entropy

MDPI

Article

The Case for Shifting the Rényi Entropy

Francisco J. Valverde-Albacete † and **Carmen Peláez-Moreno** *,†

Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain; fva@tsc.uc3m.es

* Correspondence: carmen@tsc.uc3m.es; Tel.: +34-91-624-8771

† These authors contributed equally to this work.

Received: 15 November 2018; Accepted: 7 January 2019; Published: 9 January 2019

check for updates

Abstract: We introduce a variant of the Rényi entropy definition that aligns it with the well-known Hölder mean: in the new formulation, the r-th order Rényi Entropy is the logarithm of the inverse of the r-th order Hölder mean. This brings about new insights into the relationship of the Rényi entropy to quantities close to it, like the information potential and the partition function of statistical mechanics. We also provide expressions that allow us to calculate the Rényi entropies from the Shannon cross-entropy and the escort probabilities. Finally, we discuss why shifting the Rényi entropy is fruitful in some applications.

Keywords: shifted Rényi entropy; Shannon-type relations; generalized weighted means; Hölder means; escort distributions

1. Introduction

The suggestive framework for the description and assessment of information transmission that Shannon proposed and co-developed [1–3] soon took hold of the mind of a generation of scientists and overflowed its initial field of application, despite the cautions of the inceptor himself [4]. He had independently motivated and re-discovered the Boltzmann description for the thermodynamic entropy of a system with many micro-states [5]. His build-up of the concept starting from Hartley's measure of information using the nowadays well-known axiomatic approach created a sub-science—perhaps a science—out of three papers. For information scientists, it is difficult to shatter the intellectual chains of Shannon's entropy [5–11].

After Shannon's introduction of his re-purposing of the Boltzmann entropy to analyze communication, many generalizations of it were proposed, among which Rényi's [12], Hvarda-Charvat-Tsallis' [13] and Csiszar's [14] seem to have found the widest echo. Reviews of information measures with different points of view are [14,15].

In this paper we want to contribute to the characterization and popularization of the Rényi entropy as a proper generalization of the Shannon entropy. Rényi's suggestion was obtained after noticing some limits to the axiomatic approach [16], later better analyzed by Aczel and Daroczny [17]. His critical realisation was that there are more ways to develop the means of the individual surprisals of a collection of events, whereby he resorted to the Kolmogorov-Nagumo theory of the means [18–20]. In fact, Kolmogorov had been present in the history Information Theory from foundational issues [18], to punctual clarification [21], to his own devising of a measure of entropy-complexity. The situation concerning the theory of the means at the time is described in [22].

Rényi was quite aware that entropy is a quantity related to the averages of the information function on a probability distribution: let $X \sim P_X$ be a random variable over a set of outcomes

$\mathcal{X} = \{x_i \mid 1 \le i \le n\}$ and *pmf* P_X defined in terms of the non-null values $p_i = P_X(x_i)$. The Rényi entropy for X is defined in terms of that of P_X as $H_\alpha(X) = H_\alpha(P_X)$ by a case analysis [12]

$$\begin{aligned} \alpha \neq 1 \quad & H_\alpha(P_X) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha \\ \alpha = 1 \quad & \lim_{\alpha \to 1} H_\alpha(P_X) = H(P_X) \end{aligned} \tag{1}$$

where $H(P_X) = -\sum_{i=1}^{n} p_i \log p_i$ is the Shannon entropy [1–3]. Similarly the associated divergence when $Q \sim Q_X$ is substituted by $P \sim P_X$ on a compatible support is defined in terms of their *pmf*s $q_i = Q_X(x_i)$ and $p_i = P_X(x_i)$, respectively, as $D_\alpha(X\|Q) = D_\alpha(P_X\|Q_X)$ where

$$\begin{aligned} \alpha \neq 1 \quad & D_\alpha(P_X\|Q_X) = \frac{1}{\alpha - 1} \log \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} \\ \alpha = 1 \quad & \lim_{\alpha \to 1} D_\alpha(P_X\|Q_X) = D_{KL}\left(P_X\|Q_X\right). \end{aligned} \tag{2}$$

and $D_{KL}\left(P_X\|Q_X\right) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$ is the Kullback–Leibler divergence [23].

When trying to find the closed form for a generalization of the Shannon entropy that was compatible with all the Faddev axioms but that of linear average, Rényi found that the function $\varphi(x) = x^r$ could be used with the Kolmogorov–Nagumo average to obtain such a new form of entropy. Rather arbitrarily, he decided that the constant should be $\alpha = r + 1$, thus obtaining (1) and (2), but obscuring the relationship of the entropies of order α and the generalized power means.

We propose to shift the parameter in these definitions back to $r = \alpha - 1$ to define the *shifted Rényi entropy of order* r the value

$$\tilde{H}_r(P_X) = -\log M_r(P_X, P_X)$$

and the *shifted Rényi divergence of order* r the value

$$\tilde{D}_r(P_X\|Q_X) \; = \log M_r(P_X, \frac{P_X}{Q_X})$$

where M_r is the *r-th order weighted generalized power means* or *Hölder means* [24]:

$$M_r(\vec{w}, \vec{x}) = \left(\sum_{i=1}^{n} \frac{w_i}{\sum_k w_k} \cdot x_i^r \right)^{\frac{1}{r}}.$$

In our opinion, this shifted version may be more fruitful than the original. However, since this could be deemed equally arbitrary, in this paper we argue that this statement of the Rényi entropy greatly clarifies its role vis-a-vis the Hölder means, viz. that most of the properties and special cases of the Rényi entropy arise from similar concerns in the Hölder means. We also provide a brief picture of how the theory surrounding the Rényi entropy would be modified with this change, as well as its relationship to some other magnitudes.

2. Preliminaries

2.1. The Generalized Power Means

Recall that the *generalized power or Hölder mean of order* r is defined as

$$M_r(\vec{w}, \vec{x}) = \left(\frac{\sum_{i=1}^{n} w_i \cdot x_i^r}{\sum_k w_k} \right)^{\frac{1}{r}} = \left(\sum_{i=1}^{n} \frac{w_i}{\sum_k w_k} \cdot x_i^r \right)^{\frac{1}{r}} \tag{3}$$

By formal identification, the generalized power mean is nothing but the weighted f-mean with $f(x) = x^r$ (see Appendix A). In this paper we use the notation where the weighting vector comes first—rather than the opposite, used in [24]—to align it with formulas in information theory, e.g., divergences and cross entropies. Reference [25] provides proof that this functional mean also has the Properties 1–3 of Proposition A1 and Associativity.

The evolution of $M_r(\vec{w}, \vec{x})$ with r is also called the *Hölder path (of an $\vec{x}$)*. Important cases of this mean for historical and practical reasons are obtained by giving values to r:

- The (weighted) geometric mean when $r = 0$.

$$M_0(\vec{w}, \vec{x}) = \lim_{r \to 0} M_r(\vec{w}, \vec{x}) = \left(\Pi_{i=1}^n x_i^{w_i}\right)^{\frac{1}{\sum_k w_k}} \tag{4}$$

- The weighted arithmetic mean when $r = 1$.

$$M_1(\vec{w}, \vec{x}) = \sum_{i=1}^n \frac{w_i}{\sum_k w_k} \cdot x_i$$

- The weighted harmonic mean for $r = -1$.

$$M_{-1}(\vec{w}, \vec{x}) = \left(\sum_{i=1}^n \frac{w_i}{\sum_k w_k} \cdot x_i^{-1}\right)^{-1} = \frac{\sum_k w_k}{\sum_{i=1}^n w_i \cdot \frac{1}{x_i}}$$

- The quadratic mean for $r = 2$.

$$M_2(\vec{w}, \vec{x}) = \left(\sum_{i=1}^n \frac{w_i}{\sum_k w_k} \cdot x_i^2\right)^{\frac{1}{2}}$$

- Finally, the max- and min-means appear as the limits:

$$M_\infty(\vec{w}, \vec{x}) = \lim_{r \to \infty} M_r(\vec{w}, \vec{x}) = \max_{i=1}^n x_i$$
$$M_{-\infty}(\vec{w}, \vec{x}) = \lim_{r \to -\infty} M_r(\vec{w}, \vec{x}) = \min_{i=1}^n x_i$$

They all show the following properties:

Proposition 1 (Properties of the weighted power means). *Let $\vec{x}, \vec{w} \in (0, \infty)^n$ and $r, s \in (-\infty, \infty)$. Then, the following formal identities hold, where $\vec{x}^r$ and $\frac{1}{\vec{x}}$ are to be understood entry-wise,*

1. *(0- and 1-order homogeneity in weights and values) If $k_1, k_2 \in \mathbb{R}_{\geq 0}$, then $M_r(k_1 \cdot \vec{w}, k_2 \cdot \vec{x}) = k_1^0 \cdot k_2^1 \cdot M_r(\vec{w}, \vec{x})$.*
2. *(Order factorization) If $r \neq 0 \neq s$, then $M_{rs}(\vec{w}, \vec{x}) = (M_s(\vec{w}, (\vec{x})^r))^{1/r}$.*
3. *(Reduction to the arithmetic mean) If $r \neq 0$, then $M_r(\vec{w}, \vec{x}) = [M_1(\vec{w}, (\vec{x})^r)]^{1/r}$.*
4. *(Reduction to the harmonic mean) If $r \neq 0$, then $M_{-r}(\vec{w}, \vec{x}) = [M_{-1}(\vec{w}, (\vec{x})^r)]^{1/r} = [M_r(\vec{w}, \frac{1}{\vec{x}})]^{-1} = [M_1(\vec{w}, \frac{1}{(\vec{x})^r})]^{-1/r}$.*
5. *(Monotonicity in r) Furthermore, $\vec{x} \in [0, \infty]^n$ and $r, s \in [-\infty, \infty]$, then*

$$\min_i x_i = M_{-\infty}(\vec{w}, \vec{x}) \leq M_r(\vec{w}, \vec{x}) \leq M_\infty(\vec{w}, \vec{x}) = \max_i x_i$$

and the mean is a strictly monotonic function of r, that is $r < s$ implies $M_r(\vec{w}, \vec{x}) < M_s(\vec{w}, \vec{x})$, unless:

- *$x_i = k$ is constant, in which case $M_r(\vec{w}, \vec{x}) = M_s(\vec{w}, \vec{x}) = k$.*

- *$s \leq 0$ and some $x_i = 0$, in which case $0 = M_r(\vec{w}, \vec{x}) \leq M_s(\vec{w}, \vec{x})$.*
- *$0 \leq r$ and some $x_i = \infty$, in which case $M_r(\vec{w}, \vec{x}) \leq M_s(\vec{w}, \vec{x}) = \infty$.*

6. *(Non-null derivative) Call $\tilde{q}_r(\vec{w}, \vec{x}) = \left\{ \frac{w_k x_k^r}{\sum_i w_i x_i^r} \right\}_{k=1}^n$. Then*

$$\frac{d}{dr} M_r(\vec{w}, \vec{x}) = \frac{1}{r} \cdot M_r(\vec{w}, \vec{x}) \ln \frac{M_0(\tilde{q}_r(\vec{w}, \vec{x}), \vec{x})}{M_r(\vec{w}, \vec{x})} \tag{5}$$

Proof. Property 1 follows from the commutativity, associativity and cancellation of sums and products in $\mathbb{R}_{\geq 0}$. Property 2 follows from identification in the definition, then Properties 3 and 4 follow from it with $s = 1$ and $s = -1$ respectively. Property 5 and the special cases in it are well known and studied extensively in [24]. We will next prove property 6

$$\frac{d}{dr} M_r(\vec{w}, \vec{x}) = \frac{d}{dr} e^{\frac{1}{r} \ln\left(\sum_k \frac{w_k}{\sum_i w_i} x_k^r \right)} = M_r(\vec{w}, \vec{x}) \left(\frac{-1}{r^2} \ln \left(\sum_k \frac{w_k}{\sum_i w_i} x_k^r \right) + \frac{1}{r} \cdot \frac{\sum_k w_k x_k^r \ln x_k}{\sum_i w_i x_i^r} \right)$$

Note that if we call $\tilde{q}_r(\vec{w}, \vec{x}) = \{w'_k\}_{k=1}^n = \left\{ \frac{w_k x_k^r}{\sum_i w_i x_i^r} \right\}_{k=1}^n$, since this is a probability we may rewrite:

$$\sum_k \frac{w_k x_k^r}{\sum_i w_i x_i^r} \cdot \ln x_k = \sum_k w'_k \ln x_k = \ln \left(\prod_k x_k^{w'_k} \right) = \ln M_0(\tilde{q}_r(\vec{w}, \vec{x}), x)$$

whence

$$\begin{aligned} \frac{d}{dr} M_r(\vec{w}, \vec{x}) &= M_r(\vec{w}, \vec{x}) \left(\frac{1}{r} \cdot \ln M_0(\tilde{q}_r(\vec{w}, \vec{x}), x) - \frac{1}{r} \cdot \ln M_r(\vec{w}, \vec{x}) \right) \\ &= \frac{1}{r} \cdot M_r(\vec{w}, \vec{x}) \ln \frac{M_0(\tilde{q}_r(\vec{w}, \vec{x}), \vec{x})}{M_r(\vec{w}, \vec{x})}. \end{aligned}$$

□

Remark 1. *The distribution $\tilde{q}_r(\vec{w}, \vec{x})$ when $\vec{w} = \vec{x}$ is extremely important in the theory of generalized entropy functions, where it is called a* (shifted) escort distribution (of $\vec{w}$) *[5], and we will prove below that its importance stems, at leasts partially, from this property.*

Remark 2. *Notice that in the case where both conditions at the end of Property 1.5 hold—that is for $i \neq j$ we have $x_i = 0$ and $x_j = \infty$—then we have for $r \leq 0$, $M_r(\vec{w}, \vec{x}) = 0$ and for $0 \leq r, M_r(\vec{w}, \vec{x}) = \infty$ whence $M_r(\vec{w}, \vec{x})$ has a discontinuity at $r = 0$.*

2.2. Rényi's Entropy

Although the following material is fairly standard, it bears directly into our discussion, hence we introduce it in full.

2.2.1. Probability Spaces, Random Variables and Expectations

Shannon and Rényi set out to find how much information can be gained *on average* by a single performance of an experiment Ω under different suppositions. For that purpose, let $(\Omega, \Sigma_\Omega, P)$ be a measure space, with $\Omega = \{\omega_1, \ldots, \omega_n\}$ the set of outcomes of a random experiment, Σ_Ω the sigma-algebra of this set and measure $P : \Omega \to \mathbb{R}_{\geq 0}, P(\omega_i) = p_i, 1 \leq k \leq n$. We define the support of P, as the set of outcomes with positive probability $\text{supp}\,(P) = \{\omega \in \Omega \mid P(\omega) > 0\}$.

Let $(\mathcal{X}, \Sigma_\mathcal{X})$ be a measurable space with $\mathcal{X}$ a domain and $\Sigma_\mathcal{X}$ its sigma algebra and consider the random variable $X : \Omega \to \mathcal{X}$, that is, a measurable function so that for each set of $B \in \Sigma_\mathcal{X}$ we have $X^{-1}(B) \in \Sigma_\Omega$. Then P induces a measure P_X on $(\mathcal{X}, \Sigma_\mathcal{X})$ with $\forall x \in \Sigma_\mathcal{X}$, $P_X(x) = P(X = x) =$

$P(X^{-1}(x))$, where x is an event in $\Sigma_{\mathcal{X}}$, and $P_X(x) = \sum_{\omega_i \subseteq X^{-1}(x)} P(\omega_i)$ whereby $(\mathcal{X}, \Sigma_{\mathcal{X}}, P_X)$ becomes a measure space. We will use mostly $X \sim P_X$ to denote a random variable, instead of its measurable space. The reason for this is that since information measures are defined on distributions, this is the more fundamental notion for us.

Sometimes co-occurring random variables are defined on the same sample space and sometimes on different ones. Hence, we will need another measure space sharing the same measurable space (Ω, Σ_Ω) but different measure, $(\Omega, \Sigma_\Omega, Q)$ with $Q(\omega_i) = q_i$.

Remark 3. *Modernly, discrete distributions are sets or vectors of non-negative numbers adding up to* 1*, but Rényi developed his theory for "defective distributions", that is, with $\sum_i P(\omega_i) \neq 1$ which are better described as "positive measures". In fact, we do not need to distinguish whether P is a probability measure in the $(n-1)$-simplex $P \in \Delta^{n-1} \Leftrightarrow \sum_i P(\omega_i) = 1$ or in general a measure $P \in \mathbb{R}^n_{\geq 0}$ and nothing precludes using the latter to define entropies—while it provides a bit of generalization this is the road we will take below (see [12,26] on using incomplete distributions with $\sum_i p_i < 1$).*

2.2.2. The Approach to Rényi's Information Functions Based in Postulates

One of the most important applications of the generalized weighted means is to calculate the moments of (non-negative) random variables.

Lemma 1. *Let $X \sim P_X$ be a discrete random variable. Then the r-*th moment of *X is:*

$$E_X\{X^r\} = \sum_i p_i x_i^r = (M_r(P_X, X))^r \tag{6}$$

This is the concept that Shannon, and afterwards Rényi, used to quantify information by using the distribution as a random variable (Section 3.3).

The postulate approach to characterize Shannon's information measures can be found in Appendix B. Analogue generalized postulates lead to Rényi's information functions, but, importantly, he did not consider normalized measures, that is with $\sum_k p_k = 1$.

We follow [27] in stating the Rényi postulates:

1. The amount of information provided by a single random event x_k should be a function of its probability $P_X(x_k) = p_k$, not its value $x_k = X(\omega_k)$, $\mathfrak{I} : [0,1] \to I$ where $I \subseteq \mathbb{R}$ quantifies information.
2. This amount of information should be additive on independent events.

$$\mathfrak{I}(p,q) = \mathfrak{I}(p) + \mathfrak{I}(q) \tag{7}$$

3. The amount of information of a binary equiprobable decision is one bit.

$$\mathfrak{I}(1/2) = 1 \tag{8}$$

4. If different amounts of information occur with different probabilities the total amount of information $\mathfrak{I}$ is an *average* of the individual information amounts weighted by the probability of occurrence.

These postulates *may* lead to the following consequences:

- Postulates 1 and 2 fix *Hartley's function* as the single possible amount of information of a basic event

$$\mathfrak{I} : [0,1] \to [0,\infty], p \mapsto \mathfrak{I}(p) = -k \log p. \tag{9}$$

- Postulates 3 fixes the base of the logarithm in Hartley's formula to 2 by fixing $k = 1$. Any other value $k = 1/\log b$ fixes b as the base for the logarithm and changes the unit.

- Postulate 4 defines an average amount of information, or *entropy*, properly speaking. Its basic formula is a form of the Kolmogorov–Nagumo formula or f-mean (A2) applied to information

$$H(P_X, \varphi, \mathfrak{I}) = \varphi^{-1}\left(\Sigma_{i=1}^{n} \frac{p_i}{\sum_k p_k} \varphi(\mathfrak{I}(p_i))\right). \tag{10}$$

  Thus the "entropy" in Information Theory is, by definition, synonym with "aggregate amount of information", which departs from its physical etymology, despite the numerous analogies between both concepts.

  It has repeatedly been proven that only two forms of the function φ can actually be used in the Kolmogorov–Nagumo formula that respect the previous postulates [12,26,27]:

  - The one generating Shannon's entropy:

$$\varphi(h) = ah + b \,\text{with}\, a \neq 0, \tag{11}$$

  - That originally used by Rényi himself:

$$\varphi(h) = 2^{(1-\alpha)h}, \,\text{with}\, \alpha \in [-\infty, \infty] \setminus \{1\}. \tag{12}$$

Taking the first form (11) and plugging it into (10) leads to *Shannon's measure of information*, and taking the second form leads to *Rényi's measure of information* (1), so we actually have:

Definition 1 ([12,26])**.** *The* Rényi entropy of order α *for a discrete random variable* $X \sim P_X$*, is*

$$H_\alpha(P_X) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} \frac{p_i^\alpha}{\sum_k p_k}\right), \quad \alpha \neq 1 \qquad \lim_{\alpha \to 1} H_\alpha(P_X) = H(P_X) = -\sum_i \frac{p_i}{\sum_k p_k} \log p_i, \tag{13}$$

where the fact that Shannon's entropy is the Rényi entropy when $\alpha \to 1$ *in* (1) *is found by a continuity argument.*

Rényi also used the postulate approach to define the following quantity:

Definition 2 ([12,26])**.** *The* gain of information *or* divergence (between distributions) *when* $Y \sim P_Y$*,* $P_Y(y_i) = q_i$ *is substituted by* $X \sim P_X$*,* $P_X(x_i) = p_i$ *being continuous wrt the latter—that is, with* $\operatorname{supp} Y \subseteq \operatorname{supp} X$*—as*

$$D_\alpha(P_X \| P_Y) = \frac{1}{\alpha - 1} \log \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}, \quad \alpha \neq 1 \qquad \lim_{\alpha \to 1} D_\alpha(P_X \| P_Y) = D_{KL}(P_X \| P_Y)$$

and the fact that Kullback–Leibler's divergence emerges as the limit when $\alpha \to 1$ *follows from the same continuity argument as before. Such special cases will not be stated again, as motivated in Section 3.1.*

As in the Shannon entropy case, the rest of the quantities arising in Information Theory can be defined in terms of the generalized entropy and its divergence [23,27].

3. Results

3.1. The Shifted Rényi Entropy and Divergence

To leverage the theory of generalized means to our advantage, we start with a correction to Rényi's entropy definition: The investigation into the form of the transformation function for the Rényi entropy (12) is arbitrary in the parameter α that it chooses. In fact, we may substitute in $r = \alpha - 1$ to obtain the pair of formulas:

$$\varphi'(h) = b^{-rh} \qquad\qquad \varphi'^{-1}(p) = \frac{-1}{r}\log_b p \tag{14}$$

Definition 3. *The shifted Rényi entropy of order $r \neq 0$ for a discrete random variable $X \sim P_X$, is the Kolmogorov–Nagumo φ'-mean (10) of the information function $\mathfrak{I}_*(p) = -\ln p$ over the probability values.*

$$\tilde{H}_r\left(P_X\right) = \frac{-1}{r}\log_b\left(\sum_i \frac{p_i}{\sum_k p_k} p_i^r\right) \qquad\qquad \lim_{r\to 0}\tilde{H}_r\left(P_X\right) = H(P_X). \tag{15}$$

Note that:

- *For $r \neq 0$ this is motivated by:*

$$\tilde{H}_r\left(P_X\right) = \frac{-1}{r}\log_b\left(\sum_i \frac{p_i}{\sum_k p_k} b^{r\log_b p_i}\right) = \frac{-1}{r}\log_b\left(\sum_i \frac{p_i}{\sum_k p_k} b^{\log_b p_i^r}\right) = \frac{-1}{r}\log_b\left(\sum_i \frac{p_i}{\sum_k p_k} p_i^r\right).$$

- *For $r = 0$ we can use the linear mean $\varphi(h) = ah + b$ with inverse $\varphi^{-1}(p) = \frac{1}{a}(p - b)$ as per the standard definition, leading to Shannon's entropy.*

Remark 4. *The base of the logarithm is not important as long as it is maintained in $\varphi'(\cdot)$, $\mathfrak{I}_*(\cdot)$ and their inverses, hence we leave it implicit. For some calculations—e.g., the derivative below—we explicitly provide a particular basis—e.g., $\log_e x = \ln x$.*

The shifted divergence can be obtained in the same manner—the way that Rényi followed himself [26].

Definition 4. *The shifted Rényi divergence between two distributions $P_X(x_i) = p_i$ and $Q_X(x_i) = q_i$ with compatible support is the following quantity.*

$$\tilde{D}_r\left(P_X \| Q_X\right) = \frac{1}{r}\log\sum_i \frac{p_i}{\sum_k p_k}\left(\frac{p_i}{q_i}\right)^r \qquad\qquad \lim_{r\to 0}\tilde{D}_r\left(P_X \| Q_X\right) = D_{KL}\left(P_X \| Q_X\right). \tag{16}$$

Of course, the values of the Rényi entropy and divergence are not modified by this shifting.

Lemma 2. *The Rényi entropy and the shifted Rényi entropy produce the same value, and similarly for their respective divergences.*

Proof. if we consider a new parameter $r = \alpha - 1$ we have:

$$H_\alpha(P_X) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^n \frac{p_i^\alpha}{\sum_k p_k}\right) = \frac{-1}{r}\log\left(\sum_{i=1}^n \frac{p_i^{r+1}}{\sum_k p_k}\right) = -\frac{1}{r}\log\left(\sum_{i=1}^n \frac{p_i}{\sum_k p_k} p_i^r\right) = \tilde{H}_r(P_X).$$

and similarly for the divergence:

$$D_\alpha(P_X \| Q_X) = \frac{1}{\alpha - 1}\log\sum_{i=1}^n \frac{p_i^\alpha q_i^{1-\alpha}}{\sum_k p_k} = \frac{1}{r}\log\sum_{i=1}^n \frac{p_i^{r+1} q_i^{-r}}{\sum_k p_k} = \frac{1}{r}\log\sum_{i=1}^n \frac{p_i}{\sum_k p_k}\left(\frac{p_i}{q_i}\right)^r = \tilde{D}_r\left(P_X \| Q_X\right)$$

The Shannon entropy and Kullback–Leibler divergences are clearly the limit cases. □

3.1.1. The Case for Shifting the Rényi Entropy

So what could be the reason for the shifting? First and foremost, it is a re-alignment with the more basic concept of generalized mean.

Proposition 2. *The Shifted Rényi Entropy and Divergence are logarithmic transformations of the generalized power means:*

$$\tilde{H}_r\left(P_X\right) = \log \frac{1}{M_r(P_X, P_X)} \tag{17}$$

$$\tilde{D}_r\left(P_X \| Q_X\right) = \log M_r(P_X, \frac{P_X}{Q_X}) \tag{18}$$

Proof. Simple identification of (15) and (16) in the definition of power mean definitions (3). □

Table 1 lists the shifting of these entropies and their relation both to the means and to the original Rényi definition in the parameter α.

Table 1. Relation between the most usual weighted power means, Rényi entropies and shifted versions of them.

Mean Name	**Mean $M_r(\vec{w}, \vec{x})$**	**Shifted Entropy $\tilde{H}_r(P_X)$**	**Entropy Name**	α	r
Maximum	$\max_i x_i$	$\tilde{H}_\infty = -\log \max_i p_i$	min-entropy	∞	∞
Arithmetic	$\sum_i w_i x_i$	$\tilde{H}_1 = -\log \sum_i p_i^2$	Rényi's quadratic	2	1
Geometric	$\Pi_i x_i^{w_i}$	$\tilde{H}_0 = -\sum_i p_i \log p_i$	Shannon's	1	0
Harmonic	$(\sum_i w_i \frac{1}{x_i})^{-1}$	$\tilde{H}_{-1} = \log n$	Hartley's	0	-1
Minimum	$\min_i x_i$	$\tilde{H}_{-\infty} = -\log \min_i p_i$	max-entropy	$-\infty$	$-\infty$

Remark 5. *It is no longer necessary to make the distinction between the case $r \to 0$—Shannon's—and the rest, since the means are already defined with this caveat. This actually downplays the peculiar features of Shannon's entropy, arising from the geometric mean when $\sum_i p_i = 1$:*

$$\tilde{H}_0\left(p_x\right) = \log \frac{1}{M_0(P_X, P_X)} = -\log\left(\prod_i p_i^{p_i}\right) = -\sum_i p_i \log p_i$$

However, the prominence of the Shannon entropy will emerge once again in the context of rewriting entropies in terms of each other (Section 3.2).

Since the means are properly defined for all $r \in [-\infty, \infty]$, $\tilde{H}_r\left(P_X\right)$ is likewise properly defined for all $r \in [-\infty, \infty]$—and therefore the non-shifted version with $\alpha = r + 1$. This is probably the single strongest argument in favour of the shifting and motivates the following definition.

Definition 5 (The Rényi information spectrum)**.** *For fixed P_X we will refer to $\tilde{H}_r\left(P_X\right)$ as its* Rényi information spectrum *over parameter r.*

Also, some relationships between magnitudes are clarified in the shifted enunciation with respect to the traditional one, for instance, the relation between the Rényi entropy and divergence.

Lemma 3. *The shifted formulation makes the entropy the self-information with a change of sign:*

$$\tilde{H}_r\left(P_X\right) = \tilde{D}_{-r}\left(P_{XX} \| P_X P_X\right). \tag{19}$$

Proof. $\tilde{D}_{-r}\left(P_{XX} \| P_X P_X\right) = \tilde{D}_{-r}\left(P_X \| P_X P_X\right) = \frac{-1}{r} \log \sum_i p_i \left(\frac{p_i}{p_i p_i}\right)^{-r} = \frac{-1}{r} \log \sum_i p_i \left(\frac{1}{p_i}\right)^{-r} = \tilde{H}_r\left(P_X\right)$. □

Recall that in the common formulation, $H_\alpha(P_X) = D_{2-\alpha}(P_X \| P_X P_X)$ [23].

Another simplification is the fact that the properties of the Rényi entropy and divergence stem from those of the means, inversion and logarithm.

Proposition 3 (Properties of the Rényi spectrum of P_X)**.** *Let $r, s \in \mathbb{R} \cup \{\pm\infty\}$, and $P_X, Q_X \in \Delta^{n-1}$ where Δ^{n-1} is the simplex over the support* supp X*, with cardinal* $|\operatorname{supp} X| = n$*. Then,*

1. *(Monotonicity) The Rényi entropy is a non-increasing function of the order r.*
$$s \le r \Rightarrow \tilde{H}_s\left(P_X\right) \ge \tilde{H}_r\left(P_X\right) \tag{20}$$
2. *(Boundedness) The Rényi spectrum $\tilde{H}_r\left(P_X\right)$ is bounded by the limits*
$$\tilde{H}_{-\infty}\left(P_X\right) \ge \tilde{H}_r\left(P_X\right) \ge \tilde{H}_{\infty}\left(P_X\right) \tag{21}$$
3. *The entropy of the uniform* pmf *U_X is constant over r.*
$$\forall r \in \mathbb{R} \cup \{\pm\infty\}, \tilde{H}_r\left(U_X\right) = \log n \tag{22}$$
4. *The Hartley entropy ($r = -1$) is constant over the distribution simplex.*
$$\tilde{H}_{-1}\left(P_X\right) = \log n \tag{23}$$
5. *(Divergence from uniformity) The divergence of any distribution P_X from the uniform U_X can be written in terms of the entropies as:*
$$\tilde{D}_r\left(P_X \| U_X\right) = \tilde{H}_r\left(U_X\right) - \tilde{H}_r\left(P_X\right). \tag{24}$$
6. *(Derivative of the shifted entropy) The derivative in r of Rényi's r-th order entropy is*
$$\frac{d}{dr}\tilde{H}_r\left(P_X\right) = \frac{-1}{r^2}\tilde{D}_0\left(\tilde{q}_r(P_X) \| P_X\right) = \frac{-1}{r}\log\frac{M_0(\tilde{q}_r(P_X), P_X)}{M_r(P_X, P_X)}, \tag{25}$$
where $\tilde{q}_r(P_X) = \left\{\frac{p_i p_i^r}{\sum_k p_k p_k^r}\right\}_{i=1}^n$ for $r \in \mathbb{R} \cup \{\pm\infty\}$ are the shifted escort distributions.
7. *(Relationship with the moments of P_X) The shifted Rényi Entropy of order r is the logarithm of the inverse r-th root of the r-th moment of P_X.*
$$\tilde{H}_r\left(P_X\right) = -\frac{1}{r}\log E_{P_X}\{P_X^r\} = \log\frac{1}{\sqrt[r]{E_{P_X}\{P_X^r\}}} \tag{26}$$

Proof. Note that properties used in the following are referred to the Proposition they are stated in. Property 1 issues from Property 1.2 and Hartley's information function being order-inverting or antitone. Since the free parameter r is allowed to take values in $[-\infty, \infty]$, Property 2 follows directly from Property 1. With respect to Property 3, we have, from $U_X = 1/|\operatorname{supp} X| = 1/n$ and Property A1.3:
$$\tilde{H}_r(\frac{1}{n}) = -\log M_r(\frac{1}{n}, \frac{1}{n}) = -\log\frac{1}{n} = \log n.$$

For Property 4 we have:
$$\tilde{H}_{-1}(P_X) = -\log(\sum_i p_i \cdot p_i^{-1})^{-1} = -\log(n)^{-1} = \log n$$

While for Property 5,

$$\tilde{D}_r(P_X\|U_X) = \frac{1}{r}\log\left[\sum_i p_i\left(\frac{p_i}{u_i}\right)^r\right] = \frac{1}{r}\log\left[\sum_i p_i\left(\frac{p_i}{1/n}\right)^r\right] = \frac{1}{r}\log\left[n^r\left(\sum_i p_i p_i^r\right)\right]$$
$$= \log n + \log\left(\sum_i p_i p_i^r\right)^{1/r} = \tilde{H}_r(U_X) - \tilde{H}_r(P_X).$$

For the third term of Property 6, we have from (17) with natural logarithm, with P_X in the role both of $\vec{w}$ and $\vec{x}$

$$\frac{d}{dr}\tilde{H}_r(P_X) = \frac{d}{dr}\left(-\ln M_r(P_X,P_X)\right) = -\frac{\frac{d}{dr}M_r(P_X,P_X)}{M_r(P_X,P_X)},$$

whence the property follows directly from (5). For the first identity, though, we have:

$$\frac{d\tilde{H}_r(P_X)}{dr} = -\frac{d}{dr}\left[\frac{1}{r}\ln\sum_i p_i p_i^r\right] = -\left[\frac{-1}{r^2}\ln\sum_i p_i p_i^r + \frac{1}{r}\sum_i\frac{p_i p_i^r}{\sum_i p_i p_i^r}\ln p_i\right].$$

If we introduce the abbreviation

$$\tilde{q}_r(P_X) = \tilde{q}_r(P_X,P_X) = \{\tilde{q}_r(P_X)_i\}_{i=1}^n = \left\{\frac{p_i p_i^r}{\sum_k p_k p_k^r}\right\}_{i=1}^n \tag{27}$$

noticing that $\ln\sum_k p_k p_k^r = \sum_i \tilde{q}_r(P_X)_i \ln(\sum_k p_k p_k^r)$, since $\tilde{q}_r(P_X)$ is a distribution, and factoring out $-1/r^2$:

$$\frac{d\tilde{H}_r(P_X)}{dr} = -\frac{1}{r^2}\left[-\sum_i\tilde{q}_r(P_X)_i\ln(\sum_k p_k p_k^r) + r\left(\sum_i\tilde{q}_r(P_X)_i\ln p_i\right) \pm \sum_i\tilde{q}_r(P_X)_i\ln p_i\right]$$
$$= -\frac{1}{r^2}\left[-\sum_i\tilde{q}_r(P_X)_i\ln(\sum_k p_k p_k^r) + (r+1)\sum_i\tilde{q}_r(P_X)_i\ln p_i - \sum_i\tilde{q}_r(P_X)_i\ln p_i\right]$$
$$= -\frac{1}{r^2}\left[\sum_i\tilde{q}_r(P_X)_i\ln\frac{p_i p_i^r}{\sum_k p_k p_k^r} - \sum_i\tilde{q}_r(P_X)_i\ln p_i\right] = -\frac{1}{r^2}\sum_i\tilde{q}_r(P_X)_i\ln\frac{\tilde{q}_r(P_X)_i}{p_i}$$

and recalling the definition of the shifted divergence we have the result.

For Property 7, in particular, the *probability of any event* is a function of the random variable $P_X(x_i) = p_i$ whose r-th *moment of* P_X is

$$E_X\{P_X^r\} = \sum_i p_i p_i^r = (M_r(P_X,P_X))^r \tag{28}$$

The result follows by applying the definition of the shifted entropy in terms of the means. □

Remark 6. *In the preceding proof we have introduced the notion of* shifted escort probabilities $\tilde{q}_r(P_X)$ *acting in the shifted Rényi entropies as the analogues of the* escort probabilities *in the standard definition (see [5] and Section 2.1). This notion of shifted escort probabilities is the one requested by Property 1.6 by instantiation of variables* $\tilde{q}(P_X) = \tilde{q}(P_X,P_X)$. *But notice also that* $(\tilde{q}_r(P_X))_i = \frac{p_i p_i^r}{\sum_k p_k p_k^r} = \frac{p_i^\alpha}{\sum_k p_k^\alpha} = (q_\alpha(P_X))_i$ *is just the shifting of the traditional escort probabilities [5].*

Note that for $P_X \in \mathbb{R}^n_{\geq 0}$:

- $\tilde{q}_0(P_X)$ *is the normalization of* P_X. *In fact,* $P_X \in \Delta^{n-1}$ *if and only if we have* $\tilde{q}_0(P_X) = P_X$.
- $\tilde{q}_{-1}(P_X)(x_i) = |\operatorname{supp} P_X|^{-1}$ *if* $x_i \in \operatorname{supp} P_X$ *and* 0 *otherwise.*

- *Furthermore, if P_X has P maxima (M minima), then $\tilde{q}_\infty(P_X)$ ($\tilde{q}_{-\infty}(P_X)$) is an everywhere null distribution but at the indices where the maxima (minima) of P_X are situated:*

$$\tilde{q}_\infty(P_X)(x_i) = \begin{cases} \frac{1}{P} & x_i \in \arg\max P_X \\ 0 & \textit{otherwise} \end{cases} \qquad \tilde{q}_{-\infty}(P_X)(x_i) = \begin{cases} \frac{1}{M} & x_i \in \arg\min P_X \\ 0 & \textit{otherwise} \end{cases}$$

Another important point made clear by this relation to the means is the fact that *all positive measures have a Rényi spectrum*: although so far we conceived the origin of information to be a probability function, nothing precludes applying the same procedure to non-negative, non-normalized quantities with $\sum_x f_X(x) \neq 1$, e.g., masses, sums, amounts of energy, etc.

It is well-understood that in this situation Rényi's entropy has to be slightly modified to accept this procedure. The reason for this is Property 1.1 of the means: generalized means are 1-homogeneous in the numbers being averaged, but 0-homogeneous in the weights. In the Rényi spectrum both these roles are fulfilled by the pmf. Again the escort distributions allow us to analyze the measure:

Lemma 4. *Consider a random variable $X \sim M_X$ with non-normalized measure $M_X(x_i) = m_i$ such that $\sum_i m_i = M \neq 1$. Then the normalized probability measure $\tilde{q}_0(M_X) = \{m_i / \sum_i m_i\}_{i=1}^n$ provides a Rényi spectrum that is displaced relative to that of the measure as:*

$$\tilde{H}_r\left(M_X\right) = \tilde{H}_r\left(\tilde{q}_0(M_X)\right) - \log M. \tag{29}$$

Proof.

$$\begin{aligned}\tilde{H}_r\left(\tilde{q}_0(M_X)\right) &= -\log M_r(\tilde{q}_0(M_X), \tilde{q}_0(M_X)) = -\frac{1}{r}\log\sum_i \frac{m_i}{M}\left(\frac{m_i}{M}\right)^r = \log M - \frac{1}{r}\log\sum_i \frac{m_i}{M} m_i^r \\ &= \log M - \log M_r(M_X, M_X) = \log M + \tilde{H}_r\left(M_X\right)\end{aligned}$$

□

Remark 7. *When $M \geq 1, -\log M \leq 0$ with equality for $M = 1$ and that if $M < 1$ then $-\log M > 0$. This last was the original setting Rényi envisioned and catered for in the definitions, but nothing precludes the extension provided by Lemma 4. In this paper, although P_X can be interpreted as a* pmf *in the formulas, it can also be interpreted as a mass function as in the Lemma above. However, the escort probabilities are always* pmfs.

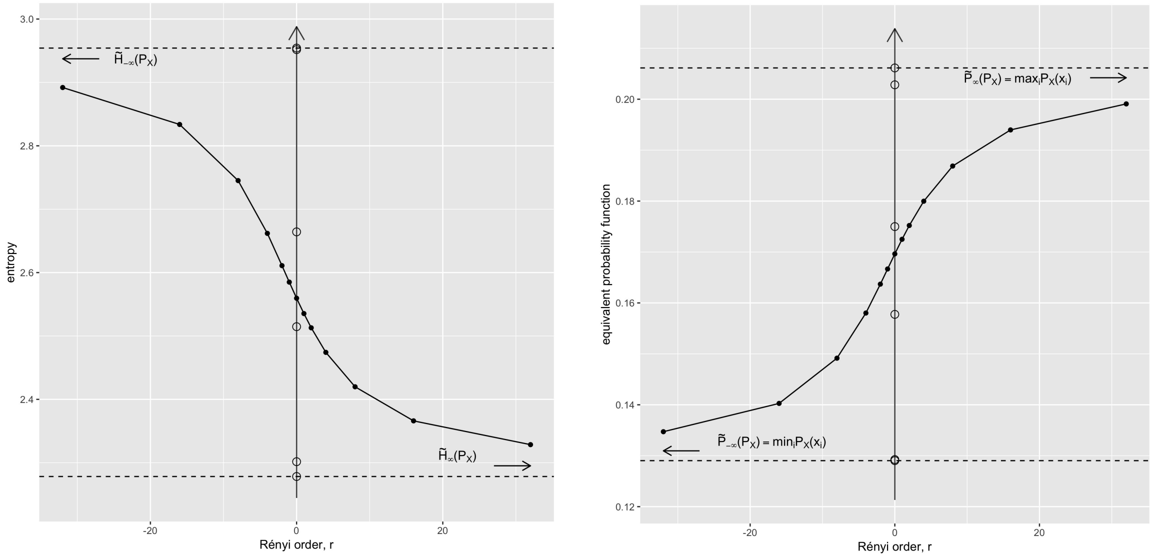
Example 1. *This example uses the UCB admission data from [28]. We analyze the distribution of admissions with count vector $M_X = [933\ 585\ 918\ 792\ 584\ 714]^\top$ and probabilities $\tilde{q}_0(M_X) \approx [0.21\ 0.13\ 0.20\ 0.17\ 0.13\ 0.16]^\top$. The names of the departments are not important, due to the symmetry property. Figure 1a shows the Rényi Spectrum extrapolated from a sample of some orders which include $r \in \{-\infty, -1, 0, 1, \infty\}$.*

3.1.2. Shifting Other Concepts Related to the Entropies

Other entropy-related concepts may also be shifted. In particular, the cross-entropy has an almost direct translation.

Definition 6. *The shifted Rényi cross-entropy of order $r \in [-\infty, \infty]$ between two distributions $P_X(x_i) = p_i$ and $Q_X(x_i) = q_i$ with compatible support is*

$$\tilde{X}_r\left(P_X \| Q_X\right) = \log \frac{1}{M_r(P_X, Q_X)} \tag{30}$$



(**a**) (**b**)

Figure 1. Rényi spectrum (**a**) and equivalent probability function (**b**)—also Hölder path—of $\tilde{q}_0(M_X)$, the probability distribution of a simple mass measure M_X with $n = 6$ (see Section 3.3). The values of the self-information (left) and probability (right) of the original distribution are shown at $r = 0$ also (hollow circles). Only 5 values seem to exist because the maximal information (minimal probability) is almost superposed on a second value.

Note that the case-based definition is redundant: the Shannon cross-entropy appears as $\tilde{X}_0\left(P_X \| Q_X\right) = -\log M_0(P_X, Q_X) = -\sum_i \frac{p_i}{\sum_k p_k} \log q_i$, while for $r \neq 0$ we have $\tilde{X}_r\left(P_X \| Q_X\right) = -\frac{1}{r} \log \sum_i p_i \left(\frac{p_i}{q_i}\right)^r$ by virtue of the definition of the means again.

Perhaps the most fundamental magnitude is the cross-entropy since it is easy to see that:

Lemma 5. *In the shifted formulation both the entropy and the divergence are functions of the cross-entropy:*

$$\tilde{H}_r\left(P_X\right) = \tilde{X}_r\left(P_X \| P_X\right) \qquad \tilde{D}_r\left(P_X \| Q_X\right) = \tilde{X}_{-r}\left(P_X \| Q_X / P_X\right) \tag{31}$$

Proof. The first equality is by comparison of definitions, while the second comes from:

$$\tilde{D}_r\left(P_X \| Q_X\right) = \frac{1}{r} \log \sum_i p_i \left(\frac{p_i}{q_i}\right)^r = -\frac{1}{-r} \log \sum_i p_i \left(\frac{q_i}{p_i}\right)^{-r} = \tilde{X}_{-r}\left(P_X \| Q_X / P_X\right)$$

□

Note that if we accept the standard criterion in Shannon's entropy $0 \times \log \frac{1}{0} = 0 \times \infty = 0$ then the previous expression for the cross-entropy is defined even if $p_i = 0$.

3.2. Writing Rényi Entropies in Terms of Each Other

Not every expression valid in the case of Shannon's entropies can be translated into Rényi entropies: recall from the properties of the Kullback–Leibler divergence its expression in terms of the Shannon entropy and cross-entropy. We have:

$$\tilde{D}_0\left(P_X \| Q_X\right) = -\tilde{H}_0\left(P_X\right) + \tilde{X}_0\left(P_X \| Q_X\right), \tag{32}$$

but, in general, $\tilde{D}_r\left(P_X \| Q_X\right) \neq -\tilde{H}_r\left(P_X\right) + \tilde{X}_r\left(P_X \| Q_X\right)$.

However, the shifting sometimes helps in obtaining "derived expressions". In particular, the (shifted) escort probabilities are ubiquitous in expressions dealing with Rényi entropies and divergences, and allow us to discover the deep relationships between their values for different r's.

Lemma 6. *Let $r \in \mathbb{R} \cup \{\pm\infty\}$, $P_X \in \Delta^{n-1}$ where Δ^{n-1} is the simplex over the support* supp X*. Then,*

$$\tilde{H}_r(P_X) = \frac{1}{r}\tilde{D}_0(\tilde{q}_r(P_X)\|P_X) + \tilde{X}_0(\tilde{q}_r(P_X)\|P_X) \tag{33}$$

$$\tilde{H}_r(P_X) = \frac{-1}{r}\tilde{H}_0(\tilde{q}_r(P_X)) + \frac{r+1}{r}\tilde{X}_0(\tilde{q}_r(P_X)\|P_X) \tag{34}$$

Proof. First, from the definitions of shifted Rényi entropy and cross-entropy and Property 3.6 we have:

$$\frac{-1}{r^2}\tilde{D}_0(\tilde{q}_r(P_X)\|P_X) = \frac{1}{r}\left[\tilde{H}_r(P_X) - \tilde{X}_0(\tilde{q}_r(P_X)\|P_X)\right]$$

Solving for $\tilde{H}_r(P_X)$ obtains the first result. By applying (32) to $\tilde{q}_r(P_X)$ and P_X we have:

$$\tilde{D}_0(\tilde{q}_r(P_X)\|P_X) = -\tilde{H}_0(\tilde{q}_r(P_X)) + \tilde{X}_0(\tilde{q}_r(P_X)\|P_X)\,. \tag{35}$$

and putting this into (33) obtains the second result.

Another way is to prove it is from the definition of

$$\tilde{H}_0(\tilde{q}_r(P_X)) = -\sum_i \frac{p_i p_i^r}{\sum_k p_k p_k^r}\log\frac{p_i p_i^r}{\sum_k p_k p_k^r} = \sum_i \tilde{q}_r(P_X)\log\left(\sum_k p_k p_k^r\right) - \sum_i \tilde{q}_r(P_X)\log p_i^{r+1}$$
$$= \log\left(\sum_k p_k p_k^r\right) - (r+1)\sum_i \tilde{q}_r(P_X)\log p_i = -r\tilde{H}_r(P_X) + (r+1)\tilde{X}_0(\tilde{q}_r(P_X)\|P_X)$$

and reorganize to obtain (34). Again inserting the definition of the Shannon divergence in terms of the cross-entropy (35), into (34) and reorganizing we get (33). □

On other occasions, using the shifted version does not help in simplifying expressions. For instance *skew symmetry* looks in the standard case as $D_\alpha(P_X\|Q_X) = \frac{\alpha}{1-\alpha}D_{1-\alpha}(Q_X\|P_X)$, for any $0 < \alpha < 1$ ([23], Proposition 2). In the shifted case we have the slightly more general expression for $r \neq 0$:

Lemma 7. *When Q_X is substituted by P_X, both probability distributions, on a compatible support, then:*

$$\tilde{D}_r(P_X\|Q_X) = -\frac{r+1}{r}\cdot\tilde{D}_{-(r+1)}(Q_X\|P_X) \tag{36}$$

Proof. By easy rewriting of the divergence $\tilde{D}_{-(r+1)}(Q_X\|P_X)$. □

3.3. Quantities Around the Shifted Rényi Entropy

On the one hand, the existence of Hartley's information function (9) ties up information values to probabilities and *vice-versa*. On the other, Rényi's averaging function and its inverse (14) also transform probabilities into information values and *vice-versa*. In this section we explore the relationship between certain quantities generated by these functions, probabilities and entropies.

3.3.1. The Equivalent Probability Function

Recall that, due to Hartley's function, from every average measure of information, an equivalent *average* probability emerges. To see this in a more general light, first define the extension to Hartley's information function to non-negative numbers $\mathfrak{I}_*(\cdot) : [0,\infty] \to [-\infty,\infty]$ as $\mathfrak{I}_*(p) = -\ln p$. This is one-to-one from $[0,\infty]$ and total onto $[-\infty,\infty]$, with inverse $(\mathfrak{I}_*)^{-1}(h) = e^{-h}$ for $h \in [-\infty,\infty]$.

Definition 7. *Let $X \sim P_X$ with Rényi spectrum $\tilde{H}_r(P_X)$. Then the* equivalent probability function of *$\tilde{P}_r(P_X)$ is the Hartley inverse of $\tilde{H}_r(P_X)$ over all values of $r \in [-\infty, \infty]$*

$$\tilde{P}_r(P_X) = (\mathfrak{I}_*)^{-1}(\tilde{H}_r(P_X)) \tag{37}$$

Remark 8. *The equivalent probability function for a fixed probability distribution P_X is a function of parameter r—like the Rényi entropy—whose values are probabilities—in the sense that it produces values in $[0,1]$—but it is* not *a probability distribution.*

Analogously, due to the extended definition of the Hartley information, this mechanism, when operating on a mass measure M_X, generates and equivalent mass function *$\tilde{P}_r(M_X)$, which is* not *a mass measure.*

Lemma 8. *Let $X \sim P_X$. The* equivalent probability function *$\tilde{P}_r(P_X)$ is the Hölder path of the probability function P_X (as a set of numbers) using the same probability function as weights.*

$$\tilde{P}_r(P_X) = M_r(P_X, P_X) \tag{38}$$

Proof. From the definition, using b as the basis chosen for the logarithm in the information function.

$$\tilde{P}_r(P_X) = (\mathfrak{I}_*)^{-1}(\tilde{H}_r(P_X)) = b^{-\tilde{H}_r(P_X)} = b^{\log_b M_r(P_X,P_X)} = M_r(P_X, P_X)$$

□

Note that by Remark 8 these means apply, in general, to sets of non-negative numbers and not only to the probabilities in a distribution, given their homogeneity properties. In the light of Lemma 8, the following properties of the equivalent probability function are a corollary of those of the weighted generalized power means of Proposition 1 in Section 2.1.

Corollary 1. *Let $X \sim P_X$ be a random variable with equivalent probability function $\tilde{P}_r(P_X)$. Then:*

1. *For all $r \in [-\infty, \infty]$, there holds that*

$$\min_k p_k = \tilde{P}_{-\infty}(P_X) \leq \tilde{P}_r(P_X) \leq \max_k p_k = \tilde{P}_{\infty}(P_X) \tag{39}$$

2. *If $P_X \equiv U_X$ the uniform over the same* supp *P_X, then $\forall k, \forall r \in [-\infty, \infty]$, $p_k = \tilde{P}_r(U_X) = \frac{1}{|\operatorname{supp} P_X|}$.*
3. *if $P_X \equiv \delta_X^k$ the Kroneker delta centered on $x_k = X(\omega_k)$, then $\tilde{P}_r\left(\delta_X^k\right) = u(r)$ where $u(r)$ is the step function.*

Proof. Claims 1 and 2 issue directly from the properties of the entropies and the inverse to the logarithm. The last claims follows from Remark 2. □

And so, in their turn, the properties of Rényi entropy can be proven from those of the equivalent probability function and Hartley's generalized information function.

An interesting property might help recovering P_X from the equivalent probability function:

Lemma 9. *Let $X \sim P_X$ be a random variable with equivalent probability function $\tilde{P}_r(P_X)$. Then: for every p_k in P_X there exists an $r_k \in [-\infty, \infty]$ such that $p_k = \tilde{P}_{r_k}(P_X)$.*

Proof. This follows from the continuity of the means with respect to its parameters $\vec{w}$ and $\vec{x}$. □

So if we could actually find those values $r_k, 1 \leq k \leq n$ which return $p_k = \tilde{P}_r(P_X)$ we would be able to retrieve P_X by sampling $\tilde{P}_{r_k}(P_X)$ in the appropriate values $P_X = \{\tilde{P}_{r_k}(P_X)\}_{k=1}^n$. Since $n \geq 2$ we know that at least two of these values are $r = \pm\infty$ retrieving the value of the highest and lowest probabilities for $k = 1$ and $k = n$ when they are sorted by increasing probability value.

$$\begin{tikzcd}
 & \tilde{P}_r\left(P_X\right) \arrow[ld, "\mathfrak{I}^*"] \arrow[dd, bend left, ".^{1/r}"] \\
\tilde{H}_r\left(P_X\right) \arrow[ru, bend left, "(\mathfrak{I}^*)^{-1}"] \arrow[rd, bend left, "\varphi'"] & \\
 & \tilde{V}_r\left(P_X\right) \arrow[lu, bend left, "\varphi'^{-1}"] \arrow[uu, bend left, ".^{r}"]
\end{tikzcd}$$

(**a**)

$$\begin{tikzcd}
 & \mathbb{R}_{\geq 0} \arrow[ld, "\mathfrak{I}^*"] \arrow[dd, bend left, ".^{1/r}"] \\
\mathbb{H} \arrow[ru, bend left, "(\mathfrak{I}^*)^{-1}"] \arrow[rd, bend left, "\varphi'"] & \\
 & \mathbb{R}^{r}_{\geq 0} \arrow[lu, bend left, "\varphi'^{-1}"] \arrow[uu, bend left, ".^{r}"]
\end{tikzcd}$$

(**b**)

Figure 2. Schematics of relationship between some magnitudes in the text and their domains of definition (see Section 3.4.5). (**a**) Between entropy-related quantities; (**b**) Between entropy-related domains.

Example 2 (Continued). *Figure 1b shows the equivalent probability function of the example in the previous section. The dual monotone behaviour with respect to that of the Rényi spectrum is clearly observable. We have also plotted over the axis at $r = 0$ the original probabilities of the distribution to set it in the context of the properties in Corollary 1 and Lemma 9.*

3.3.2. The Information Potential

In the context of Information Theoretic Learning (ITL) the information potential is an important quantity ([29], Chapter 2).

Definition 8. *Let $X \sim P_X$. Then the* information potential $\tilde{V}_r\left(P_X\right)$ *is*

$$\tilde{V}_r\left(P_X\right) = E_{P_X}\{P_X^r\} = \sum_i \frac{p_i}{\sum_k p_k} p_i^r \tag{40}$$

Note that the original definition of the information potential was presented in terms of parameter α and for distributions with $\sum_k p_k = 1$ in which case $V_\alpha(P_X) = \tilde{V}_r\left(P_X\right)$. Now, recall the conversion function in (14) $\varphi'(h) = b^{-rh}$. The next lemma is immediate using it on (26).

Lemma 10. *Let $X \sim P_X$. The information potential is the φ' image of the shifted Rényi entropy*

$$\tilde{V}_r\left(P_X\right) = \varphi'(\tilde{H}_r\left(P_X\right)) = b^{-r\tilde{H}_r(P_X)} \tag{41}$$

Proof. $\tilde{V}_r\left(P_X\right) = b^{-r\tilde{H}_r(P_X)} = b^{\log_b(\sum_i \frac{p_i}{\sum_k p_k} p_i^r)} = \sum_i \frac{p_i}{\sum_k p_k} p_i^r = E_{P_X}\{P_X^r\}$ □

Incidentally, (28) gives the relation of the information potential and the generalized weighted means.

Remark 9. *The quantity in the right-hand side of* (40) *is also the normalizing factor or* partition function *of the moments of the distribution and, as such, appears explicitly in the definition of the escort probabilities* (27). *Usually other partition functions appear in the estimation of densities based on overt, e.g., maximum entropy [6], or in covert information criteria—e.g., Ising models [5].*

3.3.3. Summary

Table 2 offers a summary of the quantities mentioned above and their relationships, while the domain diagram in Figure 2 summarizes the actions of these functions to obtain the shifted Rényi entropy. A similar diagram is, of course, available for the standard entropy, using φ with the α parameter.

Table 2. Quantities around the shifted Rényi entropy of a discrete distribution P_X.

Quantity in Terms of...	**Rényi Entropy**	**Gen. Hölder Mean**	**Information Potential**	**Distribution**
Rényi entropy	$\tilde{H}_r\,(P_X)$	$-\log M_r(P_X, P_X)$	$\frac{-1}{r}\log \tilde{V}_r\,(P_X)$	$\frac{-1}{r}\log\left(\sum_i \frac{p_i}{\sum_k p_k} p_i^r\right)$
Gen. Hölder mean	$\exp(-\tilde{H}_r\,(P_X))$	$M_r(P_X, P_X)$	$(\tilde{V}_r\,(P_X))^{\frac{1}{r}}$	$\left(\sum_i \frac{p_i}{\sum_k p_k} p_i^r\right)^{\frac{1}{r}}$
Information potential	$\exp(-r\tilde{H}_r\,(P_X))$	$M_r(P_X, P_X)^r$	$\tilde{V}_r\,(P_X) = E_{P_X}\{P_X^r\}$	$\sum_i \frac{p_i}{\sum_k p_k} p_i^r$

Note that these quantities have independent motivation: this is historically quite evident in the case of the means [24], and the Rényi information [12] and little bit less so in the case of the information potential which arose in the context of ITL [29], hence motivated by a desire to make Rényi's entropies more useful. Both quantities are generated from/generate entropy by means of independently motivated functions, Hartley's transformation (9) and Rényi's transformation (14), respectively.

Following the original axiomatic approach it would seem we first transform the probabilities into entropies using Hartley's function and then we use the φ' function to work out an average of these using the Kolmogorov–Nagumo formula. But due to the formulas for the information potential and the equivalent probability function we know that this is rather a composition of transformations, than a forward backward moving between entropies and probabilities. It is clear that the Hartley function and Rényi's choice of averaging function are special for entropies, from the postulate approach to their definition.

3.4. Discussion

A number of decisions taken in the paper might seem arbitrary. In the following, we try to discuss these issues as well as alternatives left for future work.

3.4.1. Other Reparameterization of the Rényi Entropy

Not only the parameter, but also de sign of the parameter is somewhat arbitrary in the form of (12). If we choose $r' = 1 - \alpha$ another generalization evolves that is, in a sense, symmetrical to the shifted Rényi entropy we have presented above, since $r' = -r$. This may be better or worse for the general formulas describing entropy, etc., but presents the problem that it no longer aligns with Shannon's original choice of sign. The $r = 0$ order Rényi entropy would actually be Boltzmann's, negative entropy or *negentropy* [30] and perhaps more suitable for applications in Thermodynamics [5].

Yet another formulation suggests the use of $\alpha = 1/2$, equivalently $r = -1/2$ as the origin of the parameter [31]. From our perspective, this suggests that the origin of the Rényi entropy can be chosen adequately in each application.

3.4.2. Rényi Measures and the Means

The usefulness of the (weighted) means in relation to information-theoretic concerns was already noted and explored in [32]. However, the relationship is not in there explicitly set out in terms of the identity of the Rényi entropies and logarithmic, weighted means of probabilities but rather as a part of establishing bounds for different quantities for discrete channel characterization.

A more direct approach is found in [33] that, inspired by [32], decides to generalize several results from there and other authors concerning the Rényi entropies, divergences and the Rényi centers of a set of distributions. Unlike our proposal, this deep work adheres to the standard definition of Rényi entropies of order α and avoids the issue of negative orders. The focus here is in coding and channel theorems, while ours is a re-definition of the mathematical concept to make similarities with weighted means transparent, yet evident.

3.4.3. Other Magnitudes around the Rényi Entropy

Sometimes the p-norm is used as a magnitude related to the Rényi entropy much as the information potential [29] or directly seeing the relationship with the definition [5].

Definition 9. *For a set of non-negative numbers $\vec{x} = [x_i]_{i=1}^n \in [0, \infty)^n$ the p-norm, with $0 \leq p \leq \infty$ is*

$$\|\vec{x}\|_p = \left(\sum_i x_i^p \right)^{\frac{1}{p}} \tag{42}$$

A more general definition involves both positive and negative components for $\vec{x}$, as in normed real spaces, but this is not relevant to our purposes for non-negative measures.

The p-norm has the evident problem that it is only defined for positive p whereas (14) proves that negative orders are meaningful and, indeed, interesting. A prior review of results for the negative orders can be found in [23].

We believe this is yet one more advantage of the shifting of the Rényi order: that the relation with the equivalent probability function and the information potential—the moments of the distribution—are properly highlighted.

3.4.4. Redundancy of the Rényi Entropy

Lemma 6 proves that Rényi entropies are very redundant in the sense that given its value for a particular r_0 the rest can be written in terms of those entropies with different, but systematically related, r order (see Section 3.4.4).

In particular, Equations (33) and (34) in Lemma 6, and (31) in Lemma 5 allow us to use a good estimator of Shannon's entropy to estimate the Rényi entropies and related magnitudes for all orders, special or not. Three interesting possibilities for this rewriting are:

- *That everything can be written in terms of $r = 0$, e.g., in terms of Shannon's entropy.* This is made possible by the existence of estimators for Shannon's entropy and divergence.
- *That everything can be written in terms of a finite $r \neq 0$, e.g., $r = 1$.* This is possible by means of Properties 1.3 and 1.4 of the generalized power means. The work in [29] is pointing this way (perhaps including also $r = -1$, aka Hartley's) capitalizing on the fact that Rényi's entropy for data is well estimated for $r = 1$, equivalently $\alpha = 2$ ([29], Section 2.6).
- *That everything can be written in terms of the extreme values of the entropy, e.g., $r = \pm\infty$.* This is suggested by Properties 3.1 and 3.2. Supposing we had a way to estimate either $\tilde{H}_{-\infty}(P_X)$ or $\tilde{H}_{\infty}(P_X)$. Then by a divide-and-conquer type of approach it would be feasible to extract all the probabilities of a distribution out of its Rényi entropy function.

3.4.5. The Algebra of Entropies

Technically, the completed non-negative reals $\mathbb{R}_{\geq 0}$, where the means are defined, carry a complete positive semifield structure [34]. This is an algebra similar to a real-valued field but the inverse operation to addition, e.g., subtraction, is missing.

There are some technicalities involving writing the results of the operations of the extremes of the semifields—e.g., multiplication of 0 and ∞—and this makes writing closed expressions for the means with extreme values of $\vec{w}$ or $\vec{x}$ complicated. A sample of this is the plethora of conditions on Property 1.5. An extended notation, pioneered by Moreau [35], is however capable of writing a closed expression for the means [36].

Furthermore, taking (minus) logarithms and raising to a real power are isomorphism of semifields, so that the Rényi entropies inhabit a different positive semifield structure [36]. The graph of these isomorphic structures can be seen in Figure 2b. This means that some of the intuitions about operating with entropies are misguided. We believe that failing to give a meaning to the Rényi entropies with negative orders might have been caused by this.

3.4.6. Shifted Rényi Entropies on Continuous Distributions

The treatment we use here may be repeated on continuous measures, but the definitions of Shannon [10,21] and Rényi [26] entropies in such case run into technical difficulties solved, typically, by a process of discretization [27].

Actually we believe that the shifting would also help in this process: a form for the generalized weighted continuous means was long ago established [20] and technically solved by a change of concept and Lebesgue–Stieltjes integration instead of summation ([24], Ch. VI).

Our preliminary analyses show that the relationship with the means given by (17) also holds, and this would mean that the shifting—in aligning the Rényi entropies with the (generalized weighted) continuous means—leverages the theoretical support of the latter to sustain the former.

Definition 10 (Continuous weighted f-mean)**.** *Let $\Phi(\xi)$ be a measure and let f be a monotonic function of ξ with inverse f^{-1}. Then a continuous version of* (A2) *is:*

$$M_f(\Phi, \xi) = f^{-1}\left\{\int f(\xi) d\Phi(\xi)\right\}$$

understood as a Lebesgue–Stieltjes integral.

This definition was already proposed by De Finetti [20] based upon the works of Bonferroni and Kolmogorov and thoroughly developed in ([24], Ch. VI) in connection to the discrete means. With $f(x) = x^r$ the continuous Hölder means $M_r(\Phi, \xi)$ appear. Furthermore De Finetti found ([20], #8) that the form of the f continuous, monotone function f must be

$$f(x) = a\int \gamma(x)dx + b \qquad\qquad \text{for arbitrary } a, b (a \neq 0)$$

similar to what Rényi found later for the Shannon entropy.

It is easy to see that an analogue definition of the shifted Rényi entropy but for a continuous probability density p_X with $dp_X(x) = p_X(x)dx$ [5,27] is

$$\tilde{h}(p_X) = \frac{-1}{r}\log \int p_X(x) p_X^r(x) dx = -\log M_r(p_X, p_X) \tag{43}$$

again with the distribution acting as weight and averaged quantity. Compare this to one of the standard forms of the *differential Rényi entropy* [23]:

$$h(p_X) = \frac{1}{1-\alpha}\ln \int p_X^\alpha(x)dx$$

The investigation of the properties of (43) is left pending for future work, though.

3.4.7. Pervasiveness of Rényi Entropies

Apart from the evident applications to signal processing and communications [29], physics [5] and cognition [11], the Rényi entropy is a measure of diversity in several disciplines [37]. We believe that, if its applicability comes from the same properties stemming from the means that we have explored in this paper as applied to positive distributions—e.g., of wealth in a population, or energy in a community—, then the expression to be used is (29).

4. Conclusions

In this paper we have advocated for the shifting of the traditional Rényi entropy order from a parameter α to $r = \alpha - 1$. The shifting of the Rényi entropy and divergence is motivated by a number of results:

- It aligns them with the power means and explains the apparition of the escort probabilities. Note that the importance of the escort probabilities is justified independently of their link to the means in the shifted version of entropy [5].
- It highlights the Shannon entropy $r = 0$ in the role of the "origin" of entropy orders, just as the geometric means is a particular case of the weighted averaged means. This consideration is enhanced by the existence of a formula allowing us to rewrite every other order as a combination of Shannon entropies and cross entropies of escort probabilities of the distribution.
- The shifting of the Rényi entropy aligns it with the moments of the distribution, thus enabling new insights into the moments' problem.
- It makes the relation between the divergence and the entropy more "symmetrical".
- It highlights the "information spectrum" quality of the Rényi entropy measure for fixed P_X.

The shifting might or might not be justified by applications. If the concept of the means is relevant in the application, we recommend the shifted formulation.

Author Contributions: Conceptualization, F.J.V.-A. and C.P.-M.; Formal analysis, F.J.V.-A. and C.P.-M.; Funding acquisition, C.P.-M.; Investigation, F.J.V.-A. and C.P.-M.; Methodology, F.J.V.-A. and C.P.-M.; Software, F.J.V.-A.; Supervision, C.P.-M.; Validation, F.J.V.-A. and C.P.-M.; Visualization, F.J.V.-A. and C.P.-M.; Writing—original draft, F.J.V.-A. and C.P.-M.; Writing—review & editing, F.J.V.-A. and C.P.-M.

Funding: This research was funded by the Spanish Government-MinECo projects TEC2014-53390-P and TEC2017-84395-P.

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

Appendix A. The Kolmogorov-Mean and the Kolmogorov–Nagumo Formula

The following is well known since [18–20,24].

Definition A1. *Given an invertible real function $f : \mathbb{R} \to \mathbb{R}$ the Kolmogorov–Nagumo mean of a set of non-negative numbers $\vec{x} = [x_i]_{i=1}^n \in [0, \infty)^n$ is*

$$KN_f(\vec{x}) = f^{-1}(\sum_{i=1}^{n} \frac{1}{n} f(x_i)). \tag{A1}$$

Definition A1 is an instance of the following formula to work out the *weighted f-mean* with a set of finite, non-negative weights, $\vec{w} \in [0, \infty)$

$$M_f(\vec{w}, \vec{x}) = f^{-1}(\sum_{i=1}^{n} \frac{w_i}{\sum_k w_k} f(x_i)). \tag{A2}$$

Our interest in (A2) lies in the fact that Shannon's and Rényi's entropies can be seen as special cases of it, which makes its properties especially interesting.

Proposition A1 (Properties of the Kolmogorov–Nagumo means). *Let $\vec{x}, \vec{w} \in [0, \infty)^n$. The following conditions hold if and only if there is a strictly monotonic and continuous function f such that* (A1) *holds.*

1. *Continuity and strict monotonicity in all coordinates.*
2. (Symmetry or permutation invariance) *Let σ be a permutation, then $M_f(\vec{w}, \vec{x}) = M_f(\sigma(\vec{w}), \sigma(\vec{x}))$.*

3. (Reflexivity) *The mean of a series of constants is the constant itself:*

$$M_f(\vec{w}, \{k\}_{i=1}^n) = k$$

4. (Blocking) *The computation of the mean can be split into computations of equal size sub-blocks.*
5. (Associativity) *Replacing a k-subset of the x with their partial mean in the same multiplicity does not change the overall mean.*

For a minimal axiomatization, Blocking and Associativity are redundant. A review of the axiomatization of these and other properties can be found in [22].

Appendix B. The Approach to Shannon's Information Functions Based in Postulates

It is important to recall that Shannon set out to define the *amount of information*, discarding any *notion of information* itself. Both concepts should be distinguished clearly for methodological reasons, but can be ignored for applications that deal only with quantifying information.

Recall the Faddeev postulates for the generalization of Shannon's entropy ([26], Chap. IX. §2):

1. The *amount of information* $H(P)$ of a sequence $P = [p_k]_{k=1}^n$ of n numbers is a symmetric function of this set of values $H(P) = H(\sigma(P)) = H(\{p_k\}_{k=1}^n)$, where σ is any permutation of n-elements.
2. $H(\{p, 1-p\})$ is a continuous function of $p, 0 \leq p \leq 1$.
3. $H(\{\frac{1}{2}, \frac{1}{2}\}) = 1$.
4. The following relation holds:

$$H(\{p_1, p_2, \ldots, p_n\}) = H(\{p_1 + p_2, \ldots, p_n\}) + (p_1 + p_2) H(\{\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\}) \tag{A3}$$

These postulates lead to Shannon's entropy for $X \sim P_X$ with binary logarithm [26]

$$H(P_X) = E_{P_X}\{-\log P_X\} = -\sum_k p_k \log p_k \tag{A4}$$

References

1. Shannon, C.; Weaver, W. *A mathematical Model of Communication*; The University of Illinois Press: Champaign, IL, USA, 1949.
2. Shannon, C.E.A mathematical theory of communication. Parts I and II. *Bell Syst. Tech. J.* **1948**, *XXVII*, 379–423. [CrossRef]
3. Shannon, C.E. A mathematical theory of communication. Part III. *Bell Syst. Tech. J.* **1948**, *XXVII*, 623–656. [CrossRef]
4. Shannon, C. The bandwagon. *IRE Trans. Inf. Theory* **1956**, *2*, 3. [CrossRef]
5. Beck, C.; Schögl, F. *Thermodynamics of Chaotic Systems: An Introduction*; Cambridge University Press: Cambridge, UK, 1995.
6. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 1996.
7. Mayoral, M.M. Rényi's entropy as an index of diversity in simple-stage cluster sampling. *Inf. Sci.* **1998**, *105*, 101–114. [CrossRef]
8. MacKay, D.J.C. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
9. Csiszár, I.; Shields, P.C. Information Theory and Statistics: A Tutorial. *Found. Trends Commun. Inf. Theory* **2004**, *1*, 417–528. [CrossRef]
10. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006.
11. Sayood, K. Information Theory and Cognition: A Review. *Entropy* **2018**, *20*, 706. [CrossRef]

12. Rényi, A. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20–30 July 1960; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
13. Havrda, J.; Charvát, F. Quantification method of classification processes. Concept of structural a-entropy. *Kybernetika* **1967**, *3*, 30–35.
14. Csiszár, I. Axiomatic Characterizations of Information Measures. *Entropy* **2008**, *10*, 261–273. [CrossRef]
15. Arndt, C. *Information Measures*, 1st ed.; Information and Its Description in Science and Engineering; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2004.
16. Rényi, A. On the Foundations of Information Theory. *Revue de l'Institut International de Statistique/Rev. Int. Stat. Inst.* **1965**, *33*, 1–14. [CrossRef]
17. Aczél, J.; Daróczy, Z. *On measures of inFormation and Their Characterizations*; Academic Press [Harcourt Brace Jovanovich, Publishers]: New York, NY, USA; London, UK, 1975.
18. Kolmogorov, A.N. Sur la notion de la moyenne. *Atti Della Accademia Nazionale dei Lincei* **1930**, *12*, 388–391.
19. Nagumo, M. Uber eine Klasse der Mittelwerte. *Jpn. J. Math. Trans. Abstr.* **1930**, *7*, 71–79. [CrossRef]
20. De Finetti, B. Sul concetto di media. *Giornale dell Istituto Italiano degli Attuari* **1931**, *II*, 369–396.
21. Kolmogorov, A. On the Shannon theory of information transmission in the case of continuous signals. *IRE Trans. Inf. Theory* **1956**, *2*, 102–108. [CrossRef]
22. Muliere, P.; Parmigiani, G. Utility and means in the 1930s. *Stat. Sci.* **1993**, *8*, 421–432. [CrossRef]
23. Van Erven, T.; Harremoës, P. Rényi divergence and Kullback-Leibler divergence. *IEEEE Trans. Inf. Theory* **2014**. [CrossRef]
24. Hardy, G.H.; Littlewood, J.E.; Pólya, G. *Inequalities*; Cambridge University Press: Cambridge, UK, 1952.
25. Kitagawa, T. On Some Class of Weighted Means. *Proc. Phys.-Math. Soc. Jpn. 3rd Ser.* **1934**, *16*, 117–126, doi:10.11429/ppmsj1919.16.0_117. [CrossRef]
26. Renyi, A. *Probability Theory*; Courier Dover Publications: Mineola, NY, USA, 1970.
27. Jizba, P.; Arimitsu, T. The world according to Rényi: thermodynamics of multifractal systems. *Ann. Phys.* **2004**, *312*, 17–59. [CrossRef]
28. Bickel, P.J.; Hammel, E.A.; O'Connell, J.W. Sex bias in graduate admissions: Data from Berkeley. *Science* **1975**, *187*, 398–403. [CrossRef]
29. Principe, J.C. *Information Theoretic Learning*; Information Science and Statistics; Springer: New York, NY, USA, 2010.
30. Brillouin, L. *Science and Information Theory*, 2nd ed.; Academic Press, Inc.: New York, NY, USA, 1962.
31. Harremoës, P. Interpretations of Rényi entropies and divergences. *Phys. A Stat. Mech. Its Appl.* **2005**, *365*, 57–62. [CrossRef]
32. Augustin, U. Noisy Channels. Ph.D. Thesis, Universität Erlangen, Erlangen, Germany, 1978.
33. Nakiboglu, B. The Rényi capacity and center. *IEEE Trans. Inf. Theory* **2018**. [CrossRef]
34. Gondran, M.; Minoux, M. *Graphs, Dioids and Semirings. New Models and Algorithms*; Operations Research Computer Science Interfaces Series; Springer: New York, NY, USA 2008.
35. Moreau, J.J. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Math. Pures Appl.* **1970**, *49*, 109–154
36. Valverde Albacete, F.J.; Peláez-Moreno, C. Entropy operates in Non-Linear Semifields. *arXiv* **2017**, arXiv:1710.04728.
37. Zhang, Z.; Grabchak, M. Entropic representation and estimation of diversity indices. *J. Nonparametr. Stat.* **2016**, *28*, 563–575. [CrossRef]