Article

On Conditional Tsallis Entropy

Andreia Teixeira [1,2,3], **André Souto** [4,5,6,*] and **Luís Antunes** [7]

1 CINTESIS—Centre for Health Technology and Services Research, Faculty of Medicine, University of Porto, 4200-450 Porto, Portugal; andreiasofiat@med.up.pt
2 MEDCIDS—Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, 4200-450 Porto, Portugal
3 ADiT-LAB, Instituto Politécnico de Viana do Castelo, Rua Escola Industrial e Comercial Nun'Álvares, 4900-347 Viana do Castelo, Portugal
4 LASIGE, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
5 Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
6 Instituto de Telecomunicações, Av. Rovisco Pais, n 1, 1049-001 Lisboa, Portugal
7 Computer Science Department, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal; lfa@dcc.fc.up.pt
* Correspondence: ansouto@fc.ul.pt

Abstract: There is no generally accepted definition for conditional Tsallis entropy. The standard definition of (unconditional) Tsallis entropy depends on a parameter α that converges to the Shannon entropy as α approaches 1. In this paper, we describe three proposed definitions of conditional Tsallis entropy suggested in the literature—their properties are studied and their values, as a function of α, are compared. We also consider another natural proposal for conditional Tsallis entropy and compare it with the existing ones. Lastly, we present an online tool to compute the four conditional Tsallis entropies, given the probability distributions and the value of the parameter α.

1. Introduction

Tsallis entropy [1], (The name Tsallis entropy used in this paper, to identify the quantity presented in Equation (3), is not consensual in the community, given that before Tsallis presented it in 1988, and as he himself acknowledges, other authors had already introduced it [2–4].) a generalization of Shannon entropy [5,6], was extensively studied by Constantino Tsallis in 1988, and provides an alternative way of dealing with several characteristics of nonextensive physical systems, given that the information about the intrinsic fluctuations in the physical system can be characterized by the nonextensivity parameter α. It can be applied to many scientific fields, such as physics [7], economics [8], computer science [9,10], and biology [11]. We refer the reader to Reference [12] for a more extensive bibliography on applications of Tsallis entropy. Furthermore, we refer the reader to Reference [13] for a survey on the most significant areas of application of the most usual entropy measures, including Shannon [6], Rényi [14], and Tsallis entropies [1–4].

It is known that, as the parameter α approaches 1, the Tsallis entropy corresponds to the Shannon entropy. Unlike for Shannon entropy, but similar to Rényi entropy (yet another generalization of Shannon entropy developed by Alfréd Rényi in 1961 [14], which also depends on a parameter α and converges to Shannon entropy when α approaches 1), there is no commonly accepted definition for the conditional Tsallis entropy: several versions have been proposed and used in the literature [15,16]. In this work, we revisit the notion of conditional Tsallis entropy by studying some natural and desirable properties in the existing proposals (see for instance References [15,16]): when $\alpha \to 1$, the usual conditional Shannon entropy should be recovered, the conditional Tsallis entropy should not exceed

the unconditional Tsallis entropy, and the conditional Tsallis entropy should have values between 0 and the maximum value of the unconditional version.

The use of entropies in different fields, especially in the field of information theory and its connection to communication, allowed the development of several useful information measures, such as mutual information, symmetry of information, and information distances. See, for example, References [17–19] for some recent work related to the aforementioned information measures.

Depending on the entropy measure used, all of these have been applied in many different areas of knowledge, such as physics [20], information theory [21,22], complexity theory [23–25], security [26–28], biology [29–32], finances [33], and medicine [34–36], among others. The conditional Tsallis entropy, as suggested in Reference [37], can be directly applied to information theory, especially coding theory. Furthermore, since Tsallis entropy can be applied in many areas (see, for example, Reference [12]), the study of conditional Tsallis entropies is quite promising. This paper analyzes several definitions of conditional Tsallis entropy, with the intent of providing the reader with a description of the properties that each approach satisfies.

Continuing from previous works [37,38], we introduce a new natural definition for conditional Tsallis entropy as a possible alternative to the existing ones. Our new proposal does not intend to be the ultimate version of conditional Tsallis entropy, but an alternative to the existing ones, with its own properties that, in settings, such as biomedical applications, might be useful for defining information distances or other significant measurements. None of the known definitions contain all of the desired properties for a conditional version. In particular, the one presented here (as it takes the maximum over the marginal distributions) does not converge to the Shannon entropy when $\alpha \to 1$—it behaves similar to a parameterized entropy, and is akin to the one proposed in Reference [38] as an alternative to Rényi's conditional entropy, another generalization of Shannon entropy.

The paper is organized as follows. In the next section, we present the definitions necessary for the rest of the paper, namely Shannon entropy and Tsallis entropy. In Section 3, we provide several definitions for the conditional Tsallis entropy in both existing literature and our proposal. In Section 4, we establish several results, comparing the definitions presented previously. In Section 5, we explore some features of each variant for the conditional Tsallis entropy. Finally, in Section 6, we present the conclusions and future work.

2. Preliminaries

In the remainder of the paper, we use the standard notation for entropies and for probability distributions according to Reference [5]. For the sake of simplicity of notation, we use the notation log for the logarithm in base 2. We call the reader's attention to the fact that, whenever we say that one entropy converges to another, it is always up to logarithmic factor that depends only on the choice of cardinality of the alphabet.

The Shannon entropy of X is the expectation of the surprise of an occurrence,

$$H(X) = -\sum_x (P(X = x) \log P(X = x)).\tag{1}$$

The conditional Shannon entropy, $H(Y|X)$, is the expectation over x of the entropy of the distribution $P(Y|X = x)$,

$$H(Y|X) = E_{x,y}\left[\log\left(\frac{1}{P(Y = y|X = x)}\right)\right].\tag{2}$$

It is easy to derive the *chain rule* $H(X,Y) = H(X) + H(Y|X)$: to get the average information contained in (X,Y), we may first get the average information contained in X, and add to it the average information of Y, given X.

The Tsallis entropy [1] was firstly introduced in [2,3] and is defined for a random variable X by:

$$T_\alpha(X) = \frac{1}{\alpha - 1}\left(1 - \sum_x P(X = x)^\alpha\right), \text{(for } \alpha > 0, \alpha \neq 1). \tag{3}$$

It is straightforward to show that, when the parameter α converges to 1, the value of the entropy converges to the Shannon entropy.

3. Conditional Tsallis Entropy: Four Definitions

We consider three definitions for conditional Tsallis entropy that already exist in the literature and introduce a new proposal. All definitions consider a positive parameter α.

Definition 1. *Let* $Z = (X, Y)$ *be a random vector. One can define the following variants of conditional Tsallis entropy:*

1. *Definition of* $T_\alpha(Y|X)$ *from Reference [15]*

$$T_\alpha(Y|X) = \sum_x P(X = x)^\alpha T_\alpha(Y|x) \tag{4}$$

$$= \frac{1}{\alpha - 1}\sum_x P(X = x)^\alpha\left(1 - \sum_y P(Y = y|X = x)^\alpha\right) \tag{5}$$

One can easily verify that $T_\alpha(X, Y) = T_\alpha(Y|X) + T_\alpha(X)$ *and, therefore, it satisfies the chain rule.*

2. *Definition of* $S_\alpha(Y|X)$ *from [16] (Definition 2.8)*

$$S_\alpha(Y|X) = \sum_x P(X = x)T_\alpha(Y|X = x) \tag{6}$$

$$= \sum_x P(X = x)\frac{1}{\alpha - 1}\left(1 - \sum_y P(Y = y|X = x)^\alpha\right) \tag{7}$$

$$= \frac{1}{\alpha - 1}\sum_x P(X = x)\left(1 - \sum_y P(Y = y|X = x)^\alpha\right). \tag{8}$$

3. *Definition of* $S'_\alpha(Y|X)$ *from [16] (Definition 2.10)*

$$S'_\alpha(Y|X) = \frac{1}{\alpha - 1}\left(\frac{1 - \sum_{x,y} P(X = x, Y = y)^\alpha}{\sum_x P(X = x)^\alpha}\right). \tag{9}$$

The first definition presented proposes that the conditional Tsallis entropy should be weighed by the probability of sampling $X = x$ with parameter α, while the second one proposes that one uniformly weighs only the probability of sampling $X = x$. Therefore, notice that for the first definition presented, the value of α largely affects the value of the conditional Tsallis entropy. The idea for the third proposal is to distribute evenly the influence of the parameter α by the entire joint distribution.

Next, we present another possible definition of the conditional Tsallis entropy. This definition is based on Definition III.6 of [38] and captures the intuitive notion of defining the conditional entropy, by taking the maximum over all possible marginal distributions. Note that this definition is analogous to an existing one for the Rényi entropy; however, as we will show later, this proposal does not satisfy some of the expected basic properties.

Definition 2 (Definition of $T'_\alpha(Y|X)$).

$$T'_\alpha(Y|X) = \frac{1}{\alpha - 1} \max_x \left(1 - \sum_y \mathrm{P}(Y = y | X = x)^\alpha \right). \tag{10}$$

We opted to use different notations for the variants of the conditional Tsallis entropy in the last definition, to better distinguish them in the rest of the paper. In particular, we follow the same approach as in Reference [38].

The following expressions will be useful later.

Theorem 1. *Let $Z = (X, Y)$ be a random vector. The following identities are true:*

$$T'_\alpha(Y|X) = \max_x T_\alpha(Y|X = x) \quad \text{(for } \alpha > 1) \tag{11}$$

$$T'_\alpha(Y|X) = \min_x T_\alpha(Y|X = x) \quad \text{(for } \alpha < 1). \tag{12}$$

4. Comparison of the Definitions

We now compare the above four definitions of the conditional Tsallis entropy by comparing whether or not the definition satisfies some common properties of an entropy measure. In the next theorem, we report two simple facts with straightforward proofs. We leave the details for the interested reader to check.

Theorem 2. *For any fixed joint probability distribution $\mathrm{P}(X, Y)$,*

(i) $T_\alpha(Y|X)$, $S_\alpha(Y|X)$ and $S'_\alpha(Y|X)$, *as functions of α, are continuous and differentiable;*
(ii) $T'_\alpha(Y|X)$, *as a function of α, is continuous for all $\alpha \neq 1$.*

The following results provide the possible comparisons (in terms of values) between the proposed definitions. For the sake of organization, we split the comparison by types of entropy.

First we compare $T_\alpha(Y|X)$ with $S_\alpha(Y|X)$.

Theorem 3. *For all joint probability distributions $\mathrm{P}(X, Y)$ and for every $\alpha > 0$,*

$$\text{if } \alpha < 1: \quad S_\alpha(Y|X) \leq T_\alpha(Y|X) \tag{13}$$

$$\text{if } \alpha = 1: \quad S_\alpha(Y|X) = T_\alpha(Y|X) = H(Y|X) \tag{14}$$

$$\text{if } \alpha > 1: \quad S_\alpha(Y|X) \geq T_\alpha(Y|X). \tag{15}$$

Proof. Consider first the case $\alpha < 1$. In this case, we have that $P(X = x)^\alpha \geq P(X = x)$. Thus,

$$
\begin{aligned}
P(X = x)^\alpha \times T_\alpha(Y|X = x) &\geq P(X = x) \times T_\alpha(Y|X = x) \\
\Leftrightarrow \sum_x P(X = x)^\alpha \times T_\alpha(Y|X = x) &\geq \sum_x P(X = x) \times T_\alpha(Y|X = x) \\
\Leftrightarrow T_\alpha(Y|X) &\geq S_\alpha(Y|X).
\end{aligned}
$$

For the case $\alpha = 1$, see the proof of Theorem 8.
The case $\alpha > 1$ is similar to the previous one, but this time, the conclusion follows, since for $\alpha > 1$, $P(X = x)^\alpha \leq P(X = x)$.   □

In the next theorem we provide the comparison between $T'_\alpha(Y|X)$ and $S_\alpha(Y|X)$.

Theorem 4. *For all joint probability distributions $\mathrm{P}(X, Y)$ and for every $\alpha > 0$,*

$$\text{if } \alpha \leq 1: \quad T'_\alpha(Y|X) \leq S_\alpha(Y|X) \tag{16}$$

$$\text{if } \alpha > 1: \quad T'_\alpha(Y|X) \geq S_\alpha(Y|X). \tag{17}$$

Proof. Consider first the case $\alpha < 1$. In this case, we have that $T'_\alpha(Y|X) = \min_x T_\alpha(Y|X = x)$. So,

$$S_\alpha(Y|X) = \sum_x \left(P(X = x) \cdot T_\alpha(Y|X = x) \right) \tag{18}$$

$$\geq \sum_x \left(P(X = x) \cdot \left[\min_x T_\alpha(Y|X = x) \right] \right) \tag{19}$$

$$= \left(\min_x T_\alpha(Y|X = x) \right) \cdot \sum_x P(X = x) \tag{20}$$

$$= \min_x T_\alpha(Y|X = x) \tag{21}$$

$$= T'_\alpha(Y|X). \tag{22}$$

The proof of the case $\alpha > 1$ is similar to the previous one but this time, the conclusion follows from the fact that, for $\alpha > 1$, $T'_\alpha(Y|X) = \max_x T_\alpha(Y|X = x)$. □

As a consequence of the two previous results and the definitions, we can derive the relation between T'_α and T_α.

Corollary 1. *For all joint probability distributions* $P(X, Y)$ *and for every* $\alpha > 0$,

$$\text{if } \alpha \leq 1 : \quad T'_\alpha(Y|X) \leq T_\alpha(Y|X) \tag{23}$$
$$\text{if } \alpha > 1 : \quad T'_\alpha(Y|X) \geq T_\alpha(Y|X). \tag{24}$$

The proof follows directly from Theorems 3 and 4. Now, we derive the relation between $S'_\alpha(Y|X)$ and $T'_\alpha(Y|X)$.

Theorem 5. *For all joint probability distributions* $P(X, Y)$ *and for every* $\alpha > 0$,

$$\text{if } \alpha \leq 1 : \quad S'_\alpha(Y|X) \geq T'_\alpha(Y|X) \tag{25}$$
$$\text{if } \alpha > 1 : \quad S'_\alpha(Y|X) \leq T'_\alpha(Y|X). \tag{26}$$

Proof. Consider first the case $\alpha < 1$. Proving that $S'_\alpha(Y|X) \geq T'_\alpha(Y|X)$, by definition, is the same to prove:

$$\frac{\left(1 - \dfrac{\sum\limits_{x,y} P(X = x, Y = y)^\alpha}{\sum\limits_x P(x)^\alpha} \right)}{\alpha - 1} \geq \frac{\max\limits_x \left(1 - \sum\limits_y P(y|x)^\alpha \right)}{\alpha - 1}. \tag{27}$$

As $\alpha < 1$, we have that $\dfrac{1}{\alpha - 1} < 0$. Thus, proving Equation (27) is the same, proves that:

$$1 - \frac{\sum_{x,y} P(X = x, Y = y)^{\alpha}}{\sum_{x} P(X = x)^{\alpha}} \leq \max_{x} \left(1 - \sum_{y} P(Y = y | X = x)^{\alpha} \right)$$

$$\Leftrightarrow \frac{\sum_{x,y} P(X = x, Y = y)^{\alpha}}{\sum_{x} P(X = x)^{\alpha}} \geq \min_{x} \sum_{y} P(Y = y | X = x)^{\alpha}$$

$$\Leftrightarrow \sum_{x,y} P(X = x, Y = y)^{\alpha} \geq \left(\sum_{x} P(X = x)^{\alpha} \right) \times \min_{x} \left(\sum_{y} P(Y = y | X = x)^{\alpha} \right).$$

Now, the result follows by observing that the last inequality is true, since, for $\alpha < 1$ and for every x, we have that

$$\min_{x} \left(\sum_{y} P(Y = y | X = x)^{\alpha} \right) \leq \sum_{y} P(Y = y, X = x)^{\alpha}.$$

The case $\alpha > 1$ is proved in a similar manner. $\square$

Now, we derive the relation between $T_{\alpha}(Y|X)$ and $S'_{\alpha}(Y|X)$.

Theorem 6. *For all joint probability distributions* $P(X, Y)$ *and for every* $\alpha > 0$,

$$\text{if } \alpha \leq 1 : \quad T_{\alpha}(Y|X) \geq S'_{\alpha}(Y|X) \tag{28}$$

$$\text{if } \alpha > 1 : \quad T_{\alpha}(Y|X) \leq S'_{\alpha}(Y|X). \tag{29}$$

Proof. Consider first the case $\alpha < 1$. Thus,

$$T_{\alpha}(Y|X) \geq S'_{\alpha}(Y|X)$$

$$\Leftrightarrow \frac{1}{\alpha - 1} \sum_{x} \left(P(X = x)^{\alpha} \left(1 - \sum_{y} P(Y = y | X = x)^{\alpha} \right) \right) \geq \frac{1}{\alpha - 1} \left(1 - \frac{\sum_{x,y} P(X = x, Y = y)^{\alpha}}{\sum_{x} P(X = x)^{\alpha}} \right)$$

$$\Leftrightarrow \sum_{x} \left(P(X = x)^{\alpha} \left(1 - \sum_{y} P(Y = y | X = x)^{\alpha} \right) \right) \leq 1 - \frac{\sum_{x,y} P(X = x, Y = y)^{\alpha}}{\sum_{x} P(X = x)^{\alpha}}$$

$$\Leftrightarrow \sum_{x} P(X = x)^{\alpha} - \sum_{x} \left(P(X = x)^{\alpha} \sum_{y} \frac{P(X = x, Y = y)^{\alpha}}{P(X = x)^{\alpha}} \right) \leq 1 - \frac{\sum_{x,y} P(X = x, Y = y)^{\alpha}}{\sum_{x} P(X = x)^{\alpha}}$$

$$\Leftrightarrow \sum_{x} P(X = x)^{\alpha} - \sum_{x,y} P(X = x, Y = y)^{\alpha} \leq \frac{\sum_{x} P(X = x)^{\alpha} - \sum_{x,y} P(X = x, Y = y)^{\alpha}}{\sum_{x} P(X = x)^{\alpha}}.$$

The result follows by observing that the last inequality is true, since for $\alpha < 1$, we have that:

$$P(X = x)^{\alpha} > P(X = x) \tag{30}$$

and consequently,

$$\sum_{x} P(X = x)^{\alpha} > 1. \tag{31}$$

The proof of the case $\alpha > 1$ is similar to the previous one. □

Finally, we show that the values of S_α and S'_α are incomparable in the sense that there are probability distributions for which S_α is greater than S'_α and there are probability distributions for which S'_α is greater than S_α.

Theorem 7. *The values of $S_\alpha(Y|X)$ and of $S'_\alpha(Y|X)$ are incomparable, i.e., for each $n \geq 2$ and $\alpha \neq 1$*

$$\exists P(X,Y): \quad S_\alpha(Y|X) < S'_\alpha(Y|X) \tag{32}$$
$$\exists P(X,Y): \quad S_\alpha(Y|X) > S'_\alpha(Y|X). \tag{33}$$

Proof. For Statement (32) and $\alpha < 1$, consider the following joint probability distribution:

$X\backslash Y$	1	2
1	0.0625	0.0625
2	0.0125	0.8625

$$S_{0.25}(Y|X) \quad \approx \quad 0.513 \tag{34}$$
$$S'_{0.25}(Y|X) \quad \approx \quad 0.629 \tag{35}$$

For Statement (32) and $\alpha > 1$, consider the following joint probability distribution:

$X\backslash Y$	1	2
1	0.1125	0.0125
2	0.4375	0.4375

$$S_{2.5}(Y|X) \quad \approx \quad 0.396 \tag{36}$$
$$S'_{2.5}(Y|X) \quad \approx \quad 0.429 \tag{37}$$

For Statement (33) and $\alpha < 1$, consider the following joint probability distribution:

$X\backslash Y$	1	2
1	0.125	0
2	0.5	0.375

$$S_{0.25}(Y|X) \quad \approx \quad 0.792 \tag{38}$$
$$S'_{0.25}(Y|X) \quad \approx \quad 0.560 \tag{39}$$

Finally, for Statement (33) and $\alpha > 1$, consider the following joint probability distribution:

$X\backslash Y$	1	2
1	0.0625	0.0625
2	0.0125	0.8625

$$S_{1.25}(Y|X) \quad \approx \quad 0.125 \tag{40}$$
$$S'_{1.25}(Y|X) \quad \approx \quad 0.099 \tag{41}$$

□

5. Properties of the Conditional Tsallis Entropies

In this section, we investigate some properties of the proposals considered. In particular, we show that there are probability distributions and $\alpha \neq 1$ for which the conditional Tsallis entropies are bigger than the unconditional Tsallis entropy.

Theorem 8. *For any fixed joint probability distribution* $P(X, Y)$,

$$\lim_{\alpha \to 1} T_\alpha(Y|X) = H(Y|X) \tag{42}$$

$$\lim_{\alpha \to 1} S_\alpha(Y|X) = H(Y|X) \tag{43}$$

$$\lim_{\alpha \to 1} S'_\alpha(Y|X) = H(Y|X) \tag{44}$$

where $H(Y|X)$ *is the conditional Shannon entropy. In general, it is not true that* $\lim_{\alpha \to 1} T'_\alpha(Y|X) = H(Y|X)$.

Proof. The second equation is easy to derive directly from the definition of conditional probability and from Equation (2). Furthermore, using Equation (6) we can also easily obtain (using the previous derivation) that Equation (42) is also true.

The third equation was proven in Reference [16].

Now, it is only left to prove the last statement of the theorem, i.e., in general

$$\lim_{\alpha \to 1} T'_\alpha(Y|X) \neq H(Y|X). \tag{45}$$

From Equations (6) and (11) it is easy to check that $T_\alpha(Y|X)$ is the *expectation* over x of $T_\alpha(Y|x)$, while $T'_\alpha(Y|X)$ is the *maximum* over x of $T_\alpha(Y|x)$.

The function $T_\alpha(Y|x)$ depends on the conditional probabilities $P(Y = y|X = x)$. Therefore, there are joint probability distributions $P(X = x, Y = y)$, such that:

$$\lim_{\alpha \to 1} T'_\alpha(Y|X) \neq \lim_{\alpha \to 1} T_\alpha(Y|X) = H(Y|X). \tag{46}$$

□

Contrary to the Shannon entropy, the value of any conditional Tsallis entropy may exceed the corresponding unconditional Tsallis entropy for all proposals.

Theorem 9. *There are probability distributions* $P(X, Y)$ *and values of* α, *such that:*

$$T_\alpha(Y|X) > T_\alpha(Y) \tag{47}$$
$$S_\alpha(Y|X) > T_\alpha(Y) \tag{48}$$
$$S'_\alpha(Y|X) > T_\alpha(Y) \tag{49}$$
$$T'_\alpha(Y|X) > T_\alpha(Y). \tag{50}$$

Proof. Consider the following joint probability distribution:

$X = x \backslash Y = y$	1	2
1	0.45	0.45
2	0.1	0.0

For this distribution we have:

$$
\begin{align}
T_{0.5}(Y) &\approx 0.824 \tag{51} \\
T_{0.5}(Y|X) &\approx 1.047 \tag{52} \\
S_{0.5}(Y|X) &\approx 0.828 \tag{53} \\
T_3(Y) &\approx 0.371 \tag{54} \\
S_3'(Y|X) &\approx 0.374 \tag{55} \\
T_3'(Y|X) &\approx 0.375 \tag{56}
\end{align}
$$

$\square$

Bounds on Conditional Tsallis Entropy

As mentioned in the Introduction, one of the properties of the (conditional) Shannon entropy for discrete variables is to be bounded by the number of elements of the support of the distribution. Furthermore, it is well known that the unconditional Tsallis entropy is always between 0 and $\dfrac{m^{1-\alpha}}{1-\alpha}$, where m is the number of elements in the support of the distribution. In this subsection, we derive bounds for the conditional Tsallis entropies based on the number of elements in the support of each distribution.

Theorem 10. *Let $Z = (X,Y)$ be any joint random vector defined over sets of size m each. Then,*

$$
0 \le S_\alpha(Y|X) \le \frac{m^{1-\alpha}}{1-\alpha} \tag{57}
$$

$$
0 \le T_\alpha'(Y|X) \le \frac{m^{1-\alpha}}{1-\alpha}. \tag{58}
$$

Moreover all of these lower and upper bounds may be reached by suitable probability distributions P(X,Y).

Proof. The Inequalities (57) follow from the fact that $S_\alpha(Y|X)$ is the expectation of the unconditional Tsallis entropy.

For Inequalities (58), recall that Equation (10) can be written, for $\alpha < 1$, as Equation (12). Note that, for all x, the values $T_\alpha(Y|X = x)$ are the (unconditional) Tsallis entropies of the marginal distribution, and are all defined in a set of cardinality m.

So, by definition of T_α', for some particular x, we have $T_\alpha'(Y|X) = T_\alpha(Y|X = x)$. The case $\alpha > 1$ is similar. So, independently of α, for every probability distributions P(X) and P(Y) defined over set with m elements, we have $0 \le T_\alpha'(Y|X) \le \dfrac{m^{1-\alpha}}{1-\alpha}$, since the same bound applies for the unconditional version or any its marginal distributions. $\square$

Theorem 11. *Let $Z = (X,Y)$ be any joint random vector defined over sets of size m each. Then,*

$$
\text{if } \alpha > 1: \quad 0 \le T_\alpha(Y|X) \le \frac{m^{1-\alpha}}{1-\alpha}. \tag{59}
$$

For $\alpha < 1$, in general, the inequality does not hold.

Proof. Consider first the case $\alpha > 1$. The result follows directly from Inequalities (15) and (57).

In order to prove that the inequality does not hold for all $\alpha < 1$, consider $\alpha = 0.1$ and the following joint probability distribution:

$X = x \backslash Y = y$	1	2	3
1	1/9	1/9	0
2	1/9	1/9	1/9
3	0	1/9	1/3

Notice that $\dfrac{m^{1-\alpha}}{1-\alpha} \approx 2.987$ and $T_{0.1}(Y|X) \approx 3.371$. For any other $\alpha < 1$, one can construct similarly a joint probability distribution for which the inequality is also violated. $\square$

Theorem 12. *Let $Z = (X, Y)$ be any joint random vector defined over sets of size m each. Then,*

$$\text{if } \alpha > 1 : \quad 0 \leq S'_\alpha(Y|X) \leq \frac{m^{1-\alpha}}{1-\alpha}. \tag{60}$$

Proof. The result follows directly from the Inequalities (26) and (58). $\square$

We conjecture that the above theorem also holds for $\alpha < 1$. For example, the inequality is true for all uniform probability distribution over n variables.

We now show that, for any fixed joint probability distribution $P(X, Y)$, three of the forms of conditional Tsallis entropy studied in this paper are non-increasing functions of α. First, we state a simple theorem.

Lemma 1. *If $f_1(x), \ldots, f_m(x)$ are non-increasing real functions, then the function $\max_i (f_i(x))$ is also a non-increasing function.*

Theorem 13. *For every probability distribution $P(X, Y)$,*

1. *$T_\alpha(Y|X)$ is a non-increasing function of α.*
2. *$S_\alpha(Y|X)$ is a non-increasing function of α.*
3. *$T'_\alpha(Y|X)$ is a non-increasing function of α.*

Proof.

1. First consider the case $\alpha > 1$, and consider the function $dT_\alpha(Y|X)$, the derivative of the function $T_\alpha(Y|X)$ in order to α:

$$\frac{dT_\alpha(Y|X)}{d\alpha} = \frac{-1 + \sum\limits_x P(X = x)^\alpha}{(\alpha - 1)^2} - \frac{\sum\limits_x \left(\alpha P(Y = y | X = x)^{\alpha-1} \log \alpha \right)}{\alpha - 1}.$$

It is easy to see that, since $\alpha > 1$, $\frac{dT_\alpha(X)}{d\alpha} < 0$. Therefore, the function $T_\alpha(Y|X)$ is a non-increasing function of α.

Consider now the case $\alpha < 1$ and assume that α, α' are such that $\alpha < \alpha' < 1$. In order to prove that $T_\alpha(Y|X)$ is non-increasing we have to show that $T_\alpha(X) \geq T'_\alpha(X)$, i.e.,:

$$1 - \frac{1 - \sum_x P(X = x)^\alpha}{\alpha - 1} \geq 1 - \frac{1 - \sum_x P(X = x)^{\alpha'}}{\alpha' - 1}$$

$$\Leftrightarrow \frac{-1 + \sum_x P(X = x)^\alpha}{1 - \alpha} \geq \frac{-1 + \sum_x P(X = x)^{\alpha'}}{1 - \alpha'}$$

$$\Leftrightarrow \frac{1 - \sum_x P(X = x)^\alpha}{1 - \alpha} \leq \frac{1 - \sum_x P(X = x)^{\alpha'}}{1 - \alpha'}$$

Notice that, since $\alpha < \alpha' < 1$, then $\frac{1}{1 - \alpha} < \frac{1}{1 - \alpha'}$ and, therefore, $1 - \sum_x P(X = x)^\alpha \leq 1 - \sum_x P(X = x)^{\alpha'}$. So, the last inequality is true.

2. This part of the result follows from the fact that $S_\alpha(Y|X)$ is the expectation of unconditional Tsallis entropies; see Equation (6).

3. Suppose that $\alpha > 1$. The proof is a direct consequence of Equation (11) and Lemma 1. The case $\alpha < 1$ can be proven in a similar way. □

It is easy to show that S' does not fulfill the property of the last theorem.

Theorem 14. *There exists probability distributions (X, Y) and $\alpha < \alpha'$ for which $S'_\alpha(Y|X) \leq S'_{\alpha'}(Y|X)$.*

Proof. Consider the following joint probability distribution:

$X = x \backslash Y = y$	1	2
1	0.45	0.45
2	0.1	0.0

We have:

$$S'_{0.2}(Y|X) \approx 0.563$$
$$S'_{0.5}(Y|X) \approx 0.621.$$

□

We developed a small application that, given two probability distributions, computes the values of all conditional Tsallis entropies considered in the paper. The application is self-contained and its use is extremely simple. There are two use case examples that the reader can use in order to try the calculator. The interested reader can find it in the following link: http://gloss.di.fc.ul.pt/tryit/Tsallis (accessed on 28 October 2021).

6. Conclusions

In this paper, we studied the definitions for the conditional Tsallis entropy existing in the literature. We also considered a possible alternative definition for it. This new proposal is a natural approach to consider as a possible definition. It defines the conditional value as the maximum value of all marginal distributions. Due to this fact, and similar to what happens with the Rényi entropy, this definition was also analyzed, although it was never considered in the literature before. The relationships between the four definitions, described in this work, are summarized in Figure 1.

For $\alpha = 1$

$$S'_\alpha(Y|X) = S_\alpha(Y|X) = T_\alpha(Y|X) = H(Y|X)$$
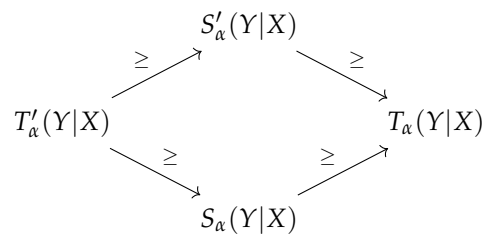
For $\alpha < 1$
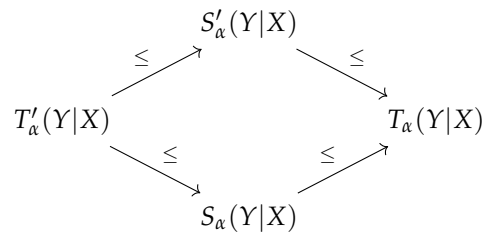
For $\alpha > 1$

Figure 1. Summary of the relations between the several proposals for the definition of conditional Tsallis entropy.

As we understand, it would be expectable that a proposal for conditional Tsallis entropy would satisfy the following properties:

1. Chain Rule;
2. Convergence to Shannon entropy as the parameter α tended to 1;
3. Its value would be between 0 and the upper bound of the unconditional version.

In Table 1, we summarize the properties that the four proposals have (we also added the property of being a non-increasing function with α). To conclude, we can say that none of the proposals fulfill all of the properties. The definition $T_\alpha(Y|X)$ is the candidate that fulfills more properties.

Table 1. Summary of the proved properties of all proposed conditional entropies. The question mark indicates that the property is not known to be fulfilled.

$f(Y\|X)$	$T_\alpha(Y\|X)$	$S_\alpha(Y\|X)$	$S'_\alpha(Y\|X)$	$T'_\alpha(Y\|X)$
Chain Rule	yes	no	no	no
$\lim\limits_{\alpha \to 1} f(Y\|X) = H(Y\|X)$	yes	yes	yes	no
$0 \le f(Y\|X) \le \dfrac{\|Y\|^{1-\alpha}}{1-\alpha}$ and $\alpha > 1$	yes	yes	yes	yes
$0 \le f(Y\|X) \le \dfrac{\|Y\|^{1-\alpha}}{1-\alpha}$ and $\alpha < 1$	no	yes	?	yes
f is non-increasing with α	yes	yes	no	yes

For future work, since all definitions focus on possible different aspects of the entropy, it would be important to consider a deeper study in this area and its possible applications, aiming to develop a theory that would emphasize the best proposal for each area, or eventually present an ultimate version for the conditional Tsallis entropy that would satisfy all of the desirable properties.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
2. Daróczy, Z. Generalized information functions. *Inf. Control* **1970**, *16*, 36–51. [CrossRef]
3. Havrda, J.; Charvat, F. Quantification method of classification processes. concept of structural α-entropy. *IEEE Trans. Inf. Theory* **1967**, *3*, 30–35.
4. Wehrl, A. General properties of entropy. *Rev. Mod. Phys.* 1978, 50, 221–260. [CrossRef]
5. Cover, T.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
6. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. 623–656. [CrossRef]
7. Tsallis, C. The Nonadditive Entropy Sq and Its Applications in Physics and Elsewhere: Some Remarks. *Entropy* **2011**, *13*, 1765–1804. [CrossRef]
8. Borland, R.O.L.; Tsallis, C. Distributions of high-frequency stock-market observables. In *Nonextensive Entropy—Interdisciplinary Applications*; Gell-Mann, M., Tsallis, C., Eds.; Oxford University Press: New York, NY, USA, 2004.
9. Ibrahim, R.W.; Darus, M. Analytic Study of Complex Fractional Tsallis' Entropy with Applications in CNNs. *Entropy* **2018**, *20*, 722. [CrossRef]
10. Mohanalin, B.; Kalra, P.K.; Kumar, N. A novel automatic microcalcification detection technique using Tsallis entropy and a type II fuzzy index *Comput. Math. Appl.* **2010**, *60*, 2426–2432.
11. Tamarit, F.A.; Cannas, S.A.; Tsallis, C. Sensitivity to initial conditions in the Bak-Sneppen model of biological evolution. *Eur. Phys. J. B* **1998**, *1*, 545–548. [CrossRef]
12. Group of Statistical Physics. Available online: http://tsallis.cat.cbpf.br/biblio.htm (accessed on 8 November 2018).
13. Ribeiro, M.; Henriques, T.; Castro, L.; Souto, A.; Antunes, L.; Costa-Santos, C.; Teixeira, A. The Entropy Universe. *Entropy* **2021**, *23*, 222. [CrossRef]
14. Rényi, A. On measures of information and entropy. *Berkeley Symp. Math. Statist. Prob.* **1961**, *1*, 547–561.
15. Furuichi, S. Information theoretical properties of Tsallis entropies. *J. Math. Phys.* **2006**, *47*, 023302. [CrossRef]
16. Manije, S.; Gholamreza, M.; Mohammad, A. Conditional Tsallis Entropy. *Cyb. Inf. Technol.* **2013**, *13*, 37–42. [CrossRef]
17. Heinrich, F.; Ramzan, F.; Rajavel, F.A.; Schmitt, A.O.; Gültas, M. MIDESP: Mutual Information-Based Detection of Epistatic SNP Pairs for Qualitative and Quantitative Phenotypes. *Biology* **2021**, *10*, 921. [CrossRef] [PubMed]
18. Oggier, F.; Datta, A.A. Renyi entropy driven hierarchical graph clustering. *PeerJ Comput. Sci.* **2021**, *7*, e366. [CrossRef]
19. Tao, M.; Wang, S.; Chen, H.; Wang, X. Information space of multi-sensor networks. *Inf. Sci.* **2021**, *565*, 128–245. [CrossRef]
20. Jozsa, R.; Schlienz, J. Distinguishability of states and von Neumann entropy. *Phys. Rev. A* **2000**, *62*, 012301. [CrossRef]
21. Hassani, H.; Unger, S.; Entezarian, M. Information content measurement of esg factors via entropy and its impact on society and security. *Information* **2021**, *12*, 391. [CrossRef]
22. Shannon, C.E. Communication theory of secrecy systems. *Bell Syst. Tech. J.* **1949**, *28*, 656–715. [CrossRef]
23. Bhotto, M.Z.A.; Antoniou, A. A new normalized minimum-error entropy algorithm with reduced computational complexity. In Proceedings of the 2009 IEEE International Symposium on Circuits and Systems, Taipei, Taiwan, 24–27 May 2009; pp. 2561–2564. [CrossRef]
24. Teixeira, A.; Matos, A.; Souto, A.; Antunes, L. Entropy measures vs. Kolmogorov complexity. *Entropy* **2011**, *13*, 595–611. [CrossRef]
25. Teixeira, A.; Souto, A.; Matos, A.; Antunes, L. Entropy measures vs. algorithmic information. In Proceedings of the 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, 13–18 June 2010; pp. 1413–1417. [CrossRef]

26. Edgar, T.; Manz, D. Chapter 2-Science and Cyber Security. In *Research Methods for Cyber Security*; Syngress: Amsterdam, The Netherlands, 2017; pp. 33–62.

27. Huang, L.; Shen, Y.; Zhang, G.; Luo, H. Information system security risk assessment based on multidimensional cloud model and the entropy theory. In Proceedings of the 2015 IEEE 5th International Conference on Electronics Information and Emergency Communication, Beijing, China, 14–16 May 2015; pp. 11–15.

28. Lu, R.; Shen, H.; Feng, Z.; Li, H.; Zhao, W.; Li, X. HTDet: A clustering method using information entropy for hardware Trojan detection. *Tsinghua Sci. Technol.* **2021**, *26*, 48–61. [CrossRef]

29. Firman, T.; Balázsi, G.; Ghosh, K. Building Predictive Models of Genetic Circuits Using the Principle of Maximum Caliber. *Biophys J.* **2017**, *113*, 2121–2130. [CrossRef]

30. Jost L. Entropy and diversity. *Oikos* **2006**, *113*, 363–375. [CrossRef]

31. Roach TNF. Use and Abuse of Entropy in Biology: A Case for Caliber. *Entropy* **2020**, *22*, 1335. [CrossRef] [PubMed]

32. Simpson E. Measurement of diversity. *Nature* **1949**, *163*, 688. [CrossRef]

33. Yin, Y.; Shang, P. Weighted permutation entropy based on different symbolic approaches for financial time series. *Phys. A Stat. Mech. Its Appl.* **2016**, *443*, 137–148. [CrossRef]

34. Castiglioni, P.; Parati, G.; Faini, A. Information-Domain Analysis of Cardiovascular Complexity: Night and Day Modulations of Entropy and the Effects of Hypertension. *Entropy* **2019**, *21*, 550. [CrossRef] [PubMed]

35. Polizzotto, N.R.; Takahashi, T.; Walker, C.P.; Cho, R.Y. Wide Range Multiscale Entropy Changes through Development. *Entropy* **2016**, *18*, 12. [CrossRef]

36. Prabhu, K.P.; Martis, R.J. Diagnosis of Schizophrenia using Kolmogorov Complexity and Sample Entropy. In Proceedings of the 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2–4 July 2020; pp. 1–4. [CrossRef]

37. Fehr, S.; Berens, S. On the Conditional Rényi Entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 6801–6810. [CrossRef]

38. Teixeira, A.; Matos, A.; Antunes, L. Conditional Rényi Entropies. *IEEE Trans. Inf. Theory* **2012**, *58*, 4273–4277. [CrossRef]