FURTHER RESULTS ON GENERALIZED CONDITIONAL ENTROPIES

ALEXEY E. $RASTEGIN^1$

Abstract. We further examine some properties of the conditional Rényi and Tsallis–Havrda–Charvát (THC) entropies. Such properties are interesting from the viewpoint of applications in studying protocols of quantum information science and foundations of quantum mechanics. In particular, we consider properties of the conditional Rényi and THC entropies with respect to conditioning on more. We also exemplify that the desired property can be violated with the conditional min-entropy. Applications of such results to the TCH entropy rate are considered. Connections between generalized conditional entropies and error probability are examined. Several relations between various conditional entropies are obtained. It is shown that such relations can be used for bounding the conditional Rényi and TCH entropies.

Mathematics Subject Classification. 94A17, 62B10, 39B62.

1. INTRODUCTION

The concept of entropy is fundamental in both information theory and statistical physics. This very fruitful idea is widely used in many theoretical and applied disciplines. The traditional entropy is known as the Shannon entropy in information theory and as the Gibbs entropy in statistical physics. Other entropic forms have been shown to be useful. The Rényi entropy [39] and the Tsallis–Havrda–Charvát entropy [24, 43] are important one-parameter extensions of the Shannon entropy. These measures are used in global thresholding approach to image processing [41].

Keywords and phrases. Rényi entropy, Tsallis–Havrda–Charvát entropy, entropy rate, index of coincidence, error probability, Fano inequality.

¹ Department of Theoretical Physics, Irkutsk State University, Gagarin Bv. 20, 664003 Irkutsk, Russia. alexrastegin@mail.ru

Applications of the Rényi entropy in quantum physics are discussed in [28]. The Tsallis–Havrda–Charvát entropy is extensively adopted in nonextensive statistical mechanics [20]. Both the mentioned entropies have been used in studying quantum entanglement and related topics [4]. An axiomatic approach to generalized entropies is reviewed in [11].

Several entropic functions are widely used in analyzing problems of theoretical informatics. One of such quantities is known as the conditional entropy [12,17]. As discussed in [26], studies of communication systems concern the interplay between the conditional entropy and error probability including the Fano inequality [17]. This issue is also related to the notion of relative entropy, or the Kullback–Leibler divergence [30]. The conditional entropy plays a key role in applying information-theoretic ideas to combinatorial problems [34]. This issue is closely related to the famous probabilistic method [1]. Generalized varieties of the conditional entropy are also of interest. In the literature, various conditional forms of the Rényi entropy [16, 22, 29] and the Tsallis–Havrda–Charvát entropy [13, 19, 35] have been considered. Many fruitful generalizations of the Kullback–Leibler divergence are covered by the family of Csiszár's f-divergences [10]. Quantum f-divergences and their role in quantum information theory are reviewed in [25].

In last years, quantum information science has shown advances in both theory [2, 15] and experiment [21]. Studies in quantum information processing inspire renewed interest to foundations of quantum theory. In particular, conceptual questions were considered in information-theoretic terms. The authors of [5] proposed entropic version of Bell's theorem. This approach has been developed for other tests related to non-locality and contextuality [8]. Such topics stimulated studies of the so-called marginal scenarios [9]. Recently, noise-disturbance trade-off relations have been formulated within the entropic approach [6]. Advantages of information-theoretic formulations are discussed in [6, 8, 37, 38]. As was shown in [38], the use of generalized entropies give new possibilities in analyzing data of tests for quantum non-locality and contextuality. To apply generalized entropies in the mentioned topics, we should previously establish certain properties. Properties of such a kind are usually not treated in typical questions of statistical physics.

The aim of the present work is to examine some properties of generalized conditional entropies. Considered properties are assumed to be used in theoretical studying protocols of quantum information science and foundations of quantum mechanics. Obtained results may also be useful in other topics including analysis of communicating systems in computer science, combinatorial problems in discrete mathematics, and characterization of multipartite systems in physics. The paper is organized as follows. In Section 2, the notation and definitions are introduced. In Section 3, we consider entropic properties related to conditioning on more and additivity. Applications to entropy rate are discussed and exemplified. In Section 4, we examine relations between generalized conditional entropies and error probability. In particular, we derive improved lower bounds in the case of TCH entropies. Relations between various versions of generalized conditional entropies are also of interest. Some of such relations are obtained in Section 5. In Section 6, we conclude the paper with a summary of results.

2. Conditional entropies and their varieties

In this section, we recall definitions of the Tsallis–Havrda–Charvát and Rényi entropies including their conditional forms. Let discrete random variable X take values on a finite set Ω_X of cardinality $\#\Omega_X$. The Tsallis–Havrda–Charvát entropy of degree $\alpha > 0 \neq 1$ is defined by [43]

$$H_{\alpha}(X) := \frac{1}{1-\alpha} \left(\sum_{x \in \Omega_X} p_X(x)^{\alpha} - 1 \right).$$
(2.1)

With the factor $(2^{1-\alpha}-1)^{-1}$ instead of $(1-\alpha)^{-1}$, this function was studied by Havrda and Charvát [24] and later by Daróczy [13]. In statistical physics, the entropy (2.1) is used due to Tsallis [43]. Following [28], we will call (2.1) the Tsallis–Havrda–Charvát (TCH) entropy.

The entropy (2.1) is concave for all $\alpha > 0$. We can rewrite this entropy as

$$H_{\alpha}(X) = -\sum_{x \in \Omega_X} p_X(x)^{\alpha} \ln_{\alpha} p_X(x) = \sum_{x \in \Omega_X} p_X(x) \ln_{\alpha} \left(\frac{1}{p_X(x)}\right) \cdot$$
(2.2)

Here, the α -logarithm is defined for $\alpha > 0 \neq 1$ and $\xi > 0$ by

$$\ln_{\alpha}(\xi) = \frac{\xi^{1-\alpha} - 1}{1-\alpha}.$$
(2.3)

In the limit $\alpha \to 1$, we obtain $\ln_{\alpha}(\xi) \to \ln \xi$ and the standard Shannon entropy

$$H_1(X) = -\sum_{x \in \Omega_X} p_X(x) \ln p_X(x) \cdot$$
(2.4)

For each $q \in [0, 1]$, the binary TCH entropy is defined as

$$h_{\alpha}(q) := -q^{\alpha} \ln_{\alpha}(q) - (1-q)^{\alpha} \ln_{\alpha}(1-q).$$
(2.5)

The maximal value $\ln_{\alpha}(\#\Omega_X)$ of the entropy (2.1) is reached with the uniform distribution. Let us introduce the index of coincidence [4,23]

$$C(X) := \sum_{x \in \Omega_X} p_X(x)^2.$$
(2.6)

Using information diagrams leads to interesting relations between the Shannon entropy and the index of coincidence [23]. In the case $\alpha = 2$, we obtain the so-called quadratic entropy

$$H_2(X) = 1 - C(X). (2.7)$$

In [44], this entropic function has been used for estimating the minimal error probability in hypotheses testing.

For $\alpha > 0 \neq 1$, the Rényi α -entropy is defined as [39]

$$R_{\alpha}(X) := \frac{1}{1-\alpha} \ln\left(\sum_{x \in \Omega_X} p_X(x)^{\alpha}\right).$$
(2.8)

This entropy is a non-increasing function of order α [39]. Other properties related to the parametric dependence are discussed in [45]. In the limit $\alpha \to \infty$, we obtain the so-called min-entropy

$$R_{\infty}(X) := -\ln(\max p_X(x)). \tag{2.9}$$

For $\alpha \neq 1$, the entropies (2.1) and (2.8) are connected as

$$(1 - \alpha)R_{\alpha}(X) = \ln(1 + (1 - \alpha)H_{\alpha}(X)).$$
(2.10)

Despite of a direct relation, the entropies (2.1) and (2.8) differ in properties. If the random variables X and Y are independent, then the entropy (2.8) has an additivity

$$R_{\alpha}(X,Y) = R_{\alpha}(X) + R_{\alpha}(Y), \qquad (2.11)$$

whereas the entropy (2.1) is pseudo-additive [19]:

$$H_{\alpha}(X,Y) = H_{\alpha}(X) + H_{\alpha}(Y) + (1-\alpha) H_{\alpha}(X) H_{\alpha}(Y).$$
(2.12)

Further, the entropy (2.1) is a concave function of probability distribution for all α . The right-hand side of (2.8) is certainly concave only for $\alpha \in (0; 1)$. Convexity properties of $R_{\alpha}(X)$ with orders $\alpha > 1$ depend on dimensionality of probabilistic vectors [3, 4]. For instance, for every $\alpha > 1$ there exist an m' such that the entropy (2.8) is neither convex nor concave for all m > m' [3]. The maximal value $\ln(\#\Omega)$ of the Rényi entropy is reached with the uniform distribution. Like the THC entropy, the entropy (2.8) recovers the Shannon entropy (2.4) in the limit $\alpha \to 1$. Quantum counterparts of the entropies (2.1) and (2.8) are discussed in [4]. Upper bounds on the entropy of a random variable with countable range are considered in [14, 33].

In the present paper, we will focus on conditional entropies. For brevity, we usually omit the symbols of the sets Ω_X and Ω_Y in entropic sums. The conditional form of entropies is widely used in information theory [12] as well as in applied disciplines. The standard conditional entropy is defined by [12]

$$H_1(X|Y) := \sum_y p_Y(y) H_1(X|y) = -\sum_x \sum_y p_{XY}(x,y) \ln p_{X|Y}(x|y). \quad (2.13)$$

Here, we use the particular function

$$H_1(X|y) = -\sum_x p_{X|Y}(x|y) \ln p_{X|Y}(x|y), \qquad (2.14)$$

and Bayes' rule $p_{X|Y}(x|y) = p_{XY}(x,y)/p_Y(y)$. In the context of quantum mechanics, the conditional entropy (2.13) was used in information-theoretic formulations of Bell's theorem [5] and noise-disturbance uncertainty relations [6].

In the literature, two kinds of the conditional THC entropy are used [19]. These forms are respectively inspired by the two expressions shown in (2.2). The first is defined as [19]

$$H_{\alpha}(X|Y) := \sum_{y} p_{Y}(y)^{\alpha} H_{\alpha}(X|y), \qquad (2.15)$$

where

$$H_{\alpha}(X|y) := \frac{1}{1-\alpha} \left(\sum_{x} p_{X|Y}(x|y)^{\alpha} - 1 \right).$$
 (2.16)

We will also use equivalent expressions

$$H_{\alpha}(X|y) = -\sum_{x} p_{X|Y}(x|y)^{\alpha} \ln_{\alpha} p_{X|Y}(x|y)$$
(2.17)

$$= \sum_{x} p_{X|Y}(x|y) \ln_{\alpha} \left(\frac{1}{p_{X|Y}(x|y)}\right) \cdot$$
(2.18)

The conditional entropy (2.15) is, up to a factor, the quantity originally introduced by Daróczy [13]. For any $\alpha > 0$, we have the chain rule written as [13]

$$H_{\alpha}(X,Y) = H_{\alpha}(X|Y) + H_{\alpha}(Y).$$
(2.19)

In [19], the chain rule (2.19) has been extended to more than two variables. Namely, it holds that

$$H_{\alpha}(X_1, X_2, \dots, X_n) = \sum_{j=1}^n H_{\alpha}(X_j | X_1, \dots, X_{j-1}).$$
(2.20)

In the case $\alpha = 1$, we have the chain rule with the standard conditional entropy (2.13). The chain rule is very essential in many derivations.

Using the particular functional (2.17), the second form of conditional THC entropy is introduced as [19]

$$\widetilde{H}_{\alpha}(X|Y) := \sum_{y} p_{Y}(y) H_{\alpha}(X|y).$$
(2.21)

It should be noted that this form of conditional entropy does not share the chain rule of usual kind [19]. We can only write the relation

$$\widetilde{H}_{\alpha}(X|Y)\left\{ \begin{array}{l} \leq, \quad \alpha \in (0;1) \\ \geq, \quad \alpha \in (1;\infty) \end{array} \right\} H_{\alpha}(X,Y) - H_{\alpha}(Y) = H_{\alpha}(X|Y).$$

$$(2.22)$$

Here, we used $p_Y(y) \leq p_Y(y)^{\alpha}$ for $\alpha \in (0; 1)$ and $p_Y(y) \geq p_Y(y)^{\alpha}$ for $\alpha \in (1; \infty)$. If $\widetilde{H}_{\alpha}(X|Y) \neq 0$ and the variable Y is not deterministic, the inequalities (2.22) are actually strict. Although the chain rule is not applicable here, the entropy (2.21) has found to be useful at least as an auxiliary quantity [19, 35]. Taking $\alpha = 2$ in (2.21), we obtain the conditional quadratic entropy

$$\widetilde{H}_{2}(X|Y) = \sum_{y} p_{Y}(y) \left(1 - \sum_{x} p_{X|Y}(x|y)^{2}\right).$$
(2.23)

This entropic function was utilized for estimating the minimal error probability [44]. In the limit $\alpha \rightarrow 1$, both the entropies (2.15) and (2.21) coincide with (2.13).

The conditional form of Rényi's entropy is used in various topics. There is no generally accepted definition of conditional Rényi entropy [42]. Its first version is defined by [7, 16, 29]

$$R_{\alpha}(X|Y) := \sum_{y} p_Y(y) R_{\alpha}(X|y), \qquad (2.24)$$

where

$$R_{\alpha}(X|y) := \frac{1}{1-\alpha} \ln\left(\sum_{x} p_{X|Y}(x|y)^{\alpha}\right).$$
(2.25)

The conditional entropy (2.24) was used in studying problems of classification [16] and interpretation [29].

The limit $\alpha \to \infty$ gives the conditional min-entropy. For given value y, we define a value

$$\hat{x}(y) := \operatorname{Arg\,max} \left\{ p_{X|Y}(x|y) : \ x \in \Omega_X \right\},\tag{2.26}$$

maximizing $p_{X|Y}(x|y)$, *i.e.*, $p_{X|Y}(x|y) \leq p_{X|Y}(\hat{x}|y)$ for all $x \in \Omega_X$. Note that a value (2.26) may be not unique. Any of such values corresponds to the standard decision in the Bayesian approach [40]. We then write

$$R_{\infty}(X|y) = -\ln p_{X|Y}(\hat{x}|y).$$
(2.27)

The conditional min-entropy $R_{\infty}(X|Y)$ is defined according to (2.24) and (2.27). Basic properties of the conditional entropy (2.24) are examined in [16,29]. However, this form of conditional entropy does not share the chain rule. Instead of the functional (2.24), the authors of [22] proposed another form written as

$$R_{\alpha}(X|Y) := R_{\alpha}(X,Y) - R_{\alpha}(Y).$$
(2.28)

Thus, the chain rule is satisfied by definition. Third version of conditional entropy is obtained from smooth Rényi entropies by removing the smoothing [42]. In the present work, we will deal with the first version (2.24) only. In fact, the particular functions (2.17) and (2.25) are clearly connected. Hence, we will obtain some useful relations between the conditional entropies (2.21) and (2.24).

A final remark concerns notation. In the following, we will often deal with functions of conditional probabilities such as $p_{Y_{n+1}|Y_1...Y_n}(y_{n+1}|y_1,...,y_n)$. To shorten formulas, we will tacitly left out subscripts and merely write $p(y_{n+1}|y_1,...,y_n)$, and so on.

3. Conditioning on more and additivity properties

In this section, we will analyze some properties of the conditional entropies (2.15), (2.21), and (2.24). First, we examine the case when the entropies

are conditioned on more variables. Then the obtained results are applied to the THC entropy rate. Additivity properties will be discussed as well. Properties considered seem to be useful in applications of generalized entropies in some questions of quantum information science. There exist several reasons for considering more general entropic forms in such a context. A utility of entropic bounds with a parametric dependence was emphasized in [31]. Such bounds may allow to find more exactly the domain of acceptable values for unknown probabilities with respect to known ones. Posing inequalities of the Bell type in terms of generalized conditional entropies, we significantly expand a class of probability distributions, for which the non-locality or contextuality are testable [38]. The mentioned approach also allow to reduce an amount of required detection efficiency. In this regard, further studies of generalized conditional entropies are of interest.

We begin with properties related to conditioning on more. For the standard conditional entropy, we have

$$H_1(X|Y_1, \dots, Y_n, Y_{n+1}) \le H_1(X|Y_1, \dots, Y_n).$$
(3.1)

For $\alpha \geq 1$, the conditional entropy (2.15) satisfies [13]

$$H_{\alpha}(X|Y) \le H_{\alpha}(X). \tag{3.2}$$

The author of [19] pointed out an immediate extension of (3.2). For $\alpha \geq 1$ and integer $n \geq 0$, we have

$$H_{\alpha}(X|Y_1, \dots, Y_n, Y_{n+1}) \le H_{\alpha}(X|Y_1, \dots, Y_n).$$
 (3.3)

For instance, relations of the form (3.1) have been used in deriving entropic Bell inequalities [5] and entropy-based approach in counting [34]. We now consider the question for the conditional entropy (2.21). The following statement takes place.

Proposition 3.1. For real $\alpha > 0$ and integer $n \ge 0$, the conditional entropy (2.21) satisfies

$$\widetilde{H}_{\alpha}(X|Y_1,\ldots,Y_n,Y_{n+1}) \le \widetilde{H}_{\alpha}(X|Y_1,\ldots,Y_n).$$
(3.4)

Proof. For brevity, we introduce the function

$$\eta_{\alpha}(\xi) := \frac{\xi^{\alpha} - \xi}{1 - \alpha} \qquad (\alpha \neq 1), \qquad \eta_{1}(\xi) := -\xi \ln \xi.$$
 (3.5)

This function is concave for all $\alpha > 0$. First, we prove the claim for n = 0. Using Jensen's inequality and $\sum_{y} p_Y(y) p_{X|Y}(x|y) = p_X(x)$, one gets

$$\widetilde{H}_{\alpha}(X|Y) = \sum_{x} \sum_{y} p_{Y}(y) \eta_{\alpha} \left(p_{X|Y}(x|y) \right) \le \sum_{x} \eta_{\alpha} \left(p_{X}(x) \right) = H_{\alpha}(X).$$
(3.6)

In the case $n \ge 1$, we write two relations

$$\sum_{y_{n+1}} p(y_{n+1}|y_1,\dots,y_n) \, p(x|y_1,\dots,y_n,y_{n+1}) = p(x|y_1,\dots,y_n) \tag{3.7}$$

and
$$\sum_{y_{n+1}} p(y_{n+1}|y_1, \dots, y_n) = 1$$
. Similarly to (3.6), for all $\alpha > 0$ we obtain

$$\sum_{y_{n+1}} p(y_{n+1}|y_1, \dots, y_n) H_{\alpha}(X|y_1, \dots, y_n, y_{n+1}) \le H_{\alpha}(X|y_1, \dots, y_n).$$
(3.8)

Multiplying (3.8) by $p(y_1, \ldots, y_n)$ and summing with respect to y_1, \ldots, y_n , we obtain the claim (3.4).

For $\alpha \geq 1$, we have $H_{\alpha}(X|Y) \leq \widetilde{H}_{\alpha}(X|Y)$ due to $p_Y(y)^{\alpha} \leq p_Y(y)$. So, the previous result (3.2) follows from (3.6). In the same manner, our relation (3.8) could lead to (3.3). According to (3.3) and (3.4), conditioning on more can only reduce the entropy. The corresponding property of the standard conditional entropy is tacitly used without explicit formulation. It is generally valid for the entropy (2.15) with $\alpha \geq 1$ and for the entropy (2.21) with $\alpha > 0$. Thus, we will rather use the quantity (2.21) in some questions such as an information-theoretic formulation of noise-disturbance trade-off relations. Indeed, the result (3.1) is essential in the argumentation given in [6].

In principle, the conditional entropy (2.15) with $\alpha \in (0; 1)$ may sometimes satisfy the inequality (3.3). Suppose that a random variable Y is deterministic, *i.e.*, one of probabilities $p_Y(y)$ is equal to 1 and other are all zero. For all $\alpha > 0$, we then obtain

$$H_{\alpha}(X|Y) = H_{\alpha}(X|Y) \le H_{\alpha}(X). \tag{3.9}$$

The relation (3.9) is a particular case of (3.3), but its scope covers $\alpha \in (0, 1)$ in our specific example. We also ask for a possible size of positive values of the difference $H_{\alpha}(X|Y) - H_{\alpha}(X)$. It is instructive to consider an example.

Example 3.2. Let us consider joint probabilities $p_{XY}(0,0) = 1/2$ and $p_{XY}(0,1) = p_{XY}(1,0) = p_{XY}(1,1) = 1/6$. For conditional entropies $p_{X|Y}(x|y)$, calculations give values $p_{X|Y}(0|0) = 3/4$, $p_{X|Y}(1|0) = 1/4$, and $p_{X|Y}(0|1) = p_{X|Y}(1|1) = 1/2$. In the case $\alpha = 1/2$, we have $H_{1/2}(X|0) = \sqrt{3} - 1 \approx 0.732$ and $H_{1/2}(X|1) = 2(\sqrt{2} - 1) \approx 0.828$, whence the conditional entropy is

$$H_{1/2}(X|Y) = \sqrt{\frac{2}{3}} H_{1/2}(X|0) + \sqrt{\frac{1}{3}} H_{1/2}(X|1) \approx 1.076.$$
(3.10)

This values is significantly larger than $H_{1/2}(X) = 2(\sqrt{2/3} + \sqrt{1/3} - 1) \approx 0.788.$

As our example shows, for $\alpha \in (0, 1)$ the conditional entropy $H_{\alpha}(X|Y)$ can essentially exceed $H_{\alpha}(X)$. In principle, this issue deserves further studies. We now consider the case of conditional Rényi's entropy (2.24).

Proposition 3.3. For $0 < \alpha \le 1$ and integer $n \ge 0$, the conditional entropy (2.24) satisfies

$$R_{\alpha}(X|Y_{1},\dots,Y_{n},Y_{n+1}) \le R_{\alpha}(X|Y_{1},\dots,Y_{n}).$$
(3.11)

74

Proof. We will assume $\alpha \neq 1$. For $\alpha \in (0, 1)$, the function $\xi \mapsto (1 - \alpha)^{-1} \ln \xi$ is both increasing and concave, whereas the function $\xi \mapsto \xi^{\alpha}$ is concave. Thus, the right-hand side of (2.8) is concave in its probabilistic vector. It follows from this property and the relation (3.7) that

$$\sum_{y_{n+1}} p(y_{n+1}|y_1,\dots,y_n) R_{\alpha}(X|y_1,\dots,y_n,y_{n+1}) \le R_{\alpha}(X|y_1,\dots,y_n).$$
(3.12)

Multiplying (3.12) by $p(y_1, \ldots, y_n)$ and summing with respect to y_1, \ldots, y_n , we obtain the claim (3.11).

According to Proposition 3.3, conditioning on more can only reduce the conditional Rényi entropy (2.24) of order $\alpha \in (0; 1)$. As pointed out in Section 2.3 of [4], the Rényi entropy (2.8) is not concave for $\alpha > \alpha_* > 1$, where α_* depends on $m = \#\Omega_X$. In fact, the binary Rényi entropy is strictly concave for $0 < \alpha \leq 2$ [3]. In the binary case, therefore, the discussed property also holds for $\alpha = 2$. Without specifying the dimensionality of probabilistic vectors, we can use (3.11) only for $0 < \alpha \leq 1$.

For some special distributions, the inequality (3.11) may be valid for orders $\alpha \in (1; \infty)$. In the case n = 0, this property reads

$$R_{\alpha}(X|Y) \le R_{\alpha}(X). \tag{3.13}$$

Let X be a random variable supported on m points with uniform distribution, *i.e.*, $p_X(x) = 1/m$ for all $x \in \Omega_X$. Then the relation (3.13) is valid for all $\alpha > 0$. The authors of [22] noticed the same behavior for the form (2.28). In the case of entropy (2.24), we actually have

$$R_{\alpha}(X|y) \le \ln m = R_{\alpha}(X). \tag{3.14}$$

Hence, we yield (3.13) after multiplying (3.14) by $p_Y(y)$ and summing with respect to $y \in \Omega_Y$. Meantime, the case of uniformly distributed X is very specific. For $\alpha \in (1; \infty)$, we will obtain an upper bound on $R_{\alpha}(X|Y)$ in terms of only the entropy $R_{\alpha}(X)$ and the cardinality $m = \#\Omega_X$. Upper bounds of such a kind will follow from the results of Section 5.

The min-entropy is widely used in cryptography as a measure of security. So, studies of its properties are of specific interest. There exist several ways to define the notion of conditional min-entropy [42]. We will consider the entropic function (2.27) solely. In general, this entropy violates the property formulated as (3.11) and (3.13). The counterexample is posed as follows.

Example 3.4. We consider the same joint probabilities $p_{XY}(x, y)$ as in Example 3.2. In this case, we obtain $R_{\infty}(X|0) = \ln(4/3)$, $R_{\infty}(X|1) = \ln 2$, and

$$R_{\infty}(X|Y) = \frac{2}{3}\ln\frac{4}{3} + \frac{1}{3}\ln 2 \approx 0.423.$$
(3.15)

At the same time, $\max p_X(x) = p_X(0) = 2/3$ and $R_{\infty}(X) = \ln(3/2) \approx 0.405$ that is less than (3.15). In other words, this example enjoys

$$R_{\infty}(X|Y) > R_{\infty}(X). \tag{3.16}$$

We can generalize a situation, in which the result (3.16) takes place. Since the second derivative of the function $\xi \mapsto -\ln \xi$ is strictly positive, this function is strictly convex. If $p_Y(y) < 1$ for all $y \in \Omega_Y$ and the probabilities $p_{X|Y}(\hat{x}|y)$ are not all equal, then the strict convexity gives

$$R_{\infty}(X|Y) = \sum_{y} p_{Y}(y) \left(-\ln p_{X|Y}(\hat{x}|y)\right) > -\ln\left(\sum_{y} p_{Y}(y) p_{X|Y}(\hat{x}|y)\right). \quad (3.17)$$

When probability distributions are such that \hat{x} can be chosen the same for all $y \in \Omega_Y$, we have

$$\sum_{y} p_Y(y) \, p_{X|Y}(\hat{x}|y) = \max p_X(x). \tag{3.18}$$

Combining this with (3.17) at once leads to the conclusion (3.16). We see, therefore, that the conditional min-entropy does not share the property considered in Propositions 3.1 and 3.3.

Since the conditional entropy (2.21) satisfies the property (3.4) for all $\alpha > 0$, it may be used in studying the TCH entropy rate. The THC entropy rate of a stochastic process $\mathbf{X} = \{X_j\}$ is defined by [19]

$$H_{\alpha}(\boldsymbol{X}) := \lim_{n \to \infty} \frac{1}{n} H_{\alpha}(X_1, \dots, X_n), \qquad (3.19)$$

whenever the limit exists. For $\alpha = 1$, it gives the standard Shannon entropy rate defined as [12]

$$H_1(\mathbf{X}) := \lim_{n \to \infty} \frac{1}{n} H_1(X_1, \dots, X_n).$$
(3.20)

The THC entropy rate (3.19) with $\alpha > 1$ certainly exists for a stationary stochastic process [19]. In this case, the entropy rate (3.19) also coincides with the limit [19]

$$\lim_{n \to \infty} H_{\alpha}(X_n | X_1, \dots, X_{n-1}).$$
(3.21)

Using the conditional entropy (2.21), we will obtain some results on the THC entropy rate for $\alpha \in (0, 1)$. For all $\alpha > 0$, we introduce another quantity for entropy rate, namely

$$\widetilde{H}'_{\alpha}(\boldsymbol{X}) := \lim_{n \to \infty} \widetilde{H}_{\alpha}(X_n | X_1, \dots, X_{n-1}), \qquad (3.22)$$

when the limit exists. In the case $\alpha = 1$, the quantity (3.22) is reduced to a quantity defined *via* the standard conditional entropy [12]:

$$\widetilde{H}'_1(\boldsymbol{X}) := \lim_{n \to \infty} H_1(X_n | X_1, \dots, X_{n-1}).$$
(3.23)

The quantities $H_1(\mathbf{X})$ and $\widetilde{H}'_1(\mathbf{X})$ correspond to two different notions of entropy rate. The first is the per symbol entropy of the *n* random variables, and the second is the conditional entropy of the last random variable given the past [12]. A similar treatment can be applied to the quantities (3.19) and (3.22). As is well known [12], for a stationary process the limits (3.20) and (3.23) exist and are equal. The following result holds for the quantity (3.22).

Proposition 3.5. Let a stochastic process $\mathbf{X} = \{X_j\}$ be stationary. For all $\alpha > 0$, the second TCH entropy rate $\widetilde{H}'_{\alpha}(\mathbf{X})$ exists. If the first TCH entropy rate $H_{\alpha}(\mathbf{X})$ exists, then it satisfies

$$H_{\alpha}(\boldsymbol{X}) \ge H'_{\alpha}(\boldsymbol{X}) \qquad (0 < \alpha < 1). \qquad (3.24)$$

$$H_{\alpha}(\boldsymbol{X}) \le H_{\alpha}'(\boldsymbol{X}) \qquad (1 < \alpha < \infty). \qquad (3.25)$$

Proof. Due to the stationarity of the process and the property (3.4), for $\alpha > 0$ we obtain

$$\widetilde{H}_{\alpha}(X_{n}|X_{1},\dots,X_{n-1}) = \widetilde{H}_{\alpha}(X_{n+1}|X_{2},\dots,X_{n})
\geq \widetilde{H}_{\alpha}(X_{n+1}|X_{1},X_{2},\dots,X_{n}).$$
(3.26)

The sequence $\{\widetilde{H}_{\alpha}(X_n|X_1,\ldots,X_{n-1})\}$ is monotonically non-increasing and contains positive elements. Thus, the limit (3.22) exists for all $\alpha > 0$.

Let us prove the claims (3.24) and (3.25). Using the chain rule (2.20), for $\alpha \in (0; 1)$ we further write

$$\frac{1}{n} H_{\alpha}(X_1, \dots, X_n) = \frac{1}{n} \sum_{j=1}^n H_{\alpha}(X_j | X_1, \dots, X_{j-1})$$
$$\geq \frac{1}{n} \sum_{j=1}^n \widetilde{H}_{\alpha}(X_j | X_1, \dots, X_{j-1}).$$
(3.27)

Due to the properties of Cesáro's mean (see, *e.g.*, Thm. 4.2.3 in [12]), the righthand side of (3.27) tends to $\widetilde{H}'_{\alpha}(\mathbf{X})$ in the limit $n \to \infty$. The left-hand side of (3.27) tends to $H_{\alpha}(\mathbf{X})$, if the latter exists. The result (3.24) is proved.

For $\alpha > 1$, the entropy rate $H_{\alpha}(\mathbf{X})$ exists and coincides with the limit (3.21). In view of the definitions (2.15) and (2.21), for $\alpha > 1$ we have

$$H_{\alpha}(X_n|X_1,\dots,X_{n-1}) \le H_{\alpha}(X_n|X_1,\dots,X_{n-1}).$$
 (3.28)

In the limit $n \to \infty$, the relation (3.28) leads to (3.25).

Thus, for a stationary process the second TCH entropy rate (3.22) exists for all $\alpha > 0$. Due to (3.24) and (3.25), this quantity is somehow related to the first form (3.19). Thus, the quantity (3.22) is a suitable measure of TCH entropy rate, at least for the range $\alpha \in (0; 1)$. Moreover, the quantity (3.22) seems to be more appropriate than (3.21). We will illustrate this opinion with an example.

 \square

Example 3.6. Let us consider a Markov chain $\{X_n\}_{n\geq 1}$ with two possible states $s \in \{1, 2\}$ and the transition matrix

$$\mathsf{P} = \begin{pmatrix} 1 - q & q \\ r & 1 - r \end{pmatrix},\tag{3.29}$$

where 0 < q < 1 and 0 < r < 1. This chain is time invariant and irreducible [12]. The stationary distribution is represented as (see, *e.g.*, Example 4.1.1 in [12]):

$$\mu(1) = \frac{r}{q+r}, \qquad \mu(2) = \frac{q}{q+r}.$$
(3.30)

Using this distribution as initial, we obtain a stationary process. Due to the Markov property, we have

$$p(x_n|x_1,...,x_{n-1}) = p(x_n|x_{n-1}).$$
 (3.31)

Hence, for arbitrary $n \ge 2$ one gets

$$H_{\alpha}(X_n|x_1,\dots,x_{n-2},1) = h_{\alpha}(q), \qquad H_{\alpha}(X_n|x_1,\dots,x_{n-2},2) = h_{\alpha}(r).$$
(3.32)

For the given value s of the variable X_{n-1} , we obviously have

$$\sum_{x_1...x_{n-2}} p(x_1,\ldots,x_{n-2},s) = \mu(s).$$
(3.33)

Due to (3.32) and (3.33), for all $\alpha > 0$ we obtain

$$\widetilde{H}'_{\alpha}(\boldsymbol{X}) = \frac{r}{q+r} h_{\alpha}(q) + \frac{q}{q+r} h_{\alpha}(r).$$
(3.34)

The right-hand side of (3.34) remains valid for $\alpha = 1$, when gives the standard entropy rate of the two-state Markov chain (see, *e.g.*, Example 4.2.1 in [12]). The following point should be noted here. In general, the two limits $n \to \infty$ and $\alpha \to 1$ are not always commuting. An instance of non-commutativity will be shown with the quantity (3.21). Nevertheless, the resulting formula (3.34) for the rate (3.22) covers the standard case $\alpha = 1$ in the considered example.

Let us consider the quantity (3.21). For the sake of simplicity, we focus on the case q = r. We then have $\mu(s) = 1/2$ and $H_1(\mathbf{X}) = h_1(q)$. Combining the Markov property with Bayes' rule leads to the following result. For integer $m \ge 2$, the corresponding joint probability

$$p(x_1, \dots, x_m) = \frac{1}{2} \prod_{j=2}^m f_j,$$
 (3.35)

where each factor f_j is either q or $\tilde{q} = 1 - q$. Taking into account the number of related choices, we have

$$\sum_{x_1...x_m} p(x_1,...,x_m)^{\alpha} = 2 \sum_{k=0}^{m-1} \binom{m-1}{k} \frac{q^{k\alpha} \,\tilde{q}^{(m-1-k)\alpha}}{2^{\alpha}} = 2^{1-\alpha} \left(q^{\alpha} + \tilde{q}^{\alpha}\right)^{m-1}.$$
(3.36)

Using this answer with m = n - 1 and the result (3.32), we finally obtain

$$H_{\alpha}(X_n|X_1,\dots,X_{n-1}) = 2^{1-\alpha} \left(q^{\alpha} + \tilde{q}^{\alpha}\right)^{n-2} h_{\alpha}(q).$$
(3.37)

As 0 < q < 1, we have $q^{\alpha} + (1 - q)^{\alpha} > 1$ for $\alpha \in (0; 1)$, and $q^{\alpha} + (1 - q)^{\alpha} < 1$ for $\alpha \in (1; \infty)$. Hence, the limit (3.21) does not exist for $\alpha \in (0; 1)$. Further, we obtain $H_{\alpha}(\mathbf{X}) = 0$ for $\alpha \in (1; \infty)$. In both the cases, it hardly tells us anything about the process. When $\alpha \to 1^+$, the quantity $H_{\alpha}(\mathbf{X})$ does not cover the standard fact $H_1(\mathbf{X}) = h_1(q)$. Here, we explicitly see non-commutativity of the limits $n \to \infty$ and $\alpha \to 1$. In opposite, the quantity (3.22) exists for all $\alpha > 0$ and also reproduces the standard result.

We complete this section with a brief discussion of additivity properties of the conditional entropies (2.15) and (2.21). Subadditivity is one of basic properties of the Shannon entropy. The joint entropy satisfies [12]

$$H_1(X,Y) \le H_1(X) + H_1(Y),$$
(3.38)

with equality if and only if the random variables are independent. It is not the case generally for the TCH entropy (2.1). For the joint distribution p(x, y) = p(x) p(y), we directly have a pseudo-additivity (2.12). The third term in the right-hand side of (2.12) is positive for $\alpha \in (0; 1)$ and negative $\alpha \in (1; \infty)$. Such properties are useful in nonextensive statistical mechanics [20]. For $\alpha \ge 1$, we can write (3.38) with the TCH α -entropies instead of the Shannon ones. That is, the TCH entropy is subadditive for $\alpha \ge 1$. A similar result holds for the conditional entropy (2.15). For $\alpha \ge 1$, one obeys [19]

$$H_{\alpha}(X,Y|Z) \le H_{\alpha}(X|Z) + H_{\alpha}(Y|Z). \tag{3.39}$$

We now consider this question for the conditional entropy (2.21).

Lemma 3.7. Let random variables X, Y, Z take values on finite sets. For $\alpha \ge 1$ and each $z \in \Omega_Z$, the entropic function (2.16) satisfies

$$H_{\alpha}(X,Y|z) \le H_{\alpha}(X|z) + H_{\alpha}(Y|z). \tag{3.40}$$

The result (3.40) can be proved similarly to the inequality (3.39). We refrain from presenting the details here. Multiplying our result (3.40) by $p_Z(z)$ and summing with respect to z, we obtain the following. For $\alpha \geq 1$, the conditional entropy (2.21) satisfies

$$H_{\alpha}(X,Y|Z) \le H_{\alpha}(X|Z) + H_{\alpha}(Y|Z). \tag{3.41}$$

Thus, the conditional entropy (2.21) shares an additivity property similarly to the conditional entropy (2.15).

4. Error probability and conditional entropies

In this section, we consider lower and upper bounds on generalized conditional entropies. Let random variables X and Y describe the input and output of a communication channel over an alphabet Ω . The conditional entropy $H_1(X|Y)$ can be treated as a measure of quality of information transmission. The classical Fano inequality is a relation between $H_1(X|Y)$ and error probability. We use random variable E with possible values from the set $\{\neg e, e\}$. The average probability of error is expressed as

$$P_{e} = \sum_{y \in \Omega} p_{Y}(y) \, p_{E|Y}(e|y), \qquad p_{E|Y}(e|y) = \sum_{\substack{x \in \Omega \\ x \neq y}} p_{X|Y}(x|y). \tag{4.1}$$

The value P_e gives an average probability that the input symbol has been mistaken through the transmission. For the standard decision (2.26), the error probability is equal to

$$\hat{P}_{e} = 1 - \sum_{y \in \Omega} p_{Y}(y) \, p_{X|Y}(\hat{x}|y).$$
(4.2)

In the context of communications, the value \hat{P}_e is attained by the so-called "maximum *a posteriori* estimator" [18].

Entropic functions are basic measures of uncertainty used in information theory. On the other hand, channel coding theorems are usually stated in terms of the error probability [12]. In this regard, relations between entropies and the error probability are of great interest [18]. The Fano inequality states that

$$H_1(X|Y) \le h_1(P_e) + P_e \ln(m-1), \tag{4.3}$$

where the binary entropy $h_1(q) = -q \ln q - (1-q) \ln(1-q)$ and $m = \#\Omega$. The fact (4.3) shows that $P_e \to 0$ implies $H_1(X|Y) \to 0$. Further, if $H_1(X|Y)$ is large then the probability of making an error in inference must be large as well.

For the conditional Rényi entropy (2.24), some inequalities of the Fano type with applications were considered in [16]. In particular, the right-hand side of (4.3) gives an upper bound on the conditional entropy (2.24) for all $\alpha > 1$. This claim follows from the fact that the function (2.25) cannot increase with growth of α . For orders $\alpha \in (0; 1)$, bounds of the Fano type can be written in terms of the error probability (4.2). The following statement takes place.

Proposition 4.1. Let $\#\Omega = m$; for $\alpha \in (0;1)$, the conditional entropy (2.24) obeys

$$R_{\alpha}(X|Y) \le \frac{1}{1-\alpha} \ln\left((1-\hat{P}_{e})^{\alpha} + (m-1)^{1-\alpha}\hat{P}_{e}^{\alpha}\right).$$
(4.4)

Proof. For the standard decision, we have $p_{E|Y}(e|y) = 1 - p_{X|Y}(\hat{x}|y)$. It follows from (Thm. 6 of [3]) that, for all $\alpha > 0 \neq 1$,

$$R_{\alpha}(X|y) \le \frac{1}{1-\alpha} \ln \left\{ \left(1 - p_{E|Y}(e|y)\right)^{\alpha} + (m-1)^{1-\alpha} p_{E|Y}(e|y)^{\alpha} \right\}.$$
 (4.5)

For $\alpha \in (0; 1)$, the function $\xi \mapsto (1 - \alpha)^{-1} \ln \xi$ is both increasing and concave, whereas the function $\xi \mapsto (1 - \xi)^{\alpha} + (m - 1)^{1-\alpha}\xi^{\alpha}$ is concave. Hence, the righthand side of (4.5) is concave in $p_{E|Y}(e|y)$. Multiplying (4.5) by $p_Y(y)$ and summing with respect to y, we get the claim (4.4) due to Jensen's inequality. \Box

For the conditional THC entropy (2.15), bounds of the Fano type were derived in [19] and improved in [36]. We shall now consider the question for the conditional entropy (2.21). Our result is formulated as follows.

Proposition 4.2. Let $\#\Omega = m$; then the conditional entropy (2.21) satisfies

$$H_{\alpha}(X|Y) \le h_{\alpha}(P_e) + P_e^{\alpha} \ln_{\alpha}(m-1)$$
 (0 < α < 1). (4.6)

$$H_{\alpha}(X|Y) \le h_{\alpha}(P_e) + P_e \ln_{\alpha}(m-1) \qquad (1 < \alpha < \infty).$$

$$(4.7)$$

Proof. We will follow the original scheme of derivation of the book [17]. It follows from Lemma 4 of [36] that

$$H_{\alpha}(X|y) \le h_{\alpha}(p_{E|Y}(e|y)) + p_{E|Y}(e|y)^{\alpha} \ln_{\alpha}(m-1),$$
(4.8)

where $\alpha > 0$. For $\alpha \in (0; 1)$, the function $\xi \mapsto \xi^{\alpha}$ is concave, whence

$$\sum_{y} p_{Y}(y) \, p_{E|Y}(e|y)^{\alpha} \le \left(\sum_{y} p_{Y}(y) \, p_{E|Y}(e|y) \right)^{\alpha} = P_{e}^{\alpha}. \tag{4.9}$$

Multiplying (4.8) by $p_Y(y)$ and summing with respect to y, we obtain the claim (4.6) due to (4.9). For $\alpha > 1$, we have $p_{E|Y}(e|y)^{\alpha} \leq p_{E|Y}(e|y)$. Substituting this into (4.8), we similarly prove the claim (4.7).

The statement of Proposition 4.2 provides Fano type inequalities for the conditional entropy (2.21). In the limit $\alpha \to 1$, the relations (4.6) and (4.7) both lead to the standard Fano inequality (4.3). The author of [19] presented the bound (4.7) with $H_{\alpha}(X|Y)$ instead of $\tilde{H}_{\alpha}(X|Y)$. So our result is stronger, since $H_{\alpha}(X|Y) \leq \tilde{H}_{\alpha}(X|Y)$ for $\alpha > 1$. As was shown in [36], for $\alpha > 1$ the conditional entropy (2.15) actually obeys a stronger bound

$$H_{\alpha}(X|Y) \le h_{\alpha}(P_e) + P_e^{\alpha} \ln_{\alpha}(m-1).$$

$$(4.10)$$

In the range $\alpha \in (0, 1)$, Fano type bounds on the conditional entropy (2.15) were derived in [36]. For a two-sided estimation of conditional entropies, lower bounds are also required. Another fact is that inequalities of Fano type involves cardinality of the set Ω [26]. Hence, results of such a kind are not applicable in the case of countably infinite alphabets. It is possible in such situations that $H_1(X|Y)$ does not tend to 0 as P_e vanishes [26]. This phenomenon is connected with an entropic discontinuity considered in [27].

Lower bounds on the standard conditional entropy (2.13) were considered by Rényi [40]. He showed that such bounds are expressed in terms of the probability

error of the standard decision. In [35], this issue was developed for all the conditional entropies (2.15), (2.21), and (2.24). In the case of finite alphabets, we have

$$\ln_{\alpha}\left(\frac{1}{1-\hat{P}_{e}}\right) \leq \widetilde{H}_{\alpha}(X|Y) \qquad (0 < \alpha \leq 2). \tag{4.11}$$

$$-\ln(1-\hat{P}_e) \le R_{\alpha}(X|Y) \qquad (0 < \alpha < \infty), \qquad (4.12)$$

For $\alpha \geq 2$, lower bounds on $\widetilde{H}_{\alpha}(X|Y)$ also involve cardinality of the alphabet [35]. Improved lower bounds on the standard conditional entropy were derived in [26]. In particular, the authors of [26] have proved that

$$(2\ln 2)\hat{P}_e \le H_1(X|Y). \tag{4.13}$$

This bound also holds for the conditional Rényi entropy $R_{\alpha}(X|Y)$ of order $\alpha \in (0; 1)$. We shall consider the question for the conditional THC entropies. Following [26], our results are based on considering entropies of truncated distributions.

Lemma 4.3. Suppose that probabilities $p_X(x)$ are arranged in non-increasing order, i.e., $p_X(i) \ge p_X(j)$ for i < j. We now define an integer

$$\ell = \left\lfloor \frac{1}{p_X(1)} \right\rfloor . \tag{4.14}$$

Let random variable W take values on $\{1, \ldots, \ell + 1\}$ with probabilities

$$p_W(w) = \begin{cases} p_X(1), & \text{if } w = 1, \dots, \ell, \\ 1 - \ell \, p_X(1), & \text{if } w = \ell + 1. \end{cases}$$
(4.15)

For all $\alpha > 0$, the TCH entropies satisfy

$$H_{\alpha}(W) \le H_{\alpha}(X). \tag{4.16}$$

Proof. It follows from the definition (4.15) that for all integer $k \ge 1$ we have [26]

$$\sum_{j=1}^{k} p_X(j) \le \sum_{j=1}^{k} p_W(j).$$
(4.17)

Thus, the probabilistic vector p_X is majorized by p_W . We also recall that if the function $\xi \mapsto g(\xi)$ is concave, then the sum $\sum_j g(\xi_j)$ is Schur-concave [32]. Combining this with concavity of the function (3.5) and the fact (4.17) immediately gives the claim (4.16).

In the case $\alpha = 1$, the result (4.16) gives a relation between the Shannon entropies, namely $H_1(W) \leq H_1(X)$. This relation was derived and further applied in [26]. Some of these results can be extended to the THC entropies. We now establish a useful lower bound on $\ln_{\alpha}(n)$.

Lemma 4.4. For all real $\alpha \in (0; 2]$ and integer $n \ge 1$, we have

$$\ln_{\alpha}(n) \ge 2\ln_{\alpha}(2)\left(1-\frac{1}{n}\right).$$
(4.18)

Proof. We first note that the inequality (4.18) is obviously saturated for n = 1. To prove (4.18) for integer $n \ge 2$, we will use the following. For all $\alpha \in (0; 2]$, one has

$$2^{1-\alpha} \ge \ln_{\alpha}(2). \tag{4.19}$$

It holds for $\alpha = 1$ due to $1 > \ln 2$. For $\alpha \neq 1$, we write the difference

$$2^{1-\alpha} - \ln_{\alpha}(2) = 2^{1-\alpha} - \frac{2^{1-\alpha} - 1}{1-\alpha} = \frac{2^{\alpha} - 2\alpha}{2^{\alpha}(1-\alpha)} \cdot$$
(4.20)

For $\alpha \in [0;1]$, we have $b(\alpha) = 2^{\alpha} - 2\alpha \ge 0$ due to $b'(\alpha) = 2^{\alpha} \ln 2 - 2 < 0$ and b(1) = 0. Hence, the right-hand side of (4.20) is non-negative for $\alpha \in (0;1)$. We further note that the function $\alpha \mapsto 2\alpha - 2^{\alpha}$ is concave and vanishes at the end points of the interval $\alpha \in [1;2]$. Thus, the right-hand side of (4.20) is non-negative for $\alpha \in (1;2]$. This completes the proof of (4.19).

We now introduce the function

$$g_{\alpha}(\xi) := \ln_{\alpha}\left(\frac{1}{\xi}\right) = \frac{\xi^{\alpha-1} - 1}{1 - \alpha},\tag{4.21}$$

including $g_1(\xi) = -\ln \xi$. We consider (4.21) on the interval $\xi \in [\varepsilon; 1/2]$ with arbitrarily small $\varepsilon > 0$ (we can take $\varepsilon = 0$ for $\alpha > 1$). By inspection of the second derivative, the function $g_{\alpha}(\xi)$ is convex for $\alpha \in (0; 2]$. By the Taylor formula with remainder written in Lagrange's form, with $\varepsilon < c < 1/2$, we have

$$g_{\alpha}(\xi) = g_{\alpha}(1/2) + g'_{\alpha}(1/2) \left(\xi - 1/2\right) + \frac{1}{2} g''_{\alpha}(c) \left(\xi - 1/2\right)^{2}$$

$$\geq \ln_{\alpha}(2) + 2^{1-\alpha} \left(1 - 2\xi\right) \geq 2 \ln_{\alpha}(2) \left(1 - \xi\right).$$
(4.22)

Here, we also used (4.19) for $\alpha \in (0; 2]$. The claim (4.22) merely says that the graph of convex $g_{\alpha}(\xi)$ goes over its tangent line drawn at the point $\xi = 1/2$. Substituting $\xi = 1/n$ with integer $n \geq 2$ into (4.22) provides the claim (4.18).

When $\alpha = 1$, the result (4.18) gives $\ln n \ge 2 \ln 2 (1 - 1/n)$ for integer $n \ge 1$. The latter was also used in [26] for estimating the Shannon entropy from below. In a similar manner, for $\alpha \in (0; 2]$ we can obtain lower bounds on the THC entropies.

Proposition 4.5. For all $\alpha \in (0; 2]$, the TCH entropy (2.2) satisfies

$$2\ln_{\alpha}(2)(1 - \max p_X(x)) \le H_{\alpha}(X). \tag{4.23}$$

Proof. Due to (4.16), we only need to prove that

$$2\ln_{\alpha}(2)(1 - \max p_X(x)) \le H_{\alpha}(W).$$

$$(4.24)$$

In the proof, we will assume that $\max p_X(x) = p_X(1)$. Let us introduce a parameter

$$\lambda := p_X(1)\,\ell(\ell+1) - \ell. \tag{4.25}$$

From the definition (4.14), we then have

$$\ell \le \frac{1}{p_X(1)} < \ell + 1, \qquad 0 < \lambda \le 1.$$
 (4.26)

Following [26], we represent the probabilistic vector p_W as a mixture of two other ones:

$$p_W(j) = \lambda p_U(j) + (1 - \lambda) p_V(j).$$
 (4.27)

Here, we mean $p_U(j) = \ell^{-1}$ for $j = \{1, \ldots, \ell\}$ and $p_U(\ell + 1) = 0$, and also $p_V(j) = (\ell + 1)^{-1}$ for $j = \{1, \ldots, \ell + 1\}$. Obviously, we have $H_\alpha(U) = \ln_\alpha(\ell)$ and $H_\alpha(V) = \ln_\alpha(\ell + 1)$. It follows from (4.27) that

$$H_{\alpha}(W) \ge \lambda \ln_{\alpha}(\ell) + (1-\lambda) \ln_{\alpha}(\ell+1)$$
(4.28)

$$\geq 2\ln_{\alpha}(2)\left(1 - \frac{\lambda + \ell}{\ell(\ell + 1)}\right). \tag{4.29}$$

Here, the step (4.28) is due to concavity of the THC entropy (2.21) for $\alpha > 0$; the step (4.29) is due to (4.18) for $\alpha \in (0; 2]$. Substituting (4.25) into (4.29) completes the proof.

For $\alpha \in (0; 2]$, the statement of Proposition 4.5 provides a lower bound on the TCH entropy in terms of the maximum probability. This result is important in own rights as well as in estimating conditional TCH entropies from below. Using (4.23), we now obtain an improved lower bound on the conditional entropy (2.21).

Proposition 4.6. For all $\alpha \in (0; 2]$, the conditional entropy (2.21) satisfies

$$2\ln_{\alpha}(2)\hat{P}_{e} \le H_{\alpha}(X|Y). \tag{4.30}$$

Proof. We first note that $\sum_{x} p_{X|Y}(x|y) = 1$. Applying (4.23) to the entropic function (2.17), for $\alpha \in (0, 2]$ we have

$$H_{\alpha}(X|y) \ge 2 \ln_{\alpha}(2) \left(1 - p_{X|Y}(\hat{x}|y)\right)$$
$$\widetilde{H}_{\alpha}(X|Y) \ge \sum_{y \in \Omega} p_{Y}(y) 2 \ln_{\alpha}(2) \left(1 - p_{X|Y}(\hat{x}|y)\right).$$
(4.31)

Recall that the value $\hat{x}(y)$ is defined in (2.26). Combining (4.31) with (4.2) completes the proof.

For $\alpha \in (0; 1)$, the results (4.30) also holds with $H_{\alpha}(X|Y)$ instead of $H_{\alpha}(X|Y)$. In the binary case, the lower bound (4.30) has already been given in [35]. It was an extension of the previous result of Rényi [40]. In the binary case, Rényi pointed out an improved lower bound on the standard conditional entropy (2.13). We now see that the lower bound (4.30) remains valid with any finite alphabet. The distinction is that Proposition 4.6 assumes $\alpha \in (0; 2]$. In the binary case, the bound (4.30) takes place for all $\alpha > 0$ [35].

5. Relations between different conditional entropies

In this section, we will derive relations between various conditional entropies. In general, the quadratic entropies (2.7) and (2.23) are relatively easy to evaluate. For many special types of quantum measurements, we can calculate or estimate the index of coincidence. For a symmetric informationally complete measurement and any quantum state, the index of coincidence was exactly found in [37]. Hence, entropic uncertainty bounds for a single measurement follow. Thus, relations between the quadratic conditional entropy and other conditional entropies could be useful. Bounds on the conditional TCH entropies are formulated in the following way.

Proposition 5.1. For $\alpha \in (0; 2]$, the conditional entropy (2.21) satisfies

$$\ln_{\alpha}\left\{\left(1-\widetilde{H}_{2}(X|Y)\right)^{-1}\right\} \leq \widetilde{H}_{\alpha}(X|Y).$$
(5.1)

For $\alpha \geq 2$, the conditional entropy (2.21) satisfies

$$\widetilde{H}_{\alpha}(X|Y) \le \ln_{\alpha} \left\{ \left(1 - \widetilde{H}_{2}(X|Y) \right)^{-1} \right\}.$$
(5.2)

Proof. Let us begin with the claim (5.1). For brevity, we introduce an analog of the index of coincidence, namely

$$C(X|y) := \sum_{x} p_{X|Y}(x|y)^2 = 1 - H_2(X|y).$$
(5.3)

The second derivative of $\ln_{\alpha}(1/\xi)$ is equal to $(2-\alpha)\xi^{\alpha-3}$ and positive for $\alpha \leq 2$. Thus, the function itself is convex. Applying Jensen's inequality to (2.18) and using (5.3), for $\alpha \in (0, 2]$ we have

$$H_{\alpha}(X|y) \ge \ln_{\alpha}\left(\frac{1}{C(X|y)}\right)$$
(5.4)

Using this inequality and Jensen's inequality again, one gives

$$\widetilde{H}_{\alpha}(X|Y) \ge \sum_{y} p_{Y}(y) \ln_{\alpha}\left(\frac{1}{C(X|y)}\right) \ge \ln_{\alpha}\left\{\left(\sum_{y} p_{Y}(y) C(X|y)\right)^{-1}\right\}.$$
 (5.5)

The right-hand side of (5.5) coincides with the left-hand side of (5.1). The above formulas are all suitable in the case $\alpha = 1$ as well.

For $\alpha \in [2; \infty)$, the function $\xi \mapsto \ln_{\alpha}(1/\xi)$ is concave. Similarly to (5.4), we then have

$$H_{\alpha}(X|y) \le \ln_{\alpha} \left(\frac{1}{C(X|y)}\right).$$
(5.6)

Substituting (5.6) into (2.21) and using Jensen's inequality again, we rewrite the inequality (5.5) in reversed direction. This gives the claim (5.2). \Box

Note that for $\alpha = 2$ both the relations (5.1) and (5.2) are reduced to a trivial identity. Recalling (2.22), we see the following. For $\alpha \in (0; 1)$, the conditional entropy $H_{\alpha}(X|Y)$ is bounded from below by the left-hand side of (5.1). For $\alpha \in [2; \infty)$, this entropy is bounded from above by the right-hand side of (5.2). Let us proceed to bounds on the conditional Rényi entropy.

Proposition 5.2. For $\alpha \in (0; 2]$, the conditional entropy (2.24) satisfies

$$-\ln\left(1 - \widetilde{H}_2(X|Y)\right) \le R_\alpha(X|Y).$$
(5.7)

For $\alpha \geq 2$, the conditional entropy (2.24) satisfies

$$R_{\alpha}(X|Y) \le \frac{m\ln m}{m-1} \widetilde{H}_2(X|Y).$$
(5.8)

where $m = \# \Omega_X$.

Proof. As the standard case $\alpha = 1$ is already known, we further assume $\alpha \neq 1$. For $\alpha \in (0; 1)$, the function $\xi \mapsto \xi^{\alpha - 1}$ is convex, whence

$$\sum_{x} p_{X|Y}(x|y) p_{X|Y}(x|y)^{\alpha-1} \ge C(X|y)^{\alpha-1}.$$
(5.9)

Since the function $\xi \mapsto (1 - \alpha)^{-1} \ln \xi$ increases for $\alpha \in (0; 1)$, the inequality (5.9) implies

$$R_{\alpha}(X|y) \ge -\ln C(X|y) = -\ln(1 - H_2(X|y)).$$
 (5.10)

Multiplying this by $p_Y(y)$ and summing with respect to y, we obtain

$$R_{\alpha}(X|Y) \ge \sum_{y} p_{Y}(y) \left\{ -\ln\left(1 - H_{2}(X|y)\right) \right\}$$

$$\ge -\ln\left(1 - \sum_{y} p_{Y}(y) H_{2}(X|y)\right).$$
(5.11)

At the last step, we used convexity of the function $\xi \mapsto -\ln(1-\xi)$. The relation (5.11) provides the claim for $\alpha \in (0; 1)$. When $\alpha \in (1; 2]$, the function $\xi \mapsto \xi^{\alpha-1}$ is concave. Instead of (5.9), therefore, we have

$$\sum_{x} p_{X|Y}(x|y) p_{X|Y}(x|y)^{\alpha-1} \le C(X|y)^{\alpha-1}.$$
(5.12)

The function $\xi \mapsto (1-\alpha)^{-1} \ln \xi$ decreases for $\alpha \in (1; 2]$. Combining this with (5.12), we obtain (5.10) and (5.11) as well.

Let us proceed to the claim (5.8). For $\alpha \in [2; \infty)$, the function $\xi \mapsto \xi^{\alpha-1}$ is convex. Hence, we again have the inequality (5.9). Combining this with decreasing of the function $\xi \mapsto (1-\alpha)^{-1} \ln \xi$ gives

$$R_{\alpha}(X|y) \le -\ln C(X|y). \tag{5.13}$$

If m positive numbers $p_{X|Y}(x|y)$ satisfy $\sum_x p_{X|Y}(x|y) = 1$, the sum C(X|y) of their squares obeys

$$m^{-1} \le C(X|y) \le 1.$$
 (5.14)

Let $f(\xi)$ be a convex function such that f(1) = 0, and let ξ be varied between ξ_0 and 1. For $\xi \in [\xi_0; 1]$, we have

$$f(\xi) \le f(\xi_0) \ \frac{1-\xi}{1-\xi_0}.$$
(5.15)

Indeed, the difference $\{f(\xi) - f(\xi_0)(1 - \xi_0)^{-1}(1 - \xi)\}$ is convex and vanishes for both the points $\xi = \xi_0$ and $\xi = 1$. So, this difference is negative in the interval $\xi \in [\xi_0; 1]$ everywhere. Using (5.15) with $f(\xi) = -\ln \xi$ and $\xi_0 = m^{-1}$, we obtain from (5.13) that

$$R_{\alpha}(X|y) \le \frac{m \ln m}{m-1} H_2(X|y).$$
 (5.16)

Multiplying this by $p_Y(y)$ and summing with respect to y, we get (5.8).

With the result (5.7), we keep in mind the following. Due to (2.23) and (5.14), the quadratic entropy $\widetilde{H}_2(X|Y)$ does not exceed 1 - 1/m. Using $\xi \leq -\ln(1 - \xi)$, for $\alpha \in (0, 2]$ the relation (5.7) gives

$$\widetilde{H}_2(X|Y) \le R_\alpha(X|Y). \tag{5.17}$$

This is a simple lower bound on the conditional Rényi entropy in terms of the quadratic conditional entropy. The result (5.8) provides an upper bound on conditional Renyi's entropies of order $\alpha \geq 2$. For the case $\alpha = 2$, we further write

$$-\ln\left(1 - \widetilde{H}_{2}(X|Y)\right) \le R_{2}(X|Y) \le \frac{m\ln m}{m-1} \,\widetilde{H}_{2}(X|Y).$$
(5.18)

It is a two-sided estimate on the conditional entropy $R_2(X|Y)$. As the right-hand side of (5.8) is independent of α , this upper bound also holds for the conditional min-entropy $R_{\infty}(X|Y)$. We will now formulate a lower bound on $R_{\infty}(X|Y)$.

Proposition 5.3. Let $\#\Omega_X = m$; then the conditional min-entropy is bounded from below as

$$\ln m - \ln\left(1 + \sqrt{m-1}\sqrt{m-1 - m\,\widetilde{H}_2(X|Y)}\right) \le R_\infty(X|Y). \tag{5.19}$$

Proof. We first note that m positive numbers $p_{X|Y}(x|y)$ satisfy the following two conditions:

$$\sum_{x} p_{X|Y}(x|y) = 1, \qquad \sum_{x} p_{X|Y}(x|y)^2 = C(X|y).$$
(5.20)

For given $y \in \Omega_Y$, the maximal probability $p_{X|Y}(\hat{x}|y)$ is then bounded from above as (see Lem. 3 of the paper [37])

$$p_{X|Y}(\hat{x}|y) \le \frac{1}{m} \left(1 + \sqrt{m-1} \sqrt{m C(X|y) - 1} \right).$$
 (5.21)

Since the function $\xi \mapsto -\ln \xi$ is convex and decreasing, we then obtain

$$R_{\infty}(X|Y) \ge -\ln\left(\sum_{y} p_{Y}(y) p_{X|Y}(\hat{x}|y)\right)$$

$$\ge \ln m - \ln\left(1 + \sqrt{m-1}\sum_{y} p_{Y}(y) \sqrt{mC(X|y) - 1}\right).$$
(5.22)

By concavity, we also obtain

$$\sum_{y} p_{Y}(y) \sqrt{m C(X|y) - 1} \le \left(m \sum_{y} p_{Y}(y) C(X|y) - 1 \right)^{1/2}.$$
 (5.23)

Combining this with (5.22) and decreasing of $\xi \mapsto -\ln \xi$ finally gives (5.19). \Box

If $\hat{H}_2(X|Y) = 0$, then the left-hand side of (5.19) vanishes. When $\hat{H}_2(X|Y)$ reaches its maximal value $\ln_2(m) = (m-1)/m$, the left-hand side of (5.19) also gives the correct value $\ln m$. The idea recalled in the above proof has been applied to deriving uncertainty bounds for any set of mutually unbiased bases [37]. Such measurement bases are used in some efficient protocols of quantum cryptography. The result (5.19) could be useful in studies of security of these protocols. Together the bounds (5.8) and (5.19) give a two-sided estimate on the conditional min-entropy. These bounds are expressed in terms of the conditional quadratic entropy (2.23) and the cardinality of support set Ω_X . We shall now proceed to other relations of such a kind. They may give convenient tools, when number of used symbols is sufficiently small, say, in the binary case.

Proposition 5.4. Let $\#\Omega_X = m$; then the conditional entropies (2.21) and (2.24) satisfy

$$R_{\alpha}(X|Y) \ge \frac{\ln m}{\ln_{\alpha}(m)} \widetilde{H}_{\alpha}(X|Y) \qquad (0 < \alpha < 1), \qquad (5.24)$$

$$R_{\alpha}(X|Y) \le \frac{\ln m}{\ln_{\alpha}(m)} \ \widetilde{H}_{\alpha}(X|Y) \qquad (1 < \alpha < \infty). \tag{5.25}$$

Proof. If m positive numbers $p_{X|Y}(x|y)$ obey the condition $\sum_{x} p_{X|Y}(x|y) = 1$, then we have

$$1 \le \sum_{x} p_{X|Y}(x|y)^{\alpha} \le m^{1-\alpha} \qquad (0 < \alpha < 1), \qquad (5.26)$$

$$m^{1-\alpha} \le \sum_{x} p_{X|Y}(x|y)^{\alpha} \le 1 \qquad (1 < \alpha < \infty). \tag{5.27}$$

Let us take powers $p_{X|Y}(x|y)^{\alpha}$ with equal weights. Then the upper bound of (5.26) and the lower bound of (5.27) follow from Jensen's inequality.

We begin with the case $\alpha \in (0; 1)$, in which the function $\xi \mapsto (1 - \alpha)^{-1} \ln \xi$ is concave. Let $g(\xi)$ be a concave function such that g(1) = 0, and let ξ be varied between 1 and ξ_1 . For $\xi \in [1; \xi_1]$, we have

$$g(\xi) \ge g(\xi_1) \frac{\xi - 1}{\xi_1 - 1}$$
 (5.28)

In fact, the difference $\{g(\xi) - g(\xi_1)(\xi_1 - 1)^{-1}(\xi - 1)\}$ is concave and vanishes for both the points $\xi = 1$ and $\xi = \xi_1$. Hence, this difference is positive in the interval $\xi \in [1; \xi_1]$ everywhere. Using (5.28) with $g(\xi) = (1 - \alpha)^{-1} \ln \xi$ and $\xi_1 = m^{1-\alpha}$, we obtain

$$R_{\alpha}(X|y) \ge \frac{\ln m}{m^{1-\alpha} - 1} \left(\sum_{x} p_{X|Y}(x|y)^{\alpha} - 1 \right) = \frac{\ln m}{\ln_{\alpha}(m)} \ H_{\alpha}(X|y).$$
(5.29)

Multiplying (5.29) by $p_Y(y)$ and summing with respect to y, we get (5.24).

In the case $\alpha \in (1; \infty)$, the function $\xi \mapsto (1 - \alpha)^{-1} \ln \xi$ is convex. Using (5.15) with $f(\xi) = (1 - \alpha)^{-1} \ln \xi$ and $\xi_0 = m^{1-\alpha}$, we obtain

$$R_{\alpha}(X|y) \le \frac{\ln m}{1 - m^{1 - \alpha}} \left(1 - \sum_{x} p_{X|Y}(x|y)^{\alpha} \right) = \frac{\ln m}{\ln_{\alpha}(m)} \ H_{\alpha}(X|y).$$
(5.30)

In a similar manner, the inequality (5.30) leads to the claim (5.25).

The statement of Proposition 5.4 describes a relationship between the conditional Rényi entropy (2.24) and the conditional entropy (2.21) of THC type. Note that both the results (5.24) and (5.25) remain valid in the limit $\alpha \to 1$. Then we have a trivial relation with the standard conditional entropy. The relation (5.25) also allows to resolve the question mentioned right after (3.14).

For $\alpha \geq 1$, we have $\widetilde{H}_{\alpha}(X|Y) \leq H_{\alpha}(X)$. Combining this with (5.25) finally gives

$$R_{\alpha}(X|Y) \le \frac{\ln m}{\ln_{\alpha}(m)} \ln_{\alpha} \left\{ \exp\left(R_{\alpha}(X)\right) \right\},$$
(5.31)

where $\alpha \in (1, \infty)$. Here, we used the relations (2.3) and (2.10). The latter immediately follows from (2.1) and (2.8). The inequality (5.31) gives an upper bound on $R_{\alpha}(X|Y)$ in terms of only the quantities $R_{\alpha}(X)$ and $m = \#\Omega_X$. This bound may be useful in cases with enough small m. For instance, it is applicable in the binary case, which is of primary importance in information theory and practice. In the limit $\alpha \to 1^+$, the inequality (5.31) gives $H_1(X|Y) \leq H_1(X)$, *i.e.*, a particular case of (3.1).

6. Concluding Remarks

We have discussed properties of the conditional entropies of the Rényi and Tsallis-Havrda-Charvát types. Such entropies are both fruitful one-parameter extensions of the Shannon entropy. In particular, considered properties may be interesting in studying protocols of quantum information science. We examined entropic properties with respect to conditioning on more. It is natural to expect that conditioning on more can only reduce the entropy. We have exemplified that the mentioned property can be violated with the conditional min-entropy. A new definition of the TCH entropy rate is introduced and compared with the definition previously given in the literature. Some advances of the proposed notion are revealed. We examined lower and upper bounds on generalized conditional entropy in terms of the error probability corresponding to the standard decision. Relations between various conditional entropies are also of interest. We have obtained some interesting inequalities of such a kind. Information-theoretic methods are widely used in computer science, discrete mathematics, and quantum physics. There is a stable interest to applications of the Rényi and THC entropies in mentioned topics. The presented results may be used in extending scope of generalized entropies.

References

- N. Alon and J.H. Spencer, The Probabilistic Method, 3rd edition. John Wiley & Sons, Hoboken (2008).
- [2] D. Bacon and W. van Dam, Recent progress in quantum algorithms. Commun. ACM 53 (2010) 84–93.
- [3] M. Ben-Bassat and J. Raviv, Rényi's entropy and error probability. *IEEE Trans. Inf. Theory* 24 (1978) 324–331.
- [4] I. Bengtsson and K. Życzkowski, Geometry of Quantum States: An Introduction to Quantum Entanglement. Cambridge University Press, Cambridge (2006).
- [5] S.L. Braunstein and C.M. Caves, Information-theoretic Bell inequalities. *Phys. Rev. Lett.* 61 (1988) 662–665.
- [6] F. Buscemi, M.J.W. Hall, M. Ozawa and M.M. Wilde, Noise and disturbance in quantum measurements: An information-theoretic approach. *Phys. Rev. Lett.* **112** (2014) 050401.
- [7] C. Cachin, Entropy measures and unconditional security in cryptography. Ph.D. thesis, Swiss Federal Institute of Technology, Zürich (1997).
- [8] R. Chaves and T. Fritz, Entropic approach to local realism and noncontextuality. *Phys. Rev.* A 85 (2012) 032113.
- [9] R. Chaves and T. Fritz, Entropic inequalities and marginal problems. *IEEE Trans. Inf. Theory* 59 (2013) 803–817.
- [10] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. 2 (1967) 299–318.
- [11] I. Csiszár, Axiomatic characterizations of information measures. Entropy 10 (2008) 261–273.
- [12] T.M. Cover and J.A. Thomas, Elements of Information Theory. John Wiley & Sons, New York (1991).
- [13] Z. Daróczy, Generalized information functions. Inform. Control 16 (1970) 36-51.
- [14] Z. Daróczy, Inequalities for some infinite series. Acta Math. Hungar. 75 (1997) 5-8.

- [15] D. de Falco and D. Tamascelli, An introduction to quantum annealing. RAIRO: ITA 45 (2011) 99–116.
- [16] D. Erdogmus and J.C. Principe, Lower and upper bounds for misclassification probability based on Rényi's information. J. VLSI Signal Process. 37 (2004) 305–317.
- [17] R.M. Fano, Transmission of Information: A Statistical Theory of Communications. MIT Press and John Wiley & Sons, New York (1961).
- [18] M. Feder and N. Merhav, Relations between entropy and error probability. *IEEE Trans. Inf. Theory* 40 (1994) 259–266.
- [19] S. Furuichi, Information-theoretical properties of Tsallis entropies. J. Math. Phys. 47 (2006) 023302.
- [20] M. Gell-Mann and C. Tsallis, Nonextensive Entropy Interdisciplinary Applications. Oxford University Press, Oxford (2004).
- [21] I.M. Georgescu, S. Ashhab and F. Nori, Quantum simulation. Rev. Mod. Phys. 86 (2014) 153–185.
- [22] L. Golshani, E. Pasha and G. Yari, Some properties of Rényi entropy and Rényi entropy rate. Inf. Sci. 179 (2009) 2426–2433.
- [23] P. Harremoës and F. Topsøe, Inequalities between entropy and index of coincidence derived from information diagrams. *IEEE Trans. Inf. Theory* 47 (2001) 2944–2960.
- [24] J. Havrda and F. Charvát, Quantification methods of classification processes: concept of structural α-entropy. *Kybernetika* 3 (1967) 30–35.
- [25] F. Hiai, M. Mosonyi, D. Petz and C. Bény, Quantum f-divergences and error correction. *Rev. Math. Phys.* 23 (2011) 691–747.
- [26] S.-W. Ho and S. Verdú, On the interplay between conditional entropy and error probability. IEEE Trans. Inf. Theory 56 (2011) 5930–5942.
- [27] S.-W. Ho and R.W. Yeung, On the discontinuity of the Shannon information measures. *IEEE Trans. Inf. Theory* 55 (2009) 5362–5374.
- [28] P. Jizba and T. Arimitsu, The world according to Rényi: thermodynamics of multifractal systems. Ann. Phys. 312 (2004) 17–59
- [29] R. Kamimura, Minimizing α-information for generalization and interpretation. Algorithmica 22 (1998) 173–197.
- [30] S. Kullback and R. A. Leibler, On information and sufficiency. Ann. Math. Stat. 22 (1951) 79–86.
- [31] H. Maassen and J.B. M. Uffink, Generalized entropic uncertainty relations. *Phys. Rev. Lett.* 60 (1988) 1103–1106.
- [32] A.W. Marshall, I. Olkin and B.C. Arnold, Inequalities: Theory of Majorization and Its Applications, 2nd edition. Springer-Verlag, New York (2011).
- [33] M. Matić, C.E.M. Pearce and J. Pečarić, On an inequality for the entropy of a probability distribution. Acta Math. Hungar. 85 (1999) 345–349.
- [34] J. Radhakrishnan, An entropy proof of Brégman's theorem. J. Comb. Theory Ser. A 77 (1997) 161–164.
- [35] A.E. Rastegin, Convexity inequalities for estimating generalized conditional entropies from below. *Kybernetika* 48 (2012) 242–253
- [36] A.E. Rastegin, Bounds of the Pinsker and Fannes types on the Tsallis relative entropy. Math. Phys. Anal. Geom. 16 (2013) 213–228.
- [37] A.E. Rastegin, Uncertainty relations for MUBs and SIC-POVMs in terms of generalized entropies. Eur. Phys. J. D 67 (2013) 269.
- [38] A.E. Rastegin, Tests for quantum contextuality in terms of q-entropies. Quantum Inf. Comput. 14 (2014) 0996–1013.
- [39] A. Rényi, On measures of entropy and information, in Proc. of the 4th Berkeley Symposium on Mathematical Statistics and Probability, edited by J. Neyman. University of California Press, Berkeley CA (1961) 547–561.
- [40] A. Rényi, Statistics and information theory. Studia Sci. Math. Hungar. 2 (1967) 249–256.

- [41] W. Shitong and F.L. Chung, Note on the equivalence relationship between Rényi-entropy based and Tsallis-entropy based image thresholding. *Pattern Recognit. Lett.* 26 (2005) 2309– 2312.
- [42] A. Teixeira, A. Matos and L. Antunes, Conditional Rényi entropies. *IEEE Trans. Inf. Theory* 58 (2012) 4273–4277.
- [43] C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics. J. Stat. Phys. 52 (1988) 479–487.
- [44] I. Vajda, Bounds of the minimal error probability on checking a finite or countable number of hypotheses. Problemy Peredači Informacii 4 (1968) 9–19, (in Russian); translated as Probl. Inf. Transm. 4 (1968) 6–14.
- [45] K. Życzkowski, Rényi extrapolation of Shannon entropy. Open Sys. Inf. Dyn. 10 (2003) 297–310, corrigendum in arXiv:quant-ph/0305062v2.

Communicated by J. Kari.

Received December 7, 2013. Accepted December 18, 2014.