Article

On the α-q-mutual information and the α-q-capacities

Velimir Ilić [1,*] and Ivan B. Djordjević [2]

[1] Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36, 11000 Beograd, Serbia; velimir.ilic@gmail.com
[2] Department of Electrical and Computer Engineering, University of Arizona, 1230 E. Speedway Blvd. Tucson, AZ 85721, USA; ivan@ece.arizona.edu
* Correspondence: velimir.ilic@gmail.com

Version May 18, 2021 submitted to Entropy

Abstract: The measures of information transfer which correspond to non-additive entropies, have intensively been studied in previous decades. The majority of work includes the ones which belongs to Sharma-Mittal entropy class, such as Renyi, Tsallis and Landsberg-Vedral entropies. All of the considerations follows the same approach mimicking some of various and mutually equivalent definitions of Shannon information measures, and the information transfer is quantified by an appropriately defined measure of mutual information, while the maximal information transfer is considered as a generalized channel capacity. However, all of the previous approaches fail down to satisfy at least one of ineluctable properties that a measure of (maximal) information transfer should satisfy, leading to counterintuitive conclusions and predicting nonphysical behavior even in the case of very simple communication channels. This paper fills the gap, by proposing new measures named the α-q mutual information and the α-q channel capacity. Beside standard Shannon approaches, a special cases of these measures include the α-mutual information and α-capacity, which are well established in the information theory literature as measures of additive Rényi information transfer, while the cases of Tsallis, Gaussian and Landsberg-Vedral entropies can also be accessed by special choices of the parameter q. It is shown that, unlike the previous definition, the α-q mutual information and the α-q capacity satisfy the set of ineluctable axioms, by which they are non-negativity, less than or equal to the input and the output non-additive entropies, they reduce to zero in the case of totally destructive channels and to the (maximal) input non-additive entropy in the case of perfect transmission. Thus, unlike the previous approaches, the proposed (maximal) information transfer measures do not manifest non-physical behaviors such as sub-capacitance or supper-capacitance, which could qualify them as appropriate measures of information transfer.

Keywords: Rényi Entropy, Tsallis entropy, Landsberg-Vedral entropy, Sharma-Mittal entropy, α-mutual Information, α-channel Capacity

1. Introduction

In the past, there was extensive work on defining the information measures which generalize the Shannon entropy [1], such as one parameter Renyi entropy [2], Tsallis entropy [3], Landsberg-Vedral entropy [4], Gaussian entropy [5], and two parameter Sharma-Mittal entropy [5,6], which reduces to former ones for special choices of the parameters. Sharma-Mittal entropy can axiomatically be founded as the unique q-additive measure [7,8] which satisfies generalized Shannon-Kihinchin axioms [9,10], and has widely been explored in different research fields starting from statistics [11], to thermodynamics [12], [13], to quantum mechanics [14], [15] to machine learning [16], [17] and to cosmology [18], [19]. Sharma-Mittal entropy has also be recognized in the field of information theory where the measures of conditional Sharma-Mittal entropy [20], Sharma-Mittal divergences [21] and Sharma-Mittal entropy rate [22] has been established and analyzed.

A considerable research has also been done in the field of communication theory in order to analyze information transmission in the presence of noise, if instead of Shannon's entropy the information is quantified with (an instances of) Sharma-Mittal entropy and, in general, the information transfer is quantified by an appropriately defined measure of mutual information, while the maximal information transfer is considered as a generalized channel capacity. Thus, after Rényi's proposal for the additive generalization of Shannon entropy [2], several different definitions for Rényi information transfer were proposed by Sibson [23], Arimoto [24], Augustin [25], Csiszar [26], Lapidoth and Pfister [27] and Tomamichel and Hayashi [28]. These measures has thoroughly been explored and their operational characterization in coding theory, hypothesis testing, cryptography and quantum information theory were established, which qualifies them as a reasonable measure of Rényi information transfer [29]. Similar attempts has also be performed in the case of non-additive entropies. Thus, starting from the work of Daroczy [30] who introduced a measure for generalized information transfer related to Tsallis entropy, several attempts followed for the measures which corresponds to non-additive particular instances of Sharma-Mittal entropy, so that the definitions for Rényi information transfer were considered in [24], [31], Tsallis information transfer were considered in [32] and Landsber-Vedral information transfer in [4] [33].

In this paper we provide a general treatment of Sharma-Mittal entropy transfer and we provide a detailed analyses of existing measures, showing that all of the definitions related to non-additive entropies fail down to satisfy at least one of ineluctable properties common to Shannon case, by which the information transfer has to be non-negative, less than input and output uncertainty, equal to the input uncertainty in the case of perfect transmission and equal to zero, in the case of totally destructive channel. Thus, breaking some of these properties implies unexpected and counterintuitive conclusions about the channels, such as achieving super-capacitance or sub-capacitance [4], which could be treated as a nonphysical behavior. As an alternative, we propose the α-q mutual information as a measure of Sharma-Mittal information transfer, being maximized with the α-q channel capacity. The α-q mutual information generalizes the α-mutual information by Arimoto [24], which is defined as a q-difference between the input Sharma-Mittal entropy and appropriately defined conditional Sharma-Mittal entropy if the output is given, while the α-q-capacity represents a generalization of Arimoto's α-capacity in the case of $q = 1$. In addition, several other instances can be obtained by specifying the values of the parameters α and q, which includes the information transfer measures for Tsallis, Landsber-Vedral and Shannon entropy, as well as the case of Gaussian entropy which was not considered before in the context of information transmission.

The paper is organized as follows. In Section 4 we review basic properties of the α-capacity. The basic properties and special instances of Sharma-Mittal entropy are listed in Section 2. The α-q mutual information and the α-q channel capacity are defined and analyzed in Section 6. In Section **??** we review the previous definitions, discuss their unphysical behavior and show that the α-q information transfer measure avoid it.

2. Sharma-Mittal entropy, conditional entropy and divergence

Let the sets of positive and nonnegative real numbers be denoted with $\mathbb{R}^+$ and $\mathbb{R}_0^+$, respectively, and let the mapping $\eta_q : \mathbb{R} \to \mathbb{R}$ be defined in

$$\eta_q(x) = \begin{cases} x, & \text{for} \quad q = 1 \\ \dfrac{2^{(1-q)\,x} - 1}{(1-q)\ln 2}, & \text{for} \quad q \neq 1 \end{cases} \tag{1}$$

so that its inverse is given in

$$\eta_q^{-1}(x) = \begin{cases} x, & \text{for} \quad q = 1 \\ \dfrac{1}{1-q}\log((1-q)x\ln 2 + 1), & \text{for} \quad q \neq 1 \end{cases}. \tag{2}$$

The mapping h_q and its inverse are increasing continuous (hence invertible) functions such that $\eta(0) = 0$. The q-logarithm and is defined in

$$\mathrm{Log}_q(x) = \eta_q(\log x) = \begin{cases} \log x, & \text{for} \quad q = 1 \\ \dfrac{x^{(1-q)} - 1}{(1-q)\ln 2}, & \text{for} \quad q \neq 1 \end{cases}, \tag{3}$$

and its inverse, the q-exponential, is defined in

$$\mathrm{Exp}_q(y) = \begin{cases} 2^y, & \text{for} \quad q = 1 \\ (1 + (1-q)y\ln 2)^{\frac{1}{1-q}} & \text{for} \quad q \neq 1 \end{cases}, \tag{4}$$

for $1 + (1-q)y\ln 2 > 0$. Using η_q, we can define the pseudo-addition operation $\oplus_q$ [7,8]

$$x \oplus_q y = \eta_q\left(\eta_q^{-1}(x) + \eta_q^{-1}(y)\right) = x + y + (1-q)xy; \quad x, y \in \mathbb{R}, \tag{5}$$

and its inverse operation, the pseudo substraction

$$x \ominus_q y = \eta_q\left(\eta_q^{-1}(x) - \eta_q^{-1}(y)\right) = \frac{x - y}{1 + (1-q)y\ln 2}; \quad x, y \in \mathbb{R} \tag{6}$$

The $\oplus_q$ can be rewritten in terms of the generalized logarithm by settings $x = \log u$ and $y = \log v$ so that

$$\mathrm{Log}_q(u \cdot v) = \mathrm{Log}_q(u) \oplus_q \mathrm{Log}_q(v); \quad u, v \in \mathbb{R}_+. \tag{7}$$

2.1. Sharma-Mittal entropy

Let the set of all n-dimensional distributions be denoted with

$$\Delta_n \equiv \left\{ (p_1, \ldots, p_n) \middle| \; p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}; \quad n > 1. \tag{8}$$

Let the function $H_n : \Delta_n \to \mathbb{R}_0^+$ satisfy the following Shannon-Khinchin axioms, for all $n \in \mathbb{N}$, $n > 1$.

[GSK1] H_n is continuous in Δ_n;

[GSK2] H_n takes its largest value for the uniform distribution, $U_n = (1/n, \ldots, 1/n) \in \Delta_n$, i.e. $H_n(P) \leq H_n(U_n)$, for any $P \in \Delta_n$;

[GSK3] H_n is expandable: $H_{n+1}(p_1, p_2, \ldots, p_n, 0) = H_n(p_1, p_2, \ldots, p_n)$ for all $(p_1, \ldots, p_n) \in \Delta_n$;

[GSK4] Let $P = (p_1, \ldots, p_n) \in \Delta_n$, $PQ = (r_{11}, r_{12}, \ldots, r_{nm}) \in \Delta_{nm}$, $n, m \in \mathbb{N}$, $n, m > 1$ such that $p_i = \sum_{j=1}^{m} r_{ij}$, and $Q_{|k} = (q_{1|k}, \ldots, q_{m|k}) \in \Delta_m$, where $q_{i|k} = r_{ik}/p_k$ and $\alpha \in \mathbb{R}_0^+$ is some fixed parameter. Then,

$$H_{nm}(PQ) = H_n(P) \oplus_q H_m(Q|P), \quad \text{where} \quad H_m(Q|P) = f^{-1}\left(\sum_{k=1}^{n} p_k^{(\alpha)} f(H_m(Q_{|k}))\right), \tag{9}$$

where f is an invertible continuous function and $P^{(\alpha)} = (p_1^{(\alpha)}, \ldots, p_n^{(\alpha)}) \in \Delta_n$ is the α-escort distribution of distribution $P \in \Delta_n$ defined in (38).

[GSK5] $H_2\left(\frac{1}{2},\frac{1}{2}\right) = \mathrm{Log}_q(1)$.

As it is shown in [9], the unique function H_n which satisfies [GSK1]-[GSK5] is Sharma-Mittal entropy [6]. In the following paragraphs, we will identify the entropy of a random variable X with the entropy of its distribution P_X and using the notation $H_{\alpha,q}(X) \equiv H_n(P_X)$, Sharma-Mittal entropy.

In the following paragraphs we assume that X and Y are discrete jointly distributed random variables taking values from a sample spaces $\{x_1,\dots,x_n\}$ and $\{y_1,\dots,y_m\}$, and distributed in accordance to $P_X \in \Delta_n$ and $P_Y \in \Delta_m$, respectively. In addition, the joint distribution of X and Y be denoted in $P_{X,Y} \in \Delta_{nm}$ and the conditional distribution of X given Y will be denoted in $P_{X|Y} = \frac{P_{X,Y}(x,y)}{P_Y(y)} \in \Delta_m$, provided that $P_Y(y) > 0$.

Thus, for a random variable which is distributed to X, Sharma-Mittal entropy can be expressed in

$$H_{\alpha,q}(X) = \frac{1}{1-q}\left(\left(\sum_x P_X(x)^\alpha\right)^{\frac{1-q}{1-\alpha}} - 1\right), \tag{10}$$

and it can equivalently be expressed as the η_q transformation of Rényi entropy as in

$$H_{\alpha,q}(X) \equiv \eta_q\left(R_\alpha(X)\right). \tag{11}$$

Sharma-Mittal entropy, for $\alpha, q \in \mathbb{R}_0^+ \setminus 1$, being continuous function of the parameters and the sums goes over the support of P_X. Thus, in the case of $q=1$, $\alpha \neq 1$, Sharma-Mittal reduces to Rényi entropy of order α [2]

$$R_\alpha(X) \equiv H_{\alpha,1}(X) = \frac{1}{1-\alpha}\log\left(\sum_x P_X(x)^\alpha\right), \tag{12}$$

which further reduces to Shannon entropy for $\alpha = 1, q = 1$, [34]

$$S(X) \equiv H_{1,1}(X) = \sum_x P_X(x)\log P_X(x), \tag{13}$$

while in the case of $q \neq 1$, $\alpha = 1$ it reduces to Gaussian entropy [5]

$$G_q(X) \equiv H_{1,q}(X) = \frac{1}{(1-q)\ln 2}\left(\prod_{i=1}^{n} P_X(x)^{P_X(x)} - 1\right) \tag{14}$$

In addition, Tsallis entropy [3] is obtained for $\alpha = q \neq 1$,

$$T_q(X) \equiv \frac{1}{(1-q)\ln 2}\left(\sum_x P_X(x)^q - 1\right) \tag{15}$$

while in the case of for $q = 2-\alpha$ it reduces to the Landsberg-Vedral entropy [4]

$$L_\alpha(X) \equiv H_{\alpha,2-\alpha}(X) = \frac{1}{(\alpha-1)\ln 2}\left(\frac{1}{\sum_x P_X(x)^\alpha} - 1\right). \tag{16}$$

3. Sharma-Mittal information transfer axioms

One of the main goals of information and communication theories is characterization and analysis of the information transfer between sender X and receiver Y, which communicate through a channel. The sender and receiver are described by probability distributions P_X and P_Y while the communication channel with the input X and the output Y is described by transition matrix $P_{Y|X}$:

$$P_{Y|X}^{(i,j)} = P_{Y|X}(y_j|x_i). \tag{17}$$

We assume that maximum likelihood detection is performed at the receiver, which is defined by the mapping $d : \{y_1, \dots, y_m\} \to \{x_1, \dots, x_n\}$ as follows:

$$d(y_j) = x_i \quad \Leftrightarrow \quad P_{Y|X}(y_j|x_i) > P_{Y|X}(y_j|x_k); \quad \text{for all } k \neq i, \tag{18}$$

assuming that the inequality in (18) is uniquely satisfied. Thus, if the input symbol x_i is sent and the output symbol y_j is received, the x_i will be detected if $x_i = d(y_j)$ and detection error will be made otherwise, and we define the error function functions $\phi : \{x_1, \dots, x_m\} \times \{y_1, \dots, y_m\} \to \{0, 1\}$ as in

$$\phi(x_i, y_j) = \begin{cases} 1, & \text{if } x_i = d(y_j) \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

as well as the detection error if a symbol x_i is sent

$$P_{err}(x_i) = \sum_{y_j} P_{Y|X}(y_j|x_i)\phi(x_i, y_j); \quad \text{for all} \quad x_i. \tag{20}$$

Totally destructive channel: A channel is said to be totally destructive if

$$P^{(i,j)}_{Y|X} = P_{Y|X}(y_j|x_i) = P_Y(y_j) = \frac{1}{m}; \quad \text{for all} \quad x_i, \tag{21}$$

i.e. if the sender X and receiver Y are described by independent random variables,

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad P_{X,Y}(x, y) = P_X(x)P_Y(y), \tag{22}$$

where the relationship of independence is denoted in $\perp\!\!\!\perp$. In this case, $\phi_i(y_j) = 1$ for all y_j and the probability of error is $P_{err}(x_i) = 1$; for all x_i, which means that a correct maximum likelihood detection is not possible.

Perfect communication channel: A channel is said to be perfect if for every x_i,

$$P^{(i,j)}_{Y|X} = P_{Y|X}(y_j|x_i) > 0, \quad \text{for at least one } y_j \tag{23}$$

and for every y_j

$$P^{(i,j)}_{Y|X} = P_{Y|X}(y_j|x_i) > 0, \quad \text{for exactly one } x_i. \tag{24}$$

Note that in this case $P_{Y|X}(y_j|x_i)$ can still take a zero value for some y_j and that $\phi_i(y_j) = 0$ for any non-zero $P_{Y|X}(y_j|x_i)$. Thus, the error probability is equal to zero $P_{err}(x_i) = 0$; for all x_i, which means that perfect detection is possible by means of a maximum likelihood detector.

Noisy channel with non-overlapping outputs: A simple example of a perfect transmission channel is the noisy channel with non-overlapping outputs, which is schematically described in Figure 1. It is 2-input $m = 2k$-output channel ($k \in \mathbb{N}$) defined by the transition matrix:

$$P_{Y|X} = \begin{bmatrix} P_{Y|X=x_1} \\ P_{Y|X=x_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{k} & \dots & \frac{1}{k} & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1}{k} & \dots & \frac{1}{k} \end{bmatrix} \tag{25}$$

(in this and in the following matrices, the symbol "$\cdots$" stands for the k-time repletion). In the case of $k = 1$ and $m = 2k = 2$, the channel reduces to the noiseless channel. Although the channel is noisy, the input can always be recovered from the output (if y_j is received and $j \leq k$, the input symbol x_1 is sent, otherwise x_2 is sent). Thus, it is expected that the information which is passed through the channel is

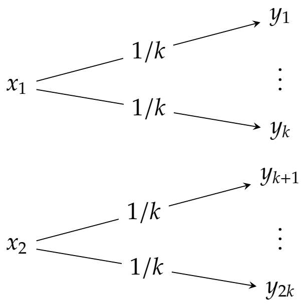

Figure 1. Noisy channel with non-overlapping outputs

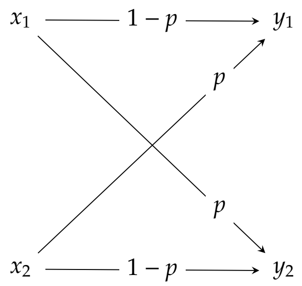

Figure 2. Binary symmetric channel

equal to the information that can be generated by the input. Note that for a channel input distributed in accordance with

$$P_X = \begin{bmatrix} P_{X=x_1} \\ P_{X=x_2} \end{bmatrix} = \begin{bmatrix} a \\ 1-a \end{bmatrix}; \quad 0 \le a \le 1, \tag{26}$$

the joint probability distribution $P_{X,Y}$ can be expressed as in:

$$P_{X,Y} = \begin{bmatrix} \frac{a}{k} & \dots & \frac{a}{k} & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1-a}{k} & \dots & \frac{1-a}{k} \end{bmatrix} \tag{27}$$

and the output distribution P_Y, which can be obtained by the summations over columns, is

$$P_Y = \left[P_{Y=y_1}, \dots, P_{Y=y_m}\right]^T = \left[\frac{a}{k}, \dots, \frac{a}{k}, \frac{1-a}{k}, \dots, \frac{1-a}{k}\right]^T. \tag{28}$$

Binary symmetric channels: The binary symmetric channel (BSC) is a two input two output channel described by the transition matrix

$$P_{Y|X} = \begin{bmatrix} P_{Y|X=x_1} \\ P_{Y|X=x_2} \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}, \tag{29}$$

which is schematically described in Figure 2. Note that for $p = \frac{1}{2}$ BSC reduces to a totally destructive channel, while in the case of $p = 0$ it reduces to a perfect channel.

3.1. Sharma-Mittal information transfer measures axioms

In this paper, we search for information theoretical measures of information transfer between sender X and receiver Y, which communicate through a channel if the information is measured with Sharma-Mittal entropy. Thus, we are interested in the information transfer measure, $I_{\alpha,q}(X,Y)$, which is called the α-q-mutual information and its maximum,

$$C = \max_{P_X} I_{\alpha,q}(Y,X), \tag{30}$$

which is called the α-q-capacity and which requires the following set of axioms to be satisfied.

$(\overline{A}_1)$ The channel cannot convey negative information, i.e.

$$C_{\alpha,q}(X,Y) \ge I_{\alpha,q}(X,Y) \ge 0. \tag{31}$$

($\overline{A}_2$) The information transfer is zero in the case of a totally destructive channel, i.e.

$$P_{Y|X}(y|x) = \frac{1}{m}, \text{ for all } x, y \quad \Rightarrow \quad I_{\alpha,q}(X,Y) = C_{\alpha,q}(X,Y) = 0, \tag{32}$$

which is compatible with the conclusion that the probability of error is one, $P_{err}(x_i) = 1$; for all x_i, in the case of a totally destructive channel.

($\overline{A}_3$) In the case of perfect transmission, the information transfer is equal to the input information, i.e.

$$X = Y \quad \Rightarrow \quad I_{\alpha,q}(X,Y) = H_{\alpha,q}(X), \quad C_{\alpha,q}(X,Y) = \mathrm{Log}_q n, \tag{33}$$

which is compatible with the conclusion that the probability of error is zero, $P_{err}(x_i) = 0$; for all x_i, in the case of a perfect transmission channel, so that all the information from the input is conveyed.

($\overline{A}_4$) The channel cannot transfer more information than it is possible to be sent, i.e.

$$I_{\alpha,q}(Y,X) \le C_{\alpha,q}(Y,X) \le \mathrm{Log}_q\, n, \tag{34}$$

which means that a channel cannot add additional information.

($\overline{A}_5$) The channel cannot transfer more information than it is possible to be received, i.e.

$$I_{\alpha,q}(Y,X) \le C_{\alpha,q}(Y,X) \le \mathrm{Log}_q\, m, \tag{35}$$

which means that a channel cannot add additional information.

($\overline{A}_6$) Consistency with the Shannon case:

$$\lim_{\alpha\to 1, q\to 1} I_{\alpha,q}(X,Y) = I(X,Y), \quad \text{and} \quad \lim_{\alpha\to 1, q\to 1} C_{\alpha,q}(X,Y) = C(X,Y) \tag{36}$$

Thus, the axioms ($\overline{A}_2$) and ($\overline{A}_3$) ensure that the information measures are compatible with the maximum likelihood detection (18)-(20). On the other hand, the axioms ($\overline{A}_1$), ($\overline{A}_4$) and ($\overline{A}_5$), prevent a situation in which a physical system conveys information in spite of going through a completely destructive channel, or in which the negative information transfer is observed, indicating that the channel adds or removes information by itself, which could be treated as non-physical behavior without an intuitive explanation. Finally, the property ($\overline{A}_6$) ensure that the information transfer measures can be considered as generalizations of corresponding Shannon measures. For these reasons, we assume that the satisfaction of the properties ($\overline{A}_1$)-($\overline{A}_5$) is mandatory for any reasonable definition of Sharma-Mittal information transfer measures.

4. The α-mutual information and the α-capacity

One of the first proposals for Rényi mutual information goes back to Arimoto [24] who considered the following definition of mutual information:

$$I_\alpha(X,Y) = \frac{\alpha}{1-\alpha} \log \left(\sum_y \left(\sum_x P_X^{(\alpha)}(x) P_{Y|X}^{\alpha}(y \mid x) \right)^{\frac{1}{\alpha}} \right) \tag{37}$$

where the escort distribution $P_{X^{(\alpha)}}$ is defined as in

$$P_X^{(\alpha)}(X) = \frac{P_X(x)^\alpha}{\sum_x P_X(x)^\alpha}, \quad \alpha > 0, \tag{38}$$

and he also invented an iterative algorithm [35] for the computation of the α-capacity which is defined form the α-mutual information:

$$C_\alpha(P_{Y_X}) = \max_{P_X} I_\alpha(X,Y). \tag{39}$$

Notably, the Arimoto's mutual information can equivalently be represented using the conditional Rényi entropy

$$R_\alpha(X|Y) = \frac{\alpha}{\alpha-1}\log_2 \sum_y P_Y(y)\left(\sum_x P_{X|Y=y}(x)^\alpha\right)^{\frac{1}{\alpha}}, \tag{40}$$

as in

$$I_\alpha(X,Y) \equiv R_\alpha(X) - R_\alpha(X|Y), \tag{41}$$

which can be interpreted as the input uncertainty reduction after the output symbols are received and in the case of $\alpha \to 1$, the previous definition reduces to Shannon case. In addition, this measure is directly related to the famous Gallager exponent,

$$E_0(\rho, P_X) = -\log\left(\sum_y\left(\sum_x P_X(x) P_{Y|X}^{\frac{1}{1+\rho}}(y \mid x)\right)^{1+\rho}\right) \tag{42}$$

which has been wieldy used for establishing of upper bound of error probability in channel coded communication systems [36] via the relationship [29]:

$$I_\alpha(X,Y) = \frac{\alpha}{1-\alpha}E_0\left(\frac{1}{\alpha}-1, P_X^{(\alpha)}\right) \tag{43}$$

In addition, in the case of $\alpha \to 1$, it reduces to

$$I_1(X,Y) = \lim_{\alpha\to 1} I_\alpha(X,Y) = I(X,Y), \tag{44}$$

where

$$I(X,Y) = \sum_{x,y} P_{X,Y}(x,y)\log\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \tag{45}$$

stands for Shannon's mutual information [37].

The α-mutual information $I_\alpha(X,Y)$ and the α-capacity $C_\alpha(P_{Y_X})$ satisfies the axioms $(\overline{A}_1)$-$(\overline{A}_6)$ for $q = 1$ and $\alpha > 0$ as stated by the following theorem, which further justify their usage as the measures of (maximal) information transfer.

Theorem 1. *The mutual information measures $I_\alpha^t(X,Y)$; $t \in \{a,s\}$ satisfy the following set of properties*

(A_1) *The channel can not convey negative information, i.e.*

$$C_\alpha^t(X,Y) \geq I_\alpha^t(X,Y) \geq 0. \tag{46}$$

(A_2) *The (maximal) information transfer is zero in the case of totally destructive channel, i.e.*

$$P_{Y|X}(y|x) = \frac{1}{m}, \text{ for all } x,y \quad \Rightarrow \quad I_\alpha^t(X,Y) = C_\alpha^t(X,Y) = 0 \tag{47}$$

(A_3) *In the case od perfect transmission, the (maximal) information transfer is equal to the (maximal) input information, i.e.*

$$X = Y \quad \Rightarrow \quad I_\alpha^t(X,Y) = R_\alpha(X), \quad C_\alpha^t(X,Y) = \log n \tag{48}$$

(A_4) *The channel can not transfer more information than it is possible to send, i.e.*

$$I^t_\alpha(Y,X) \le C^t_\alpha(Y,X) \le \log\ n; \tag{49}$$

(A_5) *The channel can not transfer more information than it is possible to receive, i.e.*

$$I^t_\alpha(Y,X) \le C^t_\alpha(Y,X) \le \log\ m. \tag{50}$$

(A_6) *Consistency with Shannon case:*

$$\lim_{\alpha\to 1} I^t_\alpha(X,Y) = I(X,Y), \quad \textit{and} \quad \lim_{\alpha\to 1} C^t_\alpha(X,Y) = C(X,Y) \tag{51}$$

Proof. As shown in [38], $R_\alpha(X|Y) \le R_\alpha(X)$, and the nonnegativity property (A_1) follows from the definition of Arimoto's mutual information (41). In addition, if $X \perp\!\!\!\perp Y$, then $P_{Y|X}(y|x) = P_Y(y)$ so that the definition (59) implies the property (A_2). Furthermore, in the case of perfect transmission channel, the mutual information (59) can be represented in

$$I_\alpha(X,Y) = \frac{\alpha}{\alpha-1}\log\frac{\sum_y\left(\sum_x P_X(x)^\alpha P^\alpha_{Y|X}(y\mid x)\right)^{\frac{1}{\alpha}}}{\left(\sum_x P^{(\alpha)}_X(x)\right)^{\frac{1}{\alpha}}} = \frac{\alpha}{\alpha-1}\log\frac{\sum_y\left(P_X(d(y))^\alpha P^\alpha_{Y|X}(y\mid d(y))\right)^{\frac{1}{\alpha}}}{\left(\sum_x P^{(\alpha)}_X(x)\right)^{\frac{1}{\alpha}}} \tag{52}$$

and since

$$\begin{gathered}\sum_y\left(P_X(d(y))^\alpha P^\alpha_{Y|X}(y\mid d(y))\right)^{\frac{1}{\alpha}} = \sum_y P_X(d(y))P_{Y|X}(y\mid d(y)) = \\ \sum_x\sum_{y:d(y)=x} P_X(d(y))P_{Y|X}(y\mid d(y)) = \sum_x P_X(x)\sum_{y:d(y)=x} P_{Y|X}(y\mid x) = 1\end{gathered} \tag{53}$$

we obtain $I_\alpha(X,Y) = R_\alpha(X)$ which proves the property (A_3). Moreover, from the definition as shown in [38], the Arimoto's conditional entropy is positive and satisfies the weak chain rule $R_\alpha(X|Y) \ge R_\alpha(X) - \log m$, so that the properties (A_4) and (A_5) follows from the definition of Arimoto's mutual informatio (41). Finally, the property (A_6) follows directly from the equation (44), and can be approved using L'Hôpital's rule, which competes the proof of the theorem.

□

5. Alternative definitions of the α-mutual information and the α-channel capacity

Since Renyi's proposal, there were several lines of the research for an appropriate definition and characterization of information transfer measures related to Rényi entropy which are established by the substitution of Rényi divergence measure,

$$D_\alpha(P||Q) = \frac{1}{\alpha-1}\log\left(\sum_x P(x)^\alpha Q(x)^{1-\alpha}\right), \tag{54}$$

instead of Kullback-Leibler one

$$D(P||Q) = D_1(P||Q) = \sum_x P(x)\log\frac{P(x)}{Q(x)}, \tag{55}$$

in some of various definitions which are equivalent in the case of Shannon information measures (45) [29]:

$$I(X,Y) = \min_{Q_Y} \mathbb{E}\left[D_\alpha\left(P_{Y|X}\|Q_Y\right)\right] = \min_{Q_Y} \mathbb{E}\left[D_\alpha\left(P_{Y|X}\|Q_Y\right)\right]$$
$$= \min_{Q_X}\min_{Q_Y} D_\alpha\left(P_{X,Y}\|Q_X Q_Y\right) = D_\alpha\left(P_{X,Y}\|P_X P_Y\right) = S(X) - S(X|Y) \tag{56}$$

where $S(X|Y)$ stands for Shannon conditional entropy.

All of these measures are consistent with Shannon case in the view of the property (A_6), but their direct usage as measures of Rényi information transfer leads to a breaking of some the properties (A_1)-(A_5), which justifies the usage of Arimoto's measures from the previous section as appropriate ones in the context of this work. In this section, we review the alternative definitions.

5.1. Information transfer measures by Sibson

An alternative approaches based on Rényi divergence can were proposed by Sibson [39], who introduced

$$J^s_\alpha(X;Y) = \min_{Q_Y} D_\alpha\left(P_{Y|X}P_X\|Q_Y P_X\right), \tag{57}$$

which can be represented as in [40]

$$J^s_\alpha(X,Y) = \frac{\alpha}{\alpha-1}\log\left(\sum_y\left(\sum_x P_X(x)P^\alpha_{Y|X}(y \mid x)\right)^{\frac{1}{\alpha}}\right) \tag{58}$$

and in the discrete setting can be related with Gallager exponent as in [29]:

$$J^s_\alpha(X,Y) = \frac{\alpha}{1-\alpha}E_0\left(\frac{1}{\alpha}-1, P_X\right) \tag{59}$$

which differs from Arimoto's definition (59) since in this case the escort distribution does not participate in the error exponent, but an ordinary one. However, in the case of perfect channel for which $X = Y$, the conditional distribution $P^\alpha_{Y|X}(y \mid x) = 1$ for $x = y$ and zero otherwise, so that Sibson's measure (58) reduces to $R_{1/\alpha}$, thus breaking the axiom (A_3). This disadvantage can be overcame by reparametrization $\alpha \leftrightarrow 1/\alpha$ so that $J^s_{1/\alpha}(X,Y)$ is used as a measure of Rényi information transfer, and the properties of the resulting measure can be considered in a similar manner as in the case of Arimoto's one.

5.2. Information transfer measures by Augustin and Csiszar

An alternative definition of Rényi mutual information was also presented by Augustin [25], and latter Csiszar [26], who defined

$$I^{ac}_\alpha(X;Y) = \min_{Q_Y} \mathbb{E}\left[D_\alpha\left(P_{Y|X}\|Q_Y\right)\right], \tag{60}$$

However, in the case of perfect transmission, for which $X = Y$, the measure reduces to Shannon entropy

$$I^{ac}_\alpha(X;Y) = S(X) \tag{61}$$

which breaks the axiom (A_3).

5.3. Information transfer measures by Lapidoth, Pfister, Tomamichel and Hayashi

Similar obstacle as in the case of Augustin-Csiszar measure can be observed in the case of the mutual information which was considered be Lapidoth and Pfister [27] and Tomamichel and Hayashi [28] who proposed

$$I_\alpha^{lpth}(X;Y) = \min_{Q_X}\min_{Q_Y} D_\alpha\left(P_{X,Y}\|Q_X Q_Y\right). \tag{62}$$

As shown in [27] (Lemma 11) if $X = Y$, then

$$J_\alpha(X;Y) = \begin{cases} \frac{\alpha}{1-\alpha} R_\infty(X) & \text{if } \alpha \in \left[0, \frac{1}{2}\right], \\ R_{\frac{\alpha}{2\alpha-1}}(X) & \text{if } \alpha > \frac{1}{2} \\ R_{\frac{1}{2}}(X) & \text{if } \alpha = \infty \end{cases} \tag{63}$$

so that the axiom (A_3) is broken in this case, as well.

Remark 1. *Despite the difference between the definitions of information transfer, in the discrete setting, the alternative definitions discussed above reaches the same maximum over the set of input probability distributions, P_X, [26], [41] and their operational characterization of these measures, their further properties, and their relevance in source and channel coding, hypothesis testing, cryptography and quantum information theory can be found in [40], [29], [42], [43], [44], [45], [46], [47], [48].*

5.4. Information transfer measures by Chapeau-Blondeau, Delahaies, Rousseau, Tridenski, Zamir, Ingber and Harremoes

Chapeau-Blondeau, Delahaies and Rousseau [31], and independently Tridenski, Zamir and Ingber [49] and by Harremoes [50] defined Rény mutual information using Rényi divergence (54), so that the mutual information defined using the Rényi divergence

$$J_\alpha^c(X,Y) = D_\alpha\left(P_{X,Y}\|P_X P_Y\right) \tag{64}$$

for $\alpha > 0$ and $\alpha \neq 1$, while in the case of $\alpha = 1$ it reduces to Shannon mutual information. Note that the ordinal definition can correspond only to a Rényi entropy of order $2-\alpha$ since in the case of $X = Y$ it reduces to $I_\alpha^c(X,Y) = R_{2-\alpha}$ (see also [50]), which can be overcame by the reparametrization $\alpha = 2 - q$, similarly as in the case of Sibson's measure. This measure have been discussed in the past with the various operational characterizations, and could be also considered as a measure of information transfer, although the satisfaction of all of the axioms (A_1)-(A_6) is not self-evident for a general channels.

5.5. Information transfer measures by Jizba, Kleinert and Shefaat

Finally, we mention the definition by Jizba, Kleinert and Shefaat [51],

$$J_\alpha^j(X,Y) \equiv R_\alpha(X) - R_\alpha^j(X|Y). \tag{65}$$

which is defined in the same manner as in Arimoto's case (41), but with another choice of conditional Rényi entropy

$$R_\alpha^j(X|Y) = \frac{1}{1-\alpha}\log\sum_x P_X^{(\alpha)}(x) 2^{(1-\alpha)R_\alpha(X|Y=y)}, \tag{66}$$

which arises from Generalized Shannon-Khinchin axiom GSK4 if the pseudo-additivity in the equation (9) is restricted to an ordinary addition, at which case the GSK axioms uniquely determine Rényi entropy [52]. However, despite a wide applicability in modeling of causality and financial time series, this mutual information can take negative values which breaks the axiom (A_1), which is assumed to be mandatory in this work. For further discussion physicalism of negative mutual information in the domain of financial time series analysis, the reader is referred to [51]

6. The α-q mutual information and the α-q capacity

In the past several attempts were done to define an appropriate channel capacity measure which corresponds to instances of Sharma-Mittal entropy class. All of them follow a similar recipe by which the channel capacity is defined as in (30), as a maximum of appropriately defined mutual information $I_{\alpha,q}$. However, all of the classes consider only a special cases of Sharma-Mittal entropy and, all of them fails down to satisfy at least one of the properties $(\overline{A}_1)$-$(\overline{A}_5)$ that a information transfer has to satisfy, as we will discuss Section 7.

In this section we propose a general measures of the α-q mutual information and the α-q capacity by the requirement that the axioms (A_1)-(A_5) are satisfied, which could qualify them as appropriate measures of information transfer, without non-physical properties. The special instances of the α-q (maximal) information transfer measures are also discussed and the analytic expressions for ninary symmetric channel are provided.

6.1. The α-q information transfer measures and its instances

The α-q-mutual information (41) is defined using the q-substraction defined in (6), as follows:

$$I^{a}_{\alpha,q}(X,Y) = H_{\alpha,q}(X) \ominus_q H_{\alpha,q}(X|Y), \tag{67}$$

where we introduced the conditional Sharma-Mittal entropy $H_{\alpha,q}(Y|X)$ as in

$$H_{\alpha,q}(X|Y) = \eta_q\left(R_\alpha(X|Y)\right) = \frac{1}{(1-q)\ln 2}\left(\left(\sum_y P_Y(y)\left(\sum_x P_{X|Y=y}(x)^\alpha\right)^{\frac{1}{\alpha}}\right)^{\frac{\alpha(1-q)}{\alpha-1}} - 1\right) \tag{68}$$

and $R_\alpha(X|Y)$ stands for Arimoto's definition of conditional Rényi entropy (40).The expression (67) can also be obtained if the mapping η_q is applied to the both sides of the equality (41) by which the Arimoto's mutual information is defined, so that we establish the relationship

$$I^{a}_{\alpha,q}(X,Y) = \eta_q\left(I_\alpha(X,Y)\right) = \eta_q\left(\frac{\alpha}{1-\alpha}\log\left(\sum_y\left(\sum_x P_X^{(\alpha)}(x)P^{\alpha}_{Y|X}(y\mid x)\right)^{\frac{1}{\alpha}}\right)\right) \tag{69}$$

and can be represented using Gallager error exponent (42) as in

$$I^{a}_{\alpha,q}(X,Y) = \eta_q\left(\frac{\alpha}{1-\alpha}E_0\left(\frac{1}{\alpha}-1,P_X^{(\alpha)}\right)\right) = \frac{1}{(1-q)\ln 2}\left(2^{\frac{\alpha(1-q)}{1-\alpha}E_0\left(\frac{1}{\alpha}-1,P_X^{(\alpha)}\right)}-1\right). \tag{70}$$

Arimoto's α-q-capacity is now defined in

$$C_{\alpha,q} = \max_{P_X} I_{\alpha,q}(X,Y) \tag{71}$$

and using the fact that η_q is increasing it can be related with corresponding α-capacity as in

$$C_{\alpha,q} = \max_{P_X} I^{a}_{\alpha,q}(X,Y) = \max_{P_X}\eta_q\left(I_\alpha(X,Y)\right) = \eta_q\left(\max_{P_X} I_\alpha(X,Y)\right) = \eta_q\left(C^{a}_{\alpha}(X,Y)\right). \tag{72}$$

In the same manner as in the case of the α-capacity, using the expressions (44), (45) and (1), for $\alpha = 1$ both Arimoto's and Sibson's α-q mutual information reduces to

$$I_{1,q} = \frac{1}{(1-q)\ln 2}\left(\prod_{x,y} 2^{P_{X,Y}(x,y)\log\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}} - 1\right)$$
$$= \frac{1}{(1-q)\ln 2}\left(\prod_{x,y}\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right)^{P_{X,Y}(x,y)} - 1\right) \tag{73}$$

the α-q channel capacity is given in

$$C_{1,q} = \max_{P_X}\left(\frac{1}{(1-q)\ln 2}\left(\prod_{x,y}\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right)^{P_{X,Y}(x,y)} - 1\right)\right) \tag{74}$$

and these measures can serve as (maximal) information transfer measures corresponding to Gaussian entropy which was not considered before in the context of information transmission. Naturally, if in addition $q \to 1$, the measures reduce to Shannon's mutual information and Shannon capacity [37].

Additional special cases of the α-q (maximal) information transfer includes the α-mutual information (41) and the α-capacity (39), which are obtained for $q = 1$, while the measures which corresponds to Tsallis entropy can be obtained for $q = \alpha$ and the ones which corresponds to Landsberg-Vedral entropy for $q = 2 - \alpha$. These special instances are listed in Table 1.

As discussed in Section 7, previously considered information measures cover only particular special cases and break at least one of the axioms $(\overline{A}_1)$-$(\overline{A}_5)$, which lead to unexpected and counterintuitive conclusions about the channels, such as negative information transfer and achieving super-capacitance or sub-capacitance [4], which could be treated as a nonphysical behavior. On the other hand, apart from the generality, the α-q information transfer measures proposed in this work overcame the disadvantages which could qualify them as appropriate measures, as stated in the following theorem.

Theorem 2. *The α-q information transfer measures $I_{\alpha,q}(X,Y)$ and $C_{\alpha,q}$ satisfy the set of the axioms $(\overline{A}_1)$-$(\overline{A}_6)$.*

Proof. The proof is straightforward application of the mapping η_q to the equations in the α-mutual information properties (A_1)-(A_5), while the $(\overline{A}_6)$ follows from the above discussion. □

Remark 2. *Note that the symmetry $I_{\alpha,q}(X,Y) = I_{\alpha,q}(Y,X)$ does not hold in general neither in the case of Arimoto's nor Sibson's type defined Sharma-Mittal mutual information [53], [54] and if the mutual information is defined so that the symmetry is preserved, the property $(\overline{A}_1)$ might be broken. In addition, the alternative definition of the mutual information, $I_{\alpha,q}(Y,X) = H_{\alpha,q}(Y) - H_{\alpha,q}(Y|X)$, which uses ordinary substraction operator instead of $\ominus_q$ operation, can also be introduced, but in this case the property $(\overline{A}_4)$ might not hold in general, as discussed in Section 7.*

6.2. The α-q capacity of binary symmetric channels

As shown by Cai and Verdú [42], the α-mutual information of Arimoto's type I_α are maximized for the uniform distribution $P_X = (1/2, 1/2)$, and Arimoto's α-capacity has the value

$$C_\alpha(BSC) = 1 - r_\alpha(p), \tag{75}$$

where the binary entropy function r_α is defined

$$r_\alpha(p) = R_\alpha(p, 1-p) = \frac{1}{1-\alpha}\log(p^\alpha + (1-p)^\alpha), \tag{76}$$

for $\alpha > 0$, $\alpha \neq 1$ while in the limit of $\alpha \to 1$, the expression (76) reduces to the well known result for the Shannon's capacity (see Fano [55])

$$C_1(BSC) = \lim_{\alpha \to 1} C_\alpha(BSC) = 1 + p \log p + (1-p)\log(1-p). \tag{77}$$

The analytic expressions for the α-q-capacities of binary symmetric channel can be obtained from the expressions (72) and (75), so that

$$C_{\alpha,q}(BSC) = \eta_q\left(C_\alpha(BSC)\right) = \frac{1}{(1-q)\ln 2}\left(2^{1-q}\left(p^\alpha + (1-p)^\alpha\right)^{-\frac{1-q}{1-\alpha}} - 1\right) \tag{78}$$

and in the case of $q = 1$ it reduces to the case of Rényi entropy, while in the case of $\alpha = 1$, to the case of Gaussian entropy (75)

$$C_{1,q}(BSC) = \frac{1}{(1-q)\ln 2}\left(2p^p(1-p)^{1-p} - 1\right) \tag{79}$$

The analytic expressions for BSC α-q capacities for another instances can straightforwardly be obtained by specifying the values of the parameters and the instances are listed in Table 1 and the plots of the BSC α-q-capacities which corresponds to Gausian and Tsallis entropies are shown in Figure 3 and Figure 4.

The α-q-capacity (78) can equivalently be expressed in

$$C_{\alpha,q}(BSC) = \mathrm{Log}_q\, 2 \ominus_q h_{\alpha,q}(p), \tag{80}$$

where Sharma-Mittal binary entropy function is defined in

$$h_{\alpha,q}(p) = H_{\alpha,q}(p, 1-p) = \frac{1}{1-q}\left(\left(p^\alpha + (1-p)^\alpha\right)^{\frac{1-q}{1-\alpha}} - 1\right), \tag{81}$$

which reduces to Rényi binary entropy function

$$h_{\alpha,1}(p) = \lim_{q \to 1} h_{\alpha,q}(p) = R_\alpha(p, 1-p) = \frac{1}{1-\alpha}\log\left(p^\alpha + (1-p)^\alpha\right), \tag{82}$$

in the case of $q = 1$ and to Gaussian binary entropy function

$$h_{1,q}(p) = \lim_{\alpha \to 1} h_{\alpha,q}(p) R_\alpha(p, 1-p) = \frac{1}{1-\alpha}\log\left(p^\alpha + (1-p)^\alpha\right), \tag{83}$$

in the case of $\alpha = 1$. The expression (80) can be interpreted similarly as in the Shannon case. Thus, for a BSC channel with input X and output Y can be modeled with an input-output relation $Y = X \oplus Z$ where $\oplus$ stands for modulo 2 sum and Z is channel noise taking values from $\{1, 0\}$ and is distributed in accordance to $(p, 1-p)$. If we measure the information which is lost per bit during transmission with the Sharma-Mittal entropy $H_{\alpha,q}(Z) = h_\alpha(p)$, then $C_{\alpha,q}$ stands for useful information left over for every bit of information received.

7. An overview of the previous approaches for Sharma-Mittal information transfer measures

In this section, we review the previous attempts for the definition of Sharma-Mittal information transfer measures which are defined from the basic requirement of consistency with Shannon measure as given by the axiom $(\overline{A}_6)$. However, as we show in the following paragraphs, all of them break at least one of the axioms $(\overline{A}_1)$-$(\overline{A}_5)$, which are satisfied in the case of the α-q (maximal) information transfer measures (67) and (71), in accordance to the discussion from Section 6.

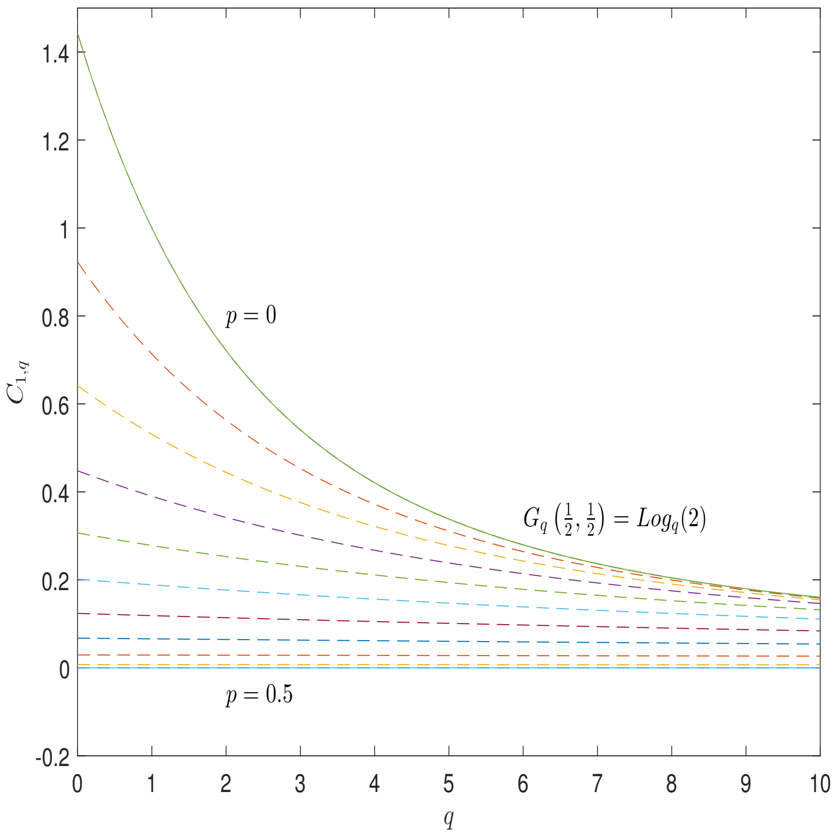

Figure 3. The α-q-capacity of BSC for Gaussian entropy (the case of $\alpha = 1$) as a function of q for various values of the channel parameter p from 0.5 (totally destructive channel) to 0 (perfect transmission). All of the curves lies between 0 and $\text{Log}_q 2$, which is the maximum value of the Gaussian entropy.

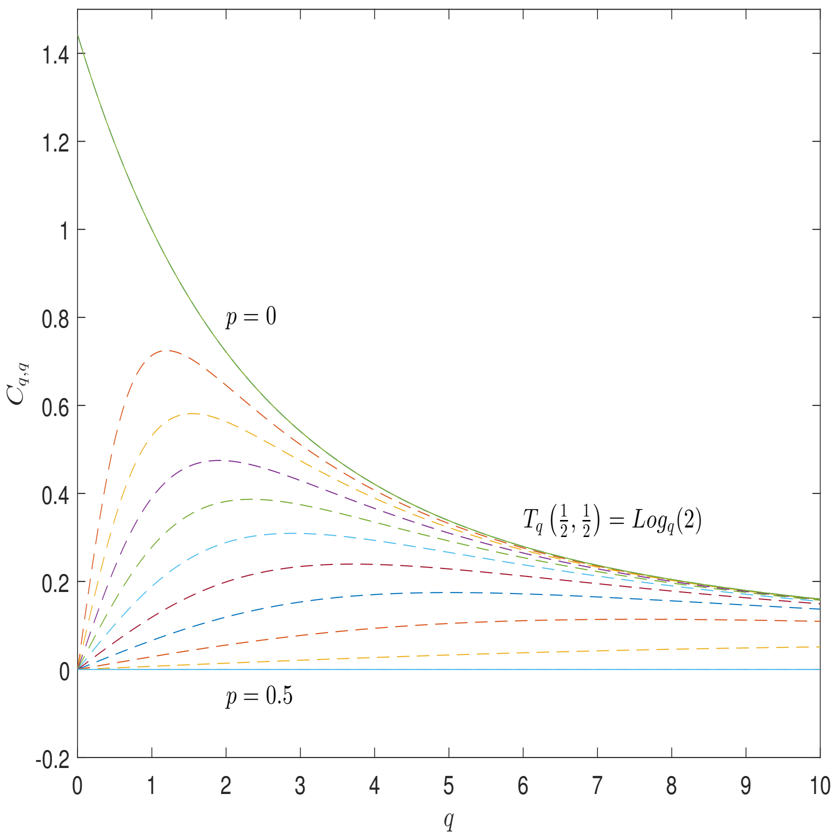

Figure 4. The α-q-capacity of BSC for Tsallis entropy (the case of $\alpha = q$) as a function of q for various values of the channel parameter p from 0.5 (totally destructive channel) to 0 (perfect transmission). All of the curves lies between 0 and $\text{Log}_q 2$, which is the maximum value of the Tsallis entropy.

$H_{\alpha,q}$	$I_{\alpha,q}$	$C_{\alpha,q}$
R_α $q=1$	$\frac{\alpha}{1-\alpha}E_0\left(\frac{1}{\alpha}-1,P_X^{(\alpha)}\right)$	$1-\frac{\log(p^\alpha+(1-p)^\alpha)}{1-\alpha}$
R_α $q=1$	$\frac{\alpha}{1-\alpha}E_0\left(\frac{1}{\alpha}-1,P_X^{(\alpha)}\right)$	$1-\frac{\log(p^\alpha+(1-p)^\alpha)}{1-\alpha}$
T_α $q=\alpha$	$\frac{1}{(1-\alpha)\ln 2}\left(2^{\alpha E_0\left(\frac{1}{\alpha}-1,P_X^{(\alpha)}\right)}-1\right)$	$\frac{1}{(1-\alpha)\ln 2}\left(2^{1-\alpha}(p^{\frac{1}{\alpha}}+(1-p)^{\frac{1}{\alpha}})^{-\alpha}-1\right)$
L_α $q=2-\alpha$	$\frac{1}{(\alpha-1)\ln 2}\left(2^{-\alpha E_0\left(\frac{1}{\alpha}-1,P_X^{(\alpha)}\right)}-1\right)$	$\frac{1}{(1-\alpha)\ln 2}\left(2^{1-\alpha}(p^{\alpha}+(1-p)^{\alpha})^{-1}-1\right)$
G_q $\alpha=1$	$\frac{1}{(1-q)\ln 2}\left(\prod_{x,y}\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right)^{P_{X,Y}(x,y)}-1\right)$	$\frac{1}{(1-q)\ln 2}\left(2^{1-q}p^{(1-q)p}(1-p)^{(1-q)(1-p)}-1\right)$
$E_0(\rho,P_X)=-\log\left(\sum_y\left(\sum_x P_X(x)P_{Y\mid X}^{\frac{1}{1+\rho}}(y\mid x)\right)^{1+\rho}\right)$		

Table 1. The instances of the α-q-mutual information for different values of the parameters and corresponding expressions for the BSC α-q-capacities.

221 *7.1. Daróczy's capacity*

The first considerations of generalized channel capacities and generalized mutual information for the q-entropy goes back to Daróczy [30], who introduced the generalized of conditional entropy

$$T_q^D(Y|X)=\sum_x P_X^q(x)T_q(Y|X=x), \tag{84}$$

where the row entropies are defined as in

$$T_q(Y|X=x)=\frac{1}{(1-q)\log(2)}\left(\sum_x P_{Y|X}(y|x)^q-1\right). \tag{85}$$

and the mutual information is defined as in

$$J_{\alpha,q}^d(X,Y)=T_q(Y)-T_q^D(Y|X), \tag{86}$$

However, in the case of totally destructive channel, $X \perp\!\!\!\perp Y$, $P_{Y|X}(y|x)=P_Y(y)$, $T_q(Y|X=x)=T_q(Y)$ and

$$T_q(Y|X)=T_q(Y)\sum_x P_X(x)^q \tag{87}$$

so that

$$J_{\alpha,q}^d(X,Y)=T_q(Y)\left(1-\sum_x P_X(x)^q\right)=\left(1-\sum_x P_X(x)^q\right)\mathrm{Log}_q m. \tag{88}$$

This expression is zero for an input probability distribution $P_X = (1, 0, \ldots, 0)$ and its permutations, but, in general, it is negative for $q < 1$, positive for $q > 1$ and 0 only for $q = 1$, so that the axiom $(\overline{A}_2)$ is broken (see Figure 5). As a result, the channel capacity which is defined in accordance to (30), is zero for $q \leq 1$, and positive for $q > 1$, which is illustrated in Figure 6 by the example of BSC for which the Daroczy's channel capacity can be computed as in [30], [56]

$$\overline{C}_{BSC}^{D} = \frac{1-2^{1-q}}{q-1} - \frac{2^{-q}}{q-1}[1-(1-p)^q - p^q]. \tag{89}$$

In the same Figure, we plotted the graph for the α-q channel capacities proposed in this work, and all of them remain zero in the case of totally destructive BSC as expected.

7.2. Yamano capacities

Similar problems as above arise in the case of mutual information and corresponding capacity measures considered by Yamano [33] who addressed the information transmission which is characterized by Landsberg-Vedral entropy L_q, given in (16).

Thus, the first proposal is based on the mutual information of the form

$$J_q^{Y_1}(X,Y) = L_q(X) + L_q(Y) - L_q(X,Y) \tag{90}$$

where the joint entropy is defined in

$$L_q(X,Y) = \frac{1}{q-1}\left(\frac{1}{\sum_{x,y} P_{X,Y}(x,y)^q} - 1\right). \tag{91}$$

However, in the case of fully destructive channel, $P_Y(y) = 1/m$ and $P_{X,Y}(x,y) = P_X(x)/m$, so that

$$J_q^{Y_1}(X,Y) = \frac{1}{q-1}\left(\frac{1}{\sum_x P_X(x)^q} - 1\right) + \frac{1}{q-1}\left(m^{q-1} - 1\right) - \frac{1}{q-1}\left(m^{q-1}\frac{1}{\sum_x P_X(x)^q} - 1\right) \tag{92}$$

which can be simplified to

$$J_q^{Y_1}(X,Y) = \frac{1-m^{q-1}}{q-1}\left(\frac{1}{\sum_x P_X(x)^q} - 1\right) \tag{93}$$

Similarly as in the case of Daroczy's capacity this expression is zero for an input probability distribution $P_X = (1, 0, \ldots, 0)$ and its permutations, but, in general, it is negative for $q > 1$, positive for $q < 1$ and 0 only for $q = 1$, so that the axiom $(\overline{A}_2)$ is broken (see Figure 5). In Figure 6 we illustrated the Yamano channel capacity as a function of the parameter q, in the case of two input channel with $P_X = [a, 1-a]$, the channel capacity is zero for $q > 1$ (which is obtained for $P_X = [1, 0]$), and

$$C_{BSC}^{Y} = \frac{1}{q-1}\left(2^q - 1 - 2^{2q-2}\right), \tag{94}$$

for $q > 1$ (which is obtained for $P_X = [1/2, 1/2]$). In the same Figure, we plotted the graph for the α-q channel capacities proposed in this work, and, as before, all of them remain zero in the case of totally destructive BSC as expected.

Further attempts were done in [33], where the mutual information is defined following in an analogous manner to (64) and (64), with the generalized divergence measure introduced in [57]. Thus, the alternative measure for mutual information is defined in

$$J_q^{Y}(X,Y) = \frac{1}{(1-q)\ln 2}\frac{1}{\sum_{x,y} P_{X,Y}^q(x,y)}\left[1 - \sum_{x,y} P_{X,Y}(x,y)\left(\frac{P_X(x)P_Y(y)}{P_{X,Y}(x,y)}\right)^{1-q}\right]. \tag{95}$$

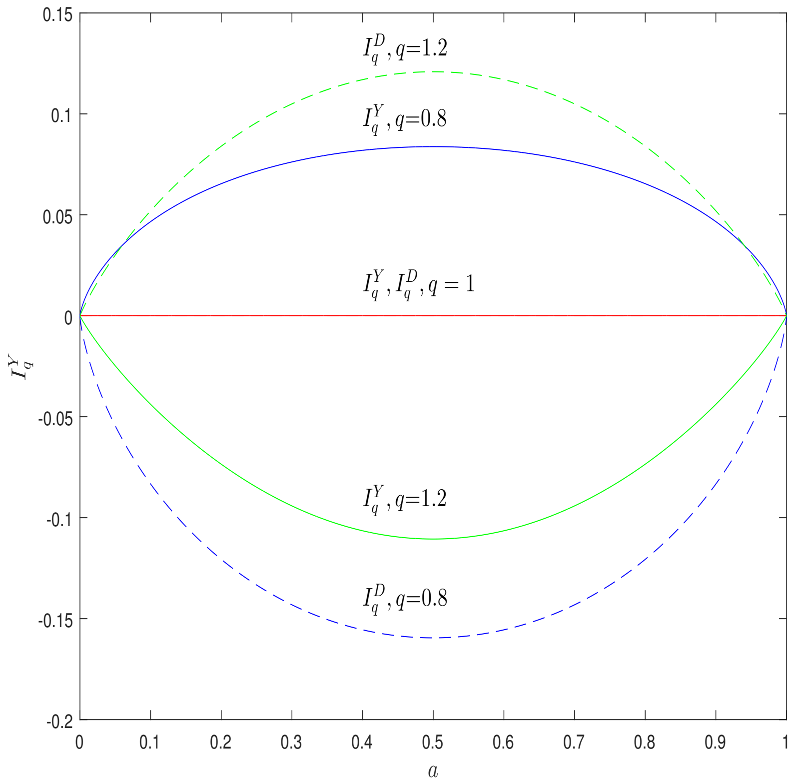

Figure 5. Daróczy's (solid lines) and Yamano's (dashed lines) mutual information in the case of totally destructive BSC as functions of the input distribution parameter a, $P_X = [a, 1-a]^T$ for different values of q, getting negative values for $q < 1$ and $q > 1$, respectively, breaking the axioms $(\overline{A}_1)$ and $(\overline{A}_2)$. The α-q-mutual information is zero and satisfies $(\overline{A}_1)$ and $(\overline{A}_2)$

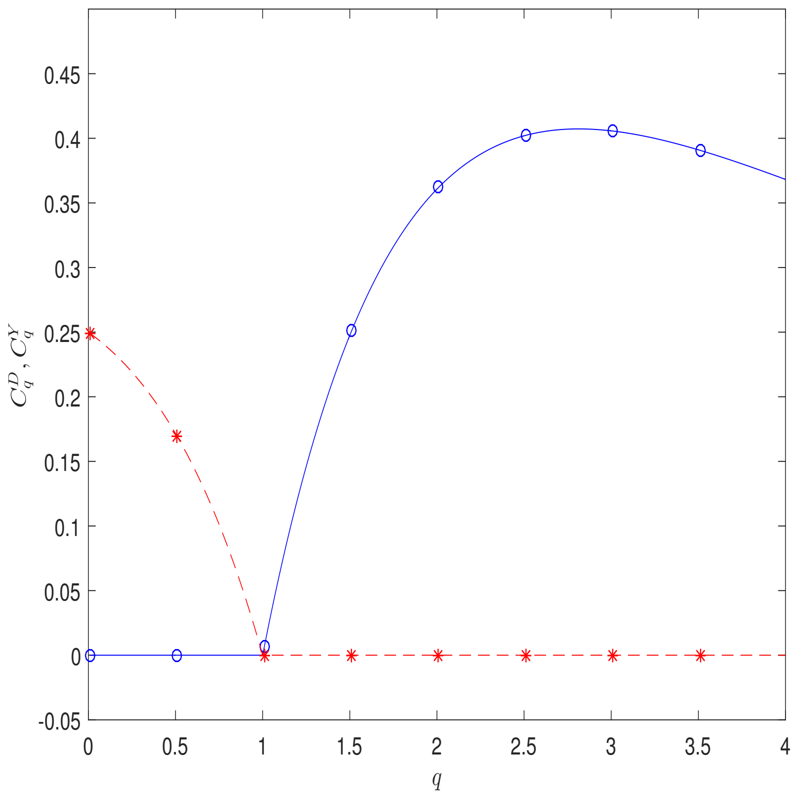

Figure 6. Daróczy's (solid lines) and Yamano's (dashed lines) capacities in the case of totally destructive BSC as functions of the parameter q. In the regions of $q < 1$ and $q > 1$, respectively, corresponding negative mutual information are maximized for $P_X = [1, 0]^T$ (zero capacity) having the positive values outside the regions and breaking the axiom $(\overline{A}_2)$. The α-q-capacity is zero and satisfies $(\overline{A}_2)$.

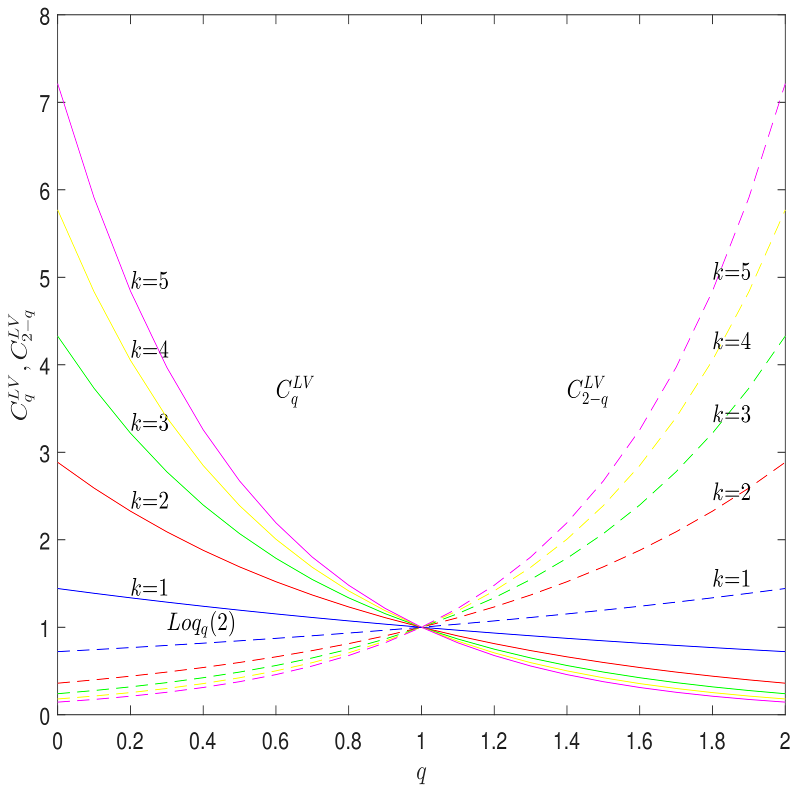

Figure 7. Landsberg-Vedral capacities for Landsberg-Vedral (solid lines) and Tsallis (dashed lines) entropies in the case of (perfect) noisy channel with non-overlapping outputs with m outputs as functions of q, for different values of m. The axiom $(\overline{A}_4)$ is broken for all $m > 2$ and satisfied in the case of corresponding α-q-capacities.

However, in the case of the simplest perfect communication channel for which $X = Y$, the mutual information reduces to

$$J_q^Y(X,Y) = \frac{1}{(1-q)\ln 2}\frac{1-\sum_x P_X(x)^{2-q}}{\sum_x P_X(x)^q} \neq L_q(X) \tag{96}$$

which breaks the axiom $(\overline{A}_3)$.

7.3. Landsber-Vedral capacities

To avoid these problems, Landsberg and Vedral [4] proposed the mutual information measure and related channel capacities for Sharma-Mittal entropy class $H_{\alpha,q}$, particularly considering the choice of $q = \alpha$, which corresponds to Tsallis entropy, $q = 2 - \alpha$ and the case of $q = 1$ which corresponds to Rényi entropy.

$$I(Y,X) = H_{\alpha,q}(Y) - H_{\alpha,q}^{LV}(Y|X), \tag{97}$$

where the conditional entropy $H_{\alpha,q}^{LV}(Y|X)$ is defined as in

$$H_{\alpha,q}^{LV}(Y|X) = \sum_x P_X(x) H_{\alpha,q}(Y|X=x) \tag{98}$$

and

$$H_{\alpha,q}(Y|X=x)^{LV} = \frac{1}{1-q}\left(\left(\sum_y P_{Y|X=x}(y|x)^\alpha\right)^{\frac{1-q}{1-\alpha}} - 1\right), \tag{99}$$

Although this definition bear some similarities with the α-q mutual information proposed in formula (67), several key differences can be observed. The first one, it characterizes the information transfer as the output uncertainty reduction, after the input symbols are known, instead of input uncertainty reduction, after the output symbols are known (41). In addition, it uses the ordinary $-$ operation, instead of the $\ominus_q$ one. Also note that the definition of conditional entropy (98) in generally differs from the definition proposed in (68).

The definition (97) successfully the issue the axiom $(\overline{A}_2)$ which appear in the case of Daroczy capacity since in the case of totally destructive channel ($X \perp\!\!\!\perp Y$), $P_{Y|X}(y|x) = P_Y(y)$ and $L_q(Y|X = x) = L_q(Y)$ and $L_q(Y|X) = L_q(Y)$, so that $I^{lv}_{\alpha,q}(X,Y) = 0$. However, the problems remains with the axiom $(\overline{A}_5)$, which can be observed in the case of noisy channel with non-overlapping outputs if the number of channel inputs is lower than the number of the channel outputs $n < m$. Indeed, in the case of noisy channel with non-overlapping outputs given by the transition matrix (25) both of the row entropies $L_q(Y|X = x)$ has the same value which is independent of x

$$H_{\alpha,q}(Y|X = x) = \frac{k^{1-q} - 1}{(q-1)\ln 2} = \mathrm{Log}_q k; \quad \text{for} \quad x = x_1, x_2, \tag{100}$$

and the maximal value of Landsberg-Vedral mutual information (97) is obtained only by maximizing $H_{\alpha,q}(Y)$ over P_X, which is achieved if X is uniformly distributed, since in this cae Y is uniformly distributed, as well ($a = \frac{1}{2}$ in (26)), so that the maximal value of the output entropy is $H_{\alpha,q}(Y) = \mathrm{Log}_q(2k)$ and the mutual information is maximized for.

$$C = \mathrm{Log}_q(2k) - \mathrm{Log}_q(k) \tag{101}$$

which is greater than $\mathrm{Log}_q(2)$ for $k \geq 2$, i.e. for $m \geq 4$ outputs, so that the axiom $(\overline{A}_5)$ is broken, which is illustrated in Figure 7.

7.4. Chapeau-Blondeau - Delahaies - Rousseau capacities

Following the similar approach as in Section (5.4), Chapeau-Blondeau, Delahaies and Rousseau considered the definition of mutual information which corresponds to Tsallis entropy using Tsallis divergence,

$$D_{q,q}(P||Q) = \frac{1}{q-1}\left(\sum_x P(x)^q Q(x)^{1-q} - 1\right), \tag{102}$$

and can be written in

$$\begin{aligned} I^{ct}_\alpha(X,Y) &= D_{\alpha,\alpha}\left(P_{X,Y}\|P_X P_Y\right) = \eta_\alpha\left(D_\alpha\left(P_{X,Y}\|P_X P_Y\right)\right) \\ &= \frac{1}{1-\alpha}\left(1 - \sum_{x,y} P_{X,Y}(x,y)^\alpha P_X(x)^{1-\alpha} P_Y(y)^{1-\alpha}\right). \end{aligned} \tag{103}$$

However, this definition is not directly applicable as a measure of information transfer fo Tsallis entropy with index q, since in the case of $X = Y$ it reduces to $I^c_q(X,Y) = T_{2-\alpha}$, requries the reparametrization $\alpha = 2 - q$, similarly as in the Section 5.4, while the satisfaction of the axioms $(\overline{A}_4)$-$(\overline{A}_5)$ is not self evident.

8. Conclusion and future work

A general treatment of Sharma-Mittal entropy transfer was provided together with the analyses of existing information transfer measures for non-additive Sharma-Mittal information transfer. It is shown that the existing definitions fails down to satisfy at least one of ineluctable properties common to Shannon case, by which the information transfer has to be non-negative, less than input and output uncertainty, equal to the input uncertainty in the case of perfect transmission and equal to zero, in the case of totally destructive channel. Thus, breaking some of these properties implies unexpected and counterintuitive conclusions about the channels, such as achieving super-capacitance or sub-capacitance [4], which could be treated as a nonphysical behavior.

In this work, alternative measures of the α-q mutual information and the α-q channel capacity were proposed so that all of ineluctable properties which are broken in the case of the Sharma-Mittal information transfer measures considered before, are satisfied, which could qualify them as physically consistent measure of information transfer. Taking into account the previous research in non-extensive statistical mechanics [3], where the linear growth of the system entropy has been recognized as an ineluctable property of physical quantity in non-extensive [58] and non-exponentially growing system [59], and taking into account the previous research from the field of information theory, where Sharma-Mittal entropy has been considered an appropriately scaling measure which provides extensive information rates [21], the α-q capacity seems to be a promising measure for characterization of information transmission in the scenarios where Shannon entropy rate diverges or disappears in infinite time limit.

Author Contributions: conceptualization, V.I. and I.Dj.; validation, V.I. and I.Dj.; formal analysis, V.I.; funding acquisition, I.Dj; project administration, I.Dj.; writing–original draft preparation, V.I.; writing–review and editing, V.I. and I.Dj .

Funding: This research was supported in part by NSF under grants 1907918 and 1828132 and by Ministry of Science and Technological Development, Republic of Serbia, Grants Nos. ON 174026 and III 044006. The APC was funded by NSF under grants 1907918 and 1828132.

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

Abbreviations

The following abbreviations are used in this manuscript:

BSC	Binary symmetric channel

References

1. Ilić, V.M.; Stanković, M.S. A unified characterization of generalized information and certainty measures. *Physica A: Statistical Mechanics and its Applications* **2014**, *415*, 229 – 239. doi:http://dx.doi.org/10.1016/j.physa.2014.07.061.
2. Renyi, A. *Probability Theory*; Dover Publications, 1970.
3. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **1988**, *52*, 479–487. doi:10.1007/BF01016429.
4. Landsberg, P.T.; Vedral, V. Distributions and channel capacities in generalized statistical mechanics. *Physics Letters A* **1998**, *247*, 211 – 217. doi:http://dx.doi.org/10.1016/S0375-9601(98)00500-3.
5. Frank, T.; Daffertshofer, A. Exact time-dependent solutions of the Renyi Fokker-Planck equation and the Fokker-Planck equations related to the entropies proposed by Sharma and Mittal. *Physica A: Statistical Mechanics and its Applications* **2000**, *285*, 351 – 366. doi:10.1016/S0378-4371(00)00178-3.
6. Sharma, B.; Mittal, D. New non-additive measures of entropy for discrete probability distributions. *Journal of mathematical sciences* **1975**, *10*, 28–40.
7. Tsallis, C. What are the numbers that experiments provide. *Quimica Nova* **1994**, *17*, 468–471.
8. Nivanen, L.; Le Méhauté, A.; Wang, Q.A. Generalized algebra within a nonextensive statistics. *Reports on Mathematical Physics* **2003**, *52*, 437–444. doi:10.1016/s0034-4877(03)80040-x.
9. Ilić, V.M.; Stanković, M.S. Generalized Shannon-Khinchin axioms and uniqueness theorem for pseudo-additive entropies. *Physica A: Statistical Mechanics and its Applications* **2014**, *411*, 138 – 145. doi:http://dx.doi.org/10.1016/j.physa.2014.05.009.
10. Jizba, P.; Korbel, J. When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics. *Physical Review E* **2020**, *101*, 042126.
11. Esteban, M.D.; Morales, D. A summary on entropy statistics. *Kybernetika* **1995**, *31*, 337–346.

12. Lenzi, E.; Scarfone, A. Extensive-like and intensive-like thermodynamical variables in generalized thermostatistics. *Physica A: Statistical Mechanics and its Applications* **2012**, *391*, 2543–2555.
13. Frank, T.; Plastino, A. Generalized thermostatistics based on the Sharma-Mittal entropy and escort mean values. *The European Physical Journal B - Condensed Matter and Complex Systems* **2002**, *30*, 543–549. doi:10.1140/epjb/e2002-00412-6.
14. Aktürk, O.Ü.; Aktürk, E.; Tomak, M. Can Sobolev inequality be written for Sharma-Mittal entropy? *International Journal of Theoretical Physics* **2008**, *47*, 3310–3320.
15. Mazumdar, S.; Dutta, S.; Guha, P. Sharma–Mittal quantum discord. *Quantum Information Processing* **2019**, *18*, 1–26.
16. Elhoseiny, M.; Elgammal, A. Generalized Twin Gaussian processes using Sharma–Mittal divergence. *Machine Learning* **2015**, *100*, 399–424.
17. Koltcov, S.; Ignatenko, V.; Koltsova, O. Estimating Topic Modeling Performance with Sharma–Mittal Entropy. *Entropy* **2019**, *21*, 660.
18. Jawad, A.; Bamba, K.; Younas, M.; Qummer, S.; Rani, S. Tsallis, Rényi and Sharma-Mittal holographic dark energy models in loop quantum cosmology. *Symmetry* **2018**, *10*, 635.
19. Ghaffari, S.; Ziaie, A.; Moradpour, H.; Asghariyan, F.; Feleppa, F.; Tavayef, M. Black hole thermodynamics in Sharma–Mittal generalized entropy formalism. *General Relativity and Gravitation* **2019**, *51*, 1–11.
20. Américo, A.; Khouzani, M.; Malacaria, P. Conditional Entropy and Data Processing: An Axiomatic Approach Based on Core-Concavity. *IEEE Transactions on Information Theory* **2020**, *66*, 5537–5547.
21. Girardin, V.; Lhote, L. Rescaling entropy and divergence rates. *IEEE Transactions on Information Theory* **2015**, *61*, 5868–5882.
22. Ciuperca, G.; Girardin, V.; Lhote, L. Computation and estimation of generalized entropy rates for denumerable Markov chains. *IEEE Transactions on information theory* **2011**, *57*, 4026–4034.
23. Sibson, R. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **1969**, *14*, 149–160.
24. Arimoto, S. Information Mesures and Capacity of Order α for Discrete Memoryless Channels. Topics in Information Theory; Csiszár, I.; Elias, P., Eds. János Bolyai Mathematical Society and North-Holland, 1977, Vol. 16, *Colloquia Mathematica Societatis János Bolyai*, pp. 493–519.
25. Augustin, U. Noisy Channels. PhD thesis, Universität Erlangen-Nürnberg, Erlangen, Germany, Peter Joslin, 1978.
26. Csiszár, I. Generalized cutoff rates and Rényi's information measures. *IEEE Transactions on information theory* **1995**, *41*, 26–34.
27. Lapidoth, A.; Pfister, C. Two measures of dependence. *Entropy* **2019**, *21*, 778.
28. Tomamichel, M.; Hayashi, M. Operational interpretation of Rényi information measures via composite hypothesis testing against product and Markov distributions. *IEEE Transactions on Information Theory* **2017**, *64*, 1064–1082.
29. Verdú, S. α-mutual information. 2015 Information Theory and Applications Workshop (ITA). IEEE, 2015, pp. 1–6.
30. Daróczy, Z. Generalized information functions. *Information and Control* **1970**, *16*, 36 – 51. doi:http://dx.doi.org/10.1016/S0019-9958(70)80040-7.
31. Chapeau-Blondeau, F.; Rousseau, D.; Delahaies, A. Renyi entropy measure of noise-aided information transmission in a binary channel. *Physical Review E* **2010**, *81*, 051112.
32. Chapeau-Blondeau, F.; Delahaies, A.; Rousseau, D. Tsallis entropy measure of noise-aided information transmission in a binary channel. *Physics Letters A* **2011**, *375*, 2211–2219.
33. Yamano, T. A possible extension of Shannon's information theory. *Entropy* **2001**, *3*, 280–292.
34. Shannon, C.E. A mathematical theory of communication. *The Bell system technical journal* **1948**, *27*, 379–423.
35. Arimoto, S. Computation of random coding exponent functions. *Information Theory, IEEE Transactions on* **1976**, *22*, 665–671.
36. Gallager, R. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory* **1965**, *11*, 3–18.
37. Cover, T.M.; Thomas, J.A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*; Wiley-Interscience, 2006.

38. Fehr, S.; Berens, S. On the conditional Rényi entropy. *Information Theory, IEEE Transactions on* **2014**, *60*, 6801–6810.
39. Sibson, R. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **1969**, *14*, 149–160.
40. Csiszár, I. Generalized cutoff rates and Renyi's information measures. *Information Theory, IEEE Transactions on* **1995**, *41*, 26–34.
41. Arimoto, S. Information measures and capacity of order α for discrete memoryless channels. *Topics in information theory* **1977**.
42. Cai, C.; Verdú, S. Conditional Rényi Divergence Saddlepoint and the Maximization of α-Mutual Information. *Entropy* **2019**, *21*, 969.
43. Hayashi, M. Quantum Wiretap Channel With Non-Uniform Random Number and Its Exponent and Equivocation Rate of Leaked Information. *IEEE Transactions on Information Theory* **2015**, *61*, 5595–5622. doi:10.1109/TIT.2015.2464215.
44. Beigi, S. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics* **2013**, *54*, 122202.
45. Gupta, M.K.; Wilde, M.M. Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity. *Communications in Mathematical Physics* **2015**, *334*, 867–887.
46. Wilde, M.M.; Winter, A.; Yang, D. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Communications in Mathematical Physics* **2014**, *331*, 593–622.
47. Hayashi, M.; Tajima, H. Measurement-based formulation of quantum heat engines. *Physical Review A* **2017**, *95*, 032132.
48. Hayashi, M.; Tomamichel, M. Correlation detection and an operational interpretation of the Rényi mutual information. *Journal of Mathematical Physics* **2016**, *57*, 102201.
49. Tridenski, S.; Zamir, R.; Ingber, A. The Ziv–Zakai–Rényi bound for joint source-channel coding. *IEEE Transactions on Information Theory* **2015**, *61*, 4293–4315.
50. Harremoës, P. Interpretations of Rényi entropies and divergences. *Physica A: Statistical Mechanics and its Applications* **2006**, *365*, 57–62.
51. Jizba, P.; Kleinert, H.; Shefaat, M. Rényi's information transfer between financial time series. *Physica A: Statistical Mechanics and its Applications* **2012**, *391*, 2971 – 2989. doi:http://dx.doi.org/10.1016/j.physa.2011.12.064.
52. Jizba, P.; Arimitsu, T. The world according to Rényi: thermodynamics of multifractal systems. *Annals of Physics* **2004**, *312*, 17 – 59. doi:10.1016/j.aop.2004.01.002.
53. Iwamoto, M.; Shikata, J. Information theoretic security for encryption based on conditional Rényi entropies. International Conference on Information Theoretic Security. Springer, 2013, pp. 103–121.
54. Ilić, V.; Djordjević, I.; Stanković, M. On a general definition of conditional Rényi entropies. Multidisciplinary Digital Publishing Institute Proceedings, 2017, Vol. 2, p. 166.
55. Fano, R.M. *Transmission of Information*; M.I.T. Press, Cambridge, Mass, 1961.
56. Ilic, V.M.; Djordjevic, I.B.; Küeppers, F. On the Daróczy-Tsallis capacities of discrete channels. *Entropy* **2015**, *20*, 2.
57. Yamano, T. Information theory based on nonadditive information content. *Physical Review E* **2001**, *63*, 046105.
58. Tsallis, C.; Gell-Mann, M.; Sato, Y. Asymptotically scale-invariant occupancy of phase space makes the entropy Sq extensive. *Proceedings of the National Academy of Sciences* **2005**, *102*, 15377–15382.
59. Korbel, J.; Hanel, R.; Thurner, S. Classification of complex systems by their sample-space scaling exponents. *New Journal of Physics* **2018**, *20*, 093007.