}.$$

We denote the integral of a measurable function f on $(\mathcal{Y}, \mathcal{Y})$ with respect to the measure ν by $\int f\nu(\mathrm{d}y)$ or $\int f(y)\nu(\mathrm{d}y)$. We denote the integral by $\int f\mathrm{d}y$ or $\int f(y)\mathrm{d}y$, as well, if it is on the real line and with respect to the Lebesgue measure. If ν is a probability measure, then we also call the integral of f with respect to ν the expectation of f or the expected value of f and denote it by $\mathbf{E}_\nu[f]$ or $\mathbf{E}_\nu[f(\mathsf{Y})]$.

While discussing the continuity of measure valued functions and functions defined on sets of measures, we use either the topology of setwise convergence or the total variation topology. The topology of setwise convergence is the topology generated by the sets of the form $\{\mu : |\mu(\mathcal{E}) - t| < \epsilon\}$ for some $\mathcal{E} \in \mathcal{Y}$, $t \in \mathbb{R}_+$, $\epsilon \in \mathbb{R}_+$; see [8, §4.7(v)] for a more detailed discussion. The total variation topology is the metric topology generated by the total variation norm. For any μ in $\mathcal{M}(\mathcal{Y})$ the total variation norm of μ is defined as

$$\|\mu\| \triangleq \sup_{\mathcal{E} \in \mathcal{Y}} \mu(\mathcal{E}) - \mu(\mathcal{Y} \setminus \mathcal{E}).$$

As a consequence of the Lebesgue decomposition theorem [30, 5.5.3] and the Radon-Nikodym theorem [30, 5.5.4] we have

$$\|\mu\| = \int \left|\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right|\nu(\mathrm{d}y) \qquad \forall \mu, \nu : \mu \prec \nu.$$

Our notation will be overloaded for certain symbols; however, the relations represented by these symbols will be clear from the context. We denote the products of topologies [30, p. 38], σ-algebras [30, p. 118], and measures [30, Thm. 4.4.4] by $\otimes$. We denote the Cartesian product of sets [30, p. 38] by $\times$. We denote the absolute value of real numbers and the size of sets by $|\cdot|$. For extended real valued functions f and g on $\mathcal{Y}$, $f \leq g$ iff $f(y) \leq g(y)$ for all $y \in \mathcal{Y}$. For measures μ and ν on $(\mathcal{Y}, \mathcal{Y})$, $\mu \leq \nu$ iff $\mu(\mathcal{E}) \leq \nu(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{Y}$.

For $x, y \in \mathbb{R}$, $x \wedge y$ is the minimum of x and y. For extended real valued functions f and g on $\mathcal{Y}$, $f \wedge g$ is the pointwise minimum of f and g. For $\mu, w \in \mathcal{M}(\mathcal{Y})$, $\mu \wedge w$ is the unique measure satisfying $\frac{\mathrm{d}\mu \wedge w}{\mathrm{d}\nu} = \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \wedge \frac{\mathrm{d}w}{\mathrm{d}\nu}$ for any ν satisfying $\mu \prec \nu$ and $w \prec \nu$. If $\mathcal{F}$ is a set of real valued functions, then $\wedge_{f \in \mathcal{F}} f$ is the extended real valued function obtained by taking the pointwise infimum of f's in $\mathcal{F}$. For a $\mathcal{U} \subset \mathcal{M}(\mathcal{Y})$ satisfying $w \leq u$ for all $u \in \mathcal{U}$ for some $w \in \mathcal{M}(\mathcal{Y})$, $\wedge_{u \in \mathcal{U}} u$ is the measure which is the infimum of $\mathcal{U}$ with respect to the partial order $\leq$. The existence of a unique infimum is guaranteed by [8, Thm. 4.7.5]. We use the symbol $\vee$ analogously to $\wedge$ but we represent maxima and suprema with it, rather than minima and infima.

II. PRELIMINARIES

We commence our discussion by defining the mean measure and analyzing it, first as a function of the order for a given prior then as a function of the prior for a given order. After that we define the Rényi information using the mean measure and analyze it as a function of the order and the prior using the analysis of the mean measure. Then we define the Rényi divergence and review those features of it that will be needed in our analysis. We conclude the current section by defining the Rényi mean and deriving an alternative expression for the Rényi information in terms of the Rényi divergence using the Rényi mean.

A. The Mean Measure

The weighted power means are generalizations of the weighted arithmetic mean. For any positive real number α and p.m.f. p on non-negative real numbers, the order α mean for the prior p is $\left(\sum_x p(x) x^\alpha\right)^{1/\alpha}$. For any prior p, the order α weighted mean is a nondecreasing and continuously differentiable function of α on $\mathbb{R}_+$. Hence we can calculate its limit as α approaches zero, or infinity, using the L'Hospital's rule [80, Thm. 5.13]:

$$\lim_{\alpha \downarrow 0} \left(\sum_x p(x) x^\alpha\right)^{1/\alpha} = \prod_x x^{p(x)}$$

$$\lim_{\alpha \uparrow \infty} \left(\sum_x p(x) x^\alpha\right)^{1/\alpha} = \max_{x:p(x)>0} x.$$

The order α mean of measures for the prior p is defined via the pointwise order α mean of their Radon-Nikodym derivatives for the prior p. In the following, we confine our discussion to the means of probability measure.

Definition 1. Let p be a p.m.f. on $\mathcal{P}(\mathcal{Y})$ and ν be a reference measure for w's with positive $p(w)$. Then *the order α mean of the Radon-Nikodym derivatives for the prior p is*[9]

$$\frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu} \triangleq \begin{cases} \prod_{w:p(w)>0} \left(\frac{\mathrm{d}w}{\mathrm{d}\nu}\right)^{p(w)} & \text{if } \alpha = 0 \\ \left(\sum_w p(w) \left(\frac{\mathrm{d}w}{\mathrm{d}\nu}\right)^\alpha\right)^{1/\alpha} & \text{if } \alpha \in \mathbb{R}_+ \quad \nu\text{-a.e.} \quad (7) \\ \max_{w:p(w)>0} \frac{\mathrm{d}w}{\mathrm{d}\nu} & \text{if } \alpha = \infty \end{cases}$$

The order α mean measure for the prior p is defined as

$$\mu_{\alpha,p}(\mathcal{E}) \triangleq \int_{\mathcal{E}} \frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu} \nu(\mathrm{d}y) \qquad \forall \mathcal{E} \in \mathcal{Y}. \quad (8)$$

In (7) and throughout this section sums of the form $\sum_w$ stands for sums of the form $\sum_{w:p(w)>0}$. In (7), w is a dummy variable used to express the elements of $\mathcal{P}(\mathcal{Y})$, i.e. probability measures on $(\mathcal{Y}, \mathcal{Y})$. The probability mass assigned to each w by p is denoted by $p(w)$. The reference measure ν is absent from the symbol for the mean measure because mean measure does not depend on the choice of the reference measure: Let $\widetilde{\mu}_{\alpha,p}$ be the mean measure obtained using a reference measure $\widetilde{\nu}$ instead of ν; then

$$\mu_{\alpha,p}(\mathcal{E}) = \widetilde{\mu}_{\alpha,p}(\mathcal{E}) \qquad \forall \alpha \in [0, \infty] \text{ and } \forall \mathcal{E} \in \mathcal{Y}.$$

This follows from a standard application of the Lebesgue decomposition theorem and the Radon-Nikodym theorem.

We are interested in the mean measure primarily as a tool to define and analyze the Rényi information. In [6, §26], Augustin introduced the mean measure and derived some of the observations we present in Lemmas 1-4, albeit for different parametrizations of the order. Augustin, however, did not define or analyze the Rényi information in [6]. Proofs of Lemmas 1-4 are presented in Appendix D.

Lemma 1. *Let p be a p.m.f. on $\mathcal{P}(\mathcal{Y})$.*

(a) $\mu_{\alpha,p} \sim \mu_{1,p}$ *and* $|\mathrm{supp}(p)|^{-\frac{1}{\alpha}} \leq \|\mu_{\alpha,p}\| \leq |\mathrm{supp}(p)|$ *for any $\alpha \in (0, \infty]$. Furthermore,* $\|\mu_{1,p}\| = 1$.

(b) $\mu_{0,p} \prec w$ *for any* $w \in \mathrm{supp}(p)$ *and* $\|\mu_{0,p}\| \leq 1$.

The main consequence of Lemma 1 is that $\mu_{\alpha,p} \prec \mu_{1,p}$ for all $\alpha \in [0, \infty]$. Hence, we can describe and analyze the mean measures via their Radon-Nikodym derivatives with respect to the order one mean measure. We build our analysis of the mean measure as a function of the order around this observation. First, we analyze $\frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\mu_{1,p}}$ as a function the order α in Lemma 2; then use the dominated convergence theorem to obtain the corresponding results for $\mu_{\alpha,p}$ in Lemma 3.

Definition 2. Let p be a p.m.f. on $\mathcal{P}(\mathcal{Y})$ and α be in $[0, \infty]$. Then the order α density for the prior p is

$$\pi_{\alpha,p} \triangleq \frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\mu_{1,p}}. \quad (9)$$

[9] For each w with positive $p(w)$, $\frac{\mathrm{d}w}{\mathrm{d}\nu}$ exists for all y except for a ν-measure zero set by the Radon-Nikodym theorem [30, 5.5.4]. Since there are only finite number of w's with positive $p(w)$, $\frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu}$ exists as a function of α from $[0, \infty]$ to $\mathbb{R}_{\geq 0}$ for all y except for a ν-measure zero set.

Note that for any p.m.f. p on $\mathcal{P}(\mathcal{Y})$, the order α density for the prior p is a $\mathcal{Y}$-measurable function from $\mathcal{Y}$ to $\mathbb{R}$ by the Radon-Nikodym theorem [30, 5.5.4].

The order α posteriors defined in the following provides us an alternative way to express $\pi_{\alpha,p}$ and its derivatives.

Definition 3. Let p be a p.m.f. on $\mathcal{P}(\mathcal{Y})$ and α be a positive real number. Then for each $y \in \mathcal{Y}$ *the order α posterior $p_{[\alpha]}$ is a p.m.f. on $\mathcal{P}(\mathcal{Y})$ given by*

$$p_{[\alpha]}(w|y) \triangleq \begin{cases} p(w)\left(\frac{\mathrm{d}w}{\mathrm{d}\mu_{\alpha,p}}\right)^{\alpha} & \text{if } p(w) > 0 \\ 0 & \text{else} \end{cases}. \quad (10)$$

The order α posterior p.m.f. $p_{[\alpha]}$ is a $\mathcal{Y}$-measurable function for each w. The order one posterior p.m.f. $p_{[1]}$ is also called the posterior p.m.f., in accordance with the usual terminology.

Lemma 2. *For any p.m.f. p on $\mathcal{P}(\mathcal{Y})$ the following statements hold for $\mu_{1,p}$-almost every y.*

(a) $\delta^{\frac{1-\alpha}{\alpha}} \leq \pi_{\alpha,p} \leq 1$ *for $\alpha \in (0,1]$ and $1 \leq \pi_{\alpha,p} \leq \delta^{\frac{1-\alpha}{\alpha}}$ for $\alpha \in [1,\infty)$ where $\delta = \min_{w:p(w)>0} p(w)$. Furthermore,*

$$\pi_{\alpha,p}(y) = \begin{cases} \prod_{w:p(w)>0}\left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{p(w)} & \alpha = 0 \\ \left(\sum_{w} p_{[1]}(w|y)^{\alpha} p(w)^{1-\alpha}\right)^{1/\alpha} & \alpha \in \mathbb{R}_+ \\ \max_{w:p(w)>0}\frac{p_{[1]}(w|y)}{p(w)} & \alpha = \infty \end{cases}.$$

$$p_{[\alpha]}(w|y) = \begin{cases} \frac{p_{[1]}(w|y)^{\alpha} p(w)^{1-\alpha}}{\pi_{\alpha,p}^{\alpha}} & \text{if } p(w) > 0 \\ 0 & \text{else} \end{cases}.$$

(b) $\pi_{\alpha,p}$ *is a smooth function of α on $\mathbb{R}_+$. Furthermore, the first two derivatives of $\pi_{\alpha,p}$ are given by*

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\pi_{\alpha,p} = \frac{\pi_{\alpha,p}}{\alpha^2}\sum_w p_{[\alpha]}(w|y)\ln\frac{p_{[\alpha]}(w|y)}{p(w)}.$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\pi_{\alpha,p} = \frac{1-\alpha}{\pi_{\alpha,p}}\left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\pi_{\alpha,p}\right)^2 - \frac{2}{\alpha}\frac{\mathrm{d}}{\mathrm{d}\alpha}\pi_{\alpha,p}$$
$$+ \frac{\pi_{\alpha,p}}{\alpha^3}\sum_w p_{[\alpha]}(w|y)\left(\ln\frac{p_{[\alpha]}(w|y)}{p(w)}\right)^2.$$

(c) $(\pi_{\alpha,p})^{\alpha}$ *is log-convex[10] in α on $\mathbb{R}_+$, i.e. for any $\beta \in (0,1)$ and $\alpha_0, \alpha_1 \in \mathbb{R}_+$*

$$(\pi_{\alpha_\beta,p})^{\alpha_\beta} \leq (\pi_{\alpha_1,p})^{\beta\alpha_1}(\pi_{\alpha_0,p})^{(1-\beta)\alpha_0}$$

where $\alpha_\beta = \beta\alpha_1 + (1-\beta)\alpha_0$. Furthermore, for $\alpha_1 \neq \alpha_0$ the inequality is strict iff there exist $w, \tilde{w} \in \mathrm{supp}(p)$ such that $\frac{p_{[1]}(w|y)}{p(w)} > \frac{p(\tilde{w}|y)}{p(\tilde{w})} > 0$.

(d) If there exists a w such that $p_{[1]}(w|y) > p(w)$, then $\pi_{\alpha,p}(y)$ is bounded, continuous, and monotone increasing in α on $[0,\infty]$, else $\pi_{\alpha,p}(y) = 1$ for all α in $[0,\infty]$.

Lemma 2 establishes the density $\pi_{\alpha,p}$ as a smooth function $\mu_{1,p}$-a.e. and provides expressions for its first two derivatives. These derivatives are $\mathcal{Y}$-measurable functions because $\pi_{\alpha,p}$

---

[10]Both of the following statements are equivalent to the log-convexity of $(\pi_{\alpha,p})^{\alpha}$ in α: "$\pi_{\frac{1}{1+\rho},p}$ is log-convex in ρ" and "For any $\beta \in [0,1]$ and $\alpha_0,\alpha_1 \in (0,\infty]$, $\pi_{\alpha_\beta,p} \leq (\pi_{\alpha_0,p})^{1-\beta}(\pi_{\alpha_1,p})^\beta$ where α_β is $\alpha_\beta = [(1-\beta)(\alpha_0)^{-1} + \beta(\alpha_1)^{-1}]^{-1}$."

---

and $p_{[\alpha]}$ are $\mathcal{Y}$-measurable. Then using their $\mu_{1,p}$-integrals we can define two mappings:

$$\mu'_{\alpha,p}(\mathcal{E}) \triangleq \int_{\mathcal{E}}(\pi'_{\alpha,p})\mu_{1,p}(\mathrm{d}y) \qquad \forall \mathcal{E} \in \mathcal{Y}, \quad (11)$$

$$\mu''_{\alpha,p}(\mathcal{E}) \triangleq \int_{\mathcal{E}}(\pi''_{\alpha,p})\mu_{1,p}(\mathrm{d}y) \qquad \forall \mathcal{E} \in \mathcal{Y} \quad (12)$$

where $\pi'_{\alpha,p}$ and $\pi''_{\alpha,p}$ are shorthands for $\frac{\mathrm{d}}{\mathrm{d}\alpha}\pi_{\alpha,p}$ and $\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\pi_{\alpha,p}$.

Note that we have not claimed that either of these mappings is defining a measure for each α. Lemma 3 given in the following establishes that fact and analyzes the mean measure $\mu_{\alpha,p}$ as a function of the order α.

Lemma 3. *For any p.m.f. p on $\mathcal{P}(\mathcal{Y})$.*

(a) $\mu_{\alpha,p}$ *is a continuous function of α from $[0,\infty]$ with its usual topology to $\mathcal{M}_0^+(\mathcal{Y})$ with the total variation topology.*

(b) $\mu'_{\alpha,p}$ *is a continuous function of α from $(0,\infty)$ with its usual topology to $\mathcal{M}_0^+(\mathcal{Y})$ with the total variation topology. Furthermore, $\frac{\mathrm{d}}{\mathrm{d}\alpha}\mu_{\alpha,p} = \mu'_{\alpha,p}$ in the sense that*

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\mu_{\alpha,p}(\mathcal{E})\Big|_{\alpha=\phi} = \mu'_{\phi,p}(\mathcal{E}) \quad \forall \mathcal{E} \in \mathcal{Y}, \quad \forall \phi \in (0,\infty).$$

(c) $\mu''_{\alpha,p}$ *is a continuous function of α from $(0,\infty)$ with its usual topology to $\mathcal{M}(\mathcal{Y})$ with the total variation topology. Furthermore, $\frac{\mathrm{d}}{\mathrm{d}\alpha}\mu'_{\alpha,p} = \mu''_{\alpha,p}$ in the sense that*

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\mu'_{\alpha,p}(\mathcal{E})\Big|_{\alpha=\phi} = \mu''_{\phi,p}(\mathcal{E}) \quad \forall \mathcal{E} \in \mathcal{Y}, \quad \forall \phi \in (0,\infty).$$

(d) $\|\mu_{\alpha,p}\|^{\alpha}$ *is a log-convex function of α on $(0,\infty)$ such that*

$$\lim_{\alpha\downarrow 0}\|\mu_{\alpha,p}\|^{\alpha} = \operatorname*{ess\,sup}_{\mu_{1,p}}\sum_{w:p_{[1]}(w|y)>0}p(w).$$

The log-convexity is strict everywhere on $(0,\infty)$, unless there exists a $\gamma \geq 1$ satisfying $\mu_{1,p}(\mathcal{A}(p,\gamma)) = 1$ for $\mathcal{A}(p,\gamma) = \{y : \frac{p_{[1]}(w|y)}{p(w)} = \gamma, \forall w : p_{[1]}(w|y) > 0\}$. If there exists such a γ, then $\|\mu_{\alpha,p}\| = \gamma^{\frac{\alpha-1}{\alpha}}$.

(e) $\|\mu_{\alpha,p}\|$ *is a continuous and nondecreasing function of α from $[0,\infty]$ to $[0,|\mathrm{supp}(p)|]$. If there exist $w, \widetilde{w}$ in $\mathrm{supp}(p)$ such that $w \neq \widetilde{w}$, then $\|\mu_{\alpha,p}\|$ is monotone increasing everywhere on $(0,\infty)$, else $\|\mu_{\alpha,p}\| = 1$ for all α in $[0,\infty]$.*

Lemma 3 described the properties of the mean measure as a function of the order for a fixed prior. Lemma 4, given in the following, describes the properties of the mean measure as a function of the prior for a fixed order.

Lemma 4. *Let $(\mathcal{Y}, \mathcal{Y})$ be a measurable space.*

(a) If $\alpha \in [0,1]$, then $\mu_{\alpha,p}$ and $\|\mu_{\alpha,p}\|$ are convex functions of p from $\mathcal{P}(\mathcal{P}(\mathcal{Y}))$ to $\mathcal{M}_0^+(\mathcal{Y})$ and $[0,1]$, respectively.

(b) If $\alpha \in [1,\infty]$, then $\mu_{\alpha,p}$ and $\|\mu_{\alpha,p}\|$ are concave functions of p from $\mathcal{P}(\mathcal{P}(\mathcal{Y}))$ to $\mathcal{M}^+(\mathcal{Y})$ and $[1,\infty)$, respectively.

(c) For any $p_1, p_2 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ such that $p_1 \neq p_2$, let $s_\wedge$, s_1 and s_2 be $s_\wedge \triangleq 2\frac{p_1 \wedge p_2}{2-\|p_1-p_2\|}$, $s_1 \triangleq 2\frac{p_1-p_1\wedge p_2}{\|p_1-p_2\|}$, and $s_2 \triangleq 2\frac{p_2-p_1\wedge p_2}{\|p_1-p_2\|}$. Then $s_\wedge, s_1, s_2 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ and

$$p_1 = (1 - \tfrac{\|p_1-p_2\|}{2})s_\wedge + \tfrac{\|p_1-p_2\|}{2}s_1,$$
$$p_2 = (1 - \tfrac{\|p_1-p_2\|}{2})s_\wedge + \tfrac{\|p_1-p_2\|}{2}s_2,$$
$$s_1 \perp s_2.$$

(d) *If $\alpha \in (0,1]$, then for any $p_1, p_2 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ we have*

$$\|\mu_{\alpha,p_1} - \mu_{\alpha,p_2}\| \leq \tfrac{1}{\alpha}\|p_1 - p_2\|.$$

Hence $\mu_{\alpha,p}$ is a Lipschitz continuous function of p for the total variation topology for $\alpha \in (0,1]$.

(e) *If $\alpha \in [1,\infty)$, then for any $p_1, p_2 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ we have*

$$\|\mu_{\alpha,p_1} - \mu_{\alpha,p_2}\| \leq (\tfrac{1}{2}\|p_1 - p_2\|)^{\frac{1}{\alpha}}\|\mu_{\alpha,s_1} - \mu_{\alpha,s_2}\|.$$

B. The Rényi Information

Definition 4. Let $\mathcal{W}$ be a subset of $\mathcal{P}(\mathcal{Y})$ and p be a p.m.f. on $\mathcal{W}$. Then *the order α Rényi information for the prior p is*

$$I_\alpha(p;\mathcal{W}) \triangleq \begin{cases} \underset{\mu_{1,p}}{\mathrm{ess\,inf}}\ \ln \frac{1}{\sum_w \mathbb{1}_{\{p_{[1]}(w|y)>0\}} p(w)} & \alpha=0 \\[2mm] \frac{\alpha}{\alpha-1}\ln\|\mu_{\alpha,p}\| & \alpha \in \mathbb{R}_+\backslash\{1\} \\[2mm] \mathbf{E}_{\mu_{1,p}}\left[\sum_w p_{[1]}(w|y)\ln\frac{p_{[1]}(w|y)}{p(w)}\right] & \alpha=1 \\[2mm] \ln\|\mu_{\infty,p}\| & \alpha=\infty \end{cases} \tag{13}$$

Sibson introduced this quantity[11] in [94] using works of Rényi [79] and Csiszár [21], [22]. Prior to [94] in [35], Gallager introduced $E_0(\rho,p)$, which is nothing but a scaled version of the Rényi information; see (3) and (4).

Note that $I_\alpha(p;\mathcal{W})$ has the same value for all $\mathcal{W}$'s for which p is in $\mathcal{P}(\mathcal{W})$. Hence, in principle, one can use $I_\alpha(p)$ rather than $I_\alpha(p;\mathcal{W})$ to denote the Rényi information. Although this unconventional symbol would be more coherent with the one we use for the mean measure, we refrain from using it for the fear of alienating readers who prefer the customary symbol. Another justification for using the conventional notation is the effect of the richness of $\mathcal{W}$ — as measured by $\sup_{p\in\mathcal{P}(\mathcal{W})} I_\alpha(p;\mathcal{W})$ — on the continuity of $I_\alpha(p;\mathcal{W})$ as a function of p, see Lemma 16-(e).

Properties of the Rényi information as a function of the order for fixed prior and as a function of the prior for fixed order are presented in Lemmas 5 and 6, respectively. Proofs of Lemmas 5 and 6 are presented in Appendix E.

Lemma 5. *For any subset $\mathcal{W}$ of $\mathcal{P}(\mathcal{Y})$ and p.m.f. p on $\mathcal{W}$, $I_\infty(p;\mathcal{W}) \leq \ln|\mathrm{supp}(p)|$ and $I_\alpha(p;\mathcal{W})$ is a non-negative continuously differentiable nondecreasing function of α on $\mathbb{R}_+$ such that*

$$I_0(p;\mathcal{W}) = \lim_{\alpha\downarrow 0} I_\alpha(p;\mathcal{W}), \tag{14}$$

$$I_\infty(p;\mathcal{W}) = \lim_{\alpha\uparrow\infty} I_\alpha(p;\mathcal{W}), \tag{15}$$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}I_\alpha(p;\mathcal{W}) = \begin{cases} \frac{\alpha}{\alpha-1}\frac{\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} - \frac{\ln\|\mu_{\alpha,p}\|}{(\alpha-1)^2} & \alpha\in\mathbb{R}_+\backslash\{1\} \\[2mm] \frac{\mu''_{1,p}(\mathcal{Y})+2\|\mu'_{1,p}\|-\|\mu'_{1,p}\|^2}{2} & \alpha=1 \end{cases}. \tag{16}$$

If $\mu_{1,p}(\mathcal{A}(p,\gamma)) = 1$ for some $\gamma \geq 1$, then $I_\alpha(p;\mathcal{W}) = \ln\gamma$ for all $\alpha \in [0,\infty]$, else $\frac{\mathrm{d}}{\mathrm{d}\alpha}I_\alpha(p;\mathcal{W}) > 0$ for all $\alpha \in \mathbb{R}_+$, where $\mathcal{A}(p,\gamma) \triangleq \{y: \frac{p_{[1]}(w|y)}{p(w)} = \gamma\ \ \forall w\ \text{with positive}\ \ p_{[1]}(w|y)\}$.

[11] Sibson called $\inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(p\circledast\mathcal{W}\|\,p\otimes q)$ "the information radius of order α" and proved that it equals to the expression given in Definition 4 in [94, Thm. 2.2]. Our presentation is different: Definition 4 does not refer to any infimum; equivalence of the alternative definition is established in Lemma 14. This is similar to the way things are, usually, handled for the mutual information: the mutual information is defined without any reference to an infimum [18, (2.28)], later it is shown to be equal to the infimum of certain Kullback-Leibler divergence [18, Lemma 10.8.1].

Using the definitions of $\mu'_{\alpha,p}$ and $\mu''_{\alpha,p}$, given in (11) and (12), together with Lemma 2-(b), we get the following two alternative expressions for the derivative of $I_\alpha(p;\mathcal{W})$ with respect to the order on $\mathbb{R}_+$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}I_\alpha(p;\mathcal{W}) = \begin{cases} \frac{1}{(\alpha-1)\alpha}\mathbf{E}_{\varpi_\alpha}\left[\ln\frac{p_{[\alpha]}(w|y)}{p(w)} - I_\alpha(p;\mathcal{W})\right] & \alpha\neq 1 \\[2mm] \frac{1}{2}\mathbf{E}_{\varpi_1}\left[\left(\ln\frac{p_{[1]}(w|y)}{p(w)} - I_1(p;\mathcal{W})\right)^2\right] & \alpha=1 \end{cases} \tag{17}$$

$$= \begin{cases} \frac{1}{(\alpha-1)^2}\mathbf{E}_{\varpi_\alpha}\left[\ln\frac{p_{[\alpha]}(w|y)\pi_{\alpha,p}}{p_{[1]}(w|y)\|\mu_{\alpha,p}\|}\right] & \alpha\neq 1 \\[2mm] \frac{1}{2}\mathbf{E}_{\varpi_1}\left[\left(\ln\frac{p_{[1]}(w|y)}{p(w)} - I_1(p;\mathcal{W})\right)^2\right] & \alpha=1 \end{cases} \tag{18}$$

where ϖ_α is a probability measure on $\mathcal{Y} \otimes 2^{\mathrm{supp}(p)}$ whose $\mathcal{Y}$ marginal is $\frac{\mu_{\alpha,p}}{\|\mu_{\alpha,p}\|}$ and whose conditional distribution is $p_{[\alpha]}$.

The continuity and the convexity properties of the Rényi information in the prior follow from the corresponding properties of the mean measure described in Lemma 4.

Lemma 6. *Let $\mathcal{W}$ be a subset of $\mathcal{P}(\mathcal{Y})$.*

(a) *If $\alpha \in [0,1)$, then $I_\alpha(p;\mathcal{W})$ is a non-negative quasi-concave function of p on $\mathcal{P}(\mathcal{W})$ that is continuous for the total variation topology on $\mathcal{P}(\mathcal{W})$.*

(b) *If $\alpha \in [1,\infty]$, then $I_\alpha(p;\mathcal{W})$ is a non-negative concave function of p on $\mathcal{P}(\mathcal{W})$.*

Gallager [35, p. 18] and Csiszár [23, Lemma 3.2] established the continuity of $I_\alpha(p;\mathcal{W})$ in p on $\mathcal{P}(\mathcal{W})$, for finite $\mathcal{W}$'s. For arbitrary $\mathcal{W}$'s, however, $I_\alpha(p;\mathcal{W})$ is continuous only for orders in $(0,1)$; for orders in $[1,\infty]$, $I_\alpha(p;\mathcal{W})$ is continuous in p on $\mathcal{P}(\mathcal{W})$ iff $\sup_{p\in\mathcal{P}(\mathcal{W})} I_\alpha(p;\mathcal{W})$ is finite, see Lemma 16-(d). The finiteness of $\sup_{p\in\mathcal{P}(\mathcal{W})} I_\alpha(p;\mathcal{W})$ also implies the uniform equicontinuity of the Rényi information, see Lemma 16-(e,f). The discontinuity of various Shannon information measures for countably infinite output sets have previously been pointed out by Ho and Yeung in [53].

C. The Rényi Divergence

Definition 5. Let w and q be two non-zero finite measures on the measurable space $(\mathcal{Y}, \mathcal{Y})$; then *the order α Rényi divergence between w and q is*

$$D_\alpha(w\|\,q) \triangleq \begin{cases} -\ln q\left(\frac{\mathrm{d}w}{\mathrm{d}\nu} > 0\right) & \alpha=0 \\[2mm] \frac{1}{\alpha-1}\ln\int\left(\frac{\mathrm{d}w}{\mathrm{d}\nu}\right)^\alpha\left(\frac{\mathrm{d}q}{\mathrm{d}\nu}\right)^{1-\alpha}\nu(\mathrm{d}y) & \alpha\in\mathbb{R}_+\backslash\{1\} \\[2mm] \int\frac{\mathrm{d}w}{\mathrm{d}\nu}\left(\ln\frac{\mathrm{d}w}{\mathrm{d}\nu} - \ln\frac{\mathrm{d}q}{\mathrm{d}\nu}\right)\nu(\mathrm{d}y) & \alpha=1 \\[2mm] \ln\,\mathrm{ess\,sup}_\nu\ \frac{\mathrm{d}w}{\mathrm{d}\nu}/\frac{\mathrm{d}q}{\mathrm{d}\nu} & \alpha=\infty \end{cases} \tag{19}$$

where ν is any measure satisfying $w\prec\nu$ and $q\prec\nu$.

The Rényi divergence is usually defined for probability measures; the inclusion of finite measures allows us to express certain observations, such as Lemma 8 given in the following, more succinctly.[12] Nonetheless, the propositions derived for the usual definition with probability measures suffice for our purposes most of the time. We appropriate all the propositions we need for our analysis, except Lemma 8, from the recent

[12] It is also convenient while studying the concept of the Rényi-Gallager information and capacity, see [69] and [70].

paper of van Erven and Harremoës [31]. The equivalence of Definition 5 and the one used by van Erven and Harremoës in [31] for probability measures follows from [31, Thm. 4-6].

Lemma 7 ([31, Thm. 3, Thm. 7]). *For all $w, q \in \mathcal{P}(\mathcal{Y})$, $D_\alpha(w \| q)$ is a nondecreasing and lower semicontinuous function of α on $[0,\infty]$ that is continuous on $[0, (1 \vee \chi_{w,q})]$ where $\chi_{w,q} \triangleq \sup\{\alpha : D_\alpha(w \| q) < \infty\}$.*

Lemma 8 is evident from the definition of Rényi divergence.

Lemma 8. *Let w, q, v be non-zero finite measures on $(\mathcal{Y}, \mathcal{Y})$ and α be an order in $[0, \infty]$.*
- *If $v \leq q$, then $D_\alpha(w \| q) \leq D_\alpha(w \| v)$.*
- *If $q = \gamma v$ for some $\gamma \in \mathbb{R}_+$ and either w is a probability measure or $\alpha \neq 1$, then $D_\alpha(w \| q) = D_\alpha(w \| v) - \ln \gamma$.*

Let w and q be two probability measures on the measurable space $(\mathcal{Y}, \mathcal{Y})$ and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{Y}$. Then the identities $w_{|\mathcal{G}}(\mathcal{E}) = w(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{G}$ and $q_{|\mathcal{G}}(\mathcal{E}) = q(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{G}$ uniquely define probability measures $w_{|\mathcal{G}}$ and $q_{|\mathcal{G}}$ on $(\mathcal{Y}, \mathcal{G})$. In the following, we denote $D_\alpha(w_{|\mathcal{G}} \| q_{|\mathcal{G}})$ by $D_\alpha^{\mathcal{G}}(w \| q)$.

Lemma 9 ([31, Thm. 9]). *For any $\alpha \in [0, \infty]$, probability measures w and q on $(\mathcal{Y}, \mathcal{Y})$ and sub-σ-algebra $\mathcal{G} \subset \mathcal{Y}$*

$$D_\alpha(w \| q) \geq D_\alpha^{\mathcal{G}}(w \| q).$$

Lemma 10 ([31, Thm. 3, Thm. 31]). *For any $\alpha \in [0, \infty]$, probability measures w and q on $(\mathcal{Y}, \mathcal{Y})$*

$$D_\alpha(w \| q) \geq \tfrac{1 \wedge \alpha}{2} \| w - q \|^2. \tag{20}$$

For orders in $(0, 1]$, the bound given in (20) is called the Pinsker's inequality; it has been proved by Csiszár [21] for $\alpha = 1$ case and by Augustin[13] [6] and Gilardoni [42] for $\alpha \in (0, 1)$ case. Furthermore the constant $\alpha/2$ is the best possible: for any $\gamma < \alpha/2$ there are probability measures w and q such that $\gamma \| w - q \|^2 > D_\alpha(w \| q)$. Determination of best lower bound on the Rényi divergence in terms of the total variation is an interesting and important problem but it is beyond the scope of the current manuscript.

Remark 1. Kullback [59], [60] bounded $D_1(w \| q)$ from below by $\| w - q \|^2/2 + \| w - q \|^4/36$. Hence, Pinsker's inequality is tight only for $\| w - q \| \approx 0$. Vajda [103] established $D_1(w \| q) \geq \ln(\frac{2+\| w-q \|}{2-\| w-q \|}) - \frac{2\| w-q \|}{2+\| w-q \|}$. Vajda's inequality is tight not only for $\| w - q \| \approx 0$ but also for $\| w - q \| \approx 2$. Fedotov, Harremoës, and Topsøe [32] determined the tight lower bound on $D_1(w \| q)$ in terms of $\| w - q \|$ in a parametric form. Gilardoni [40], [41] proved an equivalent result for f-divergences for twice differentiable f's. Gilardoni's result implies tight bounds for Rényi divergences, which are recently derived in a more explicit form by Sason [82, Prop. 1]. The core observation in the derivation of tight Vajda's inequalities is the sufficiency of the probability measures on binary alphabets. Guntuboyina, Saha, and Schiebinger [43] have recently generalized this observation considerably and explained how one can determine tight bounds on an f-divergence when its

---

[13] $\frac{\| w-q \|^2}{2} \leq \frac{1 - e^{(\alpha-1)D_\alpha(w \| q)}}{\alpha(1-\alpha)}$ for all $w, q \in \mathcal{P}(\mathcal{Y})$ and $\alpha \in [-1, 2]$ by [6, Lemma 26.5a]. This implies (20) for $\alpha \in (0, 1)$ via $e^{-x} \geq 1 - x$.

arguments are constrained in terms of other f-divergences. Recall that the total variation distance is the f-divergence for $f(x) = |x - 1|$.

Lemma 11 ([31, Thm. 12]). *For any order $\alpha \in [0, \infty]$, the order α Rényi divergence is convex in its second argument for probability measures, i.e. for all $w, q_0, q_1 \in \mathcal{P}(\mathcal{Y})$ and $\beta \in (0, 1)$ we have*

$$D_\alpha(w \| q_\beta) \leq \beta D_\alpha(w \| q_1) + (1 - \beta) D_\alpha(w \| q_0)$$

where $q_\beta = \beta q_1 + (1 - \beta) q_0$.

Lemma 12 ([31, Thm. 13]). *For any order $\alpha \in [0, \infty]$, the order α Rényi divergence is jointly quasi-convex in its arguments for probability measures, i.e. for all w_0, w_1, q_0, q_1 in $\mathcal{P}(\mathcal{Y})$ and $\beta \in (0, 1)$ we have*

$$D_\alpha(w_\beta \| q_\beta) \leq D_\alpha(w_1 \| q_1) \vee D_\alpha(w_0 \| q_0)$$

where $w_\beta = \beta w_1 + (1 - \beta) w_0$ and $q_\beta = \beta q_1 + (1 - \beta) q_0$.

Lemma 13 ([31, Thm 15]). *For any order $\alpha \in (0, \infty]$, $D_\alpha(w \| q)$ is a lower semicontinuous function of the pair of probability measures (w, q) in the topology of setwise convergence.*

The preceding lemmas discuss only the aspects of the Rényi divergence that are useful for our discussion. A more comprehensive discussion can be found in [31].

D. The Rényi Mean

We have defined the Rényi information using a closed form expression. However, the original definition of the Rényi information by Sibson is in terms of an optimization of the Rényi divergence over a set of probability measures. These two definitions are equivalent, as it has already been shown by Sibson [94, Thm. 2.2]. In the following, we establish this equivalence and briefly discuss an alternative definition of the Rényi information related to the aforementioned characterization in terms of the Rényi divergence.

Definition 6. *Let p be a p.m.f. on $\mathcal{P}(\mathcal{Y})$; then the order α Rényi mean for prior p is*

$$q_{\alpha,p} \triangleq \begin{cases} \dfrac{e^{-D_1\left(p_{[0]} \| p_{[1]}\right)} \mathbb{1}_{\{\vartheta_p(y) = \bar{\vartheta}_p\}} \mu_{1,p}}{\int e^{-D_1\left(p_{[0]} \| p_{[1]}\right)} \mathbb{1}_{\{\vartheta_p(y) = \bar{\vartheta}_p\}} \mu_{1,p}(\mathrm{d}y)} & \alpha = 0 \\ \dfrac{\mu_{\alpha,p}}{\| \mu_{\alpha,p} \|} & \alpha \in (0, \infty] \end{cases} \tag{21}$$

where $\vartheta_p(y) \triangleq \sum_w p(w) \mathbb{1}_{\{p_{[1]}(w|y) > 0\}}$, $\bar{\vartheta}_p \triangleq \operatorname{ess\,sup}_{\mu_{1,p}} \vartheta_p$, and $p_{[0]}(w|y) \triangleq \dfrac{p(w) \mathbb{1}_{\{p_{[1]}(w|y) > 0\}}}{\sum_u p(u) \mathbb{1}_{\{p(u|y) > 0\}}}$.

Then the following identity can be confirmed by substitution using (19): For any α in $(0, \infty]$, p in $\mathcal{P}(\mathcal{W})$, and q in $\mathcal{P}(\mathcal{Y})$,

$$D_\alpha(p \circledast \mathcal{W} \| p \otimes q) = D_\alpha(p \circledast \mathcal{W} \| p \otimes q_{\alpha,p}) + D_\alpha(q_{\alpha,p} \| q). \tag{22}$$

This identity was first pointed out by Sibson in [94, p. 153], then by others [25, (12)] [50, (43)] [75, (38)] [90, Lemma 3] [105, (52)]. For $\alpha = 1$ case, it had been used by Topsøe in [101], even before Sibson [94], and in [102].

On the other hand, one can also confirm by substitution that $I_\alpha(p; \mathcal{W}) = D_\alpha(p \circledast \mathcal{W} \| p \otimes q_{\alpha,p})$ for all positive values of α.

These two observations lead to the alternative characterization of the order α Rényi information in terms of the order α Rényi divergence presented in the following lemma, which is valid for all non-negative orders.

Lemma 14. *Let $\mathcal{W}$ be a subset of $\mathcal{P}(\mathcal{Y})$, p be a p.m.f. on $\mathcal{W}$, and α be an order in $[0, \infty]$; then*

$$I_\alpha(p; \mathcal{W}) = D_\alpha(p \circledast \mathcal{W} \| p \otimes q_{\alpha,p}) \tag{23}$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q) \tag{24}$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(\mu_{\alpha,p} \| q) \qquad \alpha \in (0, \infty] \setminus \{1\} \tag{25}$$

where $p \circledast \mathcal{W}$ is the probability measure on $2^{\mathrm{supp}(p)} \otimes \mathcal{Y}$ whose marginal distribution on $\mathrm{supp}(p)$ is p and whose conditional distribution is w.

Proof of Lemma 14 is presented in Appendix F. For any positive order α and prior p, the only probability measure q satisfying $D_\alpha(p \circledast \mathcal{W} \| p \otimes q) = I_\alpha(p; \mathcal{W})$ is $q_{\alpha,p}$ as a result of (22) and Lemmas 10, 14. In other words, the order α Rényi mean for prior p is the unique minimizer for the infimum given in (24) for positive orders α. For $\alpha = 0$, the order zero Rényi mean is still a minimizer by Lemma 14 but it is not necessarily the unique minimizer. Any probability measure q that is absolutely continuous in the $q_{0,p}$ satisfies $D_0(p \circledast \mathcal{W} \| p \otimes q) = I_0(p; \mathcal{W})$.

The definition of Rényi information we have adopted is not the only definition of Rényi information. The following definition is first proposed by Augustin in [6, §34] and later popularized by Csiszár [25]

$$I_\alpha^c(p; \mathcal{W}) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} \sum_w p(w) D_\alpha(w \| q). \tag{26}$$

Unlike the definition we have adopted, the one given in (26) does not have an equivalent closed form expression. But for any finite positive order α, the infimum in (26) has a unique minimizer, which is a fixed point of an operator defined using α and p, [70]. These properties were first proved by Augustin for orders between zero and one in [6]. Thus we have called the quantity defined in (26), the order α Augustin information in [69]. We present a more detailed discussion of the properties of the Augustin information and its relation to the Rényi information in [70].

Arimoto proposed a third definition for the Rényi information in [4]. Recently, Verdú has provided a discussion of the Rényi entropy and these three definitions of the Rényi information in [105].

III. THE RÉNYI CAPACITY

Definition 7. Let α be an order in $[0, \infty]$ and $\mathcal{W}$ be a subset of $\mathcal{P}(\mathcal{Y})$; then *the order α Rényi capacity of $\mathcal{W}$ is*

$$C_{\alpha,\mathcal{W}} \triangleq \sup_{p \in \mathcal{P}(\mathcal{W})} I_\alpha(p; \mathcal{W}). \tag{27}$$

Unlike the Rényi information, the Rényi capacity is not a quantity that is introduced or discussed by Sibson in [94]. In the spirit of his earlier work on f-divergences [23], Csiszár introduces it in [25]. Prior to either work, Shannon, Gallager, and Berlekamp had introduced a 'capacity', i.e. $E_0(\rho, \mathcal{W})$, using $E_0(\rho, p)$ in [88]. $E_0(\rho, \mathcal{W})$ is a scaled version of the

Rényi capacity; in particular $E_0(\rho, \mathcal{W}) = \rho C_{\frac{1}{1+\rho}, \mathcal{W}}$ for all non-zero ρ greater than minus one by (3).

Using the alternative characterization of the Rényi information given in (24), we get the following expression for the order α Rényi capacity for all α in $[0, \infty]$

$$C_{\alpha,\mathcal{W}} = \sup_{p \in \mathcal{P}(\mathcal{W})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q). \tag{28}$$

For finite orders the Rényi capacity does not have a closed form expression. The supremum given in the definition of the Rényi capacity need not to be finite, see Examples 6 and 7. Even when the supremum is finite it might not be achieved by any prior, i.e. there are $\mathcal{W}$'s for which $I_\alpha(p; \mathcal{W}) < C_{\alpha,\mathcal{W}}$ for all $p \in \mathcal{P}(\mathcal{W})$, see Examples 2 and 4. When the supremum is achieved, the optimal prior might not be unique, i.e. there are $\mathcal{W}$'s for which $I_\alpha(p_1; \mathcal{W}) = I_\alpha(p_2; \mathcal{W}) = C_{\alpha,\mathcal{W}}$ for $p_1 \neq p_2$ both of which are in $\mathcal{P}(\mathcal{W})$, see Example 3. These subtleties, however, do not constitute a serious impediment for analyzing the Rényi capacity.

In §III-A, we analyze the Rényi capacity as a function of the order. In §III-B, we determine necessary and sufficient conditions for the finiteness of the Rényi capacity and investigate the implications of the finiteness of the Rényi capacity on the continuity of the mean measure and the Rényi information.

A. The Rényi Capacity as a Function of the Order

We are interested in characterizing the behavior of the Rényi capacity as a function of the order because the operational significance of the Rényi capacity —at least for the channel coding problem and the sphere packing bound— is not through its value at a specific order but through its behavior as a function of the order. Parts (a,c,d,e,f) of Lemma 15 characterize the behavior of the Rényi capacity for an arbitrary $\mathcal{W}$ as a function of the order. In our analysis relying on the Rényi capacity some of our results might be valid only for countable or finite $\mathcal{W}$'s rather than arbitrary $\mathcal{W}$'s. Parts (b,g) of Lemma 15 are useful in such situations.[14] See the proof of [71, Corollary 2] for such a situation for the Augustin capacity.

Lemma 15. *Let $\mathcal{W}$ be a subset of $\mathcal{P}(\mathcal{Y})$.*
(a) $C_{\alpha,\mathcal{W}}$ is nondecreasing and lower semicontinuous in α on $[0, \infty]$.
(b) There exists a countable subset $\mathcal{W}'$ of $\mathcal{W}$ satisfying $C_{\alpha,\mathcal{W}'} = C_{\alpha,\mathcal{W}}$ for all $\alpha \in [0, \infty]$.
(c) $\frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}}$ is nonincreasing and continuous in α on $(0, 1)$ and $C_{\alpha,\mathcal{W}}$ is continuous in α on $(0, 1]$.
(d) $(\alpha - 1) C_{\alpha,\mathcal{W}}$ is convex in α on $(1, \infty)$.
(e) If $C_{\eta,\mathcal{W}} < \infty$ for an $\eta \in (0, 1)$, then $C_{\alpha,\mathcal{W}}$ is finite for all $\alpha \in [0, 1)$.
(f) If $C_{\eta,\mathcal{W}} < \infty$ for an $\eta \in (0, \infty]$, then $C_{\alpha,\mathcal{W}}$ is nondecreasing and continuous[15] in α on $(0, \eta)$.
(g) If $C_{\eta,\mathcal{W}} < \infty$ for an $\eta \in (0, \infty]$, then $\forall \epsilon > 0, \exists$ a finite subset $\mathcal{W}'$ of $\mathcal{W}$ such that $C_{\alpha,\mathcal{W}'} > C_{\alpha,\mathcal{W}} - \epsilon$ for all $\alpha \in [\epsilon, \eta]$.

[14] As pointwise statements, i.e. as statements for a given order, Lemma 15-(b,g) follow trivially from the definition of the Rényi capacity. They are non-trivial only because their assertions hold for all orders for the same $\mathcal{W}'$.

[15] We are unable to establish the continuity of $C_{\alpha,\mathcal{W}}$ at $\alpha = 0$ for arbitrary $\mathcal{W}$. For finite $\mathcal{W}$, Sion's minimax theorem implies the continuity of $C_{\alpha,\mathcal{W}}$ at $\alpha = 0$, see Lemma 16-(g).

The Rényi information $I_\alpha(p;\mathcal{W})$ is continuous in α for any p in $\mathcal{P}(\mathcal{W})$ by Lemma 5, however the Rényi capacity $C_{\alpha,\mathcal{W}}$ is not necessarily continuous in α. Yet, if the Rényi capacity $C_{\alpha,\mathcal{W}}$ is not continuous in α on $(0,\infty]$, then it has a very specific shape as a result of Lemma 15: there exists a $\phi \in [1,\infty)$ such that $C_{\alpha,\mathcal{W}}$ is bounded and continuous on $(0,\phi]$ and infinite on $(\phi,\infty]$. In order to see why, first note that if $C_{1/2,\mathcal{W}} = \infty$, then $C_{\alpha,\mathcal{W}} = \infty$ for all α in $(0,\infty]$ by Lemma 15-(a,e) and $C_{\alpha,\mathcal{W}}$ is continuous on $(0,\infty]$. On the other hand, if $C_{\infty,\mathcal{W}} < \infty$, then $C_{\alpha,\mathcal{W}}$ is continuous on $(0,\infty]$ by Lemma 15-(f). Hence, $C_{\alpha,\mathcal{W}}$ can fail to be continuous on $(0,\infty]$ only when $C_{1/2,\mathcal{W}} < \infty$ and $C_{\infty,\mathcal{W}} = \infty$. Let $\chi_{\mathcal{W}}$ be the set of all orders α for which $C_{\alpha,\mathcal{W}}$ is finite, i.e.

$$\chi_{\mathcal{W}} \triangleq \{\alpha \in \mathbb{R}_+ : C_{\alpha,\mathcal{W}} < \infty\}.$$

$\chi_{\mathcal{W}}$ is either of the form $(0,\phi)$ for a $\phi \in [1,\infty]$ or of the form $(0,\phi]$ for a $\phi \in [1,\infty)$ because $C_{\alpha,\mathcal{W}}$ is nondecreasing by Lemma 15-(a) and finite on $(0,1)$ by Lemma 15-(e). If $\chi_{\mathcal{W}} = (0,\phi)$ for some $\phi \in [1,\infty]$, then $C_{\alpha,\mathcal{W}}$ is continuous on $(0,\phi)$ by Lemma 15-(a,f), $C_{\alpha,\mathcal{W}}$ is infinite on $[\phi,\infty]$ by the hypothesis, and hence $C_{\alpha,\mathcal{W}}$ is continuous on $(0,\infty]$ by the pasting lemma [68, Thm. 18.3]. —Example 6 provides a $\mathcal{W}$ for each $\phi \in (1,\infty)$ such that $\chi_{\mathcal{W}} = (0,\phi)$.— Thus unless $\chi_{\mathcal{W}} = (0,\phi]$ for some $\phi \in [1,\infty)$, $C_{\alpha,\mathcal{W}}$ is continuous on $(0,\infty]$. If $\chi_{\mathcal{W}} = (0,\phi]$, then $C_{\alpha,\mathcal{W}}$ is bounded and continuous on $(0,\phi]$ and infinite on $(\phi,\infty]$. Hence the Rényi capacity has a unique discontinuity on $(0,\infty]$, which is at ϕ. —Example 7 provides a $\mathcal{W}$ for each $\phi \in [1,\infty)$ such that $C_{\alpha,\mathcal{W}}$ has its unique discontinuity at ϕ.—

Proof of Lemma 15.

(a) The pointwise supremum of a family of nondecreasing (lower semicontinuous) functions is nondecreasing (lower semicontinuous). Then $C_{\alpha,\mathcal{W}}$ is nondecreasing and lower semicontinuous in α on $[0,\infty]$ because $C_{\alpha,\mathcal{W}}$ is the pointwise supremum of the family $\{I_\alpha(p;\mathcal{W})\}_{p\in\mathcal{P}(\mathcal{W})}$ and $I_\alpha(p;\mathcal{W})$ is nondecreasing and continuous in α for each $p \in \mathcal{P}(\mathcal{W})$ by Lemma 5.

(b) The Rényi capacity is a nondecreasing and lower semi-continuous function of the order by part (a). Then

$$C_{\eta,\mathcal{W}} = \sup_{\alpha\in(0,\eta)\cap\mathbb{Q}} C_{\alpha,\mathcal{W}} \qquad \forall \eta \in (0,\infty].$$

Consequently, $C_{\alpha,\mathcal{W}'} = C_{\alpha,\mathcal{W}}$ for all α in $[0,\infty]$ if $C_{\alpha,\mathcal{W}'} = C_{\alpha,\mathcal{W}}$ for all $\alpha \in \mathbb{Q}_{\geq 0}$. Choose a sequence of p.m.f.'s $\{p^{(\alpha,i)}\}_{i\in\mathbb{Z}_+}$ satisfying $I_\alpha(p^{(\alpha,i)};\mathcal{W}) \uparrow C_{\alpha,\mathcal{W}}$ for each $\alpha \in \mathbb{Q}_{\geq 0}$. Let $\mathcal{W}'$ be $\cup_{\alpha\in\mathbb{Q}_{\geq 0}} \cup_{i\in\mathbb{Z}_+} \mathrm{supp}(p^{(\alpha,i)})$. Then $C_{\alpha,\mathcal{W}'} = C_{\alpha,\mathcal{W}}$ for all $\alpha \in \mathbb{Q}_{\geq 0}$; hence for all α in $[0,\infty]$. $\mathcal{W}'$ is countable because countable union of countable sets is countable.

(c) The definitions of $I_\alpha(p;\mathcal{W})$ and $C_{\alpha,\mathcal{W}}$ imply

$$\frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} = \sup_{p\in\mathcal{P}(\mathcal{W})} \ln \frac{1}{\|\mu_{\alpha,p}\|} \qquad \forall \alpha \in (0,1).$$

Furthermore, $\|\mu_{\alpha,p}\|$ is nondecreasing and continuous in α, by Lemma 3-(e). Then $\frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}}$ is nonincreasing and lower semicontinuous in α on $(0,1)$ because the pointwise supremum of a family of nonincreasing (lower semicontinuous) functions is nonincreasing (lower semicontinuous). Thus $\frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}}$ and $C_{\alpha,\mathcal{W}}$ are both

continuous from the right on $(0,1)$. On the other hand $C_{\alpha,\mathcal{W}}$ and $\frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}}$ are both continuous from the left on $(0,1)$ because $C_{\alpha,\mathcal{W}}$ is nondecreasing and lower semicontinuous on $(0,1)$ by part (a). Consequently, $C_{\alpha,\mathcal{W}}$ and $\frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}}$ are both continuous on $(0,1)$. Furthermore, $C_{\alpha,\mathcal{W}}$ is continuous on $(0,1]$ because $C_{\alpha,\mathcal{W}}$ is nondecreasing and lower semicontinuous by part (a).

(d) $\|\mu_{\alpha,p}\|^\alpha$ is log-convex in α by Lemma 3-(d). On the other hand, the definitions of $I_\alpha(p;\mathcal{W})$ and $C_{\alpha,\mathcal{W}}$ imply

$$(\alpha-1)C_{\alpha,\mathcal{W}} = \sup_{p\in\mathcal{P}(\mathcal{W})} \alpha \ln \|\mu_{\alpha,p}\| \quad \forall \alpha \in (1,\infty).$$

Then $(\alpha-1)C_{\alpha,\mathcal{W}}$ is convex in α because the pointwise supremum of a family of convex functions is convex.

(e) If $C_{\eta,\mathcal{W}}$ is finite, then so is $C_{\alpha,\mathcal{W}}$ for all α in $[0,\eta]$ because $C_{\alpha,\mathcal{W}}$ is nondecreasing in α by part (a). Furthermore, if $C_{\eta,\mathcal{W}}$ is finite, then so is $C_{\alpha,\mathcal{W}}$ for all α in $[\eta,1)$ because $\frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}}$ is nonincreasing in α on $(0,1)$ by part (c).

(f) $C_{\alpha,\mathcal{W}}$ is continuous in α on $(0,1]$ by part (c). Thus we only need to prove the claim for the case when $\eta > 1$ on $[1,\eta]$. We prove the continuity of $C_{\alpha,\mathcal{W}}$ in α first on $(1,\eta]$, and then from the right at $\alpha = 1$. If $C_{\eta,\mathcal{W}}$ is finite for an η in $(1,\infty)$, then $(\alpha-1)C_{\alpha,\mathcal{W}}$ is finite and convex in α on $[1,\eta]$ by parts (a) and (d). Then the continuity of $(\alpha-1)C_{\alpha,\mathcal{W}}$, and hence the continuity of $C_{\alpha,\mathcal{W}}$, in α on $(1,\eta)$ follows from [30, Thm. 6.3.3]. On the other hand $C_{\alpha,\mathcal{W}}$ is continuous from the left because $C_{\alpha,\mathcal{W}}$ is nondecreasing and lower semicontinuous in α by part (a). Hence, $C_{\alpha,\mathcal{W}}$ is continuous in α on $(1,\eta]$.

If $C_{\infty,\mathcal{W}}$ is finite, then $C_{\eta,\mathcal{W}}$ is finite for all $\eta\in\mathbb{R}_+$ by part (a) and $C_{\alpha,\mathcal{W}}$ is continuous in α on $\mathbb{R}_+$ because the continuity of a function on a collection of open set implies its continuity on their union, [68, Thm. 18.2]. This implies the continuity on $(0,\infty]$ because $C_{\alpha,\mathcal{W}}$ is nondecreasing and lower semicontinuous in α by part (a). To prove the continuity of $C_{\alpha,\mathcal{W}}$ from the right at one, we first prove that $\{I_\alpha(p;\mathcal{W})\}_{p\in\mathcal{P}(\mathcal{W})}$ is equicontinuous from the right at $\alpha = 1$. The definitions of $\mu'_{\alpha,p}$ and $I_\alpha(p;\mathcal{W})$ given in (11) and (13) and Lemma 2-(b) imply

$$I_\alpha(p;\mathcal{W}) - I_1(p;\mathcal{W}) = \frac{\alpha \ln \|\mu_{\alpha,p}\| - (\alpha-1)\|\mu'_{1,p}\|}{\alpha-1}$$

for all α in $(1,\eta]$ and p in $\mathcal{P}(\mathcal{W})$. The expression in the numerator is differentiable in α because $\|\mu_{\alpha,p}\|$ is differentiable by Lemma 3-(b). Furthermore, $\frac{d}{d\alpha}\|\mu_{\alpha,p}\| = \|\mu'_{\alpha,p}\|$ by Lemma 3-(b) and the numerator is zero at $\alpha = 1$. Then by the mean value theorem [80, 5.10], there exists a $\phi \in (1,\alpha)$ such that

$$I_\alpha(p;\mathcal{W}) - I_1(p;\mathcal{W}) = \ln \|\mu_{\phi,p}\| + \phi\frac{\|\mu'_{\phi,p}\|}{\|\mu_{\phi,p}\|} - \|\mu'_{1,p}\|.$$

The expression on the right hand side is differentiable in ϕ because $\frac{d}{d\phi}\|\mu_{\phi,p}\| = \|\mu'_{\phi,p}\|$ and $\frac{d}{d\phi}\|\mu'_{\phi,p}\| = \mu''_{\phi,p}(\mathcal{Y})$ by Lemma 3-(b,c). On the other hand, $\|\mu_{\phi,p}\| > 0$ for $\phi \in \mathbb{R}_+$ and $\|\mu_{1,p}\| = 1$ by Lemma 1-(a). Then the expression on the right hand side is zero at $\phi = 1$. Hence,

using the mean value theorem [80, 5.10] once again we can conclude that there exists a $\beta \in (1, \phi)$ such that

$$\frac{I_\alpha(p;\mathcal{W})-I_1(p;\mathcal{W})}{\phi-1} = 2\frac{\|\mu'_{\beta,p}\|}{\|\mu_{\beta,p}\|} + \beta\frac{\mu''_{\beta,p}(y)}{\|\mu_{\beta,p}\|} - \beta\frac{\|\mu'_{\beta,p}\|^2}{\|\mu_{\beta,p}\|^2}. \quad (29)$$

On the other hand, using the definition of $\mu''_{\alpha,p}$ given in (12) together with Lemma 2-(b) and $\beta > 1$ we get

$$\frac{\mu''_{\beta,p}(y)}{\|\mu_{\beta,p}\|} \le \mathbf{E}_{q_{\beta,p}}\left[\sum_w \frac{p_{[\beta]}(w|y)}{\beta^3}\ln^2\frac{p_{[\beta]}(w|y)}{p(w)}\right] - \frac{2\|\mu'_{\beta,p}\|}{\beta\|\mu_{\beta,p}\|}.$$

Then using Lemma 2-(a) and (29) we get

$$\frac{I_\alpha(p;\mathcal{W})-I_1(p;\mathcal{W})}{\phi-1} \le \mathbf{E}_{q_{\beta,p}}\left[\sum_w \frac{p_{[\beta]}(w|y)}{\beta^2}\ln^2\frac{p_{[\beta]}(w|y)}{p(w)}\right]$$
$$= \mathbf{E}_{q_{\beta,p}}\left[\sum_w p(w)\left[\frac{p_{[1]}(w|y)}{p(w)\pi_{\beta,p}}\right]^\beta\ln^2\frac{p_{[1]}(w|y)}{p(w)\pi_{\beta,p}}\right].$$

Recall that $x^\beta\ln^2 x \le (\frac{2}{\beta e})^2$ for all $x \in [0,1]$ and $\beta > 0$ and $\ln^2 x \le (\frac{2}{\epsilon e})^2 x^\epsilon$ for all $x \ge 1$ and $\epsilon > 0$. Thus

$$\frac{I_\alpha(p;\mathcal{W})-I_1(p;\mathcal{W})}{\phi-1} \le \mathbf{E}_{q_{\beta,p}}\left[(\frac{2}{\beta e})^2 + (\frac{2}{\epsilon e})^2(\frac{\pi_{\beta+\epsilon,p}}{\pi_{\beta,p}})^{\beta+\epsilon}\right].$$

Since $(\pi_{\alpha,p})^\alpha$ is log-convex in α by Lemma 2-(c),

$$(\pi_{\beta+\epsilon,p})^{\beta+\epsilon} \le (\pi_{\beta,p})^{\beta+\epsilon-1}\pi_{\frac{\beta}{1-\epsilon},p} \quad \forall\epsilon \in (0,1), \beta > 1.$$

Then using the fact that $\|\mu_{\beta,p}\| \ge \|\mu_{1,p}\| = 1$ we get

$$\frac{I_\alpha(p;\mathcal{W})-I_1(p;\mathcal{W})}{\phi-1} \le \left[(\frac{2}{\beta e})^2 + (\frac{2}{\epsilon e})^2\|\mu_{\frac{\beta}{1-\epsilon},p}\|\right].$$

Note that $\left\|\mu_{\frac{\beta}{1-\epsilon},p}\right\| \le \left\|\mu_{\frac{\alpha}{1-\epsilon},p}\right\|$ because $\|\mu_{\alpha,p}\|$ is nondecreasing in α by Lemma 3-(e). Then the definition of Rényi information, $\beta > 1$, and $\phi \in (1,\alpha)$ imply for any $\epsilon \in (0, \frac{\eta-1}{\eta})$, $\alpha \in [1, (1-\epsilon)\eta]$ and $p \in \mathcal{P}(\mathcal{W})$ that

$$I_\alpha(p;\mathcal{W}) - I_1(p;\mathcal{W}) \le \frac{8(\alpha-1)}{\epsilon^2 e^2}e^{\frac{\alpha-1+\epsilon}{\alpha}I_{\frac{\alpha}{1-\epsilon}}(p;\mathcal{W})}$$
$$\le \frac{8(\alpha-1)}{\epsilon^2 e^2}e^{\frac{\eta-1}{\eta}I_\eta(p;\mathcal{W})}. \quad (30)$$

Then for any $\epsilon \in (0, \frac{\eta-1}{\eta})$ and $\alpha \in [1, (1-\epsilon)\eta]$ we have

$$C_{\alpha,\mathcal{W}} \le \sup_{p\in\mathcal{P}(\mathcal{W})} I_1(p;\mathcal{W}) + \frac{8(\alpha-1)}{\epsilon^2 e^2}e^{\frac{\eta-1}{\eta}I_\eta(p;\mathcal{W})}$$
$$\le C_{1,\mathcal{W}} + \frac{8(\alpha-1)}{\epsilon^2 e^2}e^{\frac{\eta-1}{\eta}C_{\eta,\mathcal{W}}}.$$

Hence, $C_{\alpha,\mathcal{W}}$ is continuous from the right at $\alpha = 1$ if $C_{\eta,\mathcal{W}} < \infty$ for an $\eta > 1$.

(g) Let us first consider $\eta \in \mathbb{R}_+$ case and construct a sequence $\{\mathcal{W}_i\}_{i\in\mathbb{Z}_+}$ of finite subset of $\mathcal{W}$, such that $C_{\alpha,\mathcal{W}_i} \uparrow C_{\alpha,\mathcal{W}}$ for all $\alpha \in (0,\eta]$. Choose a $p^{(i,j)}$ in $\mathcal{P}(\mathcal{W})$ such that $I_{j2^{-i}}(p^{(i,j)};\mathcal{W}) \ge C_{j2^{-i},\mathcal{W}} - 1/2^i$ for each $i \in \mathbb{Z}_+$ and non-negative integer j not exceeding $2^i\eta$. Let $\mathcal{W}_0$ be the empty set and $\mathcal{W}_i$ be $\mathcal{W}_{i-1} \cup_{j=0}^{\lfloor 2^i\eta\rfloor} \text{supp}(p^{(i,j)})$ for each $i \in \mathbb{Z}_+$. Then

$$C_{\alpha,\mathcal{W}_i} \ge C_{\alpha,\mathcal{W}_{i-1}} \quad \forall\alpha \in [0,\infty], \ i \in \mathbb{Z}_+$$
$$C_{\alpha,\mathcal{W}_i} \ge C_{\alpha,\mathcal{W}} - 1/2^i \quad \forall\alpha \in \{\frac{0}{2^i}, \dots, \frac{\lfloor\eta 2^i\rfloor}{2^i}\}, \ i \in \mathbb{Z}_+.$$

Then $C_{\alpha,\mathcal{W}_i} \uparrow C_{\alpha,\mathcal{W}}$ for all dyadic rational numbers α less than η. Therefore $C_{\alpha,\mathcal{W}_i} \uparrow C_{\alpha,\mathcal{W}}$ for all $\alpha \in [0,\eta]$ because the Rényi capacity is nondecreasing and lower semicontinuous. Since $C_{\eta,\mathcal{W}_i} \le C_{\eta,\mathcal{W}} < \infty$, $C_{\alpha,\mathcal{W}_i}$'s and $C_{\alpha,\mathcal{W}}$ are continuous in α on $(0,\eta]$ by part (f). Then as a result of Dini's theorem [30, 2.4.10], $\{C_{\alpha,\mathcal{W}_i}\}_{i\in\mathbb{Z}_+}$

converges to $C_{\alpha,\mathcal{W}}$ uniformly on $[\epsilon,\eta]$, i.e. for all $\varepsilon > 0$, there exists an i such that $\sup_{\alpha\in[\epsilon,\eta]}|C_{\alpha,\mathcal{W}} - C_{\alpha,\mathcal{W}_j}| < \varepsilon$ for all $j > i$.

For $\eta = \infty$ case, let κ_i be the smallest integer satisfying $C_{\infty,\mathcal{W}} \le C_{\kappa_i/2^i,\mathcal{W}} + 1/2^i$ for each $i \in \mathbb{Z}_+$. We employ the construction described above for j's not exceeding κ_i rather than j's not exceeding $\lfloor 2^i\eta\rfloor$. $\quad\square$

B. Finiteness of the Rényi Capacity

If $\mathcal{W}$ is a finite set, then $\mathcal{P}(\mathcal{W})$ is compact for the total variation topology and various results relying on the compactness can be invoked while analyzing the Rényi information. For example if $\mathcal{W}$ is finite, then the compactness of $\mathcal{P}(\mathcal{W})$ and Sion's minimax theorem imply the continuity of the Rényi capacity in the order on $[0,\infty]$, see Lemma 16-(g). When $\mathcal{W}$ is an infinite set, however, $\mathcal{P}(\mathcal{W})$ is not compact. The finiteness of the Rényi capacity emerges as a shrewd substitute for the compactness of $\mathcal{P}(\mathcal{W})$ that allows us to assert the continuity of the Rényi information, see Lemma 16-(e,f).

Lemma 16-(a-d) characterize the finiteness of the order α Rényi capacity in terms of the properties of the order α mean measure or Rényi information. These equivalent conditions might be easier to confirm or reject for certain $\mathcal{W}$'s. The equicontinuity results given in Lemma 16-(e,f) imply that if $\gamma_1 \le I_\alpha(p;\mathcal{W}) \le \gamma_2$ for all $p \in \mathcal{A}$ for some α in $(0,\eta)$ and γ_1 and γ_2 in $[0, C_{\eta,\mathcal{W}}]$, then for any $\epsilon > 0$ there exists a $\delta > 0$ such that $\gamma_1 - \epsilon \le I_\phi(s;\mathcal{W}) \le \gamma_2 + \epsilon$ for all ϕ in $[\alpha-\delta, \alpha+\delta]$ and s in $\mathcal{P}(\mathcal{W})$ satisfying $\inf_{p\in\mathcal{A}}\|p - s\| \le \delta$. This observation (or its variants, which can be obtained by employing either part (e) or (f) on its own) might be helpful when we are trying to bound the Rényi information or a related function uniformly over the orders and priors through a case by case analysis on a subset of $\mathcal{P}(\mathcal{W})$ or on its neighborhoods.

Lemma 16. *Let $\mathcal{W}$ be a subset of $\mathcal{P}(\mathcal{Y})$.*

(a) *For $\alpha \in (0,1)$, $C_{\alpha,\mathcal{W}} = \infty$ iff there exists a sequence $\{p_i\}_{i\in\mathbb{Z}_+} \subset \mathcal{P}(\mathcal{W})$ such that $\lim_{i\to\infty}\|\mu_{\alpha,p_i}\| = 0$.*

(b) *For $\alpha \in (1,\infty]$, $C_{\alpha,\mathcal{W}} = \infty$ iff there exists a sequence $\{p_i\}_{i\in\mathbb{Z}_+} \subset \mathcal{P}(\mathcal{W})$ such that $\lim_{i\to\infty}\|\mu_{\alpha,p_i}\| = \infty$.*

(c) *For $\alpha \in (1,\infty)$, $C_{\alpha,\mathcal{W}} < \infty$ iff $\mu_{\alpha,p}$ is uniformly continuous in p for the total variation topology.*[16]

(d) *For $\alpha \in [1,\infty]$, $C_{\alpha,\mathcal{W}} < \infty$ iff $I_\alpha(p;\mathcal{W})$ is continuous in p on $\mathcal{P}(\mathcal{W})$ for the total variation topology.*

(e) *For $\eta \in \mathbb{R}_{\ge 0}$, if $C_{\eta,\mathcal{W}} < \infty$, then $\{I_\alpha(p;\mathcal{W})\}_{\alpha\in[0,\eta]}$ is uniformly equicontinuous,[17] in p on $\mathcal{P}(\mathcal{W})$.*

(f) *For $\eta \in \mathbb{R}_+$, if $C_{\eta,\mathcal{W}} < \infty$, then $\{I_\alpha(p;\mathcal{W})\}_{p\in\mathcal{P}(\mathcal{W})}$ is uniformly equicontinuous in α on every compact subset of $(0,\eta)$.*[18]

(g) *If $|\mathcal{W}| < \infty$, then $C_{\alpha,\mathcal{W}}$ is nondecreasing and continuous in α on $[0,\infty]$.*

---

[16]For $\alpha \in (0,1]$, $\mu_{\alpha,p}$ is uniformly continuous in p, even when $C_{\alpha,\mathcal{W}} = \infty$, because $\mu_{\alpha,p}$ is Lipschitz continuous on $\mathcal{P}(\mathcal{P}(\mathcal{Y}))$ by Lemma 4-(d).

[17]For $\alpha \in (0,1)$, Lemma 6-(a) has established the continuity of $I_\alpha(p;\mathcal{W})$ in p without assuming $C_{\alpha,\mathcal{W}}$ to be finite; but the continuity is not uniform.

[18]In order to prove the uniform equicontinuity on compact subsets of $(0,\eta)$, we prove the following stronger statement: On every compact subset of $(0,\eta)$, $\{I_\alpha(p;\mathcal{W})\}_{p\in\mathcal{P}(\mathcal{W})}$ is a family of Lipschitz continuous functions of α with a common Lipschitz constant, see (A.24).

Proof of Lemma 16 is deferred to Appendix C. For $\mathcal{W}$'s with infinite $C_{\alpha,\mathcal{W}}$, the proof of part (d) establishes the discontinuity at every p in $\mathcal{P}(\mathcal{W})$. For order one the discontinuity of $I_1(p;\mathcal{W})$ was observed by Ho and Yeung [53, Thm. 3] for a different topology for some $\mathcal{W}$. For the same topology they established the continuity of $I_1(p;\mathcal{W})$ whenever $\mathcal{Y}$ is finite [53, Corollary 8]. They, however, did not characterize the conditions for the continuity of $I_1(p;\mathcal{W})$ in their framework.

IV. THE RÉNYI CENTER

The primary focus of this section is Theorem 1, given in the following, and its applications. In §IV-A we prove Theorem 1 and discuss alternative proofs based on Sion's minimax theorem. In §IV-B we first prove a lower bound on $S_{\alpha,\mathcal{W}}(q)$, i.e. the van Erven-Harremoës bound, then we use this bound to establish the continuity of the Rényi center as a function of the order. §IV-C is composed of various applications of Theorem 1 and the van Erven-Harremoës bound.

Theorem 1. *For any $\alpha \in (0,\infty]$ and $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$*

$$C_{\alpha,\mathcal{W}} = \sup_{p \in \mathcal{P}(\mathcal{W})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q) \qquad (31)$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{W})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q) \qquad (32)$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{w \in \mathcal{W}} D_\alpha(w \| q). \qquad (33)$$

If $C_{\alpha,\mathcal{W}} < \infty$, then there exists a unique $q_{\alpha,\mathcal{W}}$ in $\mathcal{P}(\mathcal{Y})$, called the order α Rényi center, such that

$$C_{\alpha,\mathcal{W}} = \sup_{p \in \mathcal{P}(\mathcal{W})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q_{\alpha,\mathcal{W}}) \qquad (34)$$

$$= \sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha,\mathcal{W}}). \qquad (35)$$

Furthermore, for every sequence of priors $\{p_i\}_{i \in \mathbb{Z}_+}$ satisfying $\lim_{i \to \infty} I_\alpha(p_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$, the corresponding sequence of order α Rényi means $\{q_{\alpha,p_i}\}_{i \in \mathbb{Z}_+}$ is a Cauchy sequence for the total variation metric on $\mathcal{P}(\mathcal{Y})$ and $q_{\alpha,\mathcal{W}}$ is the unique limit point of that Cauchy sequence.

Theorem 1 is stated for p's that are probability mass functions on $\mathcal{W}$. However, the interpretation of the capacity as the radius implicit in (33) and (35) can be used to extend Theorem 1 to the case when p's are appropriately defined probability measures, see Theorem 3 in Appendix B.

For finite orders, neither the Rényi capacity nor the Rényi center has a closed form expression; this, however, is not the case for order infinity. The following expressions can be confirmed using the observation described in (39) by the interested reader.

$$C_{\infty,\mathcal{W}} = \ln \left\| \bigvee_{w \in \mathcal{W}} w \right\|, \qquad (36)$$

$$q_{\infty,\mathcal{W}} = \left(\bigvee_{w \in \mathcal{W}} w \right) e^{-C_{\infty,\mathcal{W}}}. \qquad (37)$$

Before presenting the proof and applications of Theorem 1, let us make a brief digression and discuss what is achieved by Theorem 1 itself.

The expression in (33) is nothing but the definition of the order α Rényi radius $S_{\alpha,\mathcal{W}}$. Hence, Theorem 1 establishes the equality of the order α Rényi capacity and the order α Rényi radius. We prefer to express the equality of $C_{\alpha,\mathcal{W}}$ and $S_{\alpha,\mathcal{W}}$ as a minimax equality because unlike the equality of

$C_{\alpha,\mathcal{W}}$ and $S_{\alpha,\mathcal{W}}$ itself, the minimax equality continues to hold in the constrained variant of the problem, see Theorem 2 of Appendix A.

Theorem 1 strengthens this minimax equality by asserting the existence of a unique Rényi center that is achieving the infimum in (32). Recall that we have already established, in Lemma 14, the existence of a unique Rényi mean $q_{\alpha,p}$ achieving the infimum in (31) for any p in $\mathcal{P}(\mathcal{W})$. The suprema in (31) and (32), however, cannot be replaced by maxima in general. Example 4 provides a $\mathcal{W}$ for which $\inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q) < C_{\alpha,\mathcal{W}}$ and $D_\alpha(p \circledast \mathcal{W} \| p \otimes q_{\alpha,\mathcal{W}}) < C_{\alpha,\mathcal{W}}$ for all p in $\mathcal{P}(\mathcal{W})$. Evidently, this subtlety exists only for infinite $\mathcal{W}$'s; for finite $\mathcal{W}$'s the compactness of $\mathcal{P}(\mathcal{W})$ and the extreme value theorem guarantees the existence of a p achieving the supremum.

The last assertion of Theorem 1, relating the problem of determining the Rényi capacity to the problem of determining the Rényi center, is important because of its potential in simplifying the problem of determining the Rényi center — defined as the unique $q_{\alpha,\mathcal{W}}$ satisfying (35).

In addition, Theorem 1 provides a necessary and sufficient condition for a prior p to satisfy $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$. That is important because we do not have a closed form expression for the order α Rényi capacity, yet occasionally the symmetries of the elements of $\mathcal{W}$ or numerical calculations suggest a prior p that might satisfy $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$.

$$I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}} \quad \text{iff} \quad S_{\alpha,\mathcal{W}}(q_{\alpha,p}) \leq I_\alpha(p;\mathcal{W}). \qquad (38)$$

In order to see why (38) holds, note that if $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$ then considering the sequence $\{p_i\}_{i \in \mathbb{Z}_+}$ where $p_i = p$ we can conclude that $q_{\alpha,p} = q_{\alpha,\mathcal{W}}$. Then $S_{\alpha,\mathcal{W}}(q_{\alpha,p}) \leq I_\alpha(p;\mathcal{W})$ by (35). On the other hand, if $S_{\alpha,\mathcal{W}}(q_{\alpha,p}) \leq I_\alpha(p;\mathcal{W})$ for some p in $\mathcal{P}(\mathcal{W})$, then $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$ by (33) because $I_\alpha(p;\mathcal{W}) \leq C_{\alpha,\mathcal{W}}$ and $S_{\alpha,\mathcal{W}} \leq S_{\alpha,\mathcal{W}}(q_{\alpha,p})$ by the definitions of Rényi capacity and center.

Following a similar reasoning one can show that $\{p_i\}_{i \in \mathbb{Z}_+}$ is optimal iff $S_{\alpha,\mathcal{W}}(\lim_{i \to \infty} q_{\alpha,p_i}) \leq \lim_{i \to \infty} I_\alpha(p_i;\mathcal{W})$. We chose the following less explicit characterization over the aforementioned one in order to avoid ensuring the convergence of probability measures formally.[19]

$$\lim_{i \to \infty} I_\alpha(p_i;\mathcal{W}) = C_{\alpha,\mathcal{W}} \text{ iff } \exists q : S_{\alpha,\mathcal{W}}(q) \leq \lim_{i \to \infty} I_\alpha(p_i;\mathcal{W}) \qquad (39)$$

where $q \in \mathcal{P}(\mathcal{Y})$ is implicit for the latter statement. We determine the Rényi capacity in Examples 1, 3, 4 using (38) and in Examples 2, 9 and in Appendix G using (39).

(34) of Theorem 1 and (22) imply that

$$D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}}) \leq C_{\alpha,\mathcal{W}} - I_\alpha(p;\mathcal{W}) \quad \forall p \in \mathcal{P}(\mathcal{W}). \qquad (40)$$

Consequently, $D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}})$ is close to zero whenever $I_\alpha(p;\mathcal{W})$ is close to $C_{\alpha,\mathcal{W}}$.

---

[19]We only need $C_{\alpha,\mathcal{W}} \leq S_{\alpha,\mathcal{W}}$, but not $C_{\alpha,\mathcal{W}} = S_{\alpha,\mathcal{W}}$, in order to deduce $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$ from $S_{\alpha,\mathcal{W}}(q_{\alpha,p}) \leq I_\alpha(p;\mathcal{W})$. The sufficiency of the conditions given in (38) and (39) for the optimality follows from the max-min inequality and the definitions of radius and capacity without invoking Theorem 1. We need Theorem 1 in order to assert their necessity.

A. Minimax Theorems and the Relative Compactness

We start by proving Theorem 1 for finite $\mathcal{W}$ case. In this case Theorem 1 can be strengthened slightly because the existence of an optimal prior is guaranteed. The optimal prior, however, is not necessarily unique, see Example 3; even then, all such p's have exactly the same Rényi mean. For finite $\mathcal{Y}$ case, Lemma 17 is well-known, though in a slightly different form, see [26, p. 128], [37, Thm. 4.5.1] for $\alpha = 1$ case and [26, p. 172], [37, Thm. 5.6.5] for $\alpha \in (0,1)$ case. [23, Thm. 3.2] of Csiszár implies Lemma 17 for α's in $\mathbb{R}_+$.

Lemma 17. *For any α in $[0,\infty]$ and finite subset $\mathcal{W}$ of $\mathcal{P}(\mathcal{Y})$, $\exists \widetilde{p} \in \mathcal{P}(\mathcal{W})$ such that $I_\alpha(\widetilde{p};\mathcal{W}) = C_{\alpha,\mathcal{W}}$. If α is in $(0,\infty]$, then $\exists ! q_{\alpha,\mathcal{W}} \in \mathcal{P}(\mathcal{Y})$ such that,*

$$D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}}) \le C_{\alpha,\mathcal{W}} - I_\alpha(p;\mathcal{W}) \quad \forall p \in \mathcal{P}(\mathcal{W}). \quad (41)$$

Hence, $q_{\alpha,\widetilde{p}} = q_{\alpha,\mathcal{W}}$ for all $\widetilde{p}$ such that $I_\alpha(\widetilde{p};\mathcal{W}) = C_{\alpha,\mathcal{W}}$.

Proof. (i) $\exists \widetilde{p} \in \mathcal{P}(\mathcal{W})$ *such that $I_\alpha(\widetilde{p};\mathcal{W}) = C_{\alpha,\mathcal{W}}$:* Since $|\mathrm{supp}(p)| \le |\mathcal{W}|$ for all $p \in \mathcal{P}(\mathcal{W})$, $C_{\alpha,\mathcal{W}} \le \ln|\mathcal{W}|$ by Lemma 5. Then $I_\alpha(p;\mathcal{W})$ is continuous on $\mathcal{P}(\mathcal{W})$ by Lemmas 6-(a) and 16-(d). Then there exists a $\widetilde{p}$ achieving the supremum by the extreme value theorem, [68, 27.4] because $\mathcal{P}(\mathcal{W})$ is compact for finite $\mathcal{W}$.

(ii) *If $I_\alpha(\widetilde{p};\mathcal{W}) = C_{\alpha,\mathcal{W}}$ for an $\alpha \in (0,\infty]$, then $D_\alpha(q_{\alpha,p} \| q_{\alpha,\widetilde{p}}) \le C_{\alpha,\mathcal{W}} - I_\alpha(p;\mathcal{W})$ for all $p \in \mathcal{P}(\mathcal{W})$:* Let $\widetilde{p} \in \mathcal{P}(\mathcal{W})$ be such that $I_\alpha(\widetilde{p};\mathcal{W}) = C_{\alpha,\mathcal{W}}$, p be any member of $\mathcal{P}(\mathcal{W})$ and $p_\imath$ be $p_\imath = \frac{\imath-1}{\imath}\widetilde{p} + \frac{1}{\imath}p$ for $\imath \in \mathbb{Z}_+$. For $\alpha = \infty$ using Lemma 14 we get

$$I_\infty(p_\imath;\mathcal{W}) = [I_\infty(\widetilde{p};\mathcal{W}) + D_\infty(q_{\infty,\widetilde{p}} \| q_{\infty,p_\imath})]$$
$$\vee [I_\infty(p;\mathcal{W}) + D_\infty(q_{\infty,p} \| q_{\infty,p_\imath})].$$

Then $D_\infty(q_{\infty,\widetilde{p}} \| q_{\infty,p_\imath}) = 0$ because $I_\infty(p_\imath;\mathcal{W}) \le C_{\infty,\mathcal{W}}$ and $I_\infty(\widetilde{p};\mathcal{W}) = C_{\infty,\mathcal{W}}$. Consequently $q_{\infty,\widetilde{p}} = q_{\infty,p_\imath}$ and $I_\infty(p_\imath;\mathcal{W}) = C_{\infty,\mathcal{W}}$. Thus

$$I_\infty(p;\mathcal{W}) + D_\infty(q_{\infty,p} \| q_{\infty,\widetilde{p}}) \le C_{\infty,\mathcal{W}}. \quad (42)$$

For $\alpha = 1$ and $\alpha \in \mathbb{R}_+ \setminus \{1\}$ we have

$$I_1(p_\imath;\mathcal{W}) = \frac{\imath-1}{\imath}[I_1(\widetilde{p};\mathcal{W}) + D_1(q_{1,\widetilde{p}} \| q_{1,p_\imath})]$$
$$+ \frac{1}{\imath}[I_1(p;\mathcal{W}) + D_1(q_{1,p} \| q_{1,p_\imath})],$$
$$I_\alpha(p_\imath;\mathcal{W}) = \frac{1}{\alpha-1}\ln\left[\frac{\imath-1}{\imath}e^{(\alpha-1)(I_\alpha(\widetilde{p};\mathcal{W})+D_\alpha(q_{\alpha,\widetilde{p}} \| q_{\alpha,p_\imath}))}\right.$$
$$\left. + \frac{1}{\imath}e^{(\alpha-1)(I_\alpha(p;\mathcal{W})+D_\alpha(q_{\alpha,p} \| q_{\alpha,p_\imath}))}\right].$$

Then using $I_\alpha(p_\imath;\mathcal{W}) \le C_{\alpha,\mathcal{W}}$, $I_\alpha(\widetilde{p};\mathcal{W}) = C_{\alpha,\mathcal{W}}$, and $D_\alpha(q_{\alpha,\widetilde{p}} \| q_{\alpha,p_\imath}) \ge 0$ we get the following identity

$$I_\alpha(p;\mathcal{W}) + D_\alpha(q_{\alpha,p} \| q_{\alpha,p_\imath}) \le C_{\alpha,\mathcal{W}}.$$

Similarly, using $I_\alpha(p_\imath;\mathcal{W}) \le C_{\alpha,\mathcal{W}}$, $I_\alpha(\widetilde{p};\mathcal{W}) = C_{\alpha,\mathcal{W}}$, $I_\alpha(p;\mathcal{W}) \ge 0$, and $D_\alpha(q_{\alpha,p} \| q_{\alpha,p_\imath}) \ge 0$ we get

$$D_\alpha(q_{\alpha,\widetilde{p}} \| q_{\alpha,p_\imath}) \le \begin{cases} \frac{1}{\alpha-1}\ln\frac{\imath-e^{(1-\alpha)C_{\alpha,\mathcal{W}}}}{\imath-1} & \alpha \in \mathbb{R}_+ \setminus \{1\} \\ \frac{C_{\alpha,\mathcal{W}}}{\imath-1} & \alpha = 1 \end{cases}.$$

Then $q_{\alpha,p_\imath} \to q_{\alpha,\widetilde{p}}$ in the total variation topology by Lemma 10. Thus

$$D_\alpha(q_{\alpha,p} \| q_{\alpha,\widetilde{p}}) \le \liminf_{\imath \to \infty} D_\alpha(q_{\alpha,p} \| q_{\alpha,p_\imath})$$

by Lemma 13. Then

$$I_\alpha(p;\mathcal{W}) + D_\alpha(q_{\alpha,p} \| q_{\alpha,\widetilde{p}}) \le C_{\alpha,\mathcal{W}} \quad \forall \alpha \in \mathbb{R}_+. \quad (43)$$

(iii) *If $\alpha \in (0,\infty]$, then $\exists ! q_{\alpha,\mathcal{W}} \in \mathcal{P}(\mathcal{Y})$ satisfying (41) such that $q_{\alpha,p} = q_{\alpha,\mathcal{W}}$ for all p with $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$:* (42), (43) and Lemma 10 implies that

$$I_\alpha(p;\mathcal{W}) + \frac{\alpha \wedge 1}{2}\|q_{\alpha,p} - q_{\alpha,\widetilde{p}}\|^2 \le C_{\alpha,\mathcal{W}}.$$

Then $q_{\alpha,\widetilde{p}} = q_{\alpha,p}$ for any p satisfying $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$. $\square$

When $\mathcal{W}$ is not a finite but an arbitrary subset of $\mathcal{P}(\mathcal{Y})$, we cannot invoke the extreme value theorem to establish the existence an optimal prior p satisfying $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$ because $\mathcal{P}(\mathcal{W})$ is not compact. Assuming $C_{\alpha,\mathcal{W}}$ to be finite, Theorem 1 recovers all assertions of Lemma 17, but the existence of an optimal prior, albeit in a weaker form.

Proof of Theorem 1. For all $p \in \mathcal{P}(\mathcal{W})$ and $q \in \mathcal{P}(\mathcal{Y})$, (19) implies $D_\alpha(p \circledast \mathcal{W} \| p \otimes q) \le \max_{w \in \mathrm{supp}(p)} D_\alpha(w \| q)$. Then considering p's satisfying $p(w) = 1$ for a w in $\mathcal{W}$ we get

$$\sup_{w \in \mathcal{W}} D_\alpha(w \| q) = \sup_{p \in \mathcal{P}(\mathcal{W})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q) \quad (44)$$

for all $q \in \mathcal{P}(\mathcal{Y})$. Note that (32) implies (33) and (34) implies (35) because of (44). Furthermore, (31) is nothing but (28) and expression on the right hand side of (31) is bounded from above by the expression in (32) as a result of max-min inequality. Thus when $C_{\alpha,\mathcal{W}}$ is infinite, (32) holds trivially. When $C_{\alpha,\mathcal{W}}$ is finite, the converse of max-min inequality, and hence (32), follows from (34). Thus, we can assume $C_{\alpha,\mathcal{W}}$ to be finite and prove the claims about $q_{\alpha,\mathcal{W}}$ in order to prove the theorem.

(i) *If $C_{\alpha,\mathcal{W}} < \infty$ and $\lim_{\imath \to \infty} I_\alpha(p_\imath;\mathcal{W}) = C_{\alpha,\mathcal{W}}$, then $\{q_{\alpha,p_\imath}\}_{\imath \in \mathbb{Z}_+}$ is a Cauchy sequence in $\mathcal{P}(\mathcal{Y})$ for the total variation metric:* For any sequence $\{p_\imath\}_{\imath \in \mathbb{Z}_+}$ of members of $\mathcal{P}(\mathcal{W})$ satisfying $\lim_{\imath \to \infty} I_\alpha(p_\imath;\mathcal{W}) = C_{\alpha,\mathcal{W}}$, let $\{\mathcal{W}_\imath\}_{\imath \in \mathbb{Z}_+}$ be a nested sequence of finite subsets of $\mathcal{W}$ defined as follows,

$$\mathcal{W}_\imath \triangleq \cup_{\jmath=1}^{\imath} \mathrm{supp}(p_\jmath).$$

Then for any $\imath \in \mathbb{Z}_+$, there exists a unique $q_{\alpha,\mathcal{W}_\imath}$ satisfying (41) by Lemma 17. Furthermore, $\mathcal{P}(\mathcal{W}_\jmath) \subset \mathcal{P}(\mathcal{W}_\imath)$ for any $\imath,\jmath \in \mathbb{Z}_+$ such that $\jmath \le \imath$. In order to bound $\|q_{\alpha,p_\jmath} - q_{\alpha,p_\imath}\|$ for positive integers $\jmath < \imath$, we use the triangle inequality for $q_{\alpha,p_\jmath}$, $q_{\alpha,p_\imath}$, and $q_{\alpha,\mathcal{W}_\imath}$:

$$\|q_{\alpha,p_\jmath} - q_{\alpha,p_\imath}\| \le \|q_{\alpha,p_\jmath} - q_{\alpha,\mathcal{W}_\imath}\| + \|q_{\alpha,p_\imath} - q_{\alpha,\mathcal{W}_\imath}\|. \quad (45)$$

Let us proceed with bounding $\|q_{\alpha,p_\jmath} - q_{\alpha,\mathcal{W}_\imath}\|$.

$$\|q_{\alpha,p_\jmath} - q_{\alpha,\mathcal{W}_\imath}\|^2 \overset{(a)}{\le} \frac{2}{\alpha \wedge 1} D_\alpha(q_{\alpha,p_\jmath} \| q_{\alpha,\mathcal{W}_\imath})$$
$$\overset{(b)}{\le} \frac{2}{\alpha \wedge 1}[C_{\alpha,\mathcal{W}_\imath} - I_\alpha(p_\jmath;\mathcal{W}_\imath)]$$
$$\overset{(c)}{\le} \frac{2}{\alpha \wedge 1}[C_{\alpha,\mathcal{W}} - I_\alpha(p_\jmath;\mathcal{W})].$$

where (a) follows from Lemma 10, (b) follows from Lemma 17 because $\widetilde{p}_\jmath \in \mathcal{P}(\mathcal{W}_\imath)$, and (c) follows from the identities $I_\alpha(p_\jmath;\mathcal{W}_\imath) = I_\alpha(p_\jmath;\mathcal{W})$ and $C_{\alpha,\mathcal{W}_\imath} \le C_{\alpha,\mathcal{W}}$. We

can obtain a similar bound on $\|q_{\alpha,p_i} - q_{\alpha,\mathcal{W}_i}\|^2$. Then $\{q_{\alpha,p_i}\}$ is a Cauchy sequence by (45).

(ii) *If $C_{\alpha,\mathcal{W}} < \infty$, then $\exists! \; q_{\alpha,\mathcal{W}}$ in $\mathcal{P}(\mathcal{Y})$ satisfying $\lim_{i\to\infty} \|q_{\alpha,\mathcal{W}} - q_{\alpha,p_i}\| = 0$ for all $\{p_i\}_{i\in\mathbb{Z}_+}$ satisfying $\lim_{i\to\infty} I_\alpha(p_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$:* Note that $\mathcal{M}(\mathcal{Y})$ is a complete metric space for the total variation metric, i.e. every Cauchy sequence has a unique limit point in $\mathcal{M}(\mathcal{Y})$, because $\mathcal{M}(\mathcal{Y})$ is a Banach space for the total variation topology [8, Thm. 4.6.1]. Then $\{q_{\alpha,p_i}\}_{i\in\mathbb{Z}_+}$ has a unique limit point $q_{\alpha,\mathcal{W}}$ in $\mathcal{M}(\mathcal{Y})$. Since $\mathcal{P}(\mathcal{Y})$ is a closed set for the total variation topology and $q_{\alpha,p_i} \in \mathcal{P}(\mathcal{Y})$ for all $i \in \mathbb{Z}_+$, the limit point $q_{\alpha,\mathcal{W}}$ is in $\mathcal{P}(\mathcal{Y})$ by [68, Thm. 2.1.3].

We have established the existence of a unique limit point for any $\{p_i\}_{i\in\mathbb{Z}_+}$ satisfying $\lim_{i\to\infty} I_\alpha(p_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$. However, we have not ruled out the possibility of distinct limit points for different sequences satisfying the constraint. Let $\{p_i\}_{i\in\mathbb{Z}_+}$ and $\{\tilde{p}_i\}_{i\in\mathbb{Z}_+}$ be two sequences satisfying $\lim_{i\to\infty} I_\alpha(p_i;\mathcal{W}) = \lim_{i\to\infty} I_\alpha(\tilde{p}_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$, with limit points $q_{\alpha,\mathcal{W}}$ and $\tilde{q}_{\alpha,\mathcal{W}}$. Let $\{\hat{p}_i\}_{i\in\mathbb{Z}_+}$ be a sequence whose elements for the odd indices are the elements of $\{p_i\}_{i\in\mathbb{Z}_+}$ and whose elements for the even indices are the elements of $\{\tilde{p}_i\}_{i\in\mathbb{Z}_+}$. Then $\lim_{i\to\infty} I_\alpha(\hat{p}_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$; consequently the sequence $\{q_{\alpha,\hat{p}_i}\}_{i\in\mathbb{Z}_+}$ is Cauchy. Thus $\{q_{\alpha,\hat{p}_i}\}_{i\in\mathbb{Z}_+}$ and all of its subsequences has the same unique limit point $\hat{q}_{\alpha,\mathcal{W}}$. Then $q_{\alpha,\mathcal{W}} = \hat{q}_{\alpha,\mathcal{W}} = \tilde{q}_{\alpha,\mathcal{W}}$.

(iii) *$q_{\alpha,\mathcal{W}}$ satisfies the equality given in (34):* For any p in $\mathcal{P}(\mathcal{W})$, let us consider a sequence $\{p_i\}_{i\in\mathbb{Z}_+}$ satisfying both $p_1 = p$ and $\lim_{i\to\infty} I_\alpha(p_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$. Then $p \in \mathcal{P}(\mathcal{W}_i)$ for all $i \in \mathbb{Z}_+$. Then using the inequality given in (41) of Lemma 17 together with (22) we get

$$D_\alpha(p\circledast\mathcal{W}\| p \otimes q_{\alpha,\mathcal{W}_i}) \leq C_{\alpha,\mathcal{W}_i} \qquad \forall i. \qquad (46)$$

Since $\mathcal{W}_i$ is a finite set, $\exists \tilde{p}_i \in \mathcal{P}(\mathcal{W}_i)$ satisfying $I_\alpha(\tilde{p}_i;\mathcal{W}_i) = C_{\alpha,\mathcal{W}_i}$ and $q_{\alpha,\tilde{p}_i} = q_{\alpha,\mathcal{W}_i}$ by Lemma 17. Then $I_\alpha(\tilde{p}_i;\mathcal{W}_i) \geq I_\alpha(p_i;\mathcal{W}_i)$ because $p_i \in \mathcal{P}(\mathcal{W}_i)$ by construction. Consequently $\lim_{i\to\infty} I_\alpha(\tilde{p}_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$. We have already established that for such a sequence $q_{\alpha,\tilde{p}_i} \to q_{\alpha,\mathcal{W}}$ in the total variation topology, and hence in the topology of setwise convergence. Then the lower semicontinuity of the Rényi divergence, i.e. Lemma 13, the identity $C_{\alpha,\mathcal{W}_i} \leq C_{\alpha,\mathcal{W}}$, and (46) imply

$$D_\alpha(p\circledast\mathcal{W}\| p \otimes q_{\alpha,\mathcal{W}}) \leq C_{\alpha,\mathcal{W}}.$$

Thus using (24) we get

$$I_\alpha(p;\mathcal{W}) \leq D_\alpha(p\circledast\mathcal{W}\| p \otimes q_{\alpha,\mathcal{W}}) \leq C_{\alpha,\mathcal{W}} \quad \forall p \in \mathcal{P}(\mathcal{W}).$$

Then (34) follows the definition of $C_{\alpha,\mathcal{W}}$.

$\square$

Theorem 1 is not just a minimax theorem, the assertions about the Rényi center are crucial. But those assertions can be derived separately, if need be. Leaving them aside, we discuss in the rest of this subsection when (32) can be proved using Sion's minimax theorem [57], [95].

Note that $\mathcal{P}(\mathcal{W})$ is compact iff $\mathcal{W}$ is a finite set and $\mathcal{P}(\mathcal{Y})$ is compact iff $\mathcal{Y}$ is a finite set. Consequently, when either $\mathcal{W}$

or $\mathcal{Y}$ is finite, (32) is an immediate consequence[20] of Sion's minimax theorem [57], [95]. When $\mathcal{W}$ and $\mathcal{Y}$ are both infinite sets, however, neither $\mathcal{P}(\mathcal{W})$ nor $\mathcal{P}(\mathcal{Y})$ is compact —for the total variation topology— and we cannot directly apply Sion's minimax theorem. Yet, it is possible to recover partial results using the concept of relative compactness. Recall that a set of points in a topological space is called relatively compact if it has a compact closure.

First note that as a result of Lemma 14

$$I_\alpha(p;\mathcal{W}) = \inf_{q\in\mathrm{cl}\mathcal{Q}_{\alpha,\mathcal{W}}} D_\alpha(p\circledast\mathcal{W}\| p \otimes q) \qquad (47)$$

for all p's in $\mathcal{P}(\mathcal{W})$ and α's in $\mathbb{R}_+$ where $\mathcal{Q}_{\alpha,\mathcal{W}}$ is the convex hull of the set of all order α Rényi means:

$$\mathcal{Q}_{\alpha,\mathcal{W}} \triangleq \mathrm{ch}\{q_{\alpha,p} : p \in \mathcal{P}(\mathcal{W})\}.$$

If $\mathcal{Q}_{\alpha,\mathcal{W}}$ is relatively compact in the topology of setwise convergence, Sion's minimax theorem imply that

$$\sup_{p\in\mathcal{P}(\mathcal{W})} \inf_{q\in\mathrm{cl}\mathcal{Q}_{\alpha,\mathcal{W}}} D_\alpha(p\circledast\mathcal{W}\| p \otimes q)$$
$$= \inf_{q\in\mathrm{cl}\mathcal{Q}_{\alpha,\mathcal{W}}} \sup_{p\in\mathcal{P}(\mathcal{W})} D_\alpha(p\circledast\mathcal{W}\| p \otimes q). \quad (48)$$

We can replace $\mathrm{cl}\mathcal{Q}_{\alpha,\mathcal{W}}$ by $\mathcal{P}(\mathcal{Y})$ in the expression on the left hand side without changing its value as a result of (47). However, that operation can decrease the value of the right hand side because $\mathrm{cl}\mathcal{Q}_{\alpha,\mathcal{W}} \subset \mathcal{P}(\mathcal{Y})$. Thus we get,

$$\sup_{p\in\mathcal{P}(\mathcal{W})} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(p\circledast\mathcal{W}\| p \otimes q)$$
$$\geq \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{P}(\mathcal{W})} D_\alpha(p\circledast\mathcal{W}\| p \otimes q).$$

The reverse inequality is the max-min inequality, which is always valid. Thus (32) holds.

A set of finite measures $\mathcal{W}$ is relatively compact in the topology of setwise convergence iff there exists a $\nu \in \mathcal{P}(\mathcal{Y})$ such that $\mathcal{W}\prec^{uni}\nu$ by a version of the Dunford-Pettis theorem [8, 4.7.25]. Using de la Vallée Poussin's characterization of the uniform integrability [8, Thm. 4.5.9] and monotonicity of the order α mean measure $\mu_{\alpha,p}$ in the order, i.e. Lemma 3-(b), we can obtain sufficient conditions for the relative compactness of $\mathcal{Q}_{\alpha,\mathcal{W}}$ in the topology of setwise convergence for any $\alpha \in \mathbb{R}_+$. As a result we get the following partial result:

Lemma 18. *Let $\mathcal{W}$ be subset of $\mathcal{P}(\mathcal{Y})$.*
(i) If $\exists\nu \in \mathcal{P}(\mathcal{Y})$ such that $\mathcal{W}\prec^{uni}\nu$ and $S_{\eta,\mathcal{W}} < \infty$ for an $\eta \in (0,1)$, then (32) holds $\forall\alpha \in (0,1)$.
(ii) If $S_{\eta,\mathcal{W}} < \infty$ for an $\eta \in [1,\infty]$, then (32) holds $\forall\alpha \in (0,\eta]$.

B. The Rényi Center as a Function of the Order

$S_{\alpha,\mathcal{W}}$ is defined as the greatest lower bound of $S_{\alpha,\mathcal{W}}(q)$. Then Theorem 1 implies, by establishing $C_{\alpha,\mathcal{W}} = S_{\alpha,\mathcal{W}}$, that

$$S_{\alpha,\mathcal{W}}(q) \geq C_{\alpha,\mathcal{W}} \qquad \forall q \in \mathcal{P}(\mathcal{Y}).$$

Van Erven and Harremoës have conjectured that a better lower bound on $S_{\alpha,\mathcal{W}}(q)$ should hold, [31, Conjecture 1]. Van Erven and Harremoës proved their claim for $\alpha = \infty$ case assuming that $\mathcal{Y}$ is countable, [31, Thm. 37]. Lemma 19 establishes

---

[20]Immediate after establishing that $D_\alpha(p\circledast\mathcal{W}\| p \otimes q)$ is upper semicontinuous and quasi-concave in p. The lower semicontinuity and the quasi-convexity of $D_\alpha(p\circledast\mathcal{W}\| p \otimes q)$ in q follow from Lemmas 11 and 13.

the van Erven-Harremoës bound for any positive order α and $\mathcal{W}$ satisfying $C_{\alpha,\mathcal{W}} < \infty$, using Theorem 1. A constrained generalization, i.e. Lemma 25, can be found in Appendix A.

Lemma 19. *For any $\alpha \in (0,\infty]$, $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$ satisfying $C_{\alpha,\mathcal{W}} < \infty$, and $q \in \mathcal{P}(\mathcal{Y})$,*

$$\sup_{w \in \mathcal{W}} D_\alpha(w \| q) \geq C_{\alpha,\mathcal{W}} + D_\alpha(q_{\alpha,\mathcal{W}} \| q). \quad (49)$$

Lemma 19 quantifies how loose $S_{\alpha,\mathcal{W}}(q)$ —defined in (1)— is as an upper bound to $C_{\alpha,\mathcal{W}}$, as surmised by van Erven and Harremoës in [31].

Proof of Lemma 19. As a result of (22) and (23) we have,

$$\sup_{\tilde{p} \in \mathcal{P}(\mathcal{W})} D_\alpha(\tilde{p} \circledast \mathcal{W} \| \tilde{p} \otimes q) \geq D_\alpha(p \circledast \mathcal{W} \| p \otimes q)$$
$$= I_\alpha(p;\mathcal{W}) + D_\alpha(q_{\alpha,p} \| q) \quad (50)$$

for all $p \in \mathcal{P}(\mathcal{W})$. Let $\{p_i\}_{i \in \mathbb{Z}_+}$ be a sequence of elements of $\mathcal{P}(\mathcal{W})$ such that $\lim_{i \to \infty} I_\alpha(p_i;\mathcal{W}) = C_{\alpha,\mathcal{W}}$. Then the sequence $\{q_{\alpha,p_i}\}_{i \in \mathbb{Z}_+}$ is a Cauchy sequence with the unique limit point $q_{\alpha,\mathcal{W}}$ by Theorem 1. Since $\{q_{\alpha,p_i}\} \to q_{\alpha,\mathcal{W}}$ in total variation topology, same convergence holds in the topology of setwise convergence because every open neighborhood in the latter includes an open neighborhood in the former by the definitions of these topologies. On the other hand, the order α Rényi divergence is lower semicontinuous for the topology of setwise convergence by Lemma 13. Thus we have

$$\liminf_{i \to \infty} [I_\alpha(p_i;\mathcal{W}) + D_\alpha(q_{\alpha,p_i} \| q)] \geq C_{\alpha,\mathcal{W}} + D_\alpha(q_{\alpha,\mathcal{W}} \| q).$$

Then (49) follows from (44) and (50). $\square$

The van Erven-Harremoës bound allows us to use the continuity of $C_{\alpha,\mathcal{W}}$ in α and Pinsker's inequality to establish the continuity of $q_{\alpha,\mathcal{W}}$ in α for the total variation topology.

Lemma 20. *For any $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$ and $\eta \in (0,\infty]$ such that $C_{\eta,\mathcal{W}} < \infty$,*

$$C_{\phi,\mathcal{W}} - C_{\alpha,\mathcal{W}} \geq D_\alpha(q_{\alpha,\mathcal{W}} \| q_{\phi,\mathcal{W}}) \quad (51)$$

for all α and ϕ satisfying $0 < \alpha < \phi \leq \eta$. Furthermore, $q_{\alpha,\mathcal{W}}$ is a continuous function of α on $(0,\eta]$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

The continuity of the Rényi center as a function of the order is important because it allows us to the interpret the Rényi centers as a transition probability from the interval on which the Rényi capacity is finite to $(\mathcal{Y},\mathcal{Y})$ and apply Augustin's method, see [73, §III-A] for a more detailed discussion.

Proof of Lemma 20. For $q = q_{\phi,\mathcal{W}}$, Lemma 19 implies

$$\sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\phi,\mathcal{W}}) \geq C_{\alpha,\mathcal{W}} + D_\alpha(q_{\alpha,\mathcal{W}} \| q_{\phi,\mathcal{W}}). \quad (52)$$

Since $D_\alpha(w \| q_{\phi,\mathcal{W}})$ is nondecreasing in α by Lemma 7,

$$D_\phi(w \| q_{\phi,\mathcal{W}}) \geq D_\alpha(w \| q_{\phi,\mathcal{W}}) \quad \forall w \in \mathcal{W}, \ \phi \in [\alpha,\eta]. \quad (53)$$

On the other hand by (35) of Theorem 1 we have

$$C_{\phi,\mathcal{W}} = \sup_{w \in \mathcal{W}} D_\phi(w \| q_{\phi,\mathcal{W}}) \quad \forall \phi \in (0,\eta]. \quad (54)$$

(51) follows from (52), (53), and (54).

Using Lemma 10 and (51) we get[21]

$$\sqrt{\tfrac{2}{\phi \wedge 1}(C_{\phi,\mathcal{W}} - C_{\alpha,\mathcal{W}})} \geq \|q_{\phi,\mathcal{W}} - q_{\alpha,\mathcal{W}}\|. \quad (55)$$

Then, for the total variation topology on $\mathcal{P}(\mathcal{Y})$, the continuity of $q_{\alpha,\mathcal{W}}$ in α follows from the continuity $C_{\alpha,\mathcal{W}}$ in α on $(0,\eta]$, i.e. Lemma 15-(f). $\square$

Lemma 20 establishes the continuity of the Rényi center in the order for the total variation topology. We suspect a much stronger statement is true.

Conjecture 1. *For any $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$ and $\eta \in (0,\infty]$. satisfying $C_{\eta,\mathcal{W}} < \infty$,*

$$\mu_{\phi,\mathcal{W}} \leq \mu_{\eta,\mathcal{W}} \qquad \forall \phi \in (0,\eta] \quad (56)$$

where $\mu_{\phi,\mathcal{W}} \triangleq e^{\frac{\phi-1}{\phi} C_{\phi,\mathcal{W}}} q_{\phi,\mathcal{W}}$ for all $\phi \in (0,\eta]$.

For any $\mathcal{W}$ using the continuity of the Rényi center in the order, one can prove that there exists a ν in $\mathcal{P}(\mathcal{Y})$ such that $\{q_{\alpha,\mathcal{W}} : C_{\alpha,\mathcal{W}} < \infty\} \prec \nu$. However, the continuity of the Rényi center as a function of the order for the total variation topology does not imply the continuity of corresponding Radon-Nikodym derivative $\frac{dq_{\alpha,\mathcal{W}}}{d\nu}$ as a function of α for ν-almost everywhere. If Conjecture 1 is correct, then it will imply the continuity of Radon-Nikodym derivative $\frac{dq_{\alpha,\mathcal{W}}}{d\nu}$ as a function of α for ν-almost everywhere.

Remark 2. The continuity in the total variation topology does not imply the continuity of the corresponding Radon-Nikodym derivative: Let the output space be the real numbers between -1 and 2, and the Radon-Nikodym derivative of q_α with respect to the Lebesgue measure l be

$$\frac{dq_\alpha}{dl} = \mathbb{1}_{\{\sin(\frac{1}{t-\alpha}) \leq y \leq \sin(\frac{1}{t-\alpha}) + |\alpha-t|\}} + \mathbb{1}_{\{0 \leq y \leq 1\}}(1 - |\alpha - t|)$$

for a $t \in (0,1)$. Evidently $\lim_{\eta \to \alpha} \|q_\alpha - q_\eta\| = 0$ for all α in $(0,1)$. But $\frac{dq_\alpha}{dl}$ is not continuous in α for any $y \in (0,1)$ at t.

C. The Unions, Cartesian Products, Closures, and More

This subsection is composed of applications of Theorem 1 and Lemma 19. Lemma 21, in the following, bounds from below and from above the Rényi capacity of a union of sets in terms of the Rényi capacities of the sets in the union. Lemma 22 establishes that the Rényi capacity of a Cartesian product is equal to the sum of the Rényi capacities of its components. Lemma 23 shows that for any positive ϵ the order α Rényi capacity of the set of w's in $\mathcal{W}$ satisfying $D_\alpha(w \| q_{\alpha,\mathcal{W}}) \geq C_{\alpha,\mathcal{W}} - \epsilon$ is equal to $C_{\alpha,\mathcal{W}}$. Lemma 24 establishes the invariance of $C_{\alpha,\mathcal{W}}$ under the closure and convexification operations on $\mathcal{W}$ and characterizes the relative compactness of $\mathcal{W}$ in terms of its Rényi capacity. Proofs of these lemmas are presented in Appendix C.

---

[21]For proving a similar continuity result in [6], instead of (51), Augustin uses the inequality given in the following —which can be proved using (13), (19), Lemma 3-(b), and Theorem 1: If either $\alpha \in [\phi,\eta]$ and $\phi < 1$ or $\alpha \in (0,\phi]$ and $\phi > 1$ then

$$D_\phi(q_{\alpha,p} \| q_{\phi,\mathcal{W}}) \leq C_{\phi,\mathcal{W}} - \tfrac{\phi}{\phi-1} \ln \|\mu_{\alpha,p}\| \qquad \forall p \in \mathcal{P}(\mathcal{W}).$$

Lemma 21. *For any $\alpha \in (0,\infty]$ and $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$ satisfying $\mathcal{W} = \cup_{\imath \in \mathcal{I}} \mathcal{W}_\imath$ for some $\mathcal{W}_\imath \subset \mathcal{P}(\mathcal{Y})$ with finite $C_{\alpha,\mathcal{W}_\imath}$'s,*

$$\sup_{\imath \in \mathcal{I}} C_{\alpha,\mathcal{W}_\imath} \leq C_{\alpha,\mathcal{W}} \leq \ln \sum_{\imath \in \mathcal{I}} e^{C_{\alpha,\mathcal{W}_\imath}}. \quad (57)$$

Furthermore,

- *$C_{\alpha,\mathcal{W}_\imath} = C_{\alpha,\mathcal{W}}$ iff $S_{\alpha,\mathcal{W}}(q_{\alpha,\mathcal{W}_\imath}) \leq C_{\alpha,\mathcal{W}_\imath}$.*
- *If $C_{\alpha,\mathcal{W}_\imath} = C_{\alpha,\mathcal{W}}$, then $q_{\alpha,\mathcal{W}} = q_{\alpha,\mathcal{W}_\imath}$.*
- *$C_{\alpha,\mathcal{W}} = \ln \sum_{\imath \in \mathcal{I}} e^{C_{\alpha,\mathcal{W}_\imath}}$ and $C_{\alpha,\mathcal{W}}$ is finite iff $\mathcal{I}$ is finite and $q_{\alpha,\mathcal{W}_\imath} \perp q_{\alpha,\mathcal{W}_\jmath}$ for all $\imath \neq \jmath$ in $\mathcal{I}$.*
- *If $\mathcal{I}$ is finite and $q_{\alpha,\mathcal{W}_\imath} \perp q_{\alpha,\mathcal{W}_\jmath}$ for all $\imath \neq \jmath$ in $\mathcal{I}$, then $q_{\alpha,\mathcal{W}} = (\sum_{\jmath \in \mathcal{I}} e^{C_{\alpha,\mathcal{W}_\jmath}})^{-1} \sum_{\imath \in \mathcal{I}} e^{C_{\alpha,\mathcal{W}_\imath}} q_{\alpha,\mathcal{W}_\imath}$.*

One might think that $q_{\alpha,\mathcal{W}_\imath} \perp q_{\alpha,\mathcal{W}_\jmath}$ iff $\mathcal{W}_\imath \perp \mathcal{W}_\jmath$. This, however, is true only for α's in $[1,\infty]$. For α's in $(0,1)$, $\mathcal{W}_\imath \perp \mathcal{W}_\jmath$ is a sufficient condition for $q_{\alpha,\mathcal{W}_\imath} \perp q_{\alpha,\mathcal{W}_\jmath}$, but it is not a necessary condition, see Examples 1 and 2. Augustin [6] is the first one to point out this subtlety and to present necessary and sufficient conditions for $C_{\alpha,\mathcal{W}} = \ln \sum_{\imath \in \mathcal{I}} e^{C_{\alpha,\mathcal{W}_\imath}}$, to the best of our knowledge. Bounds given in (57) is well known [37, p. 535, ex. 5.17]. We use the van Erven-Harremoës bound in order to characterize the necessary and sufficient conditions for $\sup_{\imath \in \mathcal{I}} C_{\alpha,\mathcal{W}_\imath} = C_{\alpha,\mathcal{W}}$ and $C_{\alpha,\mathcal{W}} = \ln \sum_{\imath \in \mathcal{I}} e^{C_{\alpha,\mathcal{W}_\imath}}$.

Let $\mathcal{I}$ be a finite set. For each $t \in \mathcal{I}$, let $(\mathcal{Y}_t, \mathcal{Y}_t)$ be a measurable space and w_t be a probability measure on $(\mathcal{Y}_t, \mathcal{Y}_t)$. Then there exists a unique product measure $\bigotimes_{t \in \mathcal{I}} w_t$ on the measurable space $(\times_{t \in \mathcal{I}} \mathcal{Y}_t, \bigotimes_{t \in \mathcal{I}} \mathcal{Y}_t)$ by [30, Thm. 8.2.2].[22] Let $\mathcal{W}_t$ be a subset of $\mathcal{P}(\mathcal{Y}_t)$ for each $t \in \mathcal{I}$. Then using the existence of a unique product measure we can map the Cartesian product of the sets $\mathcal{W}_t$ uniquely to a subset of $\mathcal{P}(\bigotimes_{t \in \mathcal{I}} \mathcal{Y}_t)$, called the product of $\mathcal{W}_t$'s. Then the Rényi capacity of the product is equal to the sum of the Rényi capacities of its components and the Rényi center of the product, whenever it exists, is equal to the product of the Rényi centers of its components. Lemma 22 asserts these observations formally.

Lemma 22. *For any finite index set $\mathcal{I}$, if $\mathcal{Y} = \times_{t \in \mathcal{I}} \mathcal{Y}_t$, $\mathcal{Y} = \bigotimes_{t \in \mathcal{I}} \mathcal{Y}_t$, and $\mathcal{W} = \{w : w = \bigotimes_{t \in \mathcal{I}} w_t : w_t \in \mathcal{W}_t\}$ for some $\mathcal{W}_t \subset \mathcal{P}(\mathcal{Y}_t)$, then*

$$C_{\alpha,\mathcal{W}} = \sum_{t \in \mathcal{I}} C_{\alpha,\mathcal{W}_t} \qquad \forall \alpha \in (0,\infty]. \quad (58)$$

Furthermore, if $C_{\alpha,\mathcal{W}} < \infty$, then $q_{\alpha,\mathcal{W}} = \bigotimes_{t \in \mathcal{I}} q_{\alpha,\mathcal{W}_t}$.

Quite frequently, the information transmission problems are analyzed on the product $\mathcal{W}$'s. Lemma 22 is instrumental when that is the case. The derivation of the sphere packing bound presented in [73, §IV] is a case in point. The additivity of the Rényi capacity for products was first reported by Gallager —in a slightly different form and for finite $\mathcal{W}$ and $\mathcal{Y}$ case— in his seminal paper [35, Thm. 5], see also [37, pp. 149-150, (5.6.59)]. Later, Augustin proved [6, Lemma 26.7a], which implies Lemma 22; see [5, Lemma 3.6] for finite $\mathcal{W}$ case.

One curious question is whether or not one can give a class of priors for which the lower bound given in (40) is not too loose. Lemma 23 answers this question in the affirmative.

Lemma 23. *For any $\alpha \in (0,\infty]$, $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$ with finite $C_{\alpha,\mathcal{W}}$, and $\epsilon \geq 0$, let $\mathcal{W}_{\alpha,\epsilon}$ be*

$$\mathcal{W}_{\alpha,\epsilon} \triangleq \{w \in \mathcal{W} : D_\alpha(w \| q_{\alpha,\mathcal{W}}) \geq C_{\alpha,\mathcal{W}} - \epsilon\}. \quad (59)$$

Then for any $\epsilon > 0$, we have $C_{\alpha,\mathcal{W}_{\alpha,\epsilon}} = C_{\alpha,\mathcal{W}}$ and[23]

$$0 \leq C_{\alpha,\mathcal{W}} - I_\alpha(p;\mathcal{W}) - D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}}) \leq \epsilon \quad (60)$$

for all p in $\mathcal{P}(\mathcal{W}_{\alpha,\epsilon})$. Furthermore, if $\mathcal{W}$ is a finite set, then $C_{\alpha,\mathcal{W}_{\alpha,0}} = C_{\alpha,\mathcal{W}}$ and (60) holds for $\epsilon = 0$.

The main conclusion of Lemma 23 is the equality $C_{\alpha,\mathcal{W}_{\alpha,\epsilon}} = C_{\alpha,\mathcal{W}}$ for positive ϵ's. This is expected for a general $\mathcal{W}$ and evident, even for $\epsilon = 0$ case, for a finite $\mathcal{W}$ because of the existence of an optimal p in $\mathcal{P}(\mathcal{W})$ for finite $\mathcal{W}$'s. One might be tempted to assume the validity of the assertions for $\epsilon = 0$ case for arbitrary $\mathcal{W}$'s. This, however, is not true; see Example 4 for a $\mathcal{W}$ for which $C_{\alpha,\mathcal{W}} > 0$ and $C_{\alpha,\mathcal{W}_{\alpha,0}} = 0$. Thus finiteness of $\mathcal{W}$ is not a superficial hypothesis for extending the claims to $\epsilon = 0$ case.

In order to apply certain technical tools, we occasionally need a given set to be closed, convex, or compact. The observations presented in Lemma 24, given in the following, can be helpful in such situations. For example, if we can prove a statement about Rényi capacity assuming $\mathcal{W}$ to be convex, then we can assert that statement for non-convex $\mathcal{W}$'s using Lemma 24-(a). Furthermore, in certain situations, calculating the Rényi capacity might be easier for the convex hull or the closure of $\mathcal{W}$ when compared to $\mathcal{W}$ itself, see Example 4. Lemma 24-(a,b) is helpful in such situations. Note that Lemma 24-(a,b) when considered together with Lemma 21 imply the equality of the Rényi centers of $\mathcal{W}$, $\text{ch}\mathcal{W}$, and $\text{cl}\mathcal{W}$ whenever one of them exists.

Lemma 24. *Let $\mathcal{W}$ be a subset of $\mathcal{P}(\mathcal{Y})$.*

(a) $C_{\alpha,\text{ch}\mathcal{W}} = C_{\alpha,\mathcal{W}}$ for all $\alpha \in (0,\infty]$ where $\text{ch}\mathcal{W}$ is the convex hull of $\mathcal{W}$ given by $\text{ch}\mathcal{W} \triangleq \{\mu_{1,p} : p \in \mathcal{P}(\mathcal{W})\}$.

(b) $C_{\alpha,\text{cl}\mathcal{W}} = C_{\alpha,\mathcal{W}}$ for all $\alpha \in (0,\infty]$ where $\text{cl}\mathcal{W}$ is the closure of $\mathcal{W}$ in the topology of setwise convergence or a stronger topology on $\mathcal{P}(\mathcal{Y})$.

(c) If $C_{\eta,\mathcal{W}} < \infty$ for an $\eta \geq 1$, then $\{\mu_{\alpha,p} : \alpha \in [0,\eta], p \in \mathcal{P}(\mathcal{W})\}$ is uniformly absolutely continuous with respect to $q_{\eta,\mathcal{W}}$ and relatively compact in both the topology of setwise convergence and the weak topology.

(d) The following four statements are equivalent:[24]

   (i) $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} = 0$.

   (ii) $\exists \mu \in \mathcal{P}(\mathcal{Y})$ such that $\mathcal{W} \prec^{uni} \mu$.

   (iii) $\mathcal{W}$ has compact closure in the topology of setwise convergence.

   (iv) $\mathcal{W}$ has compact closure in the weak topology.

Each assertion of Lemma 24 is proved using Theorem 1 together with some other observations. The invariance of $C_{\alpha,\mathcal{W}}$ under the closure and the convexification operations on $\mathcal{W}$,

---

[22]The existence of a unique product measure is guaranteed for any finite collection of σ-finite measures by [30, Thm. 4.4.4] and for any countable collection of probability measures by [30, Thm. 8.2.2].

[23]For $\alpha = \infty$, (60) is valid for a broader class of p's in particular for all p's such that $(\sum_{w \in \mathcal{W}_{\alpha,\epsilon}} p(w)) > 0$.

[24]Augustin proves the equivalence of $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} = 0$ and $\exists \mu \in \mathcal{P}(\mathcal{Y})$ such that $\mathcal{W} \prec^{uni} \mu$, using Gallager's inner bound [35, Thm. 1] and a different characterization of the relative compactness he derives in [6]. Our proof is measure theoretic and self-contained.

presented in Lemma 24-(a,b), follow from the lower semicontinuity and the quasi-convexity of the Rényi divergence in its first argument. Lemma 24-(c) follows from the monotonicity of $\mu_{\alpha,p}$ in α and de la Vallée Poussin's characterization of the uniform integrability, i.e. [8, Thm. 4.5.9].

Arguably, the most interesting observation of Lemma 24 is the following: $\lim_{\alpha\uparrow 1} \frac{\alpha-1}{\alpha} C_{\alpha,\mathcal{W}} = 0$ iff there exists a μ in $\mathcal{P}(\mathcal{Y})$ satisfying $\mathcal{W}\prec^{uni}\mu$. This characterization is important because $\mathcal{W}$ is relatively compact, i.e. has a compact closure, in the topology of setwise convergence iff there exists a μ in $\mathcal{P}(\mathcal{Y})$ satisfying $\mathcal{W}\prec^{uni}\mu$ by [8, Thm. 4.7.25]. Since the topology of set wise convergence and the weak topology have exactly the same family of sets as their compact sets by [8, Thm. 4.7.25], the uniform absolute continuity also characterizes the relative compactness in the weak topology.

Remark 3. The weak topology on $\mathcal{M}(\mathcal{Y})$ is the topology generated by all continuous linear functions from $\mathcal{M}(\mathcal{Y})$ with the total variation topology to $\mathbb{R}$ with its usual topology. Then the weak topology is weaker than the total variation topology, i.e. the initial topology. On the other hand, the topology of setwise convergence on $\mathcal{M}(\mathcal{Y})$ is the topology generated by the functions $\{f_{\mathcal{E}} : \mathcal{E} \in \mathcal{Y}\}$ where $f_{\mathcal{E}}(\mu) = \int_{\mathcal{E}} \mu(\mathrm{d}y)$ for $\mathcal{E}$ in $\mathcal{Y}$ and μ in $\mathcal{M}(\mathcal{Y})$. Since $f_{\mathcal{E}} : \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ is a continuous linear function for any $\mathcal{E} \in \mathcal{Y}$, the weak topology is stronger than the topology of setwise convergence. Nevertheless, the weak topology and the topology of setwise convergence have exactly the same class of compact sets, [8, Thm. 4.7.25].

Our use of the term weak topology is consistent with the convention used in functional analysis, see [8, pp. 281,291]. While discussing the convergence of measures, however, the term weak topology is commonly used to describe another topology. If there is a topology on $\mathcal{Y}$ and $\mathcal{Y}$ is the resulting Baire $\sigma-$algebra [8, p. 12] of the subsets of $\mathcal{Y}$, then one can interpret the space of measures as a space of linear functionals on the space of continuous and bounded functions on $\mathcal{Y}$. The weak* topology on the space of measures in this setting is often called the weak topology [8, Def. 8.1.2]. Although it is a very important and useful concept in general, the weak topology in this second sense is not relevant in our discussion because we have not assumed any topological structure on $\mathcal{Y}$.

V. Examples

The order α Rényi entropy of a binary random variable, denoted $h_\alpha(\delta)$, allows us to write certain expressions succinctly in some of the examples. For any $\delta \in [0,1]$ it is defined as

$$h_\alpha(\delta) \triangleq \begin{cases} \frac{1}{1-\alpha} \ln(\delta^\alpha + (1-\delta)^\alpha) & \alpha \neq 1 \\ \delta \ln \frac{1}{\delta} + (1-\delta) \ln \frac{1}{1-\delta} & \alpha = 1 \end{cases}. \quad (61)$$

A. Probabilities on Discrete Output Spaces

For $\alpha \geq 1$, $q_{\alpha,\mathcal{W}} \perp q_{\alpha,\mathcal{U}}$ iff $\mathcal{W} \perp \mathcal{U}$. For $\alpha \in (0,1)$, $\mathcal{W} \perp \mathcal{U}$ implies $q_{\alpha,\mathcal{W}} \perp q_{\alpha,\mathcal{U}}$ but the converse is not true, i.e. $q_{\alpha,\mathcal{W}} \perp q_{\alpha,\mathcal{U}}$ can hold even when $\mathcal{W}$ and $\mathcal{U}$ are nonsingular. Examples 1 and 2 provide such sets of probability measures.

Example 1. For a $\delta \in (0, \frac{1}{5})$, let $\mathcal{W}$ and $\mathcal{U}$ be[25]

$$\mathcal{W} = \begin{bmatrix} 1-\delta & \delta & 0 & 0 \\ \delta & 1-\delta & 0 & 0 \\ \frac{1-\delta}{2} & \frac{1-\delta}{2} & \frac{\delta}{2} & \frac{\delta}{2} \end{bmatrix} \quad \mathcal{U} = \begin{bmatrix} 0 & 0 & 1-\delta & \delta \\ 0 & 0 & \delta & 1-\delta \end{bmatrix}$$

The third member of $\mathcal{W}$ is not singular with the members of $\mathcal{U}$; thus $\mathcal{W}$ is not singular with $\mathcal{U}$. We show in the following that $q_{\alpha,\mathcal{W}} \perp q_{\alpha,\mathcal{U}}$ for all $\alpha \in (0, f^{-1}(\frac{\delta}{1-\delta})]$ where $f^{-1} : [0, \frac{1}{4}] \to [0,1]$ is the inverse of the bijective decreasing function $f(x) \triangleq (2^{1-x} - 1)^{1/x}$.

For $\mathcal{U}$ and $p = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ we have

$$I_\alpha(p;\mathcal{U}) = \ln 2 - h_\alpha(\delta) \qquad q_{\alpha,p} = \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

where $h_\alpha(\delta)$ is defined in (61). On the other hand, both u's in $\mathcal{U}$ satisfy $D_\alpha(u \| q_{\alpha,p}) = \ln 2 - h_\alpha(\delta)$. Then (38) implies that $C_{\alpha,\mathcal{U}} = \ln 2 - h_\alpha(\delta)$ and $q_{\alpha,\mathcal{U}} = \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \end{bmatrix}$.

For $\mathcal{W}$ and $\widetilde{p} = \begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix}$ we have

$$I_\alpha(\widetilde{p};\mathcal{W}) = \ln 2 - h_\alpha(\delta) \qquad q_{\alpha,\widetilde{p}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

The first two w's in $\mathcal{W}$ satisfy $D_\alpha(w \| q_{\alpha,\widetilde{p}}) = \ln 2 - h_\alpha(\delta)$. The third one satisfy $D_\alpha(w \| q_{\alpha,\widetilde{p}}) \leq \ln 2 - h_\alpha(\delta)$ if and only if $\alpha \leq f^{-1}(\frac{\delta}{1-\delta})$. Consequently, (38) implies $C_{\alpha,\mathcal{W}} = \ln 2 - h_\alpha(\delta)$ and $q_{\alpha,\mathcal{W}} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \end{bmatrix}$ for all $\alpha \in (0, f^{-1}(\frac{\delta}{1-\delta})]$.

Example 2 provides sets of probability measures that are not even pairwise disjoint but they have singular Rényi centers for all orders in $(0,1)$. Example 2 also demonstrates the possible absence of an optimal prior for infinite sets of probability measures.

Example 2. Let $(\mathcal{Y}, \mathcal{Y})$ be $(\mathbb{Z}, 2^{\mathbb{Z}})$ and let $\mathcal{W}_\imath$ be

$$\mathcal{W}_\imath = \{w^{\imath,\jmath} : \jmath \in \mathbb{Z}\}. \qquad \forall \imath \in \mathbb{Z}$$

where $w^{\imath,\jmath}(y) = (\mathbb{1}_{\{y=\imath\}} + \mathbb{1}_{\{y=\jmath\}})/2$.

For any α in $(0,1)$ and sequence $\{p_\kappa\}_{\kappa\in\mathbb{Z}_+} \subset \mathcal{P}(\mathcal{W}_\imath)$ of uniform distributions with strictly increasing support q_{α,p_κ} converges to $\mathbb{1}_{\{\cdot=\imath\}}$ in the total variation topology and $\lim_{\kappa\to\infty} I_\alpha(p_\kappa;\mathcal{W}_\imath) = \frac{\alpha \ln 2}{1-\alpha}$. Furthermore, if $q(\cdot) = \mathbb{1}_{\{\cdot=\imath\}}$ then $D_\alpha(w \| q) \leq \frac{\alpha \ln 2}{1-\alpha}$ for all $w \in \mathcal{W}_\imath$. Thus $C_{\alpha,\mathcal{W}_\imath} = \frac{\alpha \ln 2}{1-\alpha}$ and $q_{\alpha,\mathcal{W}_\imath}(\cdot) = \mathbb{1}_{\{\cdot=\imath\}}$ for all $\alpha \in (0,1)$ by (39).

Note that $\mathcal{W}_\imath$'s are not singular with one another, in fact $\mathcal{W}_\imath \cap \mathcal{W}_\jmath = \{w^{\imath,\jmath}\}$. Nonetheless, $q_{\alpha,\mathcal{W}_\imath} \perp q_{\alpha,\mathcal{W}_\jmath}$ for all α in $(0,1)$ whenever $\imath \neq \jmath$ and we can use Lemma 21 to calculate the Rényi capacity of any finite union of $\mathcal{W}_\imath$'s. For any finite set of integers $\mathcal{A}$, let $\mathcal{W}_\mathcal{A}$ be $\mathcal{W}_\mathcal{A} = \cup_{\imath\in\mathcal{A}}\mathcal{W}_\imath$; then

$$C_{\alpha,\mathcal{W}_\mathcal{A}} = \frac{\alpha \ln 2}{1-\alpha} + \ln |\mathcal{A}| \quad \text{and} \quad q_{\alpha,\mathcal{W}_\mathcal{A}}(y) = |\mathcal{A}|^{-1} \mathbb{1}_{\{y\in\mathcal{A}\}}.$$

Furthermore, for any $p \in \mathcal{P}(\mathcal{W}_\mathcal{A})$ using (22) and (23) we get

$$D_\alpha(p \circledast \mathcal{W}_\mathcal{A} \| p \otimes q_{\alpha,\mathcal{W}_\mathcal{A}}) = I_\alpha(p;\mathcal{W}_\mathcal{A}) + D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}_\mathcal{A}}).$$

Recall that $D_\alpha(p \circledast \mathcal{W}_\mathcal{A} \| p \otimes q_{\alpha,\mathcal{W}_\mathcal{A}}) \leq C_{\alpha,\mathcal{W}_\mathcal{A}}$ by Theorem 1 and $D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}_\mathcal{A}}) \geq 0$ by Lemma 10. In addition

---

[25]When $\mathcal{W}$ and $\mathcal{Y}$ are finite sets and $\mathcal{Y} = 2^{\mathcal{Y}}$, it is customary to describe $\mathcal{W}$ using a matrix. Each row corresponds to an element of $\mathcal{W}$, each column corresponds to an element of $\mathcal{Y}$ and the element on the row w and the column y is equal to $w(y)$. With a slight abuse of notation we denote the resulting matrix by $\mathcal{W}$, as well.

- $D_\alpha(p \circledast \mathcal{W}_\mathcal{A} \| p \otimes q_{\alpha,\mathcal{W}_\mathcal{A}}) < C_{\alpha,\mathcal{W}_\mathcal{A}}$ for any $p \in \mathcal{P}(\mathcal{W}_\mathcal{A})$ satisfying $p(w) > 0$ a w such that $\text{supp}(w) \subset \mathcal{A}$.
- $D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}_\mathcal{A}}) > 0$ for any $p \in \mathcal{P}(\mathcal{W}_\mathcal{A})$ satisfying $p(w) > 0$ a w such that $\text{supp}(w) \not\subset \mathcal{A}$.

Thus $I_\alpha(p;\mathcal{W}_\mathcal{A}) < C_{\alpha,\mathcal{W}_\mathcal{A}}$ for any $p \in \mathcal{P}(\mathcal{W}_\mathcal{A})$ and finite $\mathcal{A}$.

In Example 1 the optimal p satisfying $I_\alpha(p;\mathcal{W}) = C_{\alpha,\mathcal{W}}$ was unique. However, this is not the case in general as demonstrated by Example 3, given in the following.

Example 3. For a $\delta \in [0, 1/2]$, let $\mathcal{W}$ be

$$\mathcal{W} = \begin{bmatrix} \delta & \delta & 1/2-\delta & 1/2-\delta \\ 1/2-\delta & 1/2-\delta & \delta & \delta \\ \delta & 1/2-\delta & 1/2-\delta & \delta \\ 1/2-\delta & \delta & \delta & 1/2-\delta \end{bmatrix}.$$

Let p_β be $\begin{bmatrix} \beta/2 & \beta/2 & (1-\beta)/2 & (1-\beta)/2 \end{bmatrix}$ for any $\beta \in [0,1]$. Then for all α in $\mathbb{R}_+$ and β in $[0,1]$ we have

$$I_\alpha(p_\beta;\mathcal{W}) = \ln 2 - h_\alpha(2\delta) \qquad q_{\alpha,p_\beta} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}.$$

Furthermore, $D_\alpha(w \| q_{\alpha,p_\beta}) = I_\alpha(p_\beta;\mathcal{W})$ for all w in $\mathcal{W}$. Thus $I_\alpha(p_\beta;\mathcal{W}) = C_{\alpha,\mathcal{W}}$ and $q_{\alpha,\mathcal{W}} = q_{\alpha,p_\beta}$ for all β in $[0,1]$ and α in $\mathbb{R}_+$ by (38).

We have demonstrated in Example 2 that for certain infinite $\mathcal{W}$'s $I_\alpha(p;\mathcal{W}) < C_{\alpha,\mathcal{W}}$ for all p in $\mathcal{P}(\mathcal{W})$. Example 4, given in the following, demonstrates that a stronger assertion "$D_\alpha(w \| q_{\alpha,\mathcal{W}}) < C_{\alpha,\mathcal{W}}$ for all w in $\mathcal{W}$" is true for certain infinite $\mathcal{W}$'s. Hence, the claims of Lemma 23 about $\mathcal{W}_{\alpha,\epsilon}$ cannot be extended to $\epsilon = 0$ case for infinite $\mathcal{W}$'s, because for the $\mathcal{W}$ given in Example 4 $C_{\alpha,\mathcal{W}} > 0$ and $\mathcal{W}_{\alpha,0} = \emptyset$.

Example 4. Let us assume $\gamma \in (0,1)$ and $n \in \mathbb{Z}_+$. Let $\mathcal{Y}$ be $\{0,\ldots,n\}$, $\mathcal{Y}$ be $2^\mathcal{Y}$, $\mathcal{U}$ and $\mathcal{W}$ be

$$\mathcal{U} = \{w^{\delta,\jmath} : \jmath \in \{1,\ldots,n\}, \delta \in [\gamma,1]\},$$
$$\mathcal{W} = \{w^{\delta,\jmath} : \jmath \in \{1,\ldots,n\}, \delta \in (\gamma,1]\}$$

where $w^{\delta,\jmath}(y) = \mathbb{1}_{\{y=\jmath\}}(1-\delta) + \mathbb{1}_{\{y=0\}}\delta$.
Let $p \in \mathcal{P}(\mathcal{U})$ be $p(w^{\delta,\jmath}) = \frac{1}{n}\mathbb{1}_{\{\delta=\gamma\}}$. Then

$$I_\alpha(p;\mathcal{U}) = \begin{cases} \frac{\alpha}{\alpha-1}\ln\left[\gamma + (1-\gamma)n^{\frac{\alpha-1}{\alpha}}\right] & \alpha \in \mathbb{R}_+ \setminus \{1\} \\ (1-\gamma)\ln n & \alpha = 1 \end{cases},$$

$$q_{\alpha,p}(y) = \frac{\gamma\mathbb{1}_{\{y=0\}}}{\gamma+(1-\gamma)n^{1-1/\alpha}} + \sum_{\jmath=1}^{n} \frac{(1-\gamma)n^{-1/\alpha}\mathbb{1}_{\{y=\jmath\}}}{\gamma+(1-\gamma)n^{1-1/\alpha}}.$$

Furthermore, one can confirm by substitution that

$$D_\alpha(w^{\delta,\jmath} \| q_{\alpha,p}) = \begin{cases} \frac{1}{1-\alpha}\ln\frac{(1+((1/\gamma)-1)n^{1-1/\alpha})^{1-\alpha}}{\delta^\alpha+(1-\delta)^\alpha((1/\gamma)-1)^{1-\alpha}n^{1-1/\alpha}} & \alpha \neq 1 \\ \delta\ln\frac{\delta}{\gamma} + (1-\delta)\ln\frac{(1-\delta)n}{1-\gamma} & \alpha = 1 \end{cases}.$$

Then $D_\alpha(w \| q_{\alpha,p}) \leq I_\alpha(p;\mathcal{U})$ for all $w \in \mathcal{U}$ and consequently, $C_{\alpha,\mathcal{U}} = I_\alpha(p;\mathcal{U})$ and $q_{\alpha,\mathcal{U}} = q_{\alpha,p}$ by (38).
Since $\mathcal{U}$ is the closure of $\mathcal{W}$ in the topology of setwise convergence, $C_{\alpha,\mathcal{W}} = C_{\alpha,\mathcal{U}}$ for all $\alpha \in \mathbb{R}_+$ by Lemma 24-(b). Consequently, $q_{\alpha,\mathcal{W}} = q_{\alpha,\mathcal{U}}$ by Lemma 21 because $\mathcal{W} \subset \mathcal{U}$.
Then $D_\alpha(w \| q_{\alpha,\mathcal{W}}) < C_{\alpha,\mathcal{W}}$ for all w in $\mathcal{W}$ and α in $\mathbb{R}_+$. Hence $I_\alpha(p;\mathcal{W}) < C_{\alpha,\mathcal{W}}$ for all p in $\mathcal{P}(\mathcal{W})$ and α in $\mathbb{R}_+$ by Lemma 14 and $\mathcal{W}_{(\alpha,0)} = \emptyset$ for all α in $\mathbb{R}_+$ by definition.

B. Shift Invariant Families of Probabilities

The shift invariant sets of probability measures on the unit interval are relatively easy to analyze. Nevertheless, when considered as a function of the order the Rényi capacities of these sets form a diverse collection and it is relatively easy to construct examples and counterexamples for the behavior of Rényi capacity as function of the order using this family.

First we consider the set of modular shifts of a probability measure on the unit interval, which is called "channel with additive noise on the unit circle" by Agustin in [6].

Example 5. Let $\mathcal{Y}$ be $[0,1)$, $\mathcal{Y}$ be $\mathcal{B}([0,1))$, and f be a non-negative Lebesgue measurable function such that $\int_0^1 f(y)\mathrm{d}y = 1$. Then $\mathcal{W}^{[f]}$ is the set of all probability measures whose Radon-Nikodym derivatives with respect to the Lebesgue measure l is a mod one shift of f:

$$\mathcal{W}^{[f]} \triangleq \left\{ w : \frac{\mathrm{d}w}{\mathrm{d}l} = f \circ \mathtt{T}_x \text{ for some } x \in [0,1) \right\} \qquad (62)$$

where $\mathtt{T}_x(y) \triangleq y - x - \lfloor y - x \rfloor$.
Let us denote the measure whose Radon-Nikodym derivative is f by w_f. Note that $D_\alpha(w \| l) = D_\alpha(w_f \| l)$ for any w in $\mathcal{W}^{[f]}$ and $\alpha \in (0,\infty)$. Thus $\sup_{w \in \mathcal{W}^{[f]}} D_\alpha(w \| l) = D_\alpha(w_f \| l)$ for any $\alpha \in (0,\infty)$.
If $C_{\alpha,\mathcal{W}^{[f]}}$ is finite for an $\alpha \in (0,\infty]$, then $\exists! q_{\alpha,\mathcal{W}^{[f]}}$ in $\mathcal{P}(\mathcal{Y})$ such that

$$D_\alpha(w \| q_{\alpha,\mathcal{W}^{[f]}}) \leq C_{\alpha,\mathcal{W}^{[f]}} \qquad \forall w \in \mathcal{W}^{[f]}$$

by Theorem 1. On the other hand $q_{\alpha,\mathcal{W}^{[f]}} = q_s + q_{ac}$ where $q_s \perp l$ and $q_{ac} \prec l$, by the Lebesgue decomposition theorem [30, 5.5.3]. Then $D_\alpha(w \| q_{\alpha,\mathcal{W}^{[f]}}) = D_\alpha(w \| q_{ac})$ for all w in $\mathcal{W}^{[f]}$ by (19) because $w \prec l$ for all w in $\mathcal{W}^{[f]}$. Thus

$$D_\alpha(w \| {}^{q_{ac}}/_{\|q_{ac}\|}) = C_{\alpha,\mathcal{W}^{[f]}} + \ln\|q_{ac}\| \qquad \forall w \in \mathcal{W}^{[f]}.$$

If $\|q_{ac}\| < 1$, then $\sup_{w \in \mathcal{W}^{[f]}} D_\alpha(w \| {}^{q_{ac}}/_{\|q_{ac}\|}) < C_{\alpha,\mathcal{W}^{[f]}}$. This, however, is impossible because of Theorem 1. Thus $\|q_{ac}\| = 1$, $\|q_s\| = 0$ and $q_{\alpha,\mathcal{W}^{[f]}} \prec l$.
Since $q_{\alpha,\mathcal{W}^{[f]}} \prec l$, the Radon-Nikodym derivative $\frac{\mathrm{d}q_{\alpha,\mathcal{W}^{[f]}}}{\mathrm{d}l}$ exists by the Radon-Nikodym theorem [30, 5.5.4]. Since $\mathcal{W}^{[f]}$ is invariant under mod one shifts by construction, its Rényi centers need to be invariant under mods one shift, as well. Furthermore, l is invariant under mod one shifts. Hence,

$$\frac{\mathrm{d}q_{\alpha,\mathcal{W}^{[f]}}}{\mathrm{d}l} = \frac{\mathrm{d}q_{\alpha,\mathcal{W}^{[f]}}}{\mathrm{d}l} \circ \mathtt{T}_x \qquad \forall x \in [0,1)$$

Thus $\frac{\mathrm{d}q_{\alpha,\mathcal{W}^{[f]}}}{\mathrm{d}l}$ needs to be a constant. That constant is one because $q_{\alpha,\mathcal{W}^{[f]}}$ is a probability measure. Therefore $q_{\alpha,\mathcal{W}^{[f]}} = l$ and $C_{\alpha,\mathcal{W}^{[f]}} = D_\alpha(w_f \| l)$ whenever $C_{\alpha,\mathcal{W}^{[f]}}$ is finite. When it is infinite so is $D_\alpha(w_f \| l)$ by Theorem 1 because $D_\alpha(w \| l)$ equals $D_\alpha(w_f \| l)$ for all w in $\mathcal{W}^{[f]}$. Hence, $C_{\alpha,\mathcal{W}^{[f]}} = D_\alpha(w_f \| l)$, i.e.

$$C_{\alpha,\mathcal{W}^{[f]}} = \begin{cases} \frac{1}{\alpha-1}\ln\int f^\alpha(y)\mathrm{d}y & \alpha \in \mathbb{R}_+ \neq 1 \\ \int f(y)\ln f(y)\mathrm{d}y & \alpha = 1 \\ \ln\operatorname{ess\,sup}_l f(y) & \alpha = \infty \end{cases}. \qquad (63)$$

(63) is derived using the Ergodic theorem in Appendix G.

As a result of Lemma 15, $C_{\alpha,\mathcal{W}}$ is either continuous in α on $(0,\infty]$ or continuous and bounded on $(0,\phi]$ and infinite on $(\phi,\infty]$ for an $\phi \in [1,\infty)$. The following two examples

19

are special cases of Example 5 which demonstrate that the Rényi capacity can become infinite for some orders larger than one while being continuous on $(0, \infty]$ and the Rényi capacity can have a discontinuity at any order in $[1, \infty)$.

Example 6. $f_\beta(y) = (1 - \beta)y^{-\beta}$ and $\beta \in (0, 1)$.

$$C_{\alpha, \mathcal{W}^{[f_\beta]}} = \begin{cases} \frac{\alpha \ln(1-\beta) - \ln(1-\alpha\beta)}{\alpha-1} & \alpha \in [0,1) \cup (1, \beta^{-1}) \\ \frac{\beta}{1-\beta} + \ln(1-\beta) & \alpha = 1 \\ \infty & \alpha \in [\beta^{-1}, \infty] \end{cases}$$

$C_{\alpha, \mathcal{W}^{[f_\beta]}}$ is continuous on $(0, \infty]$ and monotone increasing and finite on $(0, \beta^{-1})$.

Example 7. The existence of the discontinuity is related to the integrability of $f \ln f$ and f^α because $C_{\alpha, \mathcal{W}^{[f]}} = D_\alpha(w_f \| l)$.

- If $f(y) = 2\frac{\mathbb{1}_{\{0 < y < 1/e\}}}{y(\ln\frac{1}{y})^3}$, then $C_{1, \mathcal{W}^{[f]}} = \ln 2\sqrt{e}$ and $C_{\alpha, \mathcal{W}^{[f]}} = \infty$ for all α in $(1, \infty]$.

- If $f(y) = \frac{y^{-\frac{1}{\phi}}\mathbb{1}_{\{0<y<1/e\}}}{(\ln\frac{1}{y})\int_{1-\frac{1}{\phi}}^{\infty}\frac{e^{-z}}{z}\mathrm{d}z}$ for a ϕ in $(1, \infty)$, then $C_{\phi, \mathcal{W}^{[f]}} = \frac{\ln(\phi-1)}{1-\phi} - \frac{\phi}{\phi-1}\ln\int_{1-\frac{1}{\phi}}^{\infty}\frac{e^{-z}}{z}\mathrm{d}z$ and $C_{\alpha, \mathcal{W}^{[f]}} = \infty$ for all α in $(\phi, \infty]$.

In all of the examples we have considered thus far the Rényi capacity is not only continuous but also differentiable in the order. This, however, is not the case in general.

Example 8. Let $\mathcal{F}$ be a family of non-negative Lebesgue measurable functions such that $\int f \mathrm{d}y = 1$ for all $f \in \mathcal{F}$. Then $\mathcal{W}^{[\mathcal{F}]}$ is the set of all probability measures whose Radon-Nikodym derivative is a mod one shift of an f in $\mathcal{F}$:

$$\mathcal{W}^{[\mathcal{F}]} \triangleq \left\{ w : \tfrac{\mathrm{d}w}{\mathrm{d}l} = f \circ \mathrm{T}_x \text{ for some } x \in [0,1), \ f \in \mathcal{F} \right\} \quad (64)$$

where $\mathrm{T}_x(y) \triangleq y - x - \lfloor y - x \rfloor$.

Note that $\sup_{w \in \mathcal{W}^{[\mathcal{F}]}} D_\alpha(w \| l) = \sup_{f \in \mathcal{F}} D_\alpha(w_f \| l)$ because $\mathcal{W}^{[\mathcal{F}]} = \cup_{f \in \mathcal{F}} \mathcal{W}^{[f]}$ and $D_\alpha(w \| l) = D_\alpha(w_f \| l)$ for all w in $\mathcal{W}^{[f]}$. Thus $C_{\alpha, \mathcal{W}^{[\mathcal{F}]}} \leq \sup_{f \in \mathcal{F}} D_\alpha(w_f \| l)$ by Theorem 1. On the other hand, the reverse inequality follows from (63) and Lemma 21. Thus, $C_{\alpha, \mathcal{W}^{[\mathcal{F}]}} = \sup_{f \in \mathcal{F}} D_\alpha(w_f \| l)$, i.e.

$$C_{\alpha, \mathcal{W}^{[\mathcal{F}]}} = \begin{cases} \sup_{f \in \mathcal{F}} \frac{1}{\alpha-1} \ln \int f^\alpha(y) \mathrm{d}y & \alpha \in \mathbb{R}_+ \neq 1 \\ \sup_{f \in \mathcal{F}} \int f(y) \ln f(y) \mathrm{d}y & \alpha = 1 \\ \sup_{f \in \mathcal{F}} \ln \mathrm{ess} \sup_l f(y) & \alpha = \infty \end{cases} \quad (65)$$

If $\mathcal{F} = \{2y, \frac{1}{2\sqrt{y}}\}$, then $C_{\alpha, \mathcal{W}^{[\mathcal{F}]}}$ is not differentiable at $\alpha = \frac{1}{2}$.

C. Certain Families of Poisson Point Processes

The following examples demonstrate the generality of our framework by determining the Rényi capacity of various families of Poisson point processes with integrable intensity functions, on real line.[26] Some of these families have been considered before in the context of channel coding problems,

---

[26]The analysis we present in the following can be applied to the spatial Poisson processes defined on appropriately chosen subsets of the Euclidean space without any major modification. We restrict our analysis to the one dimensional case, because even the one dimensional case has a structure that is rich enough to demonstrate the generality of our framework.

---

such as the ones in (70) and (72) in the following (see [13], [28], [55], [107], [108]), others have not been considered before, such as the ones in (69), (71), and (91).

The Poisson point processes are, sometimes, formulated and analyzed via the characterization of the interarrival times without even mentioning the Radon-Nikodym derivatives, see [38, Ch. 2]. For many applications such an approach turns out to be sufficient; as a result, the Radon-Nikodym derivatives of Poisson point processes are not as well-known as one would expect. Considering this fact, we follow the approach of Burnashev and Kutoyants in [13] and start our discussion with a brief refresher on the Radon-Nikodym derivatives of the Poisson processes.

For any $T \in \mathbb{R}_+$, let $\mathcal{X}_T$ be the set of all nondecreasing, right-continuous, integer valued functions on $(0, T]$. The sample paths of Poisson point processes are members of $\mathcal{X}_T$. Furthermore, any Poisson point process with deterministic intensity function f can be represented by a unique probability measure on the measurable space $(\mathcal{Y}, \mathcal{Y})$ for $\mathcal{Y} = \mathcal{X}_T$ when $\mathcal{Y}$ is an appropriately chosen σ-algebra.[27]

For any sample path $y \in \mathcal{Y}$, we denote the jth arrival time by $\tau_j(y)$ and the number of arrivals up to and including time t by $N_t(y)$. The probability measure associated with a Poisson process with the intensity function f is denoted by w_f. The probability measure of the Poisson point process with constant intensity γ is denoted by ν_γ. If $\gamma = 1$, we also use ν to denote ν_γ, i.e. $\nu = \nu_1$.

For any non-negative integrable function f on $(0, T]$ the associated probability measures w_f is absolutely continuous with respect to ν and the Radon-Nikodym derivative $\frac{\mathrm{d}w_f}{\mathrm{d}\nu}$ is given by,[28] [13, (2.1)], [9, VI.6.T12, p187],

$$\frac{\mathrm{d}w_f}{\mathrm{d}\nu}(y) = \left(\prod_{\tau_j(y) \leq T} f(\tau_j(y))\right) e^{\int_0^T (1-f(t))\mathrm{d}t}. \quad (66)$$

For any non-negative measurable function g, the following expression for the expectation[29] follows from (66), [13, (2.2)]:

$$\int \left(\prod_{\tau_j(y) \leq T} g(\tau_j(y))\right) w_f(\mathrm{d}y) = e^{\int_0^T (g(t)-1)f(t)\mathrm{d}t}. \quad (67)$$

An immediate consequence of (66) and (67) is the following expression for the Rényi divergence between w_f and w_g for

---

[27]One choice of $\mathcal{Y}$ that works is the Borel σ-algebra for the topology generated by the Skorokhod metric s on $\mathcal{X}_T$, denoted by $\mathcal{B}(\mathcal{X}_T, s)$. In fact, $\mathcal{B}(\mathcal{X}_T, s)$ is rich enough to express the Poisson point processes whose intensity functions are not deterministic but Markovian, i.e. the intensity at any $t \in (0, T]$ depends on the previous arrivals. Kabanov's original work [55] considers such Poisson point processes, as well.

[28]$\left(\prod_{\tau_j(y) \leq T} f(\tau_j(y))\right)$ stands for 1 for y's that do not have any arrivals.

[29]In [13], Burnashev and Kutoyants express the identities given in (66) and (67) more succinctly and elegantly, as follows:

$$\frac{\mathrm{d}w_f}{\mathrm{d}\nu}(y) = e^{\int_0^T (\ln f(t))y(\mathrm{d}t) + \int_0^T (1-f(t))\mathrm{d}t},$$

$$\int e^{\int_0^T (\ln g(t))y(\mathrm{d}t)} w_f(\mathrm{d}y) = e^{\int_0^T (g(t)-1)f(t)\mathrm{d}t}.$$

In the expressions $\int_0^T (\ln f(t))y(\mathrm{d}t)$ and $\int_0^T (\ln g(t))y(\mathrm{d}t)$, the sample path y is interpreted as a measure that is equal to the sum of Dirac delta functions located at the arrival times of the sample path y.

integrable intensity functions f and g and positive real orders:

$$D_\alpha(w_f\| w_g) = \begin{cases} \int_0^T \left(\frac{f^\alpha g^{1-\alpha}-f}{\alpha-1} - f + g\right)\mathrm{d}t & \alpha \neq 1 \\ \int_0^T \left(f\ln\frac{f}{g} - f + g\right)\mathrm{d}t & \alpha = 1 \end{cases}. \quad (68)$$

For positive real orders other than one (68) follows from (66) and (67) by substitution, via the definition of the Rényi divergence. On the other hand, $D_1(w_f\| w_g) = \lim_{\alpha\uparrow 1} D_\alpha(w_f\| w_g)$ because the Rényi divergence is continuous in order on $[0,1]$ by Lemma 7. Then the expression for $D_1(w_f\| w_g)$ follows from the dominated convergence theorem [8, 2.8.1] and the expression for $D_\alpha(w_f\| w_g)$ for $\alpha \in (0,1)$ because $\frac{x^\alpha - x}{\alpha - 1} \uparrow x\ln x$ as $\alpha \uparrow 1$ for any $x \geq 0$.

Let us proceed with defining the set of Poisson point processes we will be investigating.

Definition 8. For any $T \in \mathbb{R}_+$ and intensity levels a, ϱ, b satisfying $0 \leq a \leq \varrho \leq b \leq \infty$, let $\Lambda^{T,a,b,\varrho}$, $\Lambda^{T,a,b,\leq\varrho}$, $\Lambda^{T,a,b,\geq\varrho}$, and $\Lambda^{T,a,b}$ be the set of all Poisson point processes with $[a,b]$ valued deterministic integrable intensity functions on $(0,T]$ with an average equal to ϱ, less than or equal to ϱ, greater than or equal to ϱ, and in $[a,b]$, respectively:

$$\Lambda^{T,a,b,\varrho} \triangleq \left\{w_f : a \leq f \leq b \text{ and } \int_0^T f(t)\mathrm{d}t = T\varrho\right\}, \quad (69)$$

$$\Lambda^{T,a,b,\leq\varrho} \triangleq \left\{w_f : a \leq f \leq b \text{ and } \int_0^T f(t)\mathrm{d}t \leq T\varrho\right\}, \quad (70)$$

$$\Lambda^{T,a,b,\geq\varrho} \triangleq \left\{w_f : a \leq f \leq b \text{ and } \int_0^T f(t)\mathrm{d}t \geq T\varrho\right\}, \quad (71)$$

$$\Lambda^{T,a,b} \triangleq \left\{w_f : a \leq f \leq b\right\}. \quad (72)$$

The convention proposed in Definition 8 allows us to refer to various families of Poisson point processes without confusion. However, explicitly stating the dependence on T, a, and b is not necessary whenever the values of T, a, and b are unambiguous. When this is the case we use Λ^ϱ for $\Lambda^{T,a,b,\varrho}$, $\Lambda^{\leq\varrho}$ for $\Lambda^{T,a,b,\leq\varrho}$, $\Lambda^{\geq\varrho}$ for $\Lambda^{T,a,b,\geq\varrho}$, and Λ for $\Lambda^{T,a,b}$.

In the following, we first determine the Rényi capacity and center of $\Lambda^{T,a,b,\varrho}$, and then use these expressions to calculate the Rényi capacity and center of families described in Definition 8 and in (91).

Example 9. For any $T \in \mathbb{R}_+$, $a,b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$, and $\varrho \in [a,b]$,

$$C_{\alpha,\Lambda^\varrho} = \begin{cases} \frac{\alpha}{\alpha-1}(\zeta_{\alpha,\varrho} - \varrho)T & \alpha \neq 1 \\ \left(\frac{\varrho-a}{b-a}b\ln\frac{b}{\varrho} + \frac{b-\varrho}{b-a}a\ln\frac{a}{\varrho}\right)T & \alpha = 1 \end{cases}, \quad (73)$$

$$q_{\alpha,\Lambda^\varrho} = \nu_{\zeta_{\alpha,\varrho}}, \quad (74)$$

$$\zeta_{\alpha,\varrho} \triangleq \left(\frac{\varrho-a}{b-a}b^\alpha + \frac{b-\varrho}{b-a}a^\alpha\right)^{1/\alpha}. \quad (75)$$

An alternative expression for $C_{\alpha,\Lambda^\varrho}$ is the following:

$$C_{\alpha,\Lambda^\varrho} = \frac{\varrho-a}{b-a}D_\alpha\left(\nu_b\| \nu_{\zeta_{\alpha,\varrho}}\right) + \frac{b-\varrho}{b-a}D_\alpha\left(\nu_a\| \nu_{\zeta_{\alpha,\varrho}}\right). \quad (76)$$

If ϱ is equal to a or b, then Λ^ϱ has just one element; consequently $C_{\alpha,\Lambda^\varrho}$ is zero and the only element of Λ^ϱ is also the Rényi center. For ϱ's in (a,b), we first determine the Rényi capacity and center assuming that $\frac{\varrho-a}{b-a}$ is a rational number by giving a sequence of priors $\{p_\imath\}$ and a probability measure q satisfying $\lim_{\imath\to\infty} I_\alpha(p_\imath;\Lambda^\varrho) = S_{\alpha,\Lambda^\varrho}(q)$. Then we determine the Rényi capacity of Λ^ϱ with irrational $\frac{\varrho-a}{b-a}$

using the continuity of the resulting expression in b and the monotonicity of $C_{\alpha,\mathcal{W}}$ in $\mathcal{W}$.

There exists positive integers ℓ and n such that $\frac{\varrho-a}{b-a} = \frac{\ell}{n}$ because $\frac{\varrho-a}{b-a}$ is a rational number and $b > \varrho > a$. Then there are $\binom{n}{\ell}$ length n sequences of a's and b's with ℓ b's and $(n-\ell)$ a's. These sequences will be the building blocks for f's with positive $p_\imath(w_f)$.

For each positive integer $\imath$ let us divide the interval $(0,T]$ into $2^\imath n$ half open intervals of the form $(\frac{\jmath-1}{2^\imath n}T, \frac{\jmath}{2^\imath n}T]$ for $\jmath \in \{1,\ldots,2^\imath n\}$. Now consider f's such that:

- f is $\{a,b\}$ valued function that is constant in all intervals of the form $(\frac{\jmath-1}{2^\imath n}T, \frac{\jmath}{2^\imath n}T]$ for $\jmath \in \{1,\ldots,2^\imath n\}$.
- $\ell = \sum_{\kappa=0}^{n-1} \mathbb{1}_{\{f(\frac{n\jmath-\kappa}{2^\imath n}T)=b\}}$ for all $\jmath \in \{1,\ldots,2^\imath\}$.

For every such f corresponding w_f is in Λ^ϱ. Furthermore, there are $\binom{n}{\ell}^{(2^\imath)}$ distinct f's. The prior $p_\imath$ has equal probability mass on all w_f's with the above described f's. Then using (66) we can calculate the Radon-Nikodym derivative of $\mu_{\alpha,p_\imath}$,

$$\frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}\nu}(y) = e^{(1-\varrho)T}\left(\sum_{\kappa=1}^{\binom{n}{\ell}^{(2^\imath)}} \frac{\left(\prod_{\tau_\jmath(y)\leq T} f_\kappa(\tau_\jmath(y))\right)^\alpha}{\binom{n}{\ell}^{(2^\imath)}}\right)^{1/\alpha}.$$

For the sample paths, i.e. y's, that do not have more than one arrival in any of the intervals of the form $(\frac{\jmath-1}{2^\imath}T, \frac{\jmath}{2^\imath}T]$, one can simplify the expression for the Radon-Nikodym derivative significantly. In particular,

$$\frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}\nu}(y) = e^{(1-\varrho)T}\left(\frac{\ell b^\alpha+(n-\ell)a^\alpha}{n}\right)^{\frac{N_T(y)-N_0(y)}{\alpha}} \quad \forall y \in \mathcal{E}_\imath$$

where $N_t(y)$ is the number of arrivals on $(0,t]$ for the sample path y and $\mathcal{E}_\imath \in \mathcal{Y}$ is defined as

$$\mathcal{E}_\imath \triangleq \left\{y : \left|N_{\frac{\jmath}{2^\imath}T}(y) - N_{\frac{\jmath-1}{2^\imath}T}(y)\right| \leq 1 \; \forall \jmath \in \{1,\ldots,2^\imath\}\right\}.$$

Since $\mathcal{E}_\imath \subset \mathcal{E}_{\imath+1}$ the following holds $\forall y \in \cup_{\imath\in\mathbb{Z}_+}\mathcal{E}_\imath$

$$\lim_{\imath\to\infty} \frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}\nu}(y) = e^{(1-\varrho)T}\left(\frac{\ell b^\alpha+(n-\ell)a^\alpha}{n}\right)^{\frac{N_T(y)-N_0(y)}{\alpha}}.$$

Using the complete independence of the Poisson processes on disjoint intervals and the probability mass function of the counting process, [38, Thm. 2.2.10], [9, II.1.(1.9), p. 22], we can calculate the probability $\nu(\mathcal{E}_\imath)$:

$$\nu(\mathcal{E}_\imath) = (e^{-\frac{T}{2^\imath}} + \frac{T}{2^\imath}e^{-\frac{T}{2^\imath}})^{(2^\imath)}$$
$$= e^{-T}(1 + \frac{T}{2^\imath})^{(2^\imath)}.$$

Then $\lim_{\imath\to\infty}\nu(\mathcal{E}_\imath) = 1$ and consequently $\nu(\cup_{\imath\in\mathbb{Z}_+}\mathcal{E}_\imath) = 1$. Thus convergence on $(\cup_{\imath\in\mathbb{Z}_+}\mathcal{E}_\imath)$ implies $\nu-$a.e. convergence:

$$\frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}\nu}(y) \xrightarrow{\nu-a.e.} e^{(1-\varrho)T}\left(\frac{\ell b^\alpha+(n-\ell)a^\alpha}{n}\right)^{\frac{N_T(y)-N_0(y)}{\alpha}}.$$

On the other hand $\frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}\nu}(y) \leq e^{(1-\varrho)T}\frac{b^{N_T(y)}}{b^{N_0(y)}}$ because $f(t) \leq b$. Furthermore $\int e^{(1-\varrho)T}\frac{b^{N_T}}{b^{N_0}}\nu(\mathrm{d}y) = e^{(b-\varrho)T}$. Thus the dominated convergence theorem [8, 2.8.1] implies that

$$\lim_{\imath\to\infty}\|\mu_{\alpha,p_\imath}\| = e^{(1-\varrho)T}\int\left(\frac{\ell b^\alpha+(n-\ell)a^\alpha}{n}\right)^{\frac{N_T(y)-N_0(y)}{\alpha}}\nu(\mathrm{d}y)$$
$$= e^{\left(\left(\frac{\ell b^\alpha+(n-\ell)a^\alpha}{n}\right)^{1/\alpha}-\varrho\right)T}.$$

Thus using (13) and the fact that $\frac{\varrho-a}{b-a} = \frac{\ell}{n}$ we get

$$\lim_{\imath\to\infty} I_\alpha(p_\imath;\Lambda^\varrho) = \frac{\alpha}{\alpha-1}\left(\zeta_{\alpha,\varrho} - \varrho\right) T \qquad \forall \alpha \neq 1. \qquad (77)$$

On the other hand for any $\gamma \in \mathbb{R}_+$ and $f : (0,T] \to [a,b]$ satisfying $\int_0^T f(t)\mathrm{d}t = T\varrho$, as a result of (68)

$$
\begin{aligned}
D_\alpha(w_f \| \nu_\gamma) &= \int_0^T \left[\frac{\gamma^{1-\alpha}}{\alpha-1}f^\alpha(t) - \frac{\alpha}{\alpha-1}f(t) + \gamma\right]\mathrm{d}t \\
&\leq \int_0^T \frac{\gamma^{1-\alpha}}{\alpha-1}\left[\frac{f(t)-a}{b-a}b^\alpha + \frac{b-f(t)}{b-a}a^\alpha\right]\mathrm{d}t - \frac{\alpha\varrho T}{\alpha-1} + \gamma T \\
&= \left[\frac{\gamma^{1-\alpha}}{\alpha-1}\left(\frac{\varrho-a}{b-a}b^\alpha + \frac{b-\varrho}{b-a}a^\alpha\right) - \frac{\alpha}{\alpha-1}\varrho + \gamma\right] T \\
&= \frac{\alpha}{\alpha-1}(\zeta_{\alpha,\varrho} - \varrho)T + D_\alpha(\nu_{\zeta_{\alpha,\varrho}} \| \nu_\gamma) \qquad (78)
\end{aligned}
$$

where the inequality follows from the convexity of the function $\frac{x^\alpha}{\alpha-1}$ in x and the Jensen's inequality.

Using (77) and (78) for $\gamma = \zeta_{\alpha,\varrho}$ we can conclude that $\lim_{\imath\to\infty} I_\alpha(p_\imath;\Lambda^\varrho) = S_{\alpha,\Lambda^\varrho}(\nu_{\zeta_{\alpha,\varrho}})$. Then for α's other than one (73) follows from (39) for values of ϱ making $\frac{\varrho-a}{b-a}$ a rational number. For values of ϱ making $\frac{\varrho-a}{b-a}$ a rational number, (73) for $\alpha = 1$ case follows from the expression for $\alpha \neq 1$ case via L'Hospital's rule [80, Thm. 5.13] because the Rényi capacity is a continuous function of the order on $(0,1]$ by Lemma 15-(c).

We now prove that (73) holds for values of ϱ for which $\frac{\varrho-a}{b-a}$ is irrational. First note that $\Lambda^{T,a,b_1,\varrho} \subset \Lambda^{T,a,b_2,\varrho}$ for any T, a, ϱ, b_1, b_2 satisfying $b_1 \leq b_2$, by the definition of $\Lambda^{T,a,b,\varrho}$ given in (69). Then $C_{\alpha,\Lambda^{T,a,b_1,\varrho}} \leq C_{\alpha,\Lambda^{T,a,b_2,\varrho}}$ by definition. Then (73) holds for the case when $\frac{\varrho-a}{b-a}$ is irrational as a result of the continuity of the expression on the right hand side of (73) as a function of b for each $\alpha \in \mathbb{R}_+$.

For orders other than one (74) follows from Theorem 1 because $S_{\alpha,\Lambda^\varrho}(\nu_{\zeta_{\alpha,\varrho}}) = C_{\alpha,\Lambda^\varrho}$ by (73) and (78). In order extend (74) to $\alpha = 1$ case we invoke the continuity of Rényi center established Lemma 20.

Example 10. For any $T \in \mathbb{R}_+$, $a,b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$, and $\varrho \in [a,b]$,

$$C_{\alpha,\Lambda^{\leq\varrho}} = C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}} \qquad C_{\alpha,\Lambda^{\geq\varrho}} = C_{\alpha,\Lambda^{\varrho\vee\varrho_\alpha}} \qquad (79)$$

$$q_{\alpha,\Lambda^{\leq\varrho}} = q_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}} \qquad q_{\alpha,\Lambda^{\geq\varrho}} = q_{\alpha,\Lambda^{\varrho\vee\varrho_\alpha}} \qquad (80)$$

where $C_{\alpha,\Lambda^\varrho}$ and $q_{\alpha,\Lambda^\varrho}$ are given in (73) and (74) and ϱ_α is defined as follows:

$$\varrho_\alpha \triangleq \begin{cases} \alpha^{\frac{\alpha}{1-\alpha}}\left(\frac{b-a}{b^\alpha-a^\alpha}\right)^{\frac{1}{1-\alpha}} + \frac{ab^\alpha-ba^\alpha}{b^\alpha-a^\alpha} & \alpha \neq 1 \\ e^{-1}b^{\frac{b}{b-a}}a^{-\frac{a}{b-a}} & \alpha = 1 \end{cases}. \qquad (81)$$

Since $\Lambda^{\leq\varrho}$ is the union of Λ^γ for γ in $[a,\varrho]$, $C_{\alpha,\Lambda^{\leq\varrho}}$ equals $C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}$ iff $S_{\alpha,\Lambda^{\leq\varrho}}(q_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}) \leq C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}$ by Lemma 21.

On the other hand when considered together with the convexity of $\frac{x^\alpha-x}{\alpha-1}$ in x for $\alpha \neq 1$ case and the convexity of $x\ln x$ in x for $\alpha = 1$ case, (68) implies

$$D_\alpha(w \| \nu_s) \leq \frac{b-\gamma}{b-a}D_\alpha(\nu_a \| \nu_s) + \frac{\gamma-a}{b-a}D_\alpha(\nu_b \| \nu_s) \qquad (82)$$

for all w in Λ^γ. Furthermore, the definitions of $\zeta_{\alpha,\gamma}$ and ϱ_α given in (75) and (81) imply that

$$D_\alpha(\nu_a \| \nu_{\zeta_{\alpha,\gamma}}) \leq D_\alpha(\nu_b \| \nu_{\zeta_{\alpha,\gamma}}) \qquad \forall \gamma \in [a,\varrho_\alpha]. \qquad (83)$$

Using (82) and (83) together with the alternative expression for $C_{\alpha,\Lambda^\varrho}$ given in (76) we get

$$
\begin{aligned}
S_{\alpha,\Lambda^{\leq\varrho}}(\nu_{\zeta_{\alpha,\varrho\wedge\varrho_\alpha}}) &\leq \frac{b-\varrho\wedge\varrho_\alpha}{b-a}D_\alpha(\nu_a \| \nu_{\zeta_{\alpha,\varrho\wedge\varrho_\alpha}}) \\
&\qquad + \frac{\varrho\wedge\varrho_\alpha-a}{b-a}D_\alpha(\nu_b \| \nu_{\zeta_{\alpha,\varrho\wedge\varrho_\alpha}}) \\
&= C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}.
\end{aligned}
$$

Thus $S_{\alpha,\Lambda^{\leq\varrho}}(q_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}) \leq C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}$ follows from (74). Hence $C_{\alpha,\Lambda^{\leq\varrho}} = C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}$ and $q_{\alpha,\Lambda^{\leq\varrho}} = q_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}$ by Lemma 21.

Assertions about $\Lambda^{\geq\varrho}$ derived similarly using the following observations: $\Lambda^{\geq\varrho}$ is the union of Λ^γ for γ in $[\varrho,b]$ and

$$D_\alpha(\nu_a \| \nu_{\zeta_{\alpha,\gamma}}) \geq D_\alpha(\nu_b \| \nu_{\zeta_{\alpha,\gamma}}) \qquad \forall \gamma \in [\varrho_\alpha,b]. \qquad (84)$$

Example 11. For any $T \in \mathbb{R}_+$ and $a,b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$,

$$C_{\alpha,\Lambda} = C_{\alpha,\Lambda^\varrho}, \qquad (85)$$

$$q_{\alpha,\Lambda} = q_{\alpha,\Lambda^\varrho}, \qquad (86)$$

where $C_{\alpha,\Lambda^\varrho}$, $q_{\alpha,\Lambda^\varrho}$, ϱ_α are described in (73), (74), (81) because $\Lambda = \Lambda^{\leq b}$. By substitution we get the following more explicitly expressions:

$$C_{\alpha,\Lambda} = \begin{cases} \left(\left(\frac{\alpha(b-a)}{b^\alpha-a^\alpha}\right)^{\frac{1}{1-\alpha}} - \frac{\alpha}{\alpha-1}\frac{ab^\alpha-ba^\alpha}{b^\alpha-a^\alpha}\right)T & \alpha\neq 1 \\ \left(e^{-1}b^{\frac{b}{b-a}}a^{-\frac{a}{b-a}} - \frac{ab}{b-a}\ln\frac{b}{a}\right)T & \alpha=1 \end{cases}, \qquad (87)$$

$$q_{\alpha,\Lambda} = \nu_{\zeta_\alpha}, \qquad (88)$$

$$\zeta_\alpha \triangleq \begin{cases} \alpha^{\frac{1}{1-\alpha}}\left(\frac{b-a}{b^\alpha-a^\alpha}\right)^{\frac{1}{1-\alpha}} & \alpha\in\mathbb{R}_+\setminus\{1\} \\ e^{-1}b^{\frac{b}{b-a}}a^{-\frac{a}{b-a}} & \alpha=1 \end{cases}. \qquad (89)$$

One can also confirm $\zeta_\alpha = \zeta_{\alpha,\varrho_\alpha}$ using (75) and (81). The following expression for $C_{\alpha,\Lambda}$ is equivalent to (87):

$$C_{\alpha,\Lambda} = \frac{\varrho_\alpha-a}{b-a}D_\alpha(\nu_b \| \nu_{\zeta_\alpha}) + \frac{b-\varrho_\alpha}{b-a}D_\alpha(\nu_a \| \nu_{\zeta_\alpha}). \qquad (90)$$

In the preceding examples, we have assumed the intensity functions are bounded above by a constant; we replace this constant with an integrable function in Example 12 given in the following. Let us first give a formal definition.

Definition 9. For any T in $\mathbb{R}_+$, a in $\mathbb{R}_{\geq 0}$, and Lebesgue integrable function g on $(0,T]$ satisfying $g \geq a$, $\Lambda^{T,a,g(\cdot)}$ is the set of all Poisson point processes with deterministic intensity functions f satisfying $a \leq f \leq g$:

$$\Lambda^{T,a,g(\cdot)} \triangleq \{w_f : a \leq f(t) \leq g(t) \ \forall t \in (0,T]\}. \qquad (91)$$

Example 12. For any $T \in \mathbb{R}_+$, $a \in \mathbb{R}_{\geq 0}$, and $g \in \mathcal{L}^1(l)$ satisfying $g(t) \geq a$ for all t in $(0,T]$ we have

$$C_{\alpha,\Lambda} = \begin{cases} \int_0^T\left[\left(\frac{\alpha(g-a)}{g^\alpha-a^\alpha}\right)^{\frac{1}{1-\alpha}} - \frac{\alpha}{\alpha-1}\frac{ag^\alpha-ga^\alpha}{g^\alpha-a^\alpha}\right]\mathrm{d}t & \alpha\neq 1 \\ \int_0^T\left[e^{-1}g^{\frac{g}{g-a}}a^{-\frac{a}{g-a}} - \frac{ag}{g-a}\ln\frac{g}{a}\right]\mathrm{d}t & \alpha=1 \end{cases}, \qquad (92)$$

$$q_{\alpha,\Lambda} = w_{\zeta_\alpha}, \qquad (93)$$

$$\zeta_\alpha(t) \triangleq \begin{cases} \alpha^{\frac{1}{1-\alpha}}\left(\frac{g(t)-a}{g^\alpha(t)-a^\alpha}\right)^{\frac{1}{1-\alpha}} & \alpha\neq 1 \\ e^{-1}[g(t)]^{\frac{g(t)}{g(t)-a}}a^{-\frac{a}{g(t)-a}} & \alpha=1 \end{cases}. \qquad (94)$$

If g is a simple function, then we can apply (87) and (88) for each possible value of g, separately. Then (92) and (93) follow from Lemma 22 because simple functions can only take finite

number of distinct values.[30] On the other hand, there exists a sequence of simple functions $\{g^{(\imath)}\}_{\imath\in\mathbb{Z}_+}$ satisfying $a \leq g^{(\imath)}$ and $g^{(\imath)} \uparrow g$ for any measurable g. Evidently $C_{\alpha,\Lambda^{T,a,g^{(\imath)}(\cdot)}} \leq C_{\alpha,\Lambda^{T,a,g(\cdot)}}$ because $\Lambda^{T,a,g^{(\imath)}(\cdot)} \subset \Lambda^{T,a,g(\cdot)}$. Furthermore, if f satisfies $a \leq f \leq g$, then

$$\left| \alpha^{\frac{1}{1-\alpha}} \left(\frac{f(t)-a}{f^\alpha(t)-a^\alpha} \right)^{\frac{1}{1-\alpha}} - \frac{\alpha}{\alpha-1} \frac{af^\alpha(t)-f(t)a^\alpha}{f^\alpha(t)-a^\alpha} \right| \leq \alpha^{\frac{1}{1-\alpha}} g(t).$$

for all $t \in (0, T]$.

Then the integral on the right hand side of (92) equals to $\lim_{\imath\to\infty} C_{\alpha,\Lambda^{T,a,g^{(\imath)}(\cdot)}}$ by the dominated convergence theorem [8, 2.8.1]. Hence, it is a lower bound on $C_{\alpha,\Lambda^{T,a,g(\cdot)}}$. It is, also, an upper bound on $C_{\alpha,\Lambda^{T,a,g(\cdot)}}$ by (33) because $D_\alpha(w \| w_{\zeta_\alpha})$ is bounded from above by it for all $w \in \Lambda^{T,a,g(\cdot)}$. Then (93) follows from the uniqueness of the Rényi center.

VI. DISCUSSION

In this paper, we define and analyze the order α Rényi capacity $C_{\alpha,\mathcal{W}}$ and the order α Rényi radius $S_{\alpha,\mathcal{W}}$ for an arbitrary set of probability measures $\mathcal{W}$ on an arbitrary measurable space. Our most important contributions are proving the van Erven-Harremoës conjecture, i.e. Lemma 19, and two uniform equicontinuity results on the Rényi information, i.e. Lemma 16-(e,f). We also prove a minimax theorem, i.e. Theorem 1, which has been previously reported by Augustin in [6] in a different form and for orders between zero and two. Theorem 1 establishes not only the equality of $C_{\alpha,\mathcal{W}}$ and $S_{\alpha,\mathcal{W}}$ for any α and $\mathcal{W}$ but also the existence of a unique order α Rényi center whenever $C_{\alpha,\mathcal{W}}$ is finite. Our analysis leads to certain immediate consequences for two generalizations: $C_{\alpha,\mathcal{W},\mathcal{A}}$ defined for $\mathcal{A} \subset \mathcal{P}(\mathcal{W})$ and $C_{\alpha,W}$ defined for transition probability W. We introduce those generalizations formally and discuss the implications of our analysis on them in Appendices A and B.

Results of our analysis, also, encourage one to consider certain related problems:

- We do not assume any topological structure on the output space $\mathcal{Y}$. Although this is a strength because of the generality of our results, it is also a weakness because of the obliviousness of our analysis towards the interactions between Rényi's information measures and the topological structure of the output space. In almost all of the applications, $\mathcal{Y}$ is a Borel or Baire σ-algebra of the topological space $(\mathcal{Y}, \tau)$; usually there is an even more specific structure. In most of the applications, $\mathcal{Y}$ is the Borel σ-algebra of a complete separable metric space $(\mathcal{Y}, d)$. Thus one can define metrics other than the total variation metric on $\mathcal{W}$ and $\mathcal{P}(\mathcal{W})$ using the metric d and analyze the behavior of Rényi's information measures on the resulting topologies. Such models have already been considered in the context of the arbitrarily varying channels [24], [99] and the typicality [54], [64], [76].
- It is easy to confirm that continuity of the order α Rényi capacity as a function of the order α implies the

continuity of the corresponding f-capacity $\mathcal{C}_{f_\alpha,\mathcal{W}}$ as a function of α where $f_\alpha(x) = \frac{x^\alpha-1}{\alpha-1}$. The existence of similar, but more general, continuity results for richer classes of f-divergences with appropriate topologies is expected. What is plausible, but not evident, to us is the existence of a topology on the set of all convex f's that ensures the continuity of the corresponding f-capacities in f for all $\mathcal{W}$ on the region that f-capacities are finite. The interaction of topologies on the space of convex functions and corresponding f-capacities seems to be a fertile subject of inquiry.

- We use the definition of the Rényi information proposed by Sibson [94]. In (26) we provide the expression for the alternative definition of the Rényi information proposed by Augustin [6] and Csiszár [25]. We call this quantity the Augustin information. Theorems 1, 2, 3, and many of the other propositions have their analogues for the Augustin information, see [69], [70]. The Augustin capacity and center are of interest to us because they are better suited than the Rényi capacity and center for deriving the sphere packing bound for memoryless channels, see [69], [71].

We have avoided using information theoretic concepts such as code, channel, or rate in our discussion because we believe Rényi's information measures can and should be defined and understood on their own as measure theoretic concepts first. Rényi's information measures, however, do have operational meaning in various information transmission problems. We discuss the case of channel coding problem in [73].

APPENDIX

A. The Constrained Rényi Capacity

Definition 10. For any $\alpha \in [0, \infty]$, $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$, $\mathcal{A} \subset \mathcal{P}(\mathcal{W})$, the order α Rényi capacity of $\mathcal{W}$ for constraint set $\mathcal{A}$ is

$$C_{\alpha,\mathcal{W},\mathcal{A}} \triangleq \sup_{p\in\mathcal{A}} I_\alpha(p;\mathcal{W}). \qquad (A.1)$$

Note that $C_{\alpha,\mathcal{W},\mathcal{P}(\mathcal{W})} = C_{\alpha,\mathcal{W}}$ and $C_{\alpha,\mathcal{W},\{p\}} = I_\alpha(p;\mathcal{W})$ for any $\mathcal{W}$ and $p \in \mathcal{P}(\mathcal{W})$. Furthermore, the proof of Theorem 1 works as is for any convex $\mathcal{A}$ subset of $\mathcal{P}(\mathcal{W})$, not just $\mathcal{P}(\mathcal{W})$ itself. Thus the minimax theorem continues to hold for $C_{\alpha,\mathcal{W},\mathcal{A}}$; the alternative expression for $C_{\alpha,\mathcal{W},\mathcal{A}}$ is, however, no longer (guaranteed to be) equal to the Rényi radius.

Theorem 2. For any $\alpha \in (0, \infty]$, $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$, and convex $\mathcal{A} \subset \mathcal{P}(\mathcal{W})$,

$$C_{\alpha,\mathcal{W},\mathcal{A}} = \sup_{p\in\mathcal{A}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q) \qquad (A.2)$$

$$= \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{A}} D_\alpha(p \circledast \mathcal{W} \| p \otimes q). \qquad (A.3)$$

If $C_{\alpha,\mathcal{W},\mathcal{A}} < \infty$, then there exists a unique $q_{\alpha,\mathcal{W},\mathcal{A}}$ in $\mathcal{P}(\mathcal{Y})$, called the order α Rényi center for constraint set $\mathcal{A}$, such that

$$C_{\alpha,\mathcal{W},\mathcal{A}} = \sup_{p\in\mathcal{A}} D_\alpha(p \circledast \mathcal{W} \| p \otimes q_{\alpha,\mathcal{W},\mathcal{A}}). \qquad (A.4)$$

Furthermore, for every sequence of priors $\{p_\imath\}_{\imath\in\mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{\imath\to\infty} I_\alpha(p_\imath;\mathcal{W}) = C_{\alpha,\mathcal{W},\mathcal{A}}$, the corresponding sequence of order α Rényi means $\{q_{\alpha,p_\imath}\}_{\imath\in\mathbb{Z}_+}$ is a Cauchy sequence for the total variation metric on $\mathcal{P}(\mathcal{Y})$ and $q_{\alpha,\mathcal{W},\mathcal{A}}$ is the unique limit point of that Cauchy sequence.

---

[30] We are not overlooking the issue of contiguity for the inverse of the image of g because Example 11 holds as is for Poisson processes defined on any measurable set of Lebesgue measure T, not just the interval $(0, T]$.

A similar modification is needed for the van Erven-Harremoës bound, i.e. for Lemma 19, as well.

Lemma 25. *For any $\alpha \in (0, \infty]$, $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$, convex $\mathcal{A} \subset \mathcal{P}(\mathcal{W})$ satisfying $C_{\alpha, \mathcal{W}, \mathcal{A}} < \infty$, and $q \in \mathcal{P}(\mathcal{Y})$*

$$C_{\alpha, \mathcal{W}, \mathcal{A}} + D_\alpha(q_{\alpha, \mathcal{W}, \mathcal{A}} \| q) \le \sup_{p \in \mathcal{A}} D_\alpha(p \circledast \mathcal{W} \| p \otimes q).$$

Lemma 20 establishing the continuity of the Rényi centers in the order holds for the constrained Rényi centers. We prove it using Theorem 2 and Lemma 25 instead of Theorem 1 and Lemma 19.

B. The Rényi Capacity of Transition Probabilities

We have defined the order α Rényi information $I_\alpha(p; \mathcal{W})$ for any p.m.f. p on a set of probability measures $\mathcal{W}$. We show in the following —using the concept of transition probability and the expression for $I_\alpha(p; \mathcal{W})$ given in (24)— that for appropriately chosen σ-algebra $\mathcal{W}$, one can extend the definition of $I_\alpha(p; \mathcal{W})$ to p's that are probability measures on $(\mathcal{W}, \mathcal{W})$. Furthermore, we show that if $\mathcal{W}$ is countably separated,[31] then Theorem 1 holds for this more general case, see Theorem 3.

Definition 11. Let $(\mathcal{X}, \mathcal{X})$ and $(\mathcal{Y}, \mathcal{Y})$ be measurable spaces. Then a function $W : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ is called a transition probability (a stochastic kernel / a Markov kernel) from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$ if it satisfies the following two conditions:

  (i) For all $x \in \mathcal{X}$, the function $W(\cdot | x) : \mathcal{Y} \to [0, 1]$ is a probability measure on $(\mathcal{Y}, \mathcal{Y})$.

  (ii) For all $\mathcal{E} \in \mathcal{Y}$, the function $W(\mathcal{E} | \cdot) : \mathcal{X} \to [0, 1]$ is a $\mathcal{X}$-measurable function.

By [8, Thm. 10.7.2.], for any transition probability W and probability measure p on $(\mathcal{X}, \mathcal{X})$ there exists a unique probability measure $p \circledast W$ on $(\mathcal{X} \times \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$ satisfying

$$p \circledast W(\mathcal{E}_x \times \mathcal{E}_y) = \int_{\mathcal{E}_x} W(\mathcal{E}_y | x) p(\mathrm{d}x)$$

for all $\mathcal{E}_x \in \mathcal{X}$ and $\mathcal{E}_y \in \mathcal{Y}$. Now, we can define the order α Rényi information for p on the transition probability W.

Definition 12. For any $\alpha \in [0, \infty]$, transition probability W from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$, *the order α Rényi information for prior p is defined as*

$$I_\alpha(p; W) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast W \| p \otimes q). \quad (\mathrm{A}.5)$$

Definitions 4 and 12 are equivalent because of Lemma 14. Using the definition of $I_\alpha(p; W)$ we can define the order α Rényi capacity of a transition probability W.

Definition 13. For any $\alpha \in [0, \infty]$ and transition probability W from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$, *the order α Rényi capacity is*

$$C_{\alpha, W} \triangleq \sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha(p; W). \quad (\mathrm{A}.6)$$

[31] A σ-algebra $\mathcal{X}$ on $\mathcal{X}$ is countably separated, [8, Def. 6.5.1], if there exists an at most countable collection sets $\{\mathcal{E}_i\} \subset \mathcal{X}$ separating the points of $\mathcal{X}$. A collection $\{\mathcal{E}_i\}$ of subsets of $\mathcal{X}$ is said to be separating the points of $\mathcal{X}$, if for every pair of distinct points z and x in $\mathcal{X}$ there exists an $\mathcal{E}_i$ which includes only one of z and x. The Borel σ-algebra of any separable metric space is countably separated. The Borel σ-algebra of any separable metric space is also countably generated, i.e. it is the minimum σ-algebra of a countable collection of sets.

The analysis of the Rényi capacity for an arbitrary transition probability W is beyond the scope of this paper. However, if the σ-algebra $\mathcal{X}$ is countably separated, then we can use Theorem 1 to show that $C_{\alpha, W} = C_{\alpha, \mathcal{W}}$ for a $\mathcal{W} \subset \mathcal{P}(\mathcal{Y})$.

Theorem 3. *For any $\alpha \in (0, \infty]$ and transition probability W from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$ for a countably separated σ-algebra $\mathcal{X}$*

$$C_{\alpha, W} = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast W \| p \otimes q) \quad (\mathrm{A}.7)$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha(p \circledast W \| p \otimes q) \quad (\mathrm{A}.8)$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{w \in \mathcal{W}} D_\alpha(w \| q) \quad (\mathrm{A}.9)$$

where $\mathcal{W} \triangleq \{W(\cdot | x) : x \in \mathcal{X}\}$. If $C_{\alpha, W} < \infty$, then there exists a unique $q_{\alpha, W}$ in $\mathcal{P}(\mathcal{Y})$, called the order α Rényi center, such that

$$C_{\alpha, W} = \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha(p \circledast W \| p \otimes q_{\alpha, W}) \quad (\mathrm{A}.10)$$

$$= \sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha, W}). \quad (\mathrm{A}.11)$$

Proof of Theorem 3. Since $\mathcal{X}$ is countably separated, all singletons are in $\mathcal{X}$ by [8, Thm. 6.5.7] and $\mathcal{P}(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$. Consequently, using max-min inequality we get

$$\sup_{p \in \mathcal{P}(\mathcal{W})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast \mathcal{W} \| p \otimes q)$$
$$\le \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast W \| p \otimes q)$$
$$\le \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha(p \circledast W \| p \otimes q) \quad (\mathrm{A}.12)$$

On the other hand, for any $\alpha \in (0, \infty]$ as a result of Tonelli-Fubini theorem [30, 4.4.5] and the definition of the Rényi divergence given in (19) we have

$$D_\alpha(p \circledast W \| p \otimes q) \le \sup_{x \in \mathcal{X}} D_\alpha(W(\cdot | x) \| q)$$
$$= \sup_{w \in \mathcal{W}} D_\alpha(w \| q). \quad (\mathrm{A}.13)$$

Hence,

$$\inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha(p \circledast W \| p \otimes q)$$
$$\le \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{w \in \mathcal{W}} D_\alpha(w \| q). \quad (\mathrm{A}.14)$$

Theorem 1 and the inequalities given in (A.12), (A.13), and (A.14) imply $C_{\alpha, W} = C_{\alpha, \mathcal{W}}$ and Theorem 3 for $q_{\alpha, W} = q_{\alpha, \mathcal{W}}$. $\square$

Theorem 1 and (A.13) imply that $C_{\alpha, W} \le C_{\alpha, \mathcal{W}}$ even when $\mathcal{X}$ is not countably separated.

C. Deferred Proofs

The following parametric function allows us to write certain expressions succinctly in the proofs:

$$d(x \| z) \triangleq x \ln \tfrac{x}{z} + (1 - x) \ln \tfrac{1-x}{1-z} \quad \forall x, z \in [0, 1]. \quad (\mathrm{A}.15)$$

Proof of Lemma 16.

  (a) For any $\alpha \in (0, 1)$ the definitions of $I_\alpha(p; \mathcal{W})$ and $C_{\alpha, \mathcal{W}}$ imply $\inf_{p \in \mathcal{P}(\mathcal{W})} \|\mu_{\alpha, p}\| = e^{\frac{\alpha}{\alpha-1} C_{\alpha, \mathcal{W}}}$.

  (b) $\sup_{p \in \mathcal{P}(\mathcal{W})} \|\mu_{\alpha, p}\| = e^{\frac{\alpha}{\alpha-1} C_{\alpha, \mathcal{W}}}$ for any $\alpha \in (1, \infty)$ and $\sup_{p \in \mathcal{P}(\mathcal{W})} \|\mu_{\infty, p}\| = e^{C_{\infty, \mathcal{W}}}$ by the definitions of $I_\alpha(p; \mathcal{W})$ and $C_{\alpha, \mathcal{W}}$.

  (c) Let us first prove that if $C_{\alpha, \mathcal{W}} < \infty$, then $\mu_{\alpha, p}$ is uniformly continuous in p. Lemma 4-(e) and the triangle inequality imply

$$\|\mu_{\alpha, p_1} - \mu_{\alpha, p_2}\| \le \|p_1 - p_2\|^{\frac{1}{\alpha}} 2^{\frac{\alpha-1}{\alpha}} (\|\mu_{\alpha, s_1}\| \vee \|\mu_{\alpha, s_2}\|).$$

for all α in $[1,\infty)$ and p_1, p_2 in $\mathcal{P}(\mathcal{W})$ where s_1 and s_2 are members of $\mathcal{P}(\mathcal{W})$ determined by the decomposition given in Lemma 4-(c).

On the other hand $\|\mu_{\alpha,s}\| \leq e^{\frac{\alpha-1}{\alpha}C_{\alpha,\mathcal{W}}}$ for any s in $\mathcal{P}(\mathcal{W})$ by the proof of part (b). Thus

$$\|\mu_{\alpha,p_1} - \mu_{\alpha,p_2}\| \leq \|p_1 - p_2\|^{\frac{1}{\alpha}} e^{\frac{\alpha-1}{\alpha}(C_{\alpha,\mathcal{W}}+\ln 2)}$$

for all α in $[1,\infty)$ and p_1, p_2 in $\mathcal{P}(\mathcal{W})$ Consequently, if $C_{\alpha,\mathcal{W}}$ is finite, then $\mu_{\alpha,p}$ is uniformly continuous in p. We are left with proving that $C_{\alpha,\mathcal{W}} = \infty$ implies the absence of uniformly continuity in p for $\mu_{\alpha,p}$. For any s in $\mathcal{P}(\mathcal{W})$ let $s_\imath$ be

$$s_\imath = (1 - \tfrac{1}{\imath})s + \tfrac{1}{\imath}p_\imath \qquad \forall \imath \in \mathbb{Z}_+$$

where $p_\imath$'s are such that $\|\mu_{\alpha,p_\imath}\| \geq \imath$. The existence of such $p_\imath$'s follows from part (b). Then $\frac{\mu_{\alpha,p_\imath}}{\imath^{1/\alpha}} \leq \mu_{\alpha,s_\imath}$ by the definition of mean measure. Thus $\|\mu_{\alpha,s_\imath}\| \geq \imath^{\frac{\alpha-1}{\alpha}}$ and using the triangle inequality we get

$$\|\mu_{\alpha,s_\imath} - \mu_{\alpha,s}\| \geq \imath^{\frac{\alpha-1}{\alpha}} - \|\mu_{\alpha,s}\|.$$

On the other hand, $\|s - s_\imath\| \leq 2/\imath$ by the triangle inequality, as well. Thus $\|\mu_{\alpha,p} - \mu_{\alpha,s}\|$ is an unbounded function of p on every neighborhood of s, i.e. $\mu_{\alpha,p}$ is not continuous at $p = s$.

(d) If $C_{\alpha,\mathcal{W}}$ is infinite, there is a sequence of $\{p_\imath\}_{\imath\in\mathbb{Z}_+}$ such that $\lim_{\imath\uparrow\infty} I_\alpha(p_\imath;\mathcal{W}) = \infty$. Let $p_{\beta,\imath} = (1-\beta)p + \beta p_\imath$, for any p. Then the concavity of the order α Rényi information in the prior for α's in $[1,\infty]$, established in Lemma 6-(b), and the non-negativity of the Rényi information imply

$$I_\alpha(p_{\beta,\imath};\mathcal{W}) - I_\alpha(p;\mathcal{W}) \geq \beta(I_\alpha(p_\imath;\mathcal{W}) - I_\alpha(p;\mathcal{W}))$$

for all $\beta \in (0,1)$ and $\imath \in \mathbb{Z}_+$. On the other hand $\|p - p_{\beta,\imath}\| \leq 2\beta$. Thus $I_\alpha(p;\mathcal{W})$ is not continuous in p, whenever $C_{\alpha,\mathcal{W}}$ is infinite. The continuity of $I_\alpha(p;\mathcal{W})$ in p for the case when $C_{\alpha,\mathcal{W}}$ is finite follows from part (e).

(e) We establish the uniform equicontinuity by proving establishing the following bound

$$\sup_{\alpha\in[0,\eta]} |I_\alpha(p_2;\mathcal{W}) - I_\alpha(p_1;\mathcal{W})|$$

$$\leq \begin{cases} \ln(\frac{1}{1-\delta} \wedge \frac{e^{C_0,\mathcal{W}}}{\delta}) + \ln(1-\delta+\delta e^{C_0,\mathcal{W}}) & \eta=0 \\ \ln \frac{1-\delta+\delta e^{C_\eta,\mathcal{W}}}{\left[(1-\delta)^{\frac{1}{\eta}} + \delta^{\frac{1}{\eta}} e^{\frac{\eta-1}{\eta}C_\eta,\mathcal{W}}\right]^{\frac{\eta}{1-\eta}}} & \eta\in\mathbb{R}_+\backslash\{1\} \\ h_1(\delta) + \delta C_{1,\mathcal{W}} + \ln(1-\delta+\delta e^{C_1,\mathcal{W}}) & \eta=1 \end{cases}$$
(A.16)

where $\delta = \frac{\|p_1-p_2\|}{2}$ and $h_\alpha(\cdot)$ is defined in (61).

As a result of the decomposition given Lemma 4-(c) we can write p_1 as $p_1 = (1-\delta)s_\wedge + \delta s_1$ for some $s_\wedge$ and s_1 in $\mathcal{P}(\mathcal{W})$. Using (19), (22), and (23) we get

$$I_1(p_1;\mathcal{W}) = (1-\delta)I_1(s_\wedge;\mathcal{W}) + (1-\delta)D_1(q_{1,s_\wedge} \| q_{1,p_1})$$
$$+ \delta I_1(s_1;\mathcal{W}) + \delta D_1(q_{1,s_1} \| q_{1,p_1}). \quad \text{(A.17)}$$

Similarly for positive orders other than one we have,

$$I_\alpha(p_1;\mathcal{W}) = \frac{1}{\alpha-1}\ln\Big[(1-\delta)e^{(\alpha-1)\left[I_\alpha(s_\wedge;\mathcal{W})+D_\alpha\left(q_{\alpha,s_\wedge}\|q_{\alpha,p_1}\right)\right]}$$
$$+ \delta e^{(\alpha-1)\left[I_\alpha(s_1;\mathcal{W})+D_\alpha\left(q_{\alpha,s_1}\|q_{\alpha,p_1}\right)\right]}\Big].$$
(A.18)

Since the Rényi divergence is non-negative by Lemma 10,

$$I_\alpha(p_1;\mathcal{W}) \geq \begin{cases} (1-\delta)I_1(s_\wedge;\mathcal{W}) + \delta I_1(s_1;\mathcal{W}) & \alpha=1 \\ \frac{\ln\left[(1-\delta)e^{(\alpha-1)I_\alpha(s_\wedge;\mathcal{W})}+\delta e^{(\alpha-1)I_\alpha(s_1;\mathcal{W})}\right]}{\alpha-1} & \alpha\neq1 \end{cases}$$
$$\geq I_\alpha(s_\wedge;\mathcal{W}) - g(\delta, \alpha, I_\alpha(s_\wedge;\mathcal{W}) - I_\alpha(s_1;\mathcal{W}))$$

where the function $g(\delta, \alpha, \gamma)$ is defined for any $\delta \in [0,1]$, $\alpha \in \mathbb{R}_+$, and $\gamma \in \mathbb{R}$ as follows

$$g(\delta, \alpha, \gamma) \triangleq \begin{cases} \delta\gamma & \alpha=1 \\ \frac{1}{1-\alpha}\ln\left[(1-\delta) + \delta e^{(1-\alpha)\gamma}\right] & \alpha\neq1 \end{cases}.$$

Given δ and γ, $g(\delta, \alpha, \gamma)$ is nonincreasing[32] in α. Then

$$I_\alpha(p_1;\mathcal{W}) \geq I_\alpha(s_\wedge;\mathcal{W}) - g(\delta, 0, I_\alpha(s_\wedge;\mathcal{W}) - I_\alpha(s_1;\mathcal{W}))$$

for all α in $(0,\eta]$. Furthermore, given δ and α, $g(\delta, \alpha, \gamma)$ is nondecreasing in γ. Then using $I_\alpha(s_1;\mathcal{W}) \geq 0$, $I_\alpha(s_\wedge;\mathcal{W}) \leq I_\eta(s_\wedge;\mathcal{W})$, and $I_\eta(s_\wedge;\mathcal{W}) \leq C_{\eta,\mathcal{W}}$ we get

$$I_\alpha(p_1;\mathcal{W}) \geq I_\alpha(s_\wedge;\mathcal{W}) - g(\delta, 0, C_{\eta,\mathcal{W}}) \quad \text{(A.19)}$$

for all $\alpha \in (0,\eta]$. On the other hand, $p_2 = (1-\delta)s_\wedge + \delta s_2$ by the decomposition given in Lemma 4-(c). Then

$$(1-\delta)^{\frac{1}{\alpha}}\mu_{\alpha,s_\wedge} \leq \mu_{\alpha,p_2}$$

as a result of the definition of the mean measure. Thus

$$e^{\frac{\alpha-1}{\alpha}(I_\alpha(s_\wedge;\mathcal{W})-I_\alpha(p_2;\mathcal{W}))}(1-\delta)^{\frac{1}{\alpha}} q_{\alpha,s_\wedge} \leq q_{\alpha,p_2}$$

by (13) and (21). Applying Lemma 8 we get

$$D_\alpha(q_{\alpha,s_\wedge} \| q_{\alpha,p_2}) \leq D_\alpha\Big(q_{\alpha,s_\wedge} \| (1-\delta)^{\frac{1}{\alpha}}\mu_{\alpha,s_\wedge}\Big)$$
$$= \frac{(1-\alpha)(I_\alpha(s_\wedge;\mathcal{W})-I_\alpha(p_2;\mathcal{W}))-\ln(1-\delta)}{\alpha}$$

for all α's in $\mathbb{R}_+$. Using the corresponding upper bound on $D_\alpha(q_{\alpha,s_2} \| q_{\alpha,p_2})$ together with (A.17) and (A.18) we get the following bound for all positive real orders

$$I_\alpha(p_2;\mathcal{W}) \leq \begin{cases} (1-\delta)I_1(s_\wedge;\mathcal{W}) + \delta I_1(s_2;\mathcal{W}) + h_1(\delta) & \alpha=1 \\ \frac{\alpha\ln\left[(1-\delta)^{\frac{1}{\alpha}}e^{\frac{\alpha-1}{\alpha}I_\alpha(s_\wedge;\mathcal{W})}+\delta^{\frac{1}{\alpha}}e^{\frac{\alpha-1}{\alpha}I_\alpha(s_2;\mathcal{W})}\right]}{\alpha-1} & \alpha\neq1 \end{cases}$$
$$= I_\alpha(s_\wedge;\mathcal{W}) + f(\delta, \alpha, I_\alpha(s_2;\mathcal{W}) - I_\alpha(s_\wedge;\mathcal{W}))$$

where the function $f(\delta, \alpha, \gamma)$ is defined for any $\delta \in [0,1]$, $\alpha \in \mathbb{R}_+$, and $\gamma \in \mathbb{R}$ as follows

$$f(\delta, \alpha, \gamma) \triangleq \begin{cases} \delta\gamma + h_1(\delta) & \alpha=1 \\ \frac{\alpha}{\alpha-1}\ln\left[(1-\delta)^{\frac{1}{\alpha}} + \delta^{\frac{1}{\alpha}}e^{\frac{\alpha-1}{\alpha}\gamma}\right] & \alpha\neq1 \end{cases}.$$

[32]For any fixed (δ,γ) pair, $g(\delta, \alpha, \gamma)$ is a continuous and differentiable function of α satisfying $\frac{\partial}{\partial\alpha}g(\delta, \alpha, \gamma) \leq 0$. In particular

$$\frac{\partial}{\partial\alpha}g(\delta, \alpha, \gamma) = \frac{-1}{(1-\alpha)^2}d\Big(\frac{\delta e^{(1-\alpha)\gamma}}{(1-\delta)+\delta e^{(1-\alpha)\gamma}}\Big\|\delta\Big).$$

For any fixed (δ, γ) pair, $f(\delta, \alpha, \gamma)$ is nondecreasing[33] in α. Then for any α in $(0, \eta]$ we have

$$I_\alpha(p_2; \mathcal{W}) \leq I_\alpha(s_\wedge; \mathcal{W}) + f(\delta, \eta, I_\alpha(s_2; \mathcal{W}) - I_\alpha(s_\wedge; \mathcal{W})).$$

Furthermore, given δ and α, $f(\delta, \alpha, \gamma)$ is nondecreasing in γ. Then using $I_\alpha(s_\wedge; \mathcal{W}) \geq 0$, $I_\alpha(s_2; \mathcal{W}) \leq I_\eta(s_2; \mathcal{W})$, and $I_\eta(s_2; \mathcal{W}) \leq C_{\eta, \mathcal{W}}$ we get

$$I_\alpha(p_2; \mathcal{W}) \leq I_\alpha(s_\wedge; \mathcal{W}) + f(\delta, \eta, C_{\eta, \mathcal{W}}) \tag{A.20}$$

for all α in $(0, \eta]$. Using (A.19) and (A.20) together with the definition of the Rényi capacity given in (27) we get

$$I_\alpha(p_2; \mathcal{W}) - I_\alpha(p_1; \mathcal{W}) \leq f(\delta, \eta, C_{\eta, \mathcal{W}}) + g(\delta, 0, C_{\eta, \mathcal{W}}).$$

A lower bound on $I_\alpha(p_2; \mathcal{W}) - I_\alpha(p_1; \mathcal{W})$ can be obtained using the same arguments with the roles of p_1 and p_2 reversed. This establishes (A.16) for $\eta > 0$ and $\alpha \in (0, \eta]$. In order to establish (A.16) for $\alpha = 0$, recall the definition of the order zero Rényi information given in (13).

$$I_0(p_1; \mathcal{W}) = -\ln \operatorname*{ess\,sup}_{\mu_{1,p_1}} \left((1-\delta) \sum_{w: s_\wedge(w|y)>0} s_\wedge(w) \right.$$
$$\left. + \delta \sum_{w: s_1(w|y)>0} s_1(w) \right)$$
$$\geq -\ln \left((1-\delta) e^{-I_0(s_\wedge; \mathcal{W})} + \delta \right)$$
$$= I_0(s_\wedge; \mathcal{W}) - \ln \left(1 - \delta + \delta e^{I_0(s_\wedge; \mathcal{W})} \right).$$

Note that $I_0(s_\wedge; \mathcal{W}) \leq I_\eta(s_\wedge; \mathcal{W})$ by Lemma 5 and $I_\eta(s_\wedge; \mathcal{W}) \leq C_{\eta, \mathcal{W}}$ by definition. Then

$$I_0(p_1; \mathcal{W}) \geq I_0(s_\wedge; \mathcal{W}) - \ln \left(1 - \delta + \delta e^{C_{\eta, \mathcal{W}}} \right). \tag{A.21}$$

On the other hand,

$$I_0(p_2; \mathcal{W}) = \operatorname*{ess\,inf}_{\mu_{1,p_2}} \ln \frac{1}{\sum_{w: p_2(w|y)>0}((1-\delta)s_\wedge(w) + \delta s_2(w))}$$
$$\leq \left(\operatorname*{ess\,inf}_{\mu_{1,s_\wedge}} \ln \frac{1}{(1-\delta) \sum_{w: s_\wedge(w|y)>0} s_\wedge(w)} \right)$$
$$\wedge \left(\operatorname*{ess\,inf}_{\mu_{1,s_2}} \ln \frac{1}{\delta \sum_{w: s_2(w|y)>0} s_2(w)} \right)$$
$$= \left(I_0(s_\wedge; \mathcal{W}) + \ln \tfrac{1}{1-\delta} \right) \wedge \left(I_0(s_2; \mathcal{W}) + \ln \tfrac{1}{\delta} \right)$$

Then $I_0(s_\wedge; \mathcal{W}) \geq 0$ and $I_0(s_2; \mathcal{W}) \leq I_\eta(s_2; \mathcal{W}) \leq C_{\eta, \mathcal{W}}$ imply

$$I_0(p_2; \mathcal{W}) \leq I_0(s_\wedge; \mathcal{W}) + \left(\ln \tfrac{1}{1-\delta} \wedge \ln \tfrac{e^{C_{\eta, \mathcal{W}}}}{\delta} \right). \tag{A.22}$$

Thus using (A.21) and (A.22) we get

$$I_0(p_2; \mathcal{W}) - I_0(p_1; \mathcal{W}) \leq \ln \left(1 - \delta + \delta e^{C_{\eta, \mathcal{W}}} \right)$$
$$+ \ln \left(\tfrac{1}{1-\delta} \wedge \tfrac{e^{C_{\eta, \mathcal{W}}}}{\delta} \right). \tag{A.23}$$

A lower bound on $I_0(p_2; \mathcal{W}) - I_0(p_1; \mathcal{W})$ can be obtained using the same arguments with the roles of p_1 and p_2 reversed. Consequently, (A.16) holds for $\eta = 0$, $\alpha = 0$

[33] For any fixed (δ, γ) pair, $f(\delta, \alpha, \gamma)$ is a continuous and differentiable function of α satisfying $\frac{\partial}{\partial \alpha} f(\delta, \alpha, \gamma) \geq 0$. In particular

$$\frac{\partial}{\partial \alpha} f(\delta, \alpha, \gamma) = \frac{1}{(1-\alpha)^2} d \left(\frac{(1-\delta)^{1/\alpha}}{(1-\delta)^{1/\alpha} + \delta^{1/\alpha} e^{(1-1/\alpha)\gamma}} \middle\| 1 - \delta \right).$$

case. In order to establish (A.16) for $\eta > 0$, $\alpha = 0$ case, note that

$$\frac{\eta \ln \left[(1-\delta)^{\frac{1}{\eta}} + \delta^{\frac{1}{\eta}} e^{\frac{\eta-1}{\eta} C_{\eta, \mathcal{W}}} \right]}{\eta - 1} \geq \begin{cases} \ln \frac{1}{1-\delta} & \frac{1-\delta}{\delta e^{-C_{\eta, \mathcal{W}}}} \geq 1 \\ \ln \frac{e^{C_{\eta, \mathcal{W}}}}{\delta} & \frac{1-\delta}{\delta e^{-C_{\eta, \mathcal{W}}}} \leq 1 \end{cases}$$
$$\geq \left[\ln \tfrac{1}{1-\delta} \wedge \ln \tfrac{e^{C_{\eta, \mathcal{W}}}}{\delta} \right]$$

Thus (A.16) holds for $\eta > 0$, $\alpha = 0$ case, as well.

(f) In order to establish the uniform equicontinuity we prove the Lipschitz continuity of $\{I_\alpha(p; \mathcal{W})\}_{p \in \mathcal{P}(\mathcal{W})}$ in α on compact subsets of $(0, \eta)$ with a common Lipschitz constant: If α and ϕ in $[\epsilon, \eta - \epsilon]$ for an $\epsilon \in (0, \epsilon_\eta]$, then

$$|I_\alpha(p; \mathcal{W}) - I_\phi(p; \mathcal{W})| \leq \tfrac{\gamma_\eta}{\epsilon^2} |\alpha - \phi| \tag{A.24}$$

for all p in $\mathcal{P}(\mathcal{W})$ where ϵ_η and γ_η are defined as follows

$$\epsilon_\eta \triangleq \begin{cases} \frac{\eta}{2} & \eta \in (0, 1] \\ \frac{\eta-1}{8\eta} & \eta \in (1, \infty) \end{cases},$$
$$\gamma_\eta \triangleq \begin{cases} C_{\eta, \mathcal{W}} & \eta \in (0, 1] \\ \eta C_{\eta, \mathcal{W}} + \frac{5 e^{2 C_{\eta, \mathcal{W}}}}{2 e^2} & \eta \in (1, \infty) \end{cases}.$$

Since $\|\mu_{\alpha, p}\|^\alpha$ is a log-convex in α by Lemma 3-(d),

$$\|\mu_{\alpha, p}\|^\alpha \leq \|\mu_{\beta, p}\|^{\beta \frac{\alpha - \phi}{\beta - \phi}} \|\mu_{\phi, p}\|^{\phi \frac{\beta - \alpha}{\beta - \phi}}.$$

for any ϕ, α, β satisfying $0 < \phi < \alpha < \beta$ and $p \in \mathcal{P}(\mathcal{W})$. Let us start with $\eta \in (0, 1]$ and $\epsilon \in (0, \frac{\eta}{2}]$ case. Then for any ϕ, α, β satisfying $0 < \phi < \epsilon \leq \alpha < \beta \leq \eta - \epsilon$,

$$I_\beta(p; \mathcal{W}) - I_\alpha(p; \mathcal{W}) = \frac{1}{1-\alpha} \ln \frac{\|\mu_{\alpha, p}\|^\alpha}{\|\mu_{\beta, p}\|^{\frac{\beta(1-\alpha)}{1-\beta}}}$$
$$\leq \frac{1}{1-\alpha} \ln \frac{\|\mu_{\beta, p}\|^{\beta \frac{\alpha - \phi}{\beta - \phi}} \|\mu_{\phi, p}\|^{\phi \frac{\beta - \alpha}{\beta - \phi}}}{\|\mu_{\beta, p}\|^{\frac{\beta(1-\alpha)}{1-\beta}}}$$
$$= \frac{(\beta - \alpha)(1-\phi)}{(1-\alpha)(\beta - \phi)} (I_\beta(p; \mathcal{W}) - I_\phi(p; \mathcal{W}))$$
$$\leq \frac{\beta - \alpha}{\epsilon(\epsilon - \phi)} I_\beta(p; \mathcal{W}).$$

The above bound holds for any ϕ in $(0, \epsilon)$. Furthermore, the Rényi information is a nondecreasing function of the order by Lemma 5. Then

$$0 \leq I_\beta(p; \mathcal{W}) - I_\alpha(p; \mathcal{W}) \leq \tfrac{C_{\eta, \mathcal{W}}}{\epsilon^2} (\beta - \alpha) \tag{A.25}$$

for any p in $\mathcal{P}(\mathcal{W})$ and β, α satisfying $\epsilon \leq \alpha \leq \beta \leq \eta - \epsilon$. (A.25) establishes (A.24) for $\eta \in (0, 1]$ and $\epsilon \in (0, \frac{\eta}{2}]$ case.

We proceed with $\eta \in (1, \infty)$ and $\epsilon \in (0, \frac{\eta-1}{8\eta}]$ case. For any ϕ, α, β such that $1 + \epsilon \leq \phi < \alpha \leq \beta - \epsilon$ and p in $\mathcal{P}(\mathcal{W})$ we have

$$I_\alpha(p; \mathcal{W}) - I_\phi(p; \mathcal{W}) = \frac{1}{\alpha - 1} \ln \frac{\|\mu_{\alpha, p}\|^\alpha}{\|\mu_{\phi, p}\|^{\frac{\phi(\alpha-1)}{\phi-1}}}$$
$$\leq \frac{1}{\alpha - 1} \ln \frac{\|\mu_{\beta, p}\|^{\beta \frac{\alpha - \phi}{\beta - \phi}} \|\mu_{\phi, p}\|^{\phi \frac{\beta - \alpha}{\beta - \phi}}}{\|\mu_{\phi, p}\|^{\frac{\phi(\alpha-1)}{\phi-1}}}$$
$$= \frac{(\alpha - \phi)(\beta - 1)}{(\alpha - 1)(\beta - \phi)} (I_\beta(p; \mathcal{W}) - I_\phi(p; \mathcal{W}))$$
$$\leq \frac{(\alpha - \phi)}{\epsilon^2} \beta C_{\beta, \mathcal{W}}. \tag{A.26}$$

If $0 < \alpha - \phi < \epsilon$, then at least one of the three closed intervals $[\epsilon, 1 - \epsilon]$, $[\frac{1}{2}, \frac{5\eta-1}{4\eta}]$, $[1 + \epsilon, \eta - \epsilon]$ includes both α

26

and ϕ. When α and ϕ are in $[\epsilon, 1-\epsilon]$ we use (A.25). When α and ϕ are in $[1+\epsilon, \eta-\epsilon]$ we use (A.26). Derivation of the bound for the second interval takes some effort. Let us first finish the proof of (A.24) assuming that the bound given in (A.33) holds for the second interval. Then for any ϕ, α such that $\epsilon \le \phi \le \alpha \le (\phi+\epsilon)\wedge(\eta-\epsilon)$ we have

$$0 \le \frac{I_\alpha(p;\mathcal{W}) - I_\phi(p;\mathcal{W})}{\alpha - \phi}$$

$$\le \begin{cases} \frac{C_{\eta,\mathcal{W}}}{\epsilon^2} & \epsilon \le \phi \le \alpha \le (\phi+\epsilon)\wedge(1-\epsilon) \\ 4C_{\eta,\mathcal{W}} + \frac{160\eta^2 e^{2C_{\eta,\mathcal{W}}}}{e^2(\eta-1)^2} & \frac{1}{2} \le \phi \le \alpha \le (\phi+\epsilon)\wedge\frac{5\eta-1}{4\eta} \\ \frac{\eta C_{\eta,\mathcal{W}}}{\epsilon^2} & 1+\epsilon \le \phi \le \alpha \le (\phi+\epsilon)\wedge(\eta-\epsilon) \end{cases}$$

Thus for any α and ϕ satisfying $\epsilon \le \phi \le \alpha \le (\phi+\epsilon)\wedge(\eta-\epsilon)$ and p in $\mathcal{P}(\mathcal{W})$ we have

$$I_\alpha(p;\mathcal{W}) - I_\phi(p;\mathcal{W}) \le \frac{\alpha-\phi}{\epsilon^2}\left[\eta C_{\eta,\mathcal{W}} + \frac{5e^{2C_{\eta,\mathcal{W}}}}{2e^2}\right].$$

Note that the preceding bound is linear with a uniform constant, thus the hypothesis $\phi \le \alpha \le \phi + \epsilon$ can be removed without loss of generality. Thus (A.24) holds for $\eta \in (1,\infty)$ case for any $\epsilon \in (0,\epsilon_\eta]$, as well.

We are left with establishing the bound given in (A.33). For orders other than one, (4) and (16) imply that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W}) = \frac{1}{\alpha-1}\left[\alpha\frac{\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} - \frac{I_\alpha(p;\mathcal{W})}{\alpha}\right].$$

The expression in the brackets is differentiable in α on $\mathbb{R}_+$ because $\|\mu_{\alpha,p}\|$ is positive and $\|\mu_{\alpha,p}\|$, $\|\mu'_{\alpha,p}\|$, and $I_\alpha(p;\mathcal{W})$ are differentiable by Lemmas 1-(a), 3-(b,c), and 5. Furthermore, the expression in the brackets is equal to zero at $\alpha = 1$. Then as a result of the mean value theorem [80, 5.10] for each $\alpha \in [1/2, 1)$ there exists a $\phi \in (\alpha, 1)$ and such that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W}) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\frac{\alpha\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} - \frac{I_\alpha(p;\mathcal{W})}{\alpha}\right]\bigg|_{\alpha=\phi}. \quad \text{(A.27)}$$

Using the expressions for derivatives given in Lemmas 2-(b) and 3-(b,c) we get

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\frac{\alpha\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} = \frac{\alpha\mu''_{\alpha,p}(\mathcal{Y})}{\|\mu_{\alpha,p}\|} + \frac{\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} - \frac{\alpha\|\mu'_{\alpha,p}\|^2}{\|\mu_{\alpha,p}\|^2}$$

$$= \mathbf{E}_{q_{\alpha,p}}\left[\sum_w \frac{p_{[\alpha]}(w|y)}{\alpha^2}\left(\ln\frac{p_{[\alpha]}(w|y)}{p(w)}\right)^2\right]$$

$$+ \mathbf{E}_{q_{\alpha,p}}\left[\frac{\alpha(1-\alpha)(\pi'_{\alpha,p})^2}{(\pi_{\alpha,p})^2}\right] - \frac{\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} - \frac{\alpha\|\mu'_{\alpha,p}\|^2}{\|\mu_{\alpha,p}\|^2}$$

$$\le \mathbf{E}_{q_{\alpha,p}}\left[\frac{{\pi'_{\alpha,p}}^2}{4\pi_{\alpha,p}^2} + \sum_w \frac{p_{[\alpha]}(w|y)}{\alpha^2}\left(\ln\frac{p_{[\alpha]}(w|y)}{p(w)}\right)^2\right]$$

Then using $\sum_w p_{[\alpha]}(w|y)\frac{1}{\alpha}\ln\frac{p_{[\alpha]}(w|y)}{p(w)} = \frac{\alpha\pi'_{\alpha,p}}{\pi_{\alpha,p}}$, which follows from Lemma 2-(b), we get

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\frac{\alpha\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} \le \frac{4\alpha^2+1}{4\alpha^2}\mathbf{E}_{q_{\alpha,p}}\left[\sum_w \frac{p_{[\alpha]}(w|y)}{\alpha^2}\left(\ln\frac{p_{[\alpha]}(w|y)}{p(w)}\right)^2\right] \quad \text{(A.28)}$$

Since $I_\alpha(p;\mathcal{W})$ is differentiable and nondecreasing in α

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\frac{I_\alpha(p;\mathcal{W})}{\alpha} = -\frac{I_\alpha(p;\mathcal{W})}{\alpha^2} + \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}\alpha}I_\alpha(p;\mathcal{W})$$

$$\ge -\frac{I_\alpha(p;\mathcal{W})}{\alpha^2} \quad \text{(A.29)}$$

Using (A.27), (A.28), and (A.29) we can conclude that there exists a $\phi \in (\alpha, 1)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W}) \le 2\mathbf{E}_{q_{\phi,p}}\left[\sum_w \frac{p_{[\phi]}(w|y)}{\phi^2}\left[\ln\frac{p_{[\phi]}(w|y)}{p(w)}\right]^2\right] + 4I_\phi(p;\mathcal{W})$$

Similarly for all $\alpha \in (1,\infty)$ there exists a $\phi \in (1,\alpha)$ satisfying the same identity. Furthermore, one can confirm by substitution for the expression given in (17) that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W})\big|_{\alpha=1} \le \frac{1}{2}\mathbf{E}_{q_{1,p}}\left[\sum_w p_{[1]}(w|y)\left(\ln\frac{p_{[1]}(w|y)}{p(w)}\right)^2\right].$$

Thus there exist an $\phi \in (\frac{1}{2}, \frac{5\eta-1}{4\eta})$ such that

$$\sup_{\alpha\in[\frac{1}{2}, \frac{5\eta-1}{4\eta}]}\frac{\mathrm{d}}{\mathrm{d}\alpha}I_\alpha(p;\mathcal{W})$$

$$\le 4I_\phi(p;\mathcal{W}) + 2\mathbf{E}_{q_{\phi,p}}\left[\sum_w \frac{p_{[\phi]}(w|y)}{\phi^2}\left[\ln\frac{p_{[\phi]}(w|y)}{p(w)}\right]^2\right].$$
$$\text{(A.30)}$$

Note that $x^\phi \ln^2 x \le \frac{4}{e^2\phi^2}\mathbb{1}_{\{x\in[0,1)\}} + \frac{4x^\beta}{e^2(\beta-\phi)^2}\mathbb{1}_{\{x>1\}}$ for all $\beta > \phi$. Then using Lemma 2-(a) we get the following bound for all ϕ in $[\frac{1}{2}, \frac{5\eta-1}{4\eta}]$

$$\sum_w p_{[\phi]}(w|y)\left(\frac{1}{\phi}\ln\frac{p_{[\phi]}(w|y)}{p(w)}\right)^2$$

$$= \sum_w p(w)\left(\frac{p_{[1]}(w|y)}{p(w)\pi_{\phi,p}}\right)^\phi\left(\ln\frac{p_{[1]}(w|y)}{p(w)\pi_{\phi,p}}\right)^2$$

$$\le \frac{4}{e^2\phi^2} + \frac{4}{e^2}\left(\frac{3\eta-1}{2\eta} - \phi\right)^{-2}\sum_w p(w)\left(\frac{p_{[1]}(w|y)}{p(w)\pi_{\phi,p}}\right)^{\frac{3\eta-1}{2\eta}}$$

$$= \frac{4}{e^2\phi^2} + \frac{4}{e^2}\left(\frac{3\eta-1}{2\eta} - \phi\right)^{-2}\left(\pi_{\frac{3\eta-1}{2\eta},p}/\pi_{\phi,p}\right)^{\frac{3\eta-1}{2\eta}}$$

$$\le \frac{16}{e^2} + \frac{64}{e^2}\left(\frac{\eta}{\eta-1}\right)^2\left(\pi_{\frac{3\eta-1}{2\eta},p}/\pi_{\phi,p}\right)^{\frac{3\eta-1}{2\eta}}. \quad \text{(A.31)}$$

On the other hand $(\pi_{\alpha,p})^\alpha$ is log-convex in α by Lemma 2-(c) and $\pi_{\alpha,p}$ is nondecreasing in α by Lemma 2-(d). Thus for all ϕ in $[\frac{1}{2}, \frac{5\eta-1}{4\eta}]$ we have

$$\left(\pi_{\frac{3\eta-1}{2\eta},p}\right)^{\frac{3\eta-1}{2\eta}} \le (\pi_{\frac{2\eta\phi}{2\eta\phi-\eta+1},p})(\pi_{\phi,p})^{\phi\frac{\eta-1}{2\eta\phi}}$$

$$\le (\pi_{\eta,p})(\pi_{\phi,p})^{\frac{\eta-1}{2\eta}}. \quad \text{(A.32)}$$

Using equations (A.30), (A.31), and (A.32) we get

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W}) \le 4I_\eta(p;\mathcal{W}) + \frac{32}{e^2} + \frac{128}{e^2}\left(\frac{\eta}{\eta-1}\right)^2\frac{\|\mu_{\eta,p}\|}{\|\mu_{1/2,p}\|}$$

$$= 4I_\eta(p;\mathcal{W}) + \frac{32}{e^2} + \frac{128\eta^2}{e^2(\eta-1)^2}e^{\frac{\eta-1}{\eta}I_\eta(p;\mathcal{W})+I_\frac{1}{2}(p;\mathcal{W})}$$

for all α in $[\frac{1}{2}, \frac{5\eta-1}{4\eta}]$. Since $I_\alpha(p;\mathcal{W})$ is nondecreasing in α by Lemma 5, the definition of Rényi capacity implies

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}I_\alpha(p;\mathcal{W}) \le 4C_{\eta,\mathcal{W}} + \frac{32}{e^2} + \frac{128}{e^2}\left(\frac{\eta}{\eta-1}\right)^2 e^{2C_{\alpha,\mathcal{W}}}$$

for all α in $[\frac{1}{2}, \frac{5\eta-1}{4\eta}]$ and p in $\mathcal{P}(\mathcal{W})$. Hence,

$$I_\alpha(p;\mathcal{W}) - I_\phi(p;\mathcal{W}) \le (\alpha-\phi)\left[4C_{\eta,\mathcal{W}} + \frac{160\eta^2 e^{2C_{\eta,\mathcal{W}}}}{e^2(\eta-1)^2}\right] \quad \text{(A.33)}$$

for all ϕ, α in $[\frac{1}{2}, \frac{5\eta-1}{4\eta}]$ satisfying $\phi \le \alpha$ and p in $\mathcal{P}(\mathcal{W})$.

(g) For any $p \in \mathcal{P}(\mathcal{W})$, $I_\alpha(p;\mathcal{W})$ is nondecreasing and continuous in α on $[0,\infty]$ by Lemma 5. Then $I_\alpha(p;\mathcal{W})$

is a quasi-convex continuous function of α satisfying $I_\alpha(p;\mathcal{W}) = \inf_{\eta\in(\alpha,\infty)} I_\eta(p;\mathcal{W})$ for any p in $\mathcal{P}(\mathcal{W})$. Using the definition of $C_{\alpha,\mathcal{W}}$ we get

$$C_{\alpha,\mathcal{W}} = \sup_{p\in\mathcal{P}(\mathcal{W})} \inf_{\eta\in(\alpha,\infty)} I_\eta(p;\mathcal{W}) \quad \forall\alpha\in\mathbb{R}_{\geq 0}. \quad \text{(A.34)}$$

Since $I_\alpha(p;\mathcal{W}) \leq \ln|\mathcal{W}|$ by Lemma 5, if $\mathcal{W}$ is finite, then $C_{\alpha,\mathcal{W}}$ is finite for all $\alpha\in\mathbb{R}_{\geq 0}$ and $I_\alpha(p;\mathcal{W})$ is continuous in p on $\mathcal{P}(\mathcal{W})$ for all $\alpha\in\mathbb{R}_+$ by part (e). Furthermore $I_\alpha(p;\mathcal{W})$ is quasi-concave in p for all $\alpha\in\mathbb{R}_+$ by Lemma 6. Then we can change the order of the supremum and the infimum in (A.34) using Sion's minimax theorem, [95, Cor. 3.3], [57] because $\mathcal{P}(\mathcal{W})$ is compact.

$$C_{\alpha,\mathcal{W}} = \inf_{\eta\in(\alpha,\infty)} \sup_{p\in\mathcal{P}(\mathcal{W})} I_\eta(p;\mathcal{W})$$
$$= \inf_{\eta\in(\alpha,\infty)} C_{\eta,\mathcal{W}} \qquad \forall\alpha\in\mathbb{R}_{\geq 0}.$$

Then $C_{\alpha,\mathcal{W}}$ is continuous from the right. On the other hand $C_{\alpha,\mathcal{W}}$ is continuous from the left because it is nondecreasing and lower semicontinuous on $[0,\infty]$ by Lemma 15-(a). $\qquad\square$

Proof of Lemma 21. $\underline{\sup_{\imath\in\mathcal{T}} C_{\alpha,\mathcal{W}_\imath} \leq C_{\alpha,\mathcal{W}}}$:

$$C_{\alpha,\mathcal{W}} \overset{(a)}{\geq} S_{\alpha,\mathcal{W}_\imath}(q_{\alpha,\mathcal{W}})$$
$$\overset{(b)}{\geq} C_{\alpha,\mathcal{W}_\imath} + D_\alpha(q_{\alpha,\mathcal{W}_\imath}\|q_{\alpha,\mathcal{W}})$$
$$\overset{(c)}{\geq} C_{\alpha,\mathcal{W}_\imath} + \tfrac{\alpha\wedge 1}{2}\|q_{\alpha,\mathcal{W}_\imath} - q_{\alpha,\mathcal{W}}\|^2$$

where (a) follows from (1), Theorem 1, and $\mathcal{W}_\imath\subset\mathcal{W}$, (b) follows from Lemma 19, (c) follows from Lemma 10. Consequently, $C_{\alpha,\mathcal{W}} \geq \sup_{\imath\in\mathcal{T}} C_{\alpha,\mathcal{W}_\imath}$ and if $C_{\alpha,\mathcal{W}_\imath} = C_{\alpha,\mathcal{W}}$, then $q_{\alpha,\mathcal{W}} = q_{\alpha,\mathcal{W}_\imath}$.

- If $C_{\alpha,\mathcal{W}_\imath} = C_{\alpha,\mathcal{W}}$ and $q_{\alpha,\mathcal{W}} = q_{\alpha,\mathcal{W}_\imath}$, then $S_{\alpha,\mathcal{W}}(q_{\alpha,\mathcal{W}_\imath}) \leq C_{\alpha,\mathcal{W}_\imath}$ by Theorem 1.
- If $S_{\alpha,\mathcal{W}}(q_{\alpha,\mathcal{W}_\imath}) \leq C_{\alpha,\mathcal{W}_\imath}$, then $C_{\alpha,\mathcal{W}} \leq C_{\alpha,\mathcal{W}_\imath}$ because $C_{\alpha,\mathcal{W}} = S_{\alpha,\mathcal{W}}$ by Theorem 1 and $S_{\alpha,\mathcal{W}} \leq S_{\alpha,\mathcal{W}}(q_{\alpha,\mathcal{W}_\imath})$ by definition. Then $S_{\alpha,\mathcal{W}}(q_{\alpha,\mathcal{W}_\imath}) \leq C_{\alpha,\mathcal{W}_\imath}$ implies $C_{\alpha,\mathcal{W}} = C_{\alpha,\mathcal{W}_\imath}$ because $C_{\alpha,\mathcal{W}} \geq C_{\alpha,\mathcal{W}_\imath}$ by definition.

$\underline{C_{\alpha,\mathcal{W}} \leq \ln\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}}$: If $\mathcal{T}$ is infinite, then the inequality holds trivially because the right hand side is infinite. Thus, we will establish the inequality assuming $\mathcal{T}$ is finite. Let v be $v \triangleq \bigvee_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}} q_{\alpha,\mathcal{W}_\imath}$. Then

$$S_{\alpha,\mathcal{W}}(v/\|v\|) \overset{(a)}{=} \max_{\imath\in\mathcal{T}} S_{\alpha,\mathcal{W}_\imath}(v/\|v\|)$$
$$\overset{(b)}{\leq} \max_{\imath\in\mathcal{T}} S_{\alpha,\mathcal{W}_\imath}(q_{\alpha,\mathcal{W}_\imath}) - \ln e^{C_{\alpha,\mathcal{W}_\imath}} + \ln\|v\|$$
$$\overset{(c)}{=} \ln\|v\| \qquad\qquad\qquad\qquad \text{(A.35)}$$

where (a) follows from (1) and (2), (b) follows from Lemma 8 because $e^{C_{\alpha,\mathcal{W}_\imath}} q_{\alpha,\mathcal{W}_\imath} \leq v$, and (c) follows from Theorem 1. On the other hand, $\|v\| \leq \sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}$ by the definition of v. Then $C_{\alpha,\mathcal{W}} \leq \ln\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}$ by Theorem 1.

- If $\mathcal{T}$ is infinite, then $\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}$ is infinite. If $\mathcal{T}$ is finite but $q_{\alpha,\mathcal{W}_\imath}$ and $q_{\alpha,\mathcal{W}_\jmath}$ are not singular for some distinct $\imath$ and $\jmath$, then $C_{\alpha,\mathcal{W}} < \ln\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}$ by (A.35) and Theorem 1 because $\|v\| < \sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}$. Consequently, if $C_{\alpha,\mathcal{W}} = \ln\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}} < \infty$, then $\mathcal{T}$ is finite and $q_{\alpha,\mathcal{W}_\imath} \perp q_{\alpha,\mathcal{W}_\jmath}$ for all $\imath\neq\jmath$ in $\mathcal{T}$.

- If $\mathcal{T}$ is finite and $q_{\alpha,\mathcal{W}_\imath} \perp q_{\alpha,\mathcal{W}_\jmath}$ for all $\imath\neq\jmath$, then any $u\in\mathcal{P}(\mathcal{Y})$ can be written as $u = \sum_{\imath=0}^{|\mathcal{T}|} u_\imath$ where $u_\imath$ are finite measures such that $u_\imath\prec q_{\alpha,\mathcal{W}_\imath}$ for $\imath\in\{1,\ldots,|\mathcal{T}|\}$ and $u_0 \perp (\sum_{\imath\in\mathcal{T}} q_{\alpha,\mathcal{W}_\imath})$ by the Lebesgue decomposition theorem [30, 5.5.3]. Then using Lemmas 8 and 9, we get

$$D_\alpha(q_{\alpha,\mathcal{W}_\imath}\|u) \geq -\ln\|u_\imath\|.$$

Thus Lemma 19 implies

$$S_{\alpha,\mathcal{W}_\imath}(u) \geq C_{\alpha,\mathcal{W}_\imath} - \ln\|u_\imath\| \qquad \forall u\in\mathcal{P}(\mathcal{Y}).$$

Since $S_{\alpha,\mathcal{W}}(u) = \max_{\imath\in\mathcal{T}} S_{\alpha,\mathcal{W}_\imath}(u)$ for all u in $\mathcal{P}(\mathcal{Y})$ and $C_{\alpha,\mathcal{W}} = \inf_{u\in\mathcal{P}(\mathcal{Y})} S_{\alpha,\mathcal{W}}(u)$ by Theorem 1, we get

$$C_{\alpha,\mathcal{W}} \geq \inf_{u\in\mathcal{P}(\mathcal{Y})} \max_{\imath\in\mathcal{T}} \ln\frac{e^{C_{\alpha,\mathcal{W}_\imath}}}{\|u_\imath\|}$$
$$\geq \inf_{u\in\mathcal{P}(\mathcal{Y})} \ln\frac{\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}}{\sum_{\imath\in\mathcal{T}} \|u_\imath\|}$$
$$\geq \ln\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}.$$

Then $C_{\alpha,\mathcal{W}} = \ln\sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}}$ because we have already proved the reverse inequality. Furthermore, $q_{\alpha,\mathcal{W}} = \tilde{u}/\|\tilde{u}\|$ for $\tilde{u} = \sum_{\imath\in\mathcal{T}} e^{C_{\alpha,\mathcal{W}_\imath}} q_{\alpha,\mathcal{W}_\imath}$ by Theorem 1. because $S_{\alpha,\mathcal{W}_\imath}(\tilde{u}/\|\tilde{u}\|) = C_{\alpha,\mathcal{W}}$. $\qquad\square$

Proof of Lemma 22. By the definition of $I_\alpha(p;\mathcal{W})$ for all p satisfying $p = \bigotimes_{t\in\mathcal{T}} p_t$ for some $p_t\in\mathcal{P}(\mathcal{W}_t)$ we have

$$I_\alpha(p;\mathcal{W}) = \sum_{t\in\mathcal{T}} I_\alpha(p_t;\mathcal{W}_t) \qquad \forall\alpha\in[0,\infty]. \quad \text{(A.36)}$$

Furthermore, $\{p : p = \bigotimes_{t\in\mathcal{T}} p_t, p_t\in\mathcal{P}(\mathcal{W}_t), \forall t\in\mathcal{T}\}$ is a subset of $\mathcal{P}(\mathcal{W})$. Then

$$C_{\alpha,\mathcal{W}} \geq \sup_{p_1,p_2,\ldots,p_{|\mathcal{T}|}} \sum_{t\in\mathcal{T}} I_\alpha(p_t;\mathcal{W}_t)$$
$$= \sum_{t\in\mathcal{T}} C_{\alpha,\mathcal{W}_t} \qquad\qquad \forall\alpha\in[0,\infty]. \quad \text{(A.37)}$$

Let us proceed with proving $C_{\alpha,\mathcal{W}} \leq \sum_{t\in\mathcal{T}} C_{\alpha,\mathcal{W}_t}$. If there exists a $t\in\mathcal{T}$ such that $C_{\alpha,\mathcal{W}_t} = \infty$, then the inequality holds trivially. Else, $C_{\alpha,\mathcal{W}_t} < \infty$ for all $t\in\mathcal{T}$ and by Theorem 1 there exists a $q_{\alpha,\mathcal{W}_t}$ for each $t\in\mathcal{T}$ such that

$$D_\alpha(w_t\|q_{\alpha,\mathcal{W}_t}) \leq C_{\alpha,\mathcal{W}_t} \qquad \forall w_t\in\mathcal{W}_t.$$

Recall that all w's in $\mathcal{W}$ can be written as $w = \bigotimes_{t\in\mathcal{T}} w_t$ for some $w_t\in\mathcal{W}_t$ by the hypothesis. Then for $q \triangleq \bigotimes_{t\in\mathcal{T}} q_{\alpha,\mathcal{W}_t}$ by the definition of the Rényi divergence given (19) and Tonelli-Fubini theorem [30, 4.4.5] we have

$$D_\alpha(w\|q) = \sum_{t\in\mathcal{T}} D_\alpha(w_t\|q_{\alpha,\mathcal{W}_t})$$
$$\leq \sum_{t\in\mathcal{T}} C_{\alpha,\mathcal{W}_t} \qquad\qquad \forall w\in\mathcal{W}. \quad \text{(A.38)}$$

Then $C_{\alpha,\mathcal{W}} \leq \sum_{t\in\mathcal{T}} C_{\alpha,\mathcal{W}_t}$ by (33) Thus (58) holds and $q_{\alpha,\mathcal{W}} = q$ follows from (A.38) and Theorem 1 for the case when $C_{\alpha,\mathcal{W}} < \infty$. $\qquad\square$

Proof of Lemma 23. (19) implies that

$$D_\alpha(p\circledast\mathcal{W}\|p\otimes q_{\alpha,\mathcal{W}}) \leq \sup_{w\in\text{supp}(p)} D_\alpha(w\|q_{\alpha,\mathcal{W}}),$$
$$D_\alpha(p\circledast\mathcal{W}\|p\otimes q_{\alpha,\mathcal{W}}) \geq \inf_{w\in\text{supp}(p)} D_\alpha(w\|q_{\alpha,\mathcal{W}}).$$

On the other hand, (22) and (23) imply that

$$D_\alpha(p \circledast \mathcal{W} \| p \otimes q_{\alpha,\mathcal{W}}) = I_\alpha(p;\mathcal{W}) + D_\alpha(q_{\alpha,p} \| q_{\alpha,\mathcal{W}}).$$

Then (60) follows from (59) and Theorem 1.

Note that $C_{\alpha,\mathcal{W}_{\alpha,\epsilon}}$ is bounded above by $C_{\alpha,\mathcal{W}}$ and hence finite by definition. Thus $\mathcal{W}_{\alpha,\epsilon}$ has a unique order α Rényi center by Theorem 1. If $q_{\alpha,\mathcal{W}_{\alpha,\epsilon}} = q_{\alpha,\mathcal{W}}$, then $C_{\alpha,\mathcal{W}_{\alpha,\epsilon}} = C_{\alpha,\mathcal{W}}$ by the definition of $\mathcal{W}_{\alpha,\epsilon}$ and Theorem 1.

We show in the following by contradiction that $q_{\alpha,\mathcal{W}_{\alpha,\epsilon}}$ equals to $q_{\alpha,\mathcal{W}}$. Let $q = (1 - e^{-\epsilon/2})q_{\alpha,\mathcal{W}_{\alpha,\epsilon}} + e^{-\epsilon/2}q_{\alpha,\mathcal{W}}$. Then using Lemma 8 and (59) we get

$$\sup_{w \in \mathcal{W}\setminus\mathcal{W}_{\alpha,\epsilon}} D_\alpha(w\|q) \leq C_{\alpha,\mathcal{W}} - \epsilon/2. \tag{A.39}$$

The convexity of the Rényi divergence in its second argument, i.e. Lemma 11, and Theorem 1 imply that

$$
\begin{aligned}
&\sup_{w \in \mathcal{W}_{\alpha,\epsilon}} D_\alpha(w\|q) \\
&\leq \sup_{w \in \mathcal{W}_{\alpha,\epsilon}} \left[(1 - e^{-\epsilon/2})D_\alpha\big(w\|q_{\alpha,\mathcal{W}_{\alpha,\epsilon}}\big) + e^{-\epsilon/2}D_\alpha(w\|q_{\alpha,\mathcal{W}}) \right] \\
&\leq (1 - e^{-\epsilon/2})S_{\alpha,\mathcal{W}_{\alpha,\epsilon}}(q_{\alpha,\mathcal{W}_{\alpha,\epsilon}}) + e^{-\epsilon/2}S_{\alpha,\mathcal{W}_{\alpha,\epsilon}}(q_{\alpha,\mathcal{W}}) \\
&= (1 - e^{-\epsilon/2})C_{\alpha,\mathcal{W}_{\alpha,\epsilon}} + e^{-\epsilon/2}C_{\alpha,\mathcal{W}}. \tag{A.40}
\end{aligned}
$$

If $q_{\alpha,\mathcal{W}_{\alpha,\epsilon}} \neq q_{\alpha,\mathcal{W}}$, then $C_{\alpha,\mathcal{W}_{\alpha,\epsilon}} < C_{\alpha,\mathcal{W}}$ by Lemma 21. Then $S_{\alpha,\mathcal{W}}(q) < C_{\alpha,\mathcal{W}}$ by (A.39) and (A.40). However, this is a contradiction by Theorem 1. Thus $q_{\alpha,\mathcal{W}_{\alpha,\epsilon}} = q_{\alpha,\mathcal{W}}$ holds.

As a result of the definition of $\mathcal{W}_{\alpha,\epsilon}$, an element of $\mathcal{W}$ is in $\mathcal{W}_{\alpha,0}$ iff it is in $\mathcal{W}_{\alpha,\epsilon}$ for all $\epsilon > 0$, i.e. $\bigcap_{\epsilon>0} \mathcal{W}_{\alpha,\epsilon} = \mathcal{W}_{\alpha,0}$. Consequently, if $\mathcal{W}$ is a finite set, then $\mathcal{W}_{\alpha,\epsilon} = \mathcal{W}_{\alpha,0}$ for small enough ϵ. Then $C_{\alpha,\mathcal{W}_{\alpha,0}} = C_{\alpha,\mathcal{W}_{\alpha,\epsilon}} = C_{\alpha,\mathcal{W}}$. Furthermore, (60) holds for $\epsilon = 0$ because (60) holds for all $\epsilon > 0$.

For arbitrary $\mathcal{W}$'s, identity $\bigcap_{\epsilon>0}\mathcal{W}_{\alpha,\epsilon} = \mathcal{W}_{\alpha,0}$ does not imply that $\mathcal{W}_{\alpha,\epsilon} = \mathcal{W}_{\alpha,0}$ for some $\epsilon > 0$. $\mathcal{W}_{\alpha,0}$ can be the empty set or a non-empty set such that $C_{\alpha,\mathcal{W}_{\alpha,0}} < C_{\alpha,\mathcal{W}}$, see Example 4. $\quad\square$

Proof of Lemma 24.

(a) $C_{\alpha,\mathcal{W}} \leq C_{\alpha,\mathrm{ch}\mathcal{W}}$ by definition because $\mathcal{W} \subset \mathrm{ch}\mathcal{W}$. If $C_{\alpha,\mathcal{W}} = \infty$, then the reverse inequality $C_{\alpha,\mathrm{ch}\mathcal{W}} \leq C_{\alpha,\mathcal{W}}$ holds trivially. If $C_{\alpha,\mathcal{W}} < \infty$, then $\exists! q_{\alpha,\mathcal{W}}$ satisfying $\sup_{w \in \mathcal{W}} D_\alpha(w\|q_{\alpha,\mathcal{W}}) = C_{\alpha,\mathcal{W}}$ by Theorem 1. Then as a result of the quasi-convexity of the Rényi divergence in its first argument, i.e. Lemma 12, we have

$$
\begin{aligned}
D_\alpha(\mu_{1,p}\|q_{\alpha,\mathcal{W}}) &\leq \max_{w\in\mathrm{supp}(p)} D_\alpha(w\|q_{\alpha,\mathcal{W}}) \\
&\leq C_{\alpha,\mathcal{W}}
\end{aligned}
$$

for all $p \in \mathcal{P}(\mathcal{W})$. Then $C_{\alpha,\mathrm{ch}\mathcal{W}} \leq C_{\alpha,\mathcal{W}}$ by Theorem 1.

(b) $C_{\alpha,\mathcal{W}} \leq C_{\alpha,\mathrm{cl}\mathcal{W}}$ by definition because $\mathcal{W} \subset \mathrm{cl}\mathcal{W}$. If $C_{\alpha,\mathcal{W}} = \infty$, then the reverse inequality $C_{\alpha,\mathrm{cl}\mathcal{W}} \leq C_{\alpha,\mathcal{W}}$ holds trivially. If $C_{\alpha,\mathcal{W}} < \infty$, then $\exists! q_{\alpha,\mathcal{W}}$ satisfying $\sup_{w \in \mathcal{W}} D_\alpha(w\|q_{\alpha,\mathcal{W}}) = C_{\alpha,\mathcal{W}}$ by Theorem 1. Furthermore, for all $v \in \mathcal{P}(\mathcal{Y})$ and $\epsilon > 0$ there exists an open set $\mathcal{N}$ containing v, i.e. a neighborhood of v, such that

$$D_\alpha(v\|q_{\alpha,\mathcal{W}}) - \epsilon < D_\alpha(s\|q_{\alpha,\mathcal{W}}) \qquad \forall s \in \mathcal{N}$$

by the lower semicontinuity, i.e. Lemma 13. If $v \in \mathrm{cl}\mathcal{W}$, then every open set containing v contains a member of $\mathcal{W}$. Thus $D_\alpha(v\|q_{\alpha,\mathcal{W}}) - \epsilon < C_{\alpha,\mathcal{W}}$ for every v in $\mathrm{cl}\mathcal{W}$

and positive ϵ. Then[34] $D_\alpha(v\|q_{\alpha,\mathcal{W}}) \leq C_{\alpha,\mathcal{W}}$ for every $v \in \mathrm{cl}\mathcal{W}$ and $C_{\alpha,\mathrm{cl}\mathcal{W}} \leq C_{\alpha,\mathcal{W}}$ by Theorem 1.

The closure of $\mathcal{W}$ for a topology stronger than the topology of setwise convergence is a subset of the closure of $\mathcal{W}$ for the topology of setwise convergence and a superset of $\mathcal{W}$. Thus its Rényi capacity is bounded from below and from above by $C_{\alpha,\mathcal{W}}$.

(c) If $C_{\eta,\mathcal{W}} < \infty$, then as a result Theorem 1, Lemma 14, and (22) there exists a unique $q_{\eta,\mathcal{W}}$ satisfying

$$D_\eta(q_{\eta,p}\|q_{\eta,\mathcal{W}}) \leq C_{\eta,\mathcal{W}} - I_\eta(p;\mathcal{W}) \quad \forall p \in \mathcal{P}(\mathcal{W}).$$

If $\eta > 1$, then using the definitions of Rényi information and divergence given in (13) and (19) we get

$$\int \big(\tfrac{\mathrm{d}\mu_{\eta,p}}{\mathrm{d}\nu}\big)^\eta \big(\tfrac{\mathrm{d}q_{\eta,\mathcal{W}}}{\mathrm{d}\nu}\big)^{1-\eta}\nu(\mathrm{d}y) \leq e^{(\eta-1)C_{\eta,\mathcal{W}}} \quad \forall p \in \mathcal{P}(\mathcal{W}).$$

Since $C_{\eta,\mathcal{W}}$ is finite this implies that $\mu_{\eta,p} \prec q_{\eta,\mathcal{W}}$. On the other hand $\mu_{\alpha,p}$ is nondecreasing in α, in the sense that if $\alpha < \eta$ then $\mu_{\alpha,p} \leq \mu_{\eta,p}$, by Lemma 3-(a,b). Hence,

$$\int \big(\tfrac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}q_{\eta,\mathcal{W}}}\big)^\eta q_{\eta,\mathcal{W}}(\mathrm{d}y) \leq e^{(\eta-1)C_{\eta,\mathcal{W}}}$$

for all p in $\mathcal{P}(\mathcal{W})$ and α in $[0,\eta]$. Then $\tfrac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}q_{\eta,\mathcal{W}}}$'s are $q_{\eta,\mathcal{W}}$-integrable and the set $\{\tfrac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}q_{\eta,\mathcal{W}}} : p \in \mathcal{P}(\mathcal{W}),\ \alpha \in [0,\eta]\}$ satisfies the necessary and sufficient condition for the uniform integrability[35] determined by de la Vallée Poussin [8, Thm. 4.5.9], for the growth function $G(x) = x^\eta$. But when the reference measure is finite, the uniform integrability is equivalent to the uniform absolute continuity of the integrals and boundedness in $\mathcal{L}^1(q_{\eta,\mathcal{W}})$ by [8, Thm. 4.5.3], which in our case is nothing but the uniform absolute continuity with respect to $q_{\eta,\mathcal{W}}$ and boundedness in total variation norm for the set of all mean measures. Thus $\{\mu_{\alpha,p} : p \in \mathcal{P}(\mathcal{W}),\ \alpha \in [0,\eta]\} \prec^{uni} q_{\eta,\mathcal{W}}$.

On the other hand by [8, Thm. 4.7.25], a set of measures is uniformly absolutely continuous with respect to a finite measure and bounded in variation norm iff it has compact closure in the topology of setwise convergence. A set of measures has compact closure in the topology of setwise convergence iff it has compact closure in the weak topology by [8, Thm. 4.7.25], as well.

If $\eta = 1$, then using $x \ln x \geq -1/e$, $\|\mu_{1,p}\| = 1$, and the definition of the Rényi divergence given in (19) we get,

$$\int G\left(\tfrac{\mathrm{d}\mu_{1,p}}{\mathrm{d}q_{1,\mathcal{W}}}\right) q_{1,\mathcal{W}}(\mathrm{d}y) \leq C_{1,\mathcal{W}} - I_1(p;\mathcal{W}) + \tfrac{1}{e} + 1$$

---

[34]This observation is nothing but the definition of the continuity: A function $f : \mathcal{X} \to \mathcal{Z}$ is continuous iff for any $\mathcal{A} \subset \mathcal{X}$, $f(\mathrm{cl}\mathcal{A}) \subset \mathrm{cl}f(\mathcal{A})$ by [68, Thm. 18.1]. If we chose $\mathcal{X}$ to be $\mathcal{P}(\mathcal{Y})$ with the topology of setwise convergence, $\mathcal{Z}$ to be $(-\infty, \infty]$ with the topology generated by the sets of the form $(z, \infty]$ for $z \in \mathbb{R}$, and f to be $f(\cdot) = D_\alpha(\cdot\|q_{\alpha,\mathcal{W}})$, then the lower semicontinuity of the Rényi divergence in its first argument is equivalent to the continuity of f. On the other hand, $f(\mathcal{W}) \subset (-\infty, C_{\alpha,\mathcal{W}}]$ by Theorem 1 and $(-\infty, C_{\alpha,\mathcal{W}}]$ is a closed set for the topology we have chosen for $(-\infty, \infty]$. Thus $f(\mathrm{cl}\mathcal{W}) \subset \mathrm{cl}f(\mathcal{W}) \subset (-\infty, C_{\alpha,\mathcal{W}}]$, i.e. $D_\alpha(v\|q_{\alpha,\mathcal{W}}) \leq C_{\alpha,\mathcal{W}}$ for all $v \in \mathrm{cl}\mathcal{W}$.

[35]A set of $q_{\eta,\mathcal{W}}$-integrable functions is uniformly integrable iff it has compact closure in the weak topology of $\mathcal{L}^1(q_{\eta,\mathcal{W}})$ by Dunford-Pettis theorem [8, 4.7.18]. Thus $\{\tfrac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}q_{\eta,\mathcal{W}}} : p \in \mathcal{P}(\mathcal{W}),\ \alpha \in [0,\eta]\}$ has compact closure in the weak topology of $\mathcal{L}^1(q_{\eta,\mathcal{W}})$. Since we have chosen to work with the space of measures rather than the space of integrable functions we have stated our result in terms of relative compactness in the space of measures rather than integrable functions.

for all $p \in \mathcal{P}(\mathcal{W})$ where $G(x) = x\mathbb{1}_{\{0 \leq x < e\}} + x \ln x\mathbb{1}_{\{x \geq e\}}$. Since mean measure is a nondecreasing function of the order by Lemma 3-(a,b) and $G(x)$ is an increasing function of x, we have

$$\int G\left(\frac{d\mu_{\alpha,p}}{dq_{1,\mathcal{W}}}\right) q_{1,\mathcal{W}}(dy) \leq C_{1,\mathcal{W}} + \frac{1}{e} + 1$$

for all p in $\mathcal{P}(\mathcal{W})$ and α in $[0,1]$. The rest of the proof for $\eta = 1$ case is identical to that of $\eta > 1$ case.

(d) The equivalence of the last three statements to one another is a version of Dunford-Pettis theorem [8, 4.7.25]. Thus we will only prove the equivalence of the first two statements.

Let us first prove the direct part: if there exists a μ in $\mathcal{P}(\mathcal{Y})$ satisfying $\mathcal{W} \prec^{uni} \mu$, then $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} = 0$. Note that $C_{\alpha,\mathcal{W}} \leq \sup_{w \in \mathcal{W}} D_\alpha(w \| q)$ for all $\alpha \in (0,1)$ and $q \in \mathcal{P}(\mathcal{Y})$ by Theorem 1. Thus using (19) we get

$$\limsup_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} \leq \limsup_{\alpha \uparrow 1} \sup_{w \in \mathcal{W}} D_{1-\alpha}(\mu \| w)$$
$$\leq \limsup_{\alpha \uparrow 1} \sup_{w \in \mathcal{W}} \frac{-1}{\alpha} \ln \mathbf{E}_\mu\left[\left(\frac{dw}{d\mu}\right)^\alpha\right]. \quad \text{(A.41)}$$

Since $\mathcal{W} \prec^{uni} \mu$, for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $\mu(\mathcal{E}) \leq \delta$ for an $\mathcal{E} \in \mathcal{Y}$, then $w(\mathcal{E}) \leq \epsilon$ for all $w \in \mathcal{W}$. On the other hand $\mu(\frac{dw}{d\mu} > \frac{1}{\delta}) \leq \delta$ by Markov inequality. Hence

$$w\left(\frac{dw}{d\mu} > \frac{1}{\delta}\right) \leq \epsilon. \quad \text{(A.42)}$$

On the other hand using (19) we get

$$\mathbf{E}_\mu\left[\left(\frac{dw}{d\mu}\right)^\alpha\right] \geq \mathbf{E}_w\left[\left(\frac{dw}{d\mu}\right)^{\alpha-1}\mathbb{1}_{\{\frac{dw}{d\mu} \in (0,\frac{1}{\delta}]\}}\right]$$
$$\geq \delta^{1-\alpha}\left(1 - w\left(\frac{dw}{d\mu} > \frac{1}{\delta}\right)\right) \qquad \forall w \in \mathcal{W}.$$

Then as a result of (A.41) and (A.42) we have

$$\limsup_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} \leq \frac{1}{1-\epsilon} \qquad \forall \epsilon > 0.$$

Then $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} = 0$ because $C_{\alpha,\mathcal{W}} \geq 0$.

We are left with proving the converse statement: if $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} = 0$, then there exists a $\mu \in \mathcal{P}(\mathcal{Y})$ such that $\mathcal{W} \prec^{uni} \mu$. We start with proving the following statement about the Rényi centers: For every $\epsilon > 0$ there exists a (ϕ, δ) pair such that $\phi \in (0,1)$, $\delta \in (0,\epsilon)$, and if $q_{\phi,\mathcal{W}}(\mathcal{E}) \leq \delta$, then $w(\mathcal{E}) < \epsilon$ for all $w \in \mathcal{W}$.

For any $\epsilon > 0$ there exists a $\phi \in (0,1)$ such that $e^{\frac{\phi-1}{\phi} C_{\phi,\mathcal{W}}} > 1 - \frac{\epsilon}{2}$ because $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,\mathcal{W}} = 0$. On the other hand, $D_\alpha^{\sigma(\{\mathcal{E}\})}(w \| q_{\alpha,\mathcal{W}}) \leq C_{\alpha,\mathcal{W}}$ for any $w \in \mathcal{W}$ and $\mathcal{E} \in \mathcal{Y}$, as a result of Lemma 9 and Theorem 1. Then the above described ϕ satisfies

$$f(w(\mathcal{E}), q_{\phi,\mathcal{W}}(\mathcal{E})) \geq \left(1 - \frac{\epsilon}{2}\right)^\phi \quad \text{(A.43)}$$

for all $\mathcal{E} \in \mathcal{Y}$ and $w \in \mathcal{W}$ where the function $f(x,z)$ is defined for all $x \in [0,1]$ and $z \in [0,1]$ as

$$f(x,z) \triangleq x^\phi z^{1-\phi} + (1-x)^\phi (1-z)^{1-\phi}.$$

Given $\epsilon \in (0, 0.5)$ and the corresponding $\phi \in (0,1)$, let δ be the unique z in $(0,\epsilon)$ satisfying $f(\epsilon, z) = (1 - \epsilon/2)^\phi$. Such a z exists because $f(\epsilon, 0) = (1-\epsilon)^\phi$, $f(\epsilon, \epsilon) = 1$

and $f(\epsilon, z)$ is monotone increasing and continuous in z on $[0, \epsilon]$. On the other hand $f(x, z) < f(\epsilon, \delta) = (1 - \epsilon/2)^\phi$ for any $z \in [0, \delta)$ and $x \in [\epsilon, 1]$ because $f(x, z)$ is monotone increasing in z on $[0, x]$ for any $x \in (0, 1]$ and monotone decreasing in x on $[z, 1]$ for any $z \in [0, 1)$. Hence, using (A.43) we can conclude that if $q_{\phi,\mathcal{W}}(\mathcal{E}) < \delta$ for a $\mathcal{E} \in \mathcal{Y}$, then $w(\mathcal{E}) < \epsilon$ for all $w \in \mathcal{W}$. In the following we use this property to construct a μ such that $\mathcal{W} \prec^{uni} \mu$.

Let μ be $\sum_{\imath \in \mathbb{Z}_+} \frac{q_{\phi_\imath,\mathcal{W}}}{2^\imath}$ where $(\phi_\imath, \delta_\imath)$ is the pair associate with $\epsilon = \frac{1}{\imath}$. Then for any $\mathcal{E} \in \mathcal{Y}$ and $\imath \in \mathbb{Z}_+$, if $\mu(\mathcal{E}) \leq \frac{\delta_\imath}{2^\imath}$, then $q_{\phi_\imath,\mathcal{W}}(\mathcal{E}) \leq \delta_\imath$ and consequently $w(\mathcal{E}) \leq 1/\imath$ for all w in $\mathcal{W}$. Thus for any $\epsilon > 0$ if $\mu(\mathcal{E}) \leq \frac{\delta_{\lceil 1/\epsilon \rceil}}{2^{\lceil 1/\epsilon \rceil}}$ for an $\mathcal{E} \in \mathcal{Y}$, then $w(\mathcal{E}) < \epsilon$ for all w in $\mathcal{W}$. $\square$

REFERENCES

[1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28(1):131–142, 1966.

[2] Dan Amir. *Characterizations of Inner Product Spaces*, volume 20 of *Operator Theory: Advances and Applications*. Birkhäuser, Basel, 1 edition, 1986.

[3] E. Arikan. An inequality on guessing and its application to sequential decoding. *IEEE Transactions on Information Theory*, 42(1):99–105, Jan 1996.

[4] S. Arimoto. Information measures and capacity of order α for discrete memoryless channels. In I. Csiszár and P. Elias, editors, *Topics in Information Theory*, volume 16, pages 41–52, Keszthely, 1977. Math. Soc. J. Bolyai, North-Holland, Amsterdam.

[5] U. Augustin. Error estimates for low rate codes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(1):61–88, 1969.

[6] Udo Augustin. *Noisy Channels*. Habilitation thesis, Universität Erlangen-Nürnberg, 1978. (http://bit.ly/2ID8h7m).

[7] S. G. Bobkov and G. P. Chistyakov. Entropy power inequality for the Rényi entropy. *IEEE Transactions on Information Theory*, 61(2):708–714, Feb 2015.

[8] Vladimir I. Bogachev. *Measure Theory*. Springer-Verlag, Berlin Heidelberg, 2007.

[9] Pierre Bremaud. *Point Processes and Queues: Martingale Dynamics*. Springer Series in Statistics. Springer, New York, NY, 1981.

[10] C. Bunte and A. Lapidoth. Encoding tasks and Rényi entropy. *IEEE Transactions on Information Theory*, 60(9):5065–5076, Sept 2014.

[11] C. Bunte and A. Lapidoth. On the listsize capacity with feedback. *IEEE Transactions on Information Theory*, 60(11):6733–6748, Nov 2014.

[12] C. Bunte and A. Lapidoth. Maximum Rényi entropy rate. *IEEE Transactions on Information Theory*, 62(3):1193–1205, March 2016.

[13] M. V. Burnashev and Yu. A. Kutoyants. On the sphere-packing bound, capacity, and similar results for Poisson channels. *Problems of Information Transmission*, 35(2):95–111, 1999.

[14] Z. H. Chen, Z. H. Ma, I. Nikoufar, and S. M. Fei. Sharp continuity bounds for entropy and conditional entropy. *Science China Physics, Mechanics & Astronomy*, 60(2):020321, Feb 2017.

[15] H. C. Cheng and M. H. Hsieh. Concavity of the auxiliary function for classical-quantum channels. *IEEE Transactions on Information Theory*, 62(10):5960–5965, Oct 2016.

[16] H. C. Cheng and M. H. Hsieh. Moderate deviation analysis for classical-quantum channels and quantum hypothesis testing. *IEEE Transactions on Information Theory*, 64(2):1385–1403, Feb 2018.

[17] H. C. Cheng, M. H. Hsieh, and M Tomamichel. Quantum sphere-packing bounds with polynomial prefactors. *IEEE Transactions on Information Theory*, 65(5):2872–2898, May 2019. (arXiv:1704.05703 [quant-ph]).

[18] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, 2 edition, 2006.

[19] I. Csiszár. Eine Informationstheoretisehe Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of Hungarian Academy of Sciences Series A*, 8(1-2):85–108, 1963.

[20] I. Csiszár. A note on Jensen's inequality. *Studia Scientiarum Mathematicarum Hungarica*, 1(1-2):185–188, 1966.

[21] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2(3-4):299–318, 1967.

[22] I. Csiszár. On topological properties of f-divergences. *Studia Scientiarum Mathematicarum Hungarica*, 2(3-4):329–339, 1967.

[23] I. Csiszár. A class of measures of informativity of observation channels. *Periodica Mathematica Hungarica*, 2(1-4):191–213, 1972.

[24] I. Csiszár. Arbitrarily varying channels with general alphabets and states. *IEEE Transactions on Information Theory*, 38(6):1725–1742, Nov 1992.

[25] I. Csiszár. Generalized cutoff rates and Rényi's information measures. *IEEE Transactions on Information Theory*, 41(1):26–34, Jan 1995.

[26] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, Cambridge, UK, 2011.

[27] M. Dalai. Lower bounds on the probability of error for classical and classical-quantum channels. *IEEE Transactions on Information Theory*, 59(12):8027–8056, Dec 2013.

[28] M. Davis. Capacity and cutoff rate for Poisson-type channels. *IEEE Transactions on Information Theory*, 26(6):710–715, Nov 1980.

[29] L. D. Davisson and A. Leon-Garcia. A source matching approach to finding minimax codes. *IEEE Transactions on Information Theory*, 26(2):166–174, Mar 1980.

[30] Richard M. Dudley. *Real analysis and probability*, volume 74. Cambridge University Press, New York, NY, 2002.

[31] T. van Erven and P. Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, July 2014.

[32] A. A. Fedotov, P. Harremoës, and F. Topsøe. Refinements of Pinsker's inequality. *IEEE Transactions on Information Theory*, 49(6):1491–1498, June 2003.

[33] S. Fehr and S. Berens. On the conditional Rényi entropy. *IEEE Transactions on Information Theory*, 60(11):6801–6810, Nov 2014.

[34] S. L. Fong and V. Y. F. Tan. Strong converse theorems for classes of multimessage multicast networks: A Rényi divergence approach. *IEEE Transactions on Information Theory*, 62(9):4953–4967, Sept 2016.

[35] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, 11(1):3–18, Jan. 1965.

[36] R. G. Gallager. Source coding with side information and universal coding. Tech. Rep. LIDS-P-937, MIT Lab. Inform. Decision Syst, Cambridge, MA, September 1979. http://web.mit.edu/gallager/www/papers/paper5.pdf.

[37] Robert G. Gallager. *Information theory and reliable communication*. John Wiley & Sons, Inc., New York, NY, 1968.

[38] Robert G. Gallager. *Stochastic processes : theory for applications*. Cambridge University Press, Cambridge, 2013.

[39] A. L. Garkavi. The best possible net and the best possible cross-section of a set in a normed space. *American Mathematical Society Translations: Series 2*, 39:111–132, 1964.

[40] G. L. Gilardoni. On the minimum f-divergence for given total variation. *Comptes Rendus Mathematique*, 343(11):763 – 766, 2006.

[41] G. L. Gilardoni. Corrigendum to the note "on the minimum f-divergence for given total variation"[c. r. acad. sci. paris, ser. i 343 (2006) 763-766]. *Comptes Rendus Mathematique*, 348(5):299, 2010.

[42] G. L. Gilardoni. On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences. *IEEE Transactions on Information Theory*, 56(11):5377–5386, Nov 2010.

[43] A. Guntuboyina, S. Saha, and G. Schiebinger. Sharp inequalities for f-divergences. *IEEE Transactions on Information Theory*, 60(1):104–121, Jan 2014.

[44] A. Gushchin and D. A. Zhdanov. A minimax result for f-divergences. In *From stochastic calculus to mathematical finance. The Shiryaev Festschrift.*, pages 287–294. Berlin: Springer, 2006.

[45] E. P. Hanson and N. Datta. Tight uniform continuity bound for a family of entropies. *arXiv:1707.04249v2 [quant-ph]*, 2017.

[46] E. A. Haroutunian. Estimates of the error probability exponent for a semicontinuous memoryless channel. *Problems of Information Transmission*, 4(4):37–48, 1968.

[47] E. A. Haroutunian. Lower bound for error probability in channels with feedback. *Problems of Information Transmission*, 13(2):36–44, 1977.

[48] P. Harremoës. *private communications*, August 2016.

[49] D. Haussler. A general minimax result for relative entropy. *IEEE Transactions on Information Theory*, 43(4):1276–1280, Jul 1997.

[50] M. Hayashi and V. Y. F. Tan. Equivocations, exponents and second-order coding rates under various Rényi information measures. *IEEE Transactions on Information Theory*, 63(2):975–1005, Feb 2017.

[51] Masahito Hayashi. *Quantum Information Theory: Mathematical Foundation*. Graduate Texts in Physics. Springer-Verlag, Berlin Heidelberg, 2 edition, 2017.

[52] S. W. Ho and S Verdú. Convexity/concavity of Rényi entropy and alpha-mutual information. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 745–749, June 2015.

[53] S. W. Ho and R. W. Yeung. On the discontinuity of the Shannon information measures. *IEEE Transactions on Information Theory*, 55(12):5362–5374, Dec 2009.

[54] J. Jeon. A generalized typicality for abstract alphabets. *arXiv:1401.6728v4 [cs.IT]*, 2014.

[55] Yu. M. Kabanov. The capacity of a channel of the Poisson type. *Theory of Probability & Its Applications*, 23(1):143–147, 1978.

[56] J. H. B. Kemperman. On the Shannon capacity of an arbitrary channel. *Indagationes Mathematicae (Proceedings)*, 77(2):101–115, 1974.

[57] H. Komiya. Elementary proof for Sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.

[58] R. König and S. Wehner. A strong converse for classical channel coding using entangled inputs. *Phys. Rev. Lett.*, 103(7):070504, Aug 2009.

[59] S. Kullback. A lower bound for discrimination information in terms of variation (corresp.). *IEEE Transactions on Information Theory*, 13(1):126–127, January 1967.

[60] S. Kullback. Correction to a lower bound for discrimination information in terms of variation. *IEEE Transactions on Information Theory*, 16(5):652–652, September 1970.

[61] M. A. Kumar and I. Sason. Projection theorems for the Rényi divergence on α-convex sets. *IEEE Transactions on Information Theory*, 62(9):4924–4935, Sept 2016.

[62] M. A. Kumar and R. Sundaresan. Minimization problems based on relative α-entropy I: Forward projection. *IEEE Transactions on Information Theory*, 61(9):5063–5080, Sept 2015.

[63] M. A. Kumar and R. Sundaresan. Minimization problems based on relative α-entropy II: Reverse projection. *IEEE Transactions on Information Theory*, 61(9):5081–5095, Sept 2015.

[64] P. Mitran. On a Markov lemma and typical sequences for Polish alphabets. *IEEE Transactions on Information Theory*, 61(10):5342–5356, Oct 2015.

[65] T. Morimoto. Markov processes and the h-theorem. *Journal of the Physical Society of Japan*, 18(3):328–331, 1963.

[66] M. Mosonyi and F. Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Transactions on Information Theory*, 57(4):2474–2487, April 2011.

[67] M. Mosonyi and T. Ogawa. Strong converse exponent for classical-quantum channel coding. *Communications in Mathematical Physics*, 355(1):373–426, Oct 2017.

[68] James R. Munkres. *Topology*. Prentice Hall Inc., Upper Saddle River, NJ 07458, 2000.

[69] B. Nakiboğlu. The Augustin center and the sphere packing bound for memoryless channels. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1401–1405, Aachen, Germany, June 2017.

[70] B. Nakiboğlu. The Augustin capacity and center. *arXiv:1803.07937 [cs.IT]*, 2018.

[71] B. Nakiboğlu. The sphere packing bound for memoryless channels. *arXiv:1804.06372 [cs.IT]*, 2018.

[72] B. Nakiboğlu. The sphere packing bound for DSPCs with feedback à la Augustin. *IEEE Transactions on Communications*, 2019. DOI:10.1109/TCOMM.2019.2931302, (arXiv:1806.11531 [cs.IT]).

[73] B. Nakiboğlu. The sphere packing bound via Augustin's method. *IEEE Transactions on Information Theory*, 65(2):816–840, Feb 2019. (arXiv:1611.06924 [cs.IT]).

[74] M. Ohya, D. Petz, and N. Watanabe. On capacities of quantum channels. *Probability and Mathematical Statistics*, 17(1):179–196, 1997.

[75] Y. Polyanskiy and S. Verdú. Arimoto channel coding converse and Rényi divergence. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 1327 – 1333, Oct 2010.

[76] M. Raginsky. Empirical processes, typical sequences, and coordinated actions in standard borel spaces. *IEEE Transactions on Information Theory*, 59(3):1288–1301, March 2013.

[77] E. Ram and I. Sason. On Rényi entropy power inequalities. *IEEE Transactions on Information Theory*, 62(12):6800–6815, Dec 2016.

[78] A. E. Rastegin. Some general properties of unified entropies. *Journal of Statistical Physics*, 143(6):1120, Jun 2011.

[79] A. Rényi. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–561, 1961.

[80] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, NY, 1976.

[81] B. Ryabko. Encoding of a source with unknown but ordered probabilities. *Problems of Information Transmission*, 15(2):134–138, October 1979.

[82] I. Sason. On the Rényi divergence, joint range of relative entropies, and a channel coding theorem. *IEEE Transactions on Information Theory*, 62(1):23–34, Jan 2016.

[83] I. Sason and S. Verdú. f-divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, Nov 2016.

[84] I. Sason and S. Verdú. Arimoto-Rényi conditional entropy and bayesian M-ary hypothesis testing. *IEEE Transactions on Information Theory*, 64(1):4–25, Jan 2018.

[85] I. Sason and S. Verdú. Improved bounds on lossless source coding and guessing moments via Rényi measures. *IEEE Transactions on Information Theory*, 64(6):4323–4346, June 2018.

[86] B. Schumacher and M. D. Westmoreland. Optimal signal ensembles. *Phys. Rev. A*, 63(2):022308, Jan 2001.

[87] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal, The*, 27(3 and 4):379–423 and 623–656, July and October 1948.

[88] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. I. *Information and Control*, 10(1):65–103, 1967.

[89] Claude E Shannon and Warren Weaver. *The mathematical theory of communication*. University of Illinois Press, Urbana, Illinois, 1949.

[90] N. Sharma and N. A. Warsi. Fundamental bound on the reliability of quantum information transmission. *Phys. Rev. Lett.*, 110(8):080501, Feb 2013.

[91] O. Shayevitz. On Rényi measures and hypothesis testing. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 894–898, Austin, Texas, July 2011.

[92] A. Yu Sheverdyaev. Lower bound for error probability in a discrete memoryless channel with feedback. *Problems of Information Transmission*, 18(4):5–15, 1982.

[93] Albert N. Shiryaev. *Probability*. Springer-Verlag, New York, NY, 1995.

[94] R. Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 14(2):149–160, 1969.

[95] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[96] R. Sundaresan. Guessing under source uncertainty. *IEEE Transactions on Information Theory*, 53(1):269–287, Jan 2007.

[97] V. Y. F. Tan and M. Hayashi. Analysis of remaining uncertainties and exponents under various conditional Rényi entropies. *IEEE Transactions on Information Theory*, 64(5):3734–3755, May 2018.

[98] A. Teixeira, A. Matos, and L. Antunes. Conditional Rényi entropies. *IEEE Transactions on Information Theory*, 58(7):4273–4277, July 2012.

[99] Bui Van Thanh. Semi-continuous arbitrarily varying channels with general state constraints. *ACTA MATHEMATICA VIETNAMICA*, 17(1):31–50, 1992.

[100] M. Tomamichel and M. Hayashi. Operational interpretation of Rényi information measures via composite hypothesis testing against product and markov distributions. *IEEE Transactions on Information Theory*, 64(2):1064–1082, Feb 2018.

[101] F. Topsøe. An information theoretical identity and a problem involving capacity. *Studia Scientiarum Mathematicarum Hungarica*, 2:291–292, 1967.

[102] F. Topsøe. A new proof of a result concerning computation of the capacity for a discrete channel. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 22(2):166–168, 1972.

[103] I. Vajda. Note on discrimination information and variation (corresp.). *IEEE Transactions on Information Theory*, 16(6):771–773, November 1970.

[104] R. Venkataramanan and O. Johnson. A strong converse bound for multiple hypothesis testing, with applications to high-dimensional estimation. *Electronic Journal of Statistics*, 12(1):1126–1149, 2018.

[105] S. Verdú. α-mutual information. In *2015 Information Theory and Applications Workshop*, 2015.

[106] M. M. Wilde, A. Winter, and D. Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Communications in Mathematical Physics*, 331(2):593–622, Oct 2014.

[107] A. D. Wyner. Capacity and error exponent for the direct detection photon channel. I. *IEEE Transactions on Information Theory*, 34(6):1449–1461, Nov 1988.

[108] A. D. Wyner. Capacity and error exponent for the direct detection photon channel. II. *IEEE Transactions on Information Theory*, 34(6):1462–1471, Nov 1988.

In the following, unless specified explicitly to be otherwise all $\sum_w$, $\prod_w$, $\vee_w$, $\max_w$, $\min_w$ stand for the corresponding expression with the subscript "$w : p(w) > 0$."

D. Proofs of the Lemmas on the Mean Measure

Proof of Lemma 1.

(1-a) For any $\tilde{w}$ such that $p(\tilde{w}) > 0$ and $\alpha \in \mathbb{R}_+$, the following inequalities hold ν-a.e.

$$(p(\tilde{w}))^{1/\alpha} \frac{\mathrm{d}\tilde{w}}{\mathrm{d}\nu} \leq \left(\sum_w p(w) \left(\frac{\mathrm{d}w}{\mathrm{d}\nu} \right)^\alpha \right)^{1/\alpha} \leq \bigvee_w \frac{\mathrm{d}w}{\mathrm{d}\nu} \leq \sum_w \frac{\mathrm{d}w}{\mathrm{d}\nu}.$$

Then for any $\tilde{w}$ such that $p(\tilde{w}) > 0$, $\alpha \in (0, \infty]$, and $\mathcal{E} \in \mathcal{Y}$,

$$(p(\tilde{w}))^{\frac{1}{\alpha}} \tilde{w}(\mathcal{E}) \leq \mu_{\alpha,p}(\mathcal{E}) \leq \sum_w w(\mathcal{E}).$$

Thus for any $\alpha \in (0, \infty]$, $\mu_{\alpha,p}(\mathcal{E}) = 0$ iff $w(\mathcal{E}) = 0$ for all w such that $p(w) > 0$. Then $\mu_{1,p} \sim \mu_{\alpha,p}$ for all $\alpha \in (0, \infty]$. Note that $\|w\| = 1$ for all w in $\mathcal{P}(\mathcal{Y})$, then $\|\mu_{1,p}\| = 1$ for all p. Furthermore, there exists a $\tilde{w}$ such that $p(\tilde{w}) \geq \frac{1}{|\mathrm{supp}(p)|}$ for all p, then $|\mathrm{supp}(p)|^{-1/\alpha} \leq \|\mu_{\alpha,p}\| \leq |\mathrm{supp}(p)|$.

(1-b) As a result of the Hölder's inequality,

$$\begin{aligned} \mu_{0,p}(\mathcal{E}) &= \int_\mathcal{E} \prod_w \left(\frac{\mathrm{d}w}{\mathrm{d}\nu} \right)^{p(w)} \nu(\mathrm{d}y) \\ &\leq \prod_w \left(\int_\mathcal{E} \frac{\mathrm{d}w}{\mathrm{d}\nu} \nu(\mathrm{d}y) \right)^{p(w)} \\ &= \prod_w (w(\mathcal{E}))^{p(w)}. \end{aligned}$$

Then $\mu_{0,p}(\mathcal{E}) = 0$ whenever $w(\mathcal{E}) = 0$ and $\mu_{0,p} \prec w$ for all w such that $p(w) > 0$. Since $w(\mathcal{Y}) = 1$ for all w in $\mathcal{P}(\mathcal{Y})$, $\|\mu_{0,p}\| = \mu_{0,p}(\mathcal{Y}) \leq 1$.

$\square$

Proof of Lemma 2.

(2-a) Let us establish the expressions for $\pi_{\alpha,p}(y)$ and $p_{[\alpha]}(w|y)$, first. Note that $p_{[1]}(w|y) = p(w) \frac{\mathrm{d}w}{\mathrm{d}\mu_{1,p}}$ for all w such that $p(w) > 0$ by the definition of $p_{[\alpha]}(w|y)$ given in (10). Then the expressions for $\pi_{\alpha,p}(y)$ follows from the definitions of $\frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu}$ and $\pi_{\alpha,p}$ given in (7) and (9), respectively.

On the other hand, $w \prec \mu_{1,p}$ for all w such that $p(w) > 0$ by definition and $\mu_{\alpha,p} \sim \mu_{1,p}$ by Lemma 1. Thus,

$$\begin{aligned} \frac{\mathrm{d}w}{\mathrm{d}\mu_{\alpha,p}} &= \frac{\mathrm{d}\mu_{1,p}}{\mathrm{d}\mu_{\alpha,p}} \frac{\mathrm{d}w}{\mathrm{d}\mu_{1,p}} \\ &= \frac{1}{\pi_{\alpha,p}} \frac{p_{[1]}(w|y)}{p(w)} \qquad\qquad \forall w : p(w) > 0. \end{aligned}$$

Then the expression for $p_{[\alpha]}(w|y)$ follows from its definition given in (10).

In order to bound $\pi_{\alpha,p}$ from below and from above $\mu_{1,p}$-a.e. we use the expression for $\pi_{\alpha,p}$ we have just derived. Note that $\left(\sum_w p_{[1]}(w|y)^\alpha \right)^{1/\alpha} \geq 1$ for $\alpha \in (0, 1]$. Then

$$\begin{aligned} \left(\sum_w p_{[1]}(w|y)^\alpha p(w)^{1-\alpha} \right)^{1/\alpha} &\geq \left(\sum_w p_{[1]}(w|y)^\alpha \delta^{1-\alpha} \right)^{1/\alpha} \\ &\geq \delta^{\frac{1-\alpha}{\alpha}}. \end{aligned}$$

On the other hand, as a result of the Hölder's inequality we have

$$\begin{aligned} \left(\sum_w p_{[1]}(w|y)^\alpha p(w)^{1-\alpha} \right)^{1/\alpha} &\leq \left(\sum_w p_{[1]}(w|y) \right) \left(\sum_w p(w) \right)^{\frac{1-\alpha}{\alpha}} \\ &= 1. \end{aligned}$$

Thus $\delta^{\frac{1-\alpha}{\alpha}} \leq \pi_{\alpha,p} \leq 1$ for $\alpha \in (0, 1]$.

In order to obtain the bound for α in $[1, \infty)$, we use the identity $\left(\sum_w p_{[1]}(w|y)^\alpha \right)^{1/\alpha} \leq 1$, which is valid for all α in $[1, \infty)$, together with the reverse Hölder's inequality.

(2-b) $\frac{p_{[1]}(w|y)}{p(w)}$ is a non-negative real number for all w such that $p(w) > 0$ and $\frac{p_{[1]}(w|y)}{p(w)}$ is positive at least for one such w. Then expression for $\pi_{\alpha,p}$ given in part (a) is a smooth function[36] of α on $\mathbb{R}_+$. Identities for the derivatives of $\pi_{\alpha,p}$ follow from the chain rule and elementary rules of differentiation.

---

[36]For any positive integer K, non-negative real numbers a_i and b_i for i in $\{1, 2, \ldots, K\}$, the function $(\sum_{i=1}^K a_i b_i^\alpha)^{1/\alpha}$ is a smooth function of α on $\mathbb{R}_+$, because the exponential function and the logarithm are smooth functions and composition, sum, and product of smooth functions are also smooth.

(2-c) As a result of the Hölder's inequality we have,

$$\left(\pi_{\alpha_\beta,p}\right)^{\alpha_\beta} = \sum_w p(w) \left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{\alpha_1 \beta} \left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{\alpha_0(1-\beta)}$$

$$\leq \left(\sum_w p(w) \left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{\alpha_1}\right)^{\beta} \left(\sum_w p(w) \left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{\alpha_0}\right)^{(1-\beta)}$$

$$= (\pi_{\alpha_1,p})^{\alpha_1 \beta}(\pi_{\alpha_0,p})^{\alpha_0(1-\beta)}$$

Furthermore, the inequality is strict unless there exists a γ such that $p(w)(\frac{p_{[1]}(w|y)}{p(w)})^{\alpha_1} = \gamma p(w)(\frac{p_{[1]}(w|y)}{p(w)})^{\alpha_0}$ for all w such that $p_{[1]}(w|y) > 0$. Thus inequality is strict iff there exist $w, \tilde{w} \in \text{supp}(p)$ such that $\frac{p_{[1]}(w|y)}{p(w)} > \frac{p_{[1]}(\tilde{w}|y)}{p(\tilde{w})} > 0$.

(2-d) The continuity of $\pi_{\alpha,p}$ in α on $\mathbb{R}_+$ follows from the smoothness of $\pi_{\alpha,p}$ established in part (b). In order to show the continuity on $[0,\infty]$ we need to establish the continuity at zero and at infinity. Note that x^α is a smooth function of α for any $x \in \mathbb{R}_+$ and weighted sums of smooth functions are also smooth. Thus $(\pi_{\alpha,p})^\alpha(y)$ is a smooth function of α and we can use L'Hospital's rule [80, Thm. 5.13] for calculating the limits of $\pi_{\alpha,p}(y)$ at zero and infinity:

$$\lim_{\alpha \to 0} \left(\sum_w p(w) \left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{\alpha}\right)^{1/\alpha} = \prod_w \left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{p(w)} \tag{A.44}$$

$$\lim_{\alpha \to \infty} \left(\sum_w p(w) \left(\frac{p_{[1]}(w|y)}{p(w)}\right)^{\alpha}\right)^{1/\alpha} = \max_w \frac{p_{[1]}(w|y)}{p(w)} \tag{A.45}$$

Thus $\lim_{\alpha \to 0} \pi_{\alpha,p}(y) = \pi_{0,p}(y)$ and $\lim_{\alpha \to \infty} \pi_{\infty,p}(y) = \pi_{\infty,p}(y)$ hold $\mu_{1,p}$ almost everywhere. Thus $\pi_{\alpha,p}(y)$ is continuous on $[0,\infty]$.

On the other hand, using the Jensen's inequality and the convexity of the function $\ln 1/x$ we get,

$$\pi'_{\alpha,p} \geq -\frac{\pi_{\alpha,p}}{\alpha^2} \ln \sum_{w:p_{[1]}(w|y)>0} p(w)$$

$$\geq 0. \tag{A.46}$$

Since the function $\ln 1/x$ is strictly convex, the first inequality is strict and $\pi'_{\alpha,p}(y)$ is positive unless $p(w) = p_{[1]}(w|y)$ for all w such that $p(w) > 0$. Thus $\pi_{\alpha,p}(y)$ is monotone increasing in α unless $p(w) = p_{[1]}(w|y)$ for all w such that $p(w) > 0$. Boundedness is already established in part (a).

$\square$

Proof of Lemma 3.

(3-a) For all $y \in \mathcal{Y}$ —except for a $\mu_{1,p}$-measure zero set— density $\pi_{\alpha,p}$ is a non-negative function of α continuous on $[0,\infty]$ by Lemma 2-(d). Thus for any sequence $\{\alpha_\imath\}$ such that $\lim_{\imath \to \infty} \alpha_\imath = \alpha$ we have $\lim_{\imath \to \infty} \pi_{\alpha_\imath,p} = \pi_{\alpha,p}$ $\mu_{1,p}$-a.e. Since $\pi_{\alpha_\imath,p} \leq \pi_{\infty,p}$ by Lemma 2-(d) and $\pi_{\infty,p} \leq \frac{1}{\min_w p(w)}$ by Lemma 2-(a), we can apply the dominated convergence theorem [8, 2.8.1]. Thus $\{\pi_{\alpha_\imath,p}\} \xrightarrow{\mathcal{L}^1(\mu_{1,p})} \pi_{\alpha,p}$, i.e.

$$\lim_{\imath \to \infty} \int |\pi_{\alpha_\imath,p} - \pi_{\alpha,p}|\mu_{1,p}(\mathrm{d}y) = 0.$$

Then $\{\mu_{\alpha_\imath,p}\}$ converges to $\mu_{\alpha,p}$ in the total variation topology, for any sequence $\{\alpha_\imath\}$ such that $\lim_{\imath \to \infty} \alpha_\imath = \alpha$. Then $\mu_{\alpha,p}$ is a continuous function of α from $[0,\infty]$ with its usual topology to $\mathcal{M}_0^+(\mathcal{Y})$ with the total variation topology because $[0,\infty]$ with its usual topology is a metrizable space, see [68, Thm. 21.3].

(3-b) For $\mu'_{\alpha,p}$ defined in (11) to be a finite measure, $\pi'_{\alpha,p}$ should be a non-negative $\mu_{1,p}$-integrable function. The density $\pi'_{\alpha,p}$ is non-negative by (A.46). By the expression for $\pi'_{\alpha,p}$ given Lemma 2-(b) and the bound for $\pi'_{\alpha,p}$ given in Lemma 2-(a) we have

$$\pi'_{\alpha,p} = \frac{\pi_{\alpha,p}}{\alpha^2} \sum_w p_{[\alpha]}(w|y) \ln \frac{p_{[\alpha]}(w|y)}{p(w)}$$

$$\leq \frac{\pi_{\alpha,p}}{\alpha^2} \ln \frac{1}{\min_w p(w)}.$$

$$= \frac{1}{\alpha^2} \frac{1}{\min_w p(w)} \ln \frac{1}{\min_w p(w)}$$

Thus $\pi'_{\alpha,p}$ is bounded and $\mu'_{\alpha,p}$ is a finite measure, i.e. $\mu'_{\alpha,p} \in \mathcal{M}_0^+(\mathcal{Y})$. We can apply the dominated convergence theorem [8, 2.8.1] for $\mu'_{\alpha,p}$ as we did for $\mu_{\alpha,p}$ in part (a) in order to establish the continuity of $\mu'_{\alpha,p}$ as a function of α. Furthermore, $\frac{\mathrm{d}}{\mathrm{d}\alpha}\mu_{\alpha,p}(\mathcal{E})\big|_{\alpha=\phi} = \mu'_{\phi,p}(\mathcal{E})$ follows from the boundedness of $\pi'_{\alpha,p}$ and the definitions of $\mu'_{\alpha,p}$ and $\pi'_{\alpha,p}$ by [8, Cor. 2.8.7.(ii)] for $X = \mathcal{E}$. One can apply the Tonelli-Fubini theorem [30, 4.4.5] to obtain an equivalent result, instead of invoking [8, Cor. 2.8.7.(ii)].

(3-c) For $\mu''_{\alpha,p}$ defined in (12) to be a finite signed measure, $\pi''_{\alpha,p}$ should be a $\mu_{1,p}$-integrable function. By the expression for $\pi''_{\alpha,p}$ given in Lemma 2-(b) we have

$$-\frac{2}{\alpha^3}\left(\ln \frac{1}{\min_w p(w)}\right)\pi_{\alpha,p} \leq \pi''_{\alpha,p} \leq \left[\frac{1+\alpha}{\alpha^4}\left(\ln \frac{1}{\min_w p(w)}\right)^2 + \frac{4}{e^2\alpha^3}\right]\pi_{\alpha,p} \tag{A.47}$$

The proof of the continuity is similar to the corresponding proofs in parts (a) and (b). The identity $\frac{d}{d\alpha}\mu'_{\alpha,p}(\mathcal{E})\big|_{\alpha=\phi} = \mu''_{\phi,p}(\mathcal{E})$ follows from (A.47) by applying [8, Cor. 2.8.7.(ii)] for $X=\mathcal{E}$.

(3-d) For any $\beta \in [0,1]$ and $\alpha_0, \alpha_1 \in \mathbb{R}_+$ let α_β be $\alpha_\beta = \beta\alpha_1 + (1-\beta)\alpha_0$. Then as a result of the Hölder's inequality,

$$\int (\pi_{\alpha_1,p})^{\frac{\beta\alpha_1}{\alpha_\beta}}(\pi_{\alpha_0,p})^{\frac{(1-\beta)\alpha_0}{\alpha_\beta}}\mu_{1,p}(\mathrm{d}y) \leq (\|\mu_{\alpha_1,p}\|)^{\frac{\beta\alpha_1}{\alpha_\beta}}(\|\mu_{\alpha_0,p}\|)^{\frac{(1-\beta)\alpha_0}{\alpha_\beta}}. \tag{A.48}$$

On the other hand by Lemma 2-(c)

$$\pi_{\alpha_\beta,p} \leq (\pi_{\alpha_1,p})^{\frac{\beta\alpha_1}{\alpha_\beta}}(\pi_{\alpha_0,p})^{\frac{(1-\beta)\alpha_0}{\alpha_\beta}}. \tag{A.49}$$

Then the log-convexity of $\|\mu_{\alpha,p}\|^\alpha$ as a function of α follows from (A.48) and (A.49).

If $\mu_{1,p}(\cup_{\gamma\geq 1}\mathcal{A}(p,\gamma)) < 1$, then the log-convexity of $\|\mu_{\alpha,p}\|^\alpha$ is strict because the inequality in (A.49) is strict for y's that are not in $\cup_{\gamma\geq 1}\mathcal{A}(p,\gamma)$ by Lemma 2-(c). For $y\in\mathcal{A}(p,\gamma)$, the inequality in (A.49) is an equality and $\pi_{\alpha,p} = \gamma^{\frac{\alpha-1}{\alpha}}$ for all α. Consequently if $\mu_{1,p}(\cup_{\gamma\geq 1}\mathcal{A}(p,\gamma)) = 1$, then the log-convexity of $\|\mu_{\alpha,p}\|^\alpha$ is strict iff the inequality in (A.48) is strict. But if $\mu_{1,p}(\cup_{\gamma\geq 1}\mathcal{A}(p,\gamma)) = 1$, then the Hölder's inequality in (A.48) is strict unless there exists a $\gamma \geq 1$ such that $\mu_{1,p}(\mathcal{A}(p,\gamma)) = 1$.

We proceed with calculating the limit at zero. As a result of the expression for $\pi_{\alpha,p}$ given in part (a) we have,

$$\left(\sum_{w:p_{[1]}(w|y)>0} p(w)\right)^{\frac{\alpha-1}{\alpha}}\pi_{\alpha,p} = \left(\sum_w p_{[0]}(w|y)\left(\frac{p_{[1]}(w|y)}{p_{[0]}(w|y)}\right)^\alpha\right)^{\frac{1}{\alpha}} \quad \text{where} \quad p_{[0]}(w|y) = \frac{p(w)}{\sum_{\bar{w}:p_{[1]}(\bar{w}|y)>0}p(\bar{w})}.$$

Then using L'Hospital's rule [80, Thm. 5.13] for calculating limits and the Hölder's inequality we get,

$$\lim_{\alpha\to 0}\left(\sum_{w:p_{[1]}(w|y)>0} p(w)\right)^{\frac{\alpha-1}{\alpha}}\pi_{\alpha,p} = e^{\sum_w p_{[0]}(w|y)\ln\frac{p_{[1]}(w|y)}{p_{[0]}(w|y)}} \qquad \mu_{1,p}-a.e. \tag{A.50}$$

$$\left(\sum_{w:p_{[1]}(w|y)>0} p(w)\right)^{\frac{\alpha-1}{\alpha}}\pi_{\alpha,p} \leq 1 \qquad \forall \alpha\in(0,1),\ \mu_{1,p}-a.e. \tag{A.51}$$

The sum $\sum_{w:p_{[1]}(w|y)>0} p(w)$ is a simple function of y, i.e. its range is a finite set, because $\mathrm{supp}(p)$ has a finite number of distinct subsets. Thus the essential supremum is the maximum value of the sum with positive probability. Therefore

$$\mu_{1,p}\left(\left\{\sum_{w:p_{[1]}(w|y)>0} p(w) = \psi\right\}\right) > 0 \qquad \text{where} \qquad \psi = \operatorname*{ess\,sup}_{\mu_{1,p}}\sum_{w:p_{[1]}(w|y)>0} p(w). \tag{A.52}$$

Then using (A.50) we get

$$\lim_{\alpha\to 0}\psi^{\frac{\alpha-1}{\alpha}}\pi_{\alpha,p} = \mathbb{1}_{\{\sum_{w:p_{[1]}(w|y)>0} p(w)=\psi\}}e^{\sum_w p_{[0]}(w|y)\ln\frac{p_{[1]}(w|y)}{p_{[0]}(w|y)}} \qquad \mu_{1,p}\text{-a.e.}$$

On the other hand $\psi^{\frac{\alpha-1}{\alpha}}\pi_{\alpha,p} \leq 1$ for all $\alpha\in(0,1)$, $\mu_{1,p}$−a.e. by (A.51) and the definition of ψ given in (A.52). Thus we can apply the dominated convergence theorem [8, 2.8.1]:

$$\lim_{\alpha\to 0}\int\left|\psi^{\frac{\alpha-1}{\alpha}}\pi_{\alpha,p} - \mathbb{1}_{\{\sum_{w:p_{[1]}(w|y)>0} p(w)=\psi\}}e^{\sum_w p_{[0]}(w|y)\ln\frac{p_{[1]}(w|y)}{p_{[0]}(w|y)}}\right|\mu_{1,p}(\mathrm{d}y) = 0. \tag{A.53}$$

Consequently,

$$\lim_{\alpha\to 0}\left(\psi^{(\alpha-1)}\|\pi_{\alpha,p}\|^\alpha\right)^{\frac{1}{\alpha}} = \int\mathbb{1}_{\{\sum_{w:p_{[1]}(w|y)>0} p(w)=\psi\}}e^{\sum_w p_{[0]}(w|y)\ln\frac{p_{[1]}(w|y)}{p_{[0]}(w|y)}}\mu_{1,p}(\mathrm{d}y). \tag{A.54}$$

The right hand side of (A.54) is a real number between 0 and 1 by (A.52). Thus we have,

$$\lim_{\alpha\to 0}\psi^{\alpha-1}\|\pi_{\alpha,p}\|^\alpha = 1. \tag{A.55}$$

(3-e) $\|\mu_{\alpha,p}\| \leq |\mathrm{supp}(p)|$ by Lemma 1-(a). The continuity of $\|\mu_{\alpha,p}\|$ in α is implied by the continuity of $\mu_{\alpha,p}$ in α for the total variation topology on $\mathcal{M}_0^+(\mathcal{Y})$, proved in part (a). Furthermore, $\|\mu_{\alpha,p}\| = \mu_{\alpha,p}(\mathcal{Y})$ because $\mu_{\alpha,p} \in \mathcal{M}_0^+(\mathcal{Y})$ by part (a). In addition $\frac{d}{d\alpha}\mu_{\alpha,p}(\mathcal{Y}) \geq 0$ by part (b). Hence $\|\mu_{\alpha,p}\|$ is a nondecreasing function of α.

Let $\mathcal{E}_p$ be $\mathcal{E}_p = \{y : p_{[1]}(\cdot|y) \neq p(\cdot)\}$. Then $\forall y \in \mathcal{E}_p$, $p_{[\alpha]}(\cdot|y) \neq p(\cdot)$ and $\pi_{\alpha,p}$ is monotone increasing in α on $\mathbb{R}_{\geq 0}$. On the other hand, if there are two or more distinct w's in $\mathrm{supp}(p)$, then $\mu_{1,p}(\mathcal{E}_p) > 0$. Thus $\|\mu_{\alpha,p}\|$ is monotone increasing if there exist $w, \widetilde{w} \in \mathrm{supp}(p)$ such that $w \neq \widetilde{w}$. Else $\pi_{\alpha,p} = 1$ thus $\|\mu_{\alpha,p}\| = \int \pi_{\alpha,p} \mu_{1,p}(\mathrm{d}y) = 1$ for all $\alpha \in [0,\infty]$. $\qquad\square$

Proof of Lemma 4.

(4-a) Let us start with $\alpha = 0$ case. Since the weighted arithmetic mean of any two non-negative real numbers is greater than their weighted geometric mean, for any reference measure ν for μ_{1,p_1} and μ_{1,p_2} we have,

$$\beta \frac{\mathrm{d}\mu_{0,p_1}}{\mathrm{d}\nu} + (1-\beta)\frac{\mathrm{d}\mu_{0,p_2}}{\mathrm{d}\nu} \geq \left(\frac{\mathrm{d}\mu_{0,p_1}}{\mathrm{d}\nu}\right)^\beta \left(\frac{\mathrm{d}\mu_{0,p_2}}{\mathrm{d}\nu}\right)^{1-\beta}$$
$$= \frac{\mathrm{d}\mu_{0,p_\beta}}{\mathrm{d}\nu}.$$

For any $\alpha \in (0,1]$ the function $x^{1/\alpha}$ is convex in x. Then for any reference measure ν for μ_{α,p_1} and μ_{α,p_2} as a result of the Jensen's inequality we have,

$$\beta \frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} + (1-\beta)\frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} \geq \left(\sum_w (\beta p_1(w) + (1-\beta)p_2(w))\left(\frac{\mathrm{d}w}{\mathrm{d}\nu}\right)^\alpha\right)^{1/\alpha}$$
$$= \frac{\mathrm{d}\mu_{\alpha,p_\beta}}{\mathrm{d}\nu}.$$

$\|\mu_{\alpha,p}\|$ is convex in p because $\frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu}$ is convex in p and $\frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu}$ is non-negative.

(4-b) For $\alpha \in [1,\infty)$ the function $x^{1/\alpha}$ is concave in x. Thus the inequalities are reversed. Hence both the Radon-Nikodym derivative $\frac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu}$ and the norm $\|\mu_{\alpha,p}\|$ are concave in p.

For any reference measure ν for μ_{∞,p_1} and μ_{∞,p_2} by the definition of $\frac{\mathrm{d}\mu_{\infty,p}}{\mathrm{d}\nu}$ given in (7), we have

$$\beta \frac{\mathrm{d}\mu_{\infty,p_1}}{\mathrm{d}\nu} + (1-\beta)\frac{\mathrm{d}\mu_{\infty,p_2}}{\mathrm{d}\nu} \leq \frac{\mathrm{d}\mu_{\infty,p_\beta}}{\mathrm{d}\nu}.$$

$\|\mu_{\infty,p}\|$ is concave in p because $\frac{\mathrm{d}\mu_{\infty,p}}{\mathrm{d}\nu}$ is concave in p and $\frac{\mathrm{d}\mu_{\infty,p}}{\mathrm{d}\nu}$ is non-negative.

(4-c) Identities are confirmed using the definitions of $s_\wedge$, s_1 and s_2 by substitution. On the other hand,

$$\|p_1 - p_2\| = \|p_1 \vee p_2\| - \|p_1 \wedge p_2\|$$
$$= 2 - 2\|p_1 \wedge p_2\|.$$

Hence $s_\wedge \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$. Using the fist identity together with $s_\wedge \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ and $p_1 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ we get $s_1 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$. Similarly $s_2 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ follows from the second identity, $s_\wedge \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ and $p_1 \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$.

(4-d) Let δ be $\delta = \frac{\|p_2 - p_1\|}{2}$. For any reference measure ν for μ_{α,p_1} and μ_{α,p_2} and $\alpha \in (0,1]$,

$$\frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} = \left[(1-\delta)(\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu})^\alpha + \delta(\tfrac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu})^\alpha\right]^{1/\alpha} - \left[(1-\delta)(\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu})^\alpha + \delta(\tfrac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu})^\alpha\right]^{1/\alpha}$$
$$\leq \left[(1-\delta)(\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu})^\alpha + \delta(\tfrac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu})^\alpha\right]^{1/\alpha} - (1-\delta)^{\frac{1}{\alpha}}\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu}$$
$$\leq \left[(1-\delta)\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu} + \delta\tfrac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu}\right] - (1-\delta)^{1/\alpha}\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu}.$$

where the last inequality follows from the Jensen's inequality and the convexity of $x^{1/\alpha}$ in x for $\alpha \in (0,1]$.

We bound $\frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu}$ in a similarly way. Using these two bounds we can bound $\|\mu_{\alpha,p_1} - \mu_{\alpha,p_2}\|$ as follows

$$\|\mu_{\alpha,p_1} - \mu_{\alpha,p_2}\| = \int_{\frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} > \frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu}} (\tfrac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} - \tfrac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu})\nu(\mathrm{d}y) + \int_{\frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} > \frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu}} (\tfrac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} - \tfrac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu})\nu(\mathrm{d}y)$$
$$\leq 2\left[1 - \delta - (1-\delta)^{1/\alpha}\right]\|\mu_{\alpha,s_\wedge}\| + \delta\|\mu_{\alpha,s_1}\| + \delta\|\mu_{\alpha,s_2}\|$$
$$\leq 2\left[1 - (1-\delta)^{1/\alpha}\right]$$
$$\leq \tfrac{2}{\alpha}\delta.$$

(4-e) One can confirm using the derivative test that for any $\alpha \in [1,\infty)$, $x_0 \geq 0$ and $x_1 \geq x_2 \geq 0$ we have

$$[(1-\delta)x_0^\alpha + \delta x_1^\alpha]^{1/\alpha} - [(1-\delta)x_0^\alpha + \delta x_2^\alpha]^{1/\alpha} \leq \delta^{1/\alpha}(x_1 - x_2).$$

Then for any reference measure ν for μ_{α,p_1} and μ_{α,p_2} we have

$$\frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} = \left[(1-\delta)(\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu})^\alpha + \delta(\tfrac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu})^\alpha\right]^{1/\alpha} - \left[(1-\delta)(\tfrac{\mathrm{d}\mu_{\alpha,s_\wedge}}{\mathrm{d}\nu})^\alpha + \delta(\tfrac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu})^\alpha\right]^{1/\alpha}$$
$$\leq \delta^{1/\alpha}\left[\tfrac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu} - \tfrac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu}\right].$$

We can bound $\frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu}$ in a similarly way. On the other hand $\frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} \geq \frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu}$ iff $\frac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu} \geq \frac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu}$. Thus we can bound $\|\mu_{\alpha,p_1} - \mu_{\alpha,p_2}\|$ using the bounds on $\frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu}$ and $\frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu}$:

$$
\|\mu_{\alpha,p_1} - \mu_{\alpha,p_2}\| = \int_{\frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} > \frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu}} (\tfrac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu} - \tfrac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu})\nu(\mathrm{d}y) + \int_{\frac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} > \frac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu}} (\tfrac{\mathrm{d}\mu_{\alpha,p_2}}{\mathrm{d}\nu} - \tfrac{\mathrm{d}\mu_{\alpha,p_1}}{\mathrm{d}\nu})\nu(\mathrm{d}y)
$$

$$
\leq \delta^{1/\alpha} \int_{\frac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu} > \frac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu}} (\tfrac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu} - \tfrac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu})\nu(\mathrm{d}y) + \delta^{1/\alpha} \int_{\frac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu} > \frac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu}} (\tfrac{\mathrm{d}\mu_{\alpha,s_2}}{\mathrm{d}\nu} - \tfrac{\mathrm{d}\mu_{\alpha,s_1}}{\mathrm{d}\nu})\nu(\mathrm{d}y)
$$

$$
= \delta^{1/\alpha}\|\mu_{\alpha,s_1} - \mu_{\alpha,s_2}\|.
$$

$\square$

E. Proofs of the Lemmas on the Rényi Information

Proof of Lemma 5. $I_\infty(\alpha;\mathcal{W}) \leq \ln|\mathrm{supp}(p)|$ because $\|\mu_{\infty,p}\| \leq |\mathrm{supp}(p)|$ by Lemma 3-(e).

$\lim_{\alpha\downarrow 0} I_\alpha(p;\mathcal{W}) = I_0(p;\mathcal{W})$ follows from Lemma 3-(d) and the definition of $I_\alpha(p;\mathcal{W})$ given in (13).

$\lim_{\alpha\uparrow\infty} I_\alpha(p;\mathcal{W}) = I_\infty(p;\mathcal{W})$ follows from the continuity of $\|\mu_{\alpha,p}\|$ as a function of α at infinity, i.e. Lemma 3-(e), and the definition of $I_\alpha(p;\mathcal{W})$ given in (13).

Both $\|\mu_{\alpha,p}\|$ and $\|\mu'_{\alpha,p}\|$ are continuously differentiable on $\mathbb{R}_+$, $\frac{\mathrm{d}}{\mathrm{d}\alpha}\|\mu_{\alpha,p}\| = \|\mu'_{\alpha,p}\|$ and $\frac{\mathrm{d}}{\mathrm{d}\alpha}\|\mu'_{\alpha,p}\| = \mu''_{\alpha,p}(\mathcal{Y})$ because of Lemma 3-(a,b,c). Then as a result of its definition given in (13), $I_\alpha(p;\mathcal{W})$ is continuously differentiable in α on $(0,1)$ and $(1,\infty)$. The expression for the derivative for $\alpha \neq 1$ given in (16) follows from the chain rule.

In order to extend the continuous differentiability to $\alpha = 1$, first we establish that $I_\alpha(p;\mathcal{W})$ is continuous at $\alpha = 1$. As a result of L'Hospital's rule [80, Thm. 5.13] and Lemma 3-(b) $\lim_{\alpha\to 1}\frac{\alpha}{\alpha-1}\ln\|\mu_{\alpha,p}\| = \|\mu'_{1,p}\|$. On the other hand $\|\mu'_{1,p}\| = I_1(p;\mathcal{W})$ as a result of (11), Lemma 2-(b) and the definition of $I_1(p;\mathcal{W})$. Thus $I_\alpha(p;\mathcal{W})$ is continuous at $\alpha = 1$. Then,

$$
\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W})\big|_{\alpha=1} = \lim_{\alpha\to 1}\frac{1}{1-\alpha}\left[\|\mu'_{1,p}\| - \frac{\alpha}{\alpha-1}\ln\|\mu_{\alpha,p}\|\right].
$$

$\|\mu_{\alpha,p}\|$ and $\|\mu'_{\alpha,p}\|$ are continuously differentiable by Lemma 3-(b,c). Then using L'Hospital's rule [80, Thm. 5.13] and the identity $\frac{\mathrm{d}}{\mathrm{d}\alpha}\|\mu'_{\alpha,p}\| = \mu''_{\alpha,p}(\mathcal{Y})$ we get

$$
\lim_{\alpha\to 1}\frac{1}{1-\alpha}\left[\|\mu'_{1,p}\| - \frac{\alpha}{\alpha-1}\ln\|\mu_{\alpha,p}\|\right] = \frac{1}{2}\left[\mu''_{1,p}(\mathcal{Y}) + 2\|\mu'_{1,p}\| - \|\mu'_{1,p}\|^2\right].
$$

Hence $I_\alpha(p;\mathcal{W})$ is differentiable at $\alpha = 1$ and its derivative at $\alpha = 1$ is the one given in (16). Finally, in order to show that $\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W})$ is continuous at $\alpha = 1$ we apply L'Hospital's rule [80, Thm. 5.13] to confirm,

$$
\lim_{\alpha\to 1}\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W}) = \lim_{\alpha\to 1}\frac{\alpha(\alpha-1)\|\mu'_{\alpha,p}\| - \|\mu_{\alpha,p}\|\ln\|\mu_{\alpha,p}\|}{\|\mu_{\alpha,p}\|(\alpha-1)^2}
$$

$$
= \frac{1}{2}\left[\mu''_{1,p}(\mathcal{Y}) + 2\|\mu'_{1,p}\| - \|\mu'_{1,p}\|^2\right].
$$

As a function α on $\mathbb{R}_+$, $\alpha\ln\|\mu_{\alpha,p}\|$ is convex by Lemma 3-(d) and differentiable by Lemma 3-(b). Then $\alpha\ln\|\mu_{\alpha,p}\|$ has a tangent at each $\alpha \in \mathbb{R}_+$ and it lays above all of its tangents, i.e. for all $\alpha,\eta \in \mathbb{R}_+$ such that $\alpha \neq \eta$,

$$
\eta\ln\|\mu_{\eta,p}\| \geq \alpha\ln\|\mu_{\alpha,p}\| + \left(\ln\|\mu_{\alpha,p}\| + \frac{\alpha\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|}\right)(\eta - \alpha). \tag{A.56}
$$

Then for all $\alpha,\eta \in \mathbb{R}_+$ such that $\alpha \neq \eta$ we have

$$
\frac{\alpha}{\alpha-\eta}\frac{\|\mu'_{\alpha,p}\|}{\|\mu_{\alpha,p}\|} + \frac{\eta}{(\alpha-\eta)^2}\ln\frac{\|\mu_{\eta,p}\|}{\|\mu_{\alpha,p}\|} \geq 0. \tag{A.57}
$$

If we apply the above inequality at $\eta = 1$ we can conclude, using (16), that $\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W}) \geq 0$ for $\alpha \neq 1$. For $\alpha = 1$ using Lemma 2-(b) and Lemma 3-(c) we get

$$
\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W})\big|_{\alpha=1} = \frac{1}{2}\int\sum_w p_{[1]}(w|y)\left(\ln\frac{p_{[1]}(w|y)}{p(w)} - I_1(p;\mathcal{W})\right)^2\mu_{1,p}(\mathrm{d}y)
$$

$$
\geq 0. \tag{A.58}
$$

Thus $\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W})$ is non-negative for all $\alpha \in \mathbb{R}_+$ and $I_\alpha(p;\mathcal{W})$ is a nondecreasing function of α. Then $I_\alpha(p;\mathcal{W})$ is non-negative as well because $I_0(\alpha;\mathcal{W}) \geq -\ln\sum_w p(w) = 0$ and $\lim_{\alpha\downarrow 0} I_\alpha(p;\mathcal{W}) = I_0(\alpha;\mathcal{W})$.

If $\mu_{1,p}(\mathcal{A}(p,\gamma)) = 1$ for a γ, then $\|\mu_{\alpha,p}\| = \gamma^{\frac{\alpha-1}{\alpha}}$ for all $\alpha \in \mathbb{R}_+$ and $I_\alpha(p;\mathcal{W}) = \ln\gamma$ for all $\alpha \in [0,\infty]$, because $\pi_{\alpha,p} = \gamma^{\frac{\alpha-1}{\alpha}}$ for all $y \in \mathcal{A}(p,\gamma)$.

If there does not exist a γ such that $\mu_{1,p}(\mathcal{A}(p,\gamma)) = 1$, then the convexity of $\alpha\ln\|\mu_{\alpha,p}\|$ is strict by Lemma 3-(d) and the variance of the random variable $\ln\frac{p_{[1]}(w|y)}{p(w)}$ is positive. Thus the inequalities (A.56), (A.57), and (A.58) are strict and $\frac{\mathrm{d}}{\mathrm{d}\alpha} I_\alpha(p;\mathcal{W})$ is positive for all $\alpha \in \mathbb{R}_+$. $\square$

Proof of Lemma 6.

(6-a) Let us start with the values of α in $(0,1)$. Recall that $\|\cdot\| : \mathcal{M}^+(\mathcal{Y}) \to \mathbb{R}_+$ and $\frac{\alpha}{\alpha-1}\ln(\cdot) : \mathbb{R}_+ \to \mathbb{R}$ are continuous functions and the composition of two continuous functions is a continuous function, [68, Thm. 18.2.c]. Furthermore, the function $\|\mu_{\alpha,p}\|$ is continuous in p on $\mathcal{P}(\mathcal{P}(\mathcal{Y}))$, —and hence on $\mathcal{P}(\mathcal{W})$— for $\alpha \in (0,1)$ by Lemma 4-(d). Thus $I_\alpha(p;\mathcal{W})$ is continuous in p on $\mathcal{P}(\mathcal{W})$ for $\alpha \in (0,1)$.

For any $p_1, p_2 \in \mathcal{P}(\mathcal{W})$ and $\beta \in [0,1]$ let $p_\beta = \beta p_1 + (1-\beta)p_2$. Recall that $\|\mu_{\alpha,p}\|$ is convex in p for $\alpha \in (0,1)$ by Lemma 4-(a). Then by the definition of $I_\alpha(p;\mathcal{W})$ given in (13) we have

$$
\begin{aligned}
I_\alpha(p_\beta;\mathcal{W}) &\geq \tfrac{\alpha}{\alpha-1}\ln\left(\beta\|\mu_{\alpha,p_1}\| + (1-\beta)\|\mu_{\alpha,p_2}\|\right) \\
&\geq \tfrac{\alpha}{\alpha-1}\ln\left(\|\mu_{\alpha,p_1}\| \vee \|\mu_{\alpha,p_2}\|\right) \\
&\geq I_\alpha(p_1;\mathcal{W}) \wedge I_\alpha(p_2;\mathcal{W})
\end{aligned}
$$

Thus Rényi information is continuous and quasi-concave in p for $\alpha \in (0,1)$.

For $\alpha = 0$ case, first note that $\left|e^{-I_0(p_1;\mathcal{W})} - e^{-I_0(p_2;\mathcal{W})}\right| \leq \|p_1 - p_2\|$. Thus $e^{-I_0(p;\mathcal{W})}$ is continuous in p. Since $-\ln x$ is continuous on $\mathbb{R}_+$, $I_0(p;\mathcal{W})$ is continuous in p. In order to prove that $I_0(p;\mathcal{W})$ is quasi-concave, note that for any $\beta \in (0,1)$ and $p_1, p_2 \in \mathcal{P}(\mathcal{W})$ we have,

$$
\begin{aligned}
I_0(p_\beta;\mathcal{W}) &= -\ln\operatorname*{ess\,sup}_{\mu_{1,p_\beta}}\left[\beta\sum_{w:p_1(w|y)>0}p_1(w) + (1-\beta)\sum_{w:p_2(w|y)>0}p_2(w)\right] \\
&\geq -\ln\left[\operatorname*{ess\,sup}_{\mu_{1,p_1}}\left(\sum_{w:p_1(w|y)>0}p_1(w)\right)\bigvee \operatorname*{ess\,sup}_{\mu_{1,p_2}}\left(\sum_{w:p_2(w|y)>0}p_2(w)\right)\right] \\
&= I_0(p_1;\mathcal{W}) \wedge I_0(p_2;\mathcal{W}).
\end{aligned}
$$

(6-b) For any $p_1, p_2 \in \mathcal{P}(\mathcal{W})$ and $\beta \in [0,1]$ let $p_\beta = \beta p_1 + (1-\beta)p_2$. Recall that $\|\mu_{\alpha,p}\|$ is concave in p for $\alpha \in (1,\infty]$ by Lemma 4-(b). Then by the definition of $I_\alpha(p;\mathcal{W})$ we have

$$
\begin{aligned}
I_\alpha(p_\beta;\mathcal{W}) &\geq \tfrac{\alpha}{\alpha-1}\ln\left(\beta\|\mu_{\alpha,p_1}\| + (1-\beta)\|\mu_{\alpha,p_2}\|\right) \\
&\geq \beta\tfrac{\alpha}{\alpha-1}\ln\|\mu_{\alpha,p_1}\| + (1-\beta)\tfrac{\alpha}{\alpha-1}\ln\|\mu_{\alpha,p_2}\| \\
&= \beta I_\alpha(p_1;\mathcal{W}) + (1-\beta)I_\alpha(p_2;\mathcal{W})
\end{aligned}
$$

where the second inequality follows from the Jensen's inequality and the concavity of the logarithm function.

For $\alpha = 1$ case, note that as a result of the definition of $I_\alpha(p;\mathcal{W})$ we have

$$
\begin{aligned}
I_1(p_\beta;\mathcal{W}) &= \beta I_1(p_1;\mathcal{W}) + (1-\beta)I_1(p_2;\mathcal{W}) + \int\left[\beta\tfrac{\mathrm{d}\mu_{1,p_1}}{\mathrm{d}\mu_{1,p_\beta}}\ln\tfrac{\mathrm{d}\mu_{1,p_1}}{\mathrm{d}\mu_{1,p_\beta}} + (1-\beta)\tfrac{\mathrm{d}\mu_{1,p_2}}{\mathrm{d}\mu_{1,p_\beta}}\ln\tfrac{\mathrm{d}\mu_{1,p_2}}{\mathrm{d}\mu_{1,p_\beta}}\right]\mu_{1,p_\beta}(\mathrm{d}y) \\
&\geq \beta I_1(p_1;\mathcal{W}) + (1-\beta)I_1(p_2;\mathcal{W})
\end{aligned}
$$

where the inequality follows from $\beta\tfrac{\mathrm{d}\mu_{1,p_1}}{\mathrm{d}\mu_{1,p_\beta}} + (1-\beta)\tfrac{\mathrm{d}\mu_{1,p_2}}{\mathrm{d}\mu_{1,p_\beta}} = 1$, the convexity of the function $x\ln x$ and the Jensen's inequality. $\qquad\square$

F. Proof of the Lemma on the Rényi Mean

Proof of Lemma 14. For $\alpha = 0$, as a result of the definition of the order zero Rényi information given in (13) and the definition of the order zero Rényi divergence given in (19) we have

$$
\begin{aligned}
D_0(p\circledast\mathcal{W}\| p \otimes q) &= -\ln\int\sum_w p(w)\tfrac{\mathrm{d}q}{\mathrm{d}\nu}\mathbb{1}_{\{p_{[1]}(w|y)\frac{\mathrm{d}\mu_{1,p}}{\mathrm{d}\nu}>0\}}\nu(\mathrm{d}y) \\
&= I_0(p;\mathcal{W}) - \ln\int\frac{\sum_{w:p_{[1]}(w|y)>0}p(w)}{\operatorname*{ess\,sup}_{\mu_{1,p}}\sum_{w:p_{[1]}(w|y)>0}p(w)}\mathbb{1}_{\{\frac{\mathrm{d}\mu_{1,p}}{\mathrm{d}\nu}>0\}}\tfrac{\mathrm{d}q}{\mathrm{d}\nu}\nu(\mathrm{d}y).
\end{aligned}
$$

Then the definition of $q_{0,p}$ given in (21) implies (23) and (24).

For $\alpha \in (0,\infty]$, (23) follows from the definitions of the Rényi information, divergence, and mean given in (13), (19) (21) by substitution. Using (22) and (23) we get,

$$
D_\alpha(p\circledast\mathcal{W}\| p \otimes q) = I_\alpha(p;\mathcal{W}) + D_\alpha(q_{\alpha,p}\| q) \qquad\qquad \forall\alpha \in (0,\infty]. \qquad (A.59)
$$

On the other hand $q_{\alpha,p}$ is a probability measure by definition. Then (24) and uniqueness of Rényi mean as the minimizer follow from (23), (A.59), and Lemma 10.

The following identity and (24) imply (25).

$$
D_\alpha(p\circledast\mathcal{W}\| p \otimes q) = D_\alpha(\mu_{\alpha,p}\| q) \qquad\qquad \forall p \in \mathcal{P}(\mathcal{W}), q \in \mathcal{M}^+(\mathcal{Y}), \alpha \in (0,\infty] \setminus \{1\}.
$$

$\qquad\square$

G. The Ergodic Theorem and the Rényi Capacity

For $\mathcal{W}^{[\jmath]}$ described in Example 5 we have $D_\alpha(w\,\|\,l) = D_\alpha(w_f\,\|\,l)$ for all $w \in \mathcal{W}^{[\jmath]}$ where l is the Lebesgue measure. Thus by (33) of Theorem 1 we have

$$C_{\alpha,\mathcal{W}^{[\jmath]}} \le D_\alpha(w_f\,\|\,l).$$

We prove the reverse inequality, $C_{\alpha,\mathcal{W}^{[\jmath]}} \ge D_\alpha(w_f\,\|\,l)$, using the Birkoff-Khinchin ergodic theorem [30, 8.4.1]. In particular, we show that there exists a sequence of priors $\{p_\imath\}_{\imath \in \mathbb{Z}_+}$ such that[37] $\lim_{\imath \to \infty} I_\alpha(p_\imath;\mathcal{W}^{[\jmath]}) \ge D_\alpha(w_f\,\|\,l)$ for all $\alpha \in (0,\infty]$.

For any $\kappa \in \mathbb{Z}$ and $x \in \mathbb{R}$ let T_x^κ be the transformation resulting from κ successive applications of T_x. As a result of the definition of T_x given in (62), $\mathrm{T}_x^\kappa = \mathrm{T}_{\kappa x}$ for any $\kappa \in \mathbb{Z}$ and $x \in \mathbb{R}$. For any $g \in \mathcal{L}^1(l)$, $x \in \mathbb{R}$ and $\imath \in \mathbb{Z}_+$ let $\overline{g}_x^\imath$ be

$$\overline{g}_x^\imath(y) \triangleq \tfrac{1}{\imath} \sum_{\kappa=0}^{\imath-1} g \circ \mathrm{T}_x^\kappa(y)$$
$$= \tfrac{1}{\imath} \sum_{\kappa=0}^{\imath-1} g(y - \kappa x - \lfloor y - \kappa x \rfloor).$$

For any $x \in \mathbb{R}$, the transformation T_x is measure preserving for the measure space $(\mathcal{Y}, \mathcal{Y}, l)$; if in addition x is irrational, then T_x is ergodic. Hence $\overline{g}_x^\imath$ converges to $\int g(y)\mathrm{d}y$ l−a.e. for any $g \in \mathcal{L}^1(l)$ and $x \in \mathbb{R} \setminus \mathbb{Q}$, by the Birkoff-Khinchin Ergodic theorem [30, 8.4.1]:

$$\overline{g}_x^\imath \xrightarrow{l-a.e.} \int g(y)\mathrm{d}y \qquad\qquad \forall g \in \mathcal{L}^1(l) \text{ and } \forall x \in \mathbb{R} \setminus \mathbb{Q}. \tag{A.60}$$

Let x be an irrational number that will be fixed for the rest of the proof. For any $\imath \in \mathbb{Z}_+$, let $p_\imath$ be the prior that has equal probability mass on each probability measure corresponding to a $f \circ \mathrm{T}_x^\kappa$ for some $\kappa \in \{0, \ldots, (\imath-1)\}$. Then

$$\tfrac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l}(y) = \left(\tfrac{1}{\imath} \sum_{\kappa=0}^{\imath-1} (f^\alpha \circ \mathrm{T}_x^\kappa)(y) \right)^{1/\alpha}$$
$$= \left(\sum_{\kappa=0}^{\imath-1} \tfrac{1}{\imath} f^\alpha \left(y - \tfrac{\kappa x}{\imath} - \lfloor y - \tfrac{\kappa x}{\imath} \rfloor \right) \right)^{1/\alpha} \qquad \forall \alpha \in \mathbb{R}_+.$$

For $\alpha \in \mathbb{R}_+ \setminus \{1\}$, we calculate the limit $\lim_{\imath \to \infty} I_\alpha(p_\imath;\mathcal{W}^{[\jmath]})$ by calculating the limit $\lim_{\imath \to \infty} \|\mu_{\alpha,p_\imath}\|$. For $\alpha = 1$ and $\alpha = \infty$ the result follows from continuity arguments.

(a) $\alpha \in (0,1)$ *case:* $\int f^\alpha \mathrm{d}y \le \left(\int f(y)\mathrm{d}y \right)^\alpha = 1$ by the Jensen's inequality. Hence $f^\alpha \in \mathcal{L}^1(l)$ as a result of (A.60) we have

$$\left(\tfrac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l} \right)^\alpha \xrightarrow{l-a.e.} \int f^\alpha(y)\mathrm{d}y \qquad\qquad \Rightarrow \qquad\qquad \tfrac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l} \xrightarrow{l-a.e.} \left(\int f^\alpha(y)\mathrm{d}y \right)^{1/\alpha}.$$

For any $\epsilon > 0$ there exists a $\delta > 0$ such that if $l(\mathcal{E}) < \delta$ for a $\mathcal{E} \in \mathcal{Y}$, then $w_f(\mathcal{E}) < \epsilon$, because $w_f \prec l$. Since l is invariant under translations and $\mathcal{W}^{[\jmath]}$ is the set of all mod one translations of w_f, $\mu_{1,p}(\mathcal{E}) < \epsilon$ whenever $l(\mathcal{E}) < \delta$, as well. Then $\{\mu_{\alpha,p} : p \in \mathcal{P}(\mathcal{W}^{[\jmath]}), \alpha \in (0,1]\} \prec^{uni} l$ and $\{\tfrac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l}\}_{\imath \in \mathbb{Z}_+}$ is uniformly l−integrable because $\mu_{\alpha,p}(\mathcal{E})$ is a nondecreasing function of α for all $\mathcal{E} \in \mathcal{Y}$ by Lemma 3-(b). Since almost everywhere convergence implies convergence in measure by [8, Thm. 2.2.3], using Lebesgue-Vitali convergence theorem [8, 4.5.4], we can conclude that $\tfrac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l}$ converges to $\left(\int f^\alpha(y)\mathrm{d}y \right)^{1/\alpha}$ in $\mathcal{L}^1(l)$, as well: $\tfrac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l} \xrightarrow{\mathcal{L}^1(l)} \left(\int f^\alpha(y)\mathrm{d}y \right)^{1/\alpha}$. Then $\lim_{\imath \to \infty} \|\mu_{\alpha,p_\imath}\| = \left(\int f^\alpha(y)\mathrm{d}y \right)^{1/\alpha}$. Using the definition of Rényi information given in (13) we get

$$\lim_{\imath \to \infty} I_\alpha(p_\imath;\mathcal{W}^{[\jmath]}) = \tfrac{1}{\alpha-1} \ln \left(\int f^\alpha(y)\mathrm{d}y \right)$$
$$= D_\alpha(w_f\,\|\,l) \qquad\qquad \forall \alpha \in (0,1).$$

(b) $\alpha = 1$ *case:* The Rényi information is a nondecreasing function of the order by Lemma 5. Then

$$\liminf_{\imath \to \infty} I_1(p_\imath;\mathcal{W}^{[\jmath]}) \ge \liminf_{\imath \to \infty} I_\alpha(p_\imath;\mathcal{W}^{[\jmath]})$$
$$= D_\alpha(w_f\,\|\,l) \qquad\qquad \forall \alpha \in (0,1).$$

Since the Rényi divergence is a nondecreasing and lower semicontinuous function of the order by Lemma 7, we have

$$\liminf_{\imath \to \infty} I_1(p_\imath;\mathcal{W}^{[\jmath]}) \ge \lim_{\alpha \uparrow 1} D_\alpha(w_f\,\|\,l)$$
$$= D_1(w_f\,\|\,l).$$

(c) $\alpha \in (1,\infty)$ *case:* We analyze the finite $\int f^\alpha(y)\mathrm{d}y$ and infinite $\int f^\alpha(y)\mathrm{d}y$ cases separately.

---

[37]Finding a different sequence of priors for each order α in $(0,\infty]$ would have been sufficient for establishing $C_{\alpha,\mathcal{W}^{[\jmath]}} \ge D_\alpha(w_f\,\|\,l)$. The existence of a sequence of priors $\{p_\imath\}_{\imath \in \mathbb{Z}_+}$ such that $\lim_{\imath \to \infty} I_\alpha(p_\imath;\mathcal{W}^{[\jmath]}) = C_{\alpha,\mathcal{W}^{[\jmath]}}$ for all orders α in $\mathbb{R}_+$ allows us to assert the convexity of $(\alpha-1)C_{\alpha,\mathcal{W}^{[\jmath]}}$ in α on $\mathbb{R}_+$, rather than just $[1,\infty)$.

- If $\int f^{\alpha}(y)\mathrm{d}y < \infty$, then $f^{\alpha} \in \mathcal{L}^1(l)$ and $\frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l} \xrightarrow{l} \left(\int f^{\alpha}(y)\mathrm{d}y\right)^{1/\alpha}$ by (A.60) because almost everywhere convergence implies convergence in measure by [8, Thm. 2.2.3]. On the other hand, as a result of the concavity of the function $z^{1/\alpha}$ in z for $\alpha \in (1, \infty)$ and the Jensen's inequality we have

$$\mu_{\alpha,p_\imath}(\mathcal{E}) \leq \left(\frac{1}{\imath} \sum_{\kappa=0}^{\imath-1} \int_{\mathrm{T}_x^\kappa \mathcal{E}} f^{\alpha}(y)\mathrm{d}y\right)^{1/\alpha}.$$

Then the uniform $l-$integrability of $\frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l}$ follows from the translational invariance of l and the $l-$integrability of f^{α} following an argument similar to the one we have for $\alpha \in (0, 1)$ case. Thus using Lebesgue-Vitali convergence theorem [8, 4.5.4] and the definition of Rényi information exactly the same way we did for $\alpha \in (0, 1)$ case we get

$$\lim_{\imath \to \infty} I_\alpha(p_\imath; \mathcal{W}^{[\jmath]}) = D_\alpha(w_f \| l) \qquad \qquad \text{if } \int f^{\alpha}(y)\mathrm{d}y < \infty.$$

- If $\int f^{\alpha}\mathrm{d}y = \infty$, then we repeat the above analysis for $f \wedge \gamma$ for a $\gamma \in \mathbb{R}_+$ instead of f. As a result we get,

$$\liminf_{\imath \to \infty} I_\alpha(p_\imath; \mathcal{W}^{[\jmath]}) \geq \frac{1}{\alpha-1} \ln \left(\int (f(y) \wedge \gamma)^{\alpha}\mathrm{d}y\right) \qquad \qquad \forall \gamma \in \mathbb{R}_+$$

Note that as $\gamma \uparrow \infty$, $\int (f(y) \wedge \gamma)^{\alpha}\mathrm{d}y \uparrow \int f^{\alpha}(y)\mathrm{d}y$. Thus

$$\lim_{\imath \to \infty} I_\alpha(p_\imath; \mathcal{W}^{[\jmath]}) = \infty \qquad \qquad \text{if } \int f^{\alpha}(y)\mathrm{d}y = \infty.$$

(d) $\alpha = \infty$ *case:* Repeat the analysis for $\alpha = 1$ case by replacing $\alpha = 1$ and $(0, 1)$ by $\alpha = \infty$ and $(1, \infty)$.

We have used the ergodic theorem [30, 8.4.1] in order to be able to conduct our analysis for arbitrary measurable functions. If we restrict our attention to functions that are bounded and continuous at all but finite number of points, we can choose $p_\imath$ to be the priors that have $1/2^\imath$ probability mass on each probability measure corresponding to a $f \circ \mathrm{T}_{2^{-\imath}}^\kappa$ for $\kappa \in \{0, 1, \ldots, (2^\imath - 1)\}$. Then the identity $(\frac{\mathrm{d}\mu_{\alpha,p_\imath}}{\mathrm{d}l})^{\alpha} \xrightarrow{l-a.e.} \left(\int f^{\alpha}(y)\mathrm{d}y\right)$ is a result of Riemann integrability of f^{α} rather than the ergodicity.

We have used the Lebesgue-Vitali convergence theorem [8, 4.5.4] instead of the dominated convergence theorem [8, 2.8.1]. That is a matter of taste; one can prove the same statements using the dominated convergence theorem. First, do the analysis for $\tilde{f} = f \wedge \gamma$, and then take the limit as γ diverges to infinity.