Alexey Rastegin

Convexity inequalities for estimating generalized conditional entropies from below

Kybernetika, Vol. 48 (2012), No. 2, 242--253

Persistent URL: http://dml.cz/dmlcz/142811

KYBERNETIKA — VOLUME *48* (2012), NUMBER 2, PAGES 242–253

CONVEXITY INEQUALITIES FOR ESTIMATING GENERALIZED CONDITIONAL ENTROPIES FROM BELOW

Alexey E. Rastegin

Generalized entropic functionals are in an active area of research. Hence lower and upper bounds on these functionals are of interest. Lower bounds for estimating Rényi conditional α-entropy and two kinds of non-extensive conditional α-entropy are obtained. These bounds are expressed in terms of error probability of the standard decision and extend the inequalities known for the regular conditional entropy. The presented inequalities are mainly based on the convexity of some functions. In a certain sense, they are complementary to generalized inequalities of Fano type.

Keywords: Rényi α-entropy, non-extensive entropy of degree α, error probability, Bayesian problems, functional convexity

Classification: 94E17, 60E15, 62C10, 39B62

1. INTRODUCTION

Entropy is one of most important concepts in both the information theory and statistical physics. Entropic quantities are also interesting mathematical subjects with many attractive properties. There exist several fruitful extensions of the Shannon entropy. One of them was proposed by Rényi [21]. Another one-parameter extension is the non-extensive entropy of degree α. Although such non-extensive entropy was first discussed by Havrda and Charvát [12] and by Daróczy [4], it became widely used in statistical mechanics after the seminal work of Tsallis [26]. Entropic functions of degree α form an especially helpful class of Csiszár's f-entropies (see the review [3] and references therein).

The Fano inequality, which bounds the conditional entropy from above, is essential to prove the converse to Shannon's second theorem [2]. This inequality has been extended to both the Rényi entropy [6] and non-extensive entropy of degree α [9, 19]. However, inequalities of such a kind provide only upper bound on adopted conditional entropies [8]. In quantum information theory, corresponding upper bounds are formulated for the entropy exchange [25] and its non-extensive extension [20]. For two-sided estimating of conditional entropies, lower bounds are necessary. Corresponding lower bounds were derived by Rényi [22, 23, 24] and, independently, by other authors [1, 8, 16, 28]. These bounds are mainly given in terms of error probability of the Bayesian approach to statistical decisions.

The aim of the present work is to extend the mentioned inequalities to both the Rényi and non-extensive entropies. In particular, the presented inequalities interpolate between the Rényi result [23] and the Vajda result [28]. The paper is organized as follows. In Section 2, the main definitions and auxiliary material are given. In effect, we recall the Rényi conditional entropy and two kinds of the non-extensive conditional entropy. The case of input alphabet with arbitrary (finite) cardinality is examined in Section 3. Lower bounds of desired type are derived for the Rényi conditional entropy of order $\alpha > 0$ and for some of the non-extensive conditional entropies. Assuming the binary input, lower bounds in terms of the error probability are given for all the considered entropic functionals in Section 4.

2. PRELIMINARIES

In this section, we recall definitions of the used measures of information. In the context of communications, we pose as follows. Let random variables X and Y describe the input and output of a channel. The variables X and Y take values on the finite sets (alphabets) Ω_X and Ω_Y, respectively. By $\#\Omega_X = m$ and $\#\Omega_Y = n$, we denote their cardinalities. We should decide on the input symbols when the output symbols are known. This is a typical problem of statistical decision theory. For probability distribution $\{p(x) :\ x \in \Omega_X\}$, the Rényi entropy of order $\alpha > 0$, $\alpha \neq 1$, is defined as [21]

$$R_\alpha(X) \triangleq \frac{1}{1-\alpha} \ln \left(\sum_{x \in \Omega_X} p(x)^\alpha \right) . \tag{1}$$

This quantity is a non-increasing function of α [21]. Other properties related to the parametric dependence are discussed in [29]. The non-extensive entropy of degree $\alpha > 0$, $\alpha \neq 1$, is defined by [26]

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha} \left(\sum_{x \in \Omega_X} p(x)^\alpha - 1 \right) . \tag{2}$$

With the factor $\left(2^{1-\alpha} - 1\right)^{-1}$ instead of $(1-\alpha)^{-1}$, this entropic function was derived from several axioms by Havrda and Charvát [12]. In a physical setting, the entropy (2) was introduced by Tsallis [26]. Following the paper [13], the entropy (2) will be referred to as "THC entropy". Both the Rényi and THC entropies are widely used for studying system, which involve long-range interactions, long-time memories, or fractal structures (see [10, 13] and references therein). These measures are also useful for expressing quantum uncertainties [17, 18]. Note that the entropy (2) can be rewritten as

$$H_\alpha(X) = -\sum\nolimits_{x \in \Omega_X} p(x)^\alpha \ln_\alpha p(x) , \tag{3}$$

where the α-logarithm $\ln_\alpha z = (z^{1-\alpha} - 1)/(1-\alpha)$ is defined for $\alpha > 0$, $\alpha \neq 1$ and $z > 0$. In the limit $\alpha \to 1$, we have $\ln_\alpha z \to \ln z$ and the Shannon entropy

$$H_1(X) = -\sum\nolimits_{x \in \Omega_X} p(x) \ln p(x) . \tag{4}$$

The entropy (1) also recovers the Shannon entropy, when $\alpha \to 1$. For brevity, we will usually omit the symbol of the set Ω_X in entropic sums.

In the present paper, we will mainly deal with conditional entropies. The regular conditional entropy is defined as [2]

$$H_1(X|Y) \triangleq \sum\nolimits_y p(y)\, H_1(X|y) = -\sum\nolimits_x \sum\nolimits_y p(x,y) \ln p(x|y) \ , \tag{5}$$

where $H_1(X|y) = -\sum_x p(x|y) \ln p(x|y)$ and $p(x|y) = p(x,y)/p(y)$ in line with the Bayes rule. The Rényi conditional entropy is defined by [6, 14]

$$R_\alpha(X|Y) \triangleq \sum\nolimits_y p(y)\, R_\alpha(X|y) \ , \tag{6}$$

where

$$R_\alpha(X|y) \triangleq \frac{1}{1-\alpha}\ \ln\Bigl(\sum\nolimits_x p(x|y)^\alpha\Bigr) \ . \tag{7}$$

In the literature, the following two kinds of THC conditional entropy are used. The first is defined as [9]

$$H_\alpha(X|Y) \triangleq \sum\nolimits_y p(y)^\alpha H_\alpha(X|y) \ , \tag{8}$$

where

$$H_\alpha(X|y) \triangleq \frac{1}{1-\alpha}\ \Bigl[\sum\nolimits_x p(x|y)^\alpha - 1\Bigr] = -\sum\nolimits_x p(x|y)^\alpha \ln_\alpha p(x|y) \ . \tag{9}$$

The conditional entropy (8) is, up to a factor, the quantity introduced by Daróczy [4]. Some of its functional properties are examined in [9]. Another kind of THC conditional entropy is put by [9]

$$\widetilde{H}_\alpha(X|Y) \triangleq \sum\nolimits_y p(y)\, H_\alpha(X|y) \ . \tag{10}$$

Taking $\alpha = 2$, we obtain the quadratic conditional entropy

$$\widetilde{H}_2(X|Y) = \sum\nolimits_y p(y)\,\Bigl[1 - \sum\nolimits_x p(x|y)^2\Bigr] \ . \tag{11}$$

Just the same entropic measure has been used by Vajda [28] for estimating the minimal error probability. In the limit $\alpha \to 1$, all the three entropies (6), (8), and (10) coincide with the regular conditional entropy (5).

While entropic functions are basic measures of uncertainty used in information theory, the channel coding theorems are usually stated in terms of the error probability [2]. So, relations between entropy and error probability are of interest [8]. Fano's inequality provide an upper bound that can be stated with error probability of arbitrary statistical decision. On the other hand, lower bounds on the conditional entropy (5) are typically expressed in terms of the error of so-called "standard" decision [22, 23, 24]. For given output value y of Y, we decide always in favor of that value $\hat{x}(y)$ of X which maximizes the conditional probability $p(x|y)$, i. e.

$$\hat{x}(y) \triangleq \operatorname{Arg\,max}\bigl\{p(x|y) :\ x \in \Omega_X\bigr\} \ , \qquad p(x|y) \leq p(\hat{x}|y)\ \ \forall\, x \in \Omega_X \ . \tag{12}$$

This decision is corresponding to the Bayesian approach [5]. Applications of Bayes rules in hypothesis testing still attract much attention [15]. The expected error probability P_e and the probability of successful estimation P_s are given by the well-known expressions

$$P_e = 1 - P_s \ , \qquad P_s = \sum\nolimits_y p(y)\, p(\hat{x}|y) \ . \tag{13}$$

It has been proved in very general setting that no decision can have a smaller error probability than the standard decision. This is the Bayesian version of the fundamental Neyman–Pearson lemma [23, 24]. In the context of communications, the value P_e is attained by the so-called "maximum *a posteriori* estimator" [8]. Our aim is to obtain lower bounds on the conditional entropies (6), (8), and (10) in terms of the probabilities P_e and P_s. Convexity properties of the considered functions will be used widely. So one of the basic tools is the Jensen inequality extensively treated, e. g., in chapter III of [11].

3. THE CASE OF ARBITRARY FINITE INPUT

In a series of papers (see [22, 23, 24] and references therein), Rényi obtained lower bounds on the standard conditional entropy (5) in terms of the error of the standard decision. When the cardinality of Ω_X is not specified, there holds [16, 23]

$$H_1(X|Y) \geq -\ln(1 - P_e) \ . \tag{14}$$

It turns out that this lower bound can be extended to some of the above generalized entropies. Let us start with the Rényi conditional entropy (6).

Theorem 3.1. For all $\alpha \in (0;\infty)$, the Rényi conditional entropy satisfies

$$R_\alpha(X|Y) \geq -\ln(1 - P_e) \ . \tag{15}$$

Proof. In view of $p(x|y) \leq p(\hat{x}|y)$ and $\sum_x p(x|y) = 1$, we write down

$$\sum\nolimits_x p(x|y)^\alpha = \sum\nolimits_x p(x|y)^{\alpha-1} p(x|y) \geq p(\hat{x}|y)^{\alpha-1} \qquad (0 < \alpha < 1) \ , \tag{16}$$

$$\sum\nolimits_x p(x|y)^\alpha = \sum\nolimits_x p(x|y)^{\alpha-1} p(x|y) \leq p(\hat{x}|y)^{\alpha-1} \qquad (1 < \alpha < \infty) \ . \tag{17}$$

The function $(1-\alpha)^{-1} \ln \xi$ is increasing for $\alpha \in (0;1)$ and decreasing for $\alpha \in (1;\infty)$. Combining the former with (16) and the latter with (17), the formula (7) leads to

$$R_\alpha(X|y) \geq \frac{1}{1-\alpha} \ \ln\bigl[p(\hat{x}|y)^{\alpha-1}\bigr] = -\ln p(\hat{x}|y) \ . \tag{18}$$

Since $[-\ln \xi]$ is a convex function, we substitute (18) in the definition (6) and obtain

$$R_\alpha(X|Y) \geq \sum\nolimits_y p(y)\bigl[-\ln p(\hat{x}|y)\bigr] \geq -\ln P_s \ , \tag{19}$$

by the Jensen inequality and (13). Hence the claim (15) is provided. □

As it is noted in the paper [14], we have $R_\alpha(X|Y) \geq H_1(X|Y)$ for $\alpha \in (0;1)$ and $H_1(X|Y) \geq R_\alpha(X|Y)$ for $\alpha \in (1;\infty)$. Hence the inequality (15) for $\alpha \in (0;1)$ could be derived directly from the inequality (14). For $\alpha \in (1;\infty)$, the inequality (15) is a new result. We have presented the joint proof for both the ranges of the parameter α. Let us proceed to the THC conditional entropies with the following statement.

Theorem 3.2. For all $\alpha \in (0;\infty)$, there holds

$$H_\alpha(X|y) \geq \ln_\alpha \frac{1}{p(\hat{x}|y)} \ . \tag{20}$$

Proof. It is easy to check that $\ln_\alpha \xi = -\xi^{1-\alpha} \ln_\alpha(1/\xi)$ and $\ln_\alpha(1/\xi)$ is decreasing function of ξ for all $\alpha > 0$. Combining these points with (9) finally gives

$$H_\alpha(X|y) = \sum_x p(x|y) \ln_\alpha \frac{1}{p(x|y)} \geq \ln_\alpha \frac{1}{p(\hat{x}|y)} \ , \tag{21}$$

in view of $p(x|y) \leq p(\hat{x}|y)$ and $\sum_x p(x|y) = 1$. □

Corollary 3.3. For all $\alpha \in (0;1)$, the conditional entropy (8) satisfies

$$H_\alpha(X|Y) \geq \bigl\{\max p(y)\bigr\}^{\alpha-1} \ln_\alpha \left(\frac{1}{1-P_e}\right) \ . \tag{22}$$

For all $\alpha \in (0;2]$, the conditional entropy (10) satisfies

$$\widetilde{H}_\alpha(X|Y) \geq \ln_\alpha \left(\frac{1}{1-P_e}\right) \ . \tag{23}$$

Proof. First, we will derive the inequality (23). Calculating the second derivative

$$\frac{d^2}{d\xi^2} \ \ln_\alpha \frac{1}{\xi} = (2-\alpha)\,\xi^{\alpha-3} \ , \tag{24}$$

we see that the function $\ln_\alpha(1/\xi)$ is convex for $\alpha \leq 2$. So we substitute (20) in the definition (10) and obtain

$$\widetilde{H}_\alpha(X|Y) \geq \sum\nolimits_y p(y) \ln_\alpha \frac{1}{p(\hat{x}|y)} \geq \ln_\alpha \frac{1}{P_s} \ , \tag{25}$$

where the Jensen inequality was used. This completes the proof of (23). We now note that for $\alpha < 1$ there holds $p(y)^{\alpha-1} \geq \{\max p(y)\}^{\alpha-1}$, whence

$$H_\alpha(X|Y) \geq \bigl\{\max p(y)\bigr\}^{\alpha-1} \widetilde{H}_\alpha(X|Y) \ . \tag{26}$$

Combining this with (23) finally gives (22). □

In the limit $\alpha \to 1$, the lower bounds (15) and (23) coincide with the original Rényi inequality (14). So we have extended the Rényi inequality to the adopted generalized entropies. For $\alpha = 2$, the inequality (23) becomes

$$\widetilde{H}_2(X|Y) \geq P_e \ . \tag{27}$$

This inequality was actually deduced by Vajda [28]. Note that some bounds of Fano type with the measure $\widetilde{H}_2(X|Y)$ were also presented in the paper [28]. We see that the lower bound (23) extends the Vajda result (27) as well and interpolates between it and the Rényi inequality (14).

For $\alpha > 2$, the function $\ln_\alpha(1/\xi)$ is concave and the sign in Jensen's inequality is reversed. So, another way of estimating the sums from below is required. One of useful approaches is to fit this concave function by a linear one. Let $f(\xi)$ be concave function such that $f(1)=0$, and let ξ be varied between ξ_0 and 1. For $\xi \in [\xi_0;1]$, we claim that

$$f(\xi) \geq f(\xi_0)\, \frac{1-\xi}{1-\xi_0} \ . \tag{28}$$

Indeed, the difference $\left\{f(\xi) - f(\xi_0)(1-\xi_0)^{-1}(1-\xi)\right\}$ is concave and, by construction, vanishes for both the points $\xi=\xi_0$ and $\xi=1$. Hence the difference is positive in the interval $[\xi_0;1]$ everywhere. Using this fact, we obtain the following result.

Corollary 3.4. Suppose that $\#\Omega_X = m$. For all $\alpha \in (2;\infty)$, the conditional entropy (10) satisfies

$$\widetilde{H}_\alpha(X|Y) \geq \frac{m \ln_\alpha m}{m-1}\, P_e \ . \tag{29}$$

P r o o f. For given y, we have m probabilities $p(x|y)$ such that $\sum_x p(x|y)=1$. Hence the maximum of them, namely $p(\hat{x}|y)$, is not less than $1/m$. Using (28) with $f(\xi)=\ln_\alpha(1/\xi)$ and $\xi_0 = 1/m$, it follows from (20) that

$$H_\alpha(X|y) \geq \frac{m \ln_\alpha m}{m-1}\, \big[1 - p(\hat{x}|y)\big] \ . \tag{30}$$

Combining this with (10) and (13) finally gives (29). □

In the limit $\alpha \to 2^+$, the lower bound (29) coincides with the lower bound (23). In this regard, the bound (29) is a proper continuation of the bound (23) to values $\alpha > 2$. Besides the error probability P_e, the inequality (29) contains the number m of symbols from the set Ω_X. Of course, this number is always known from the specification. Note that for $\alpha \in (0;1)$ we have the inequality expressed purely in terms of P_e, namely

$$H_\alpha(X|Y) \geq \ln_\alpha\!\left(\frac{1}{1-P_e}\right) \ . \tag{31}$$

It follows from (22) by $\max p(y) \leq 1$ and $\alpha - 1 < 0$. For the conditional entropy $H_\alpha(X|Y)$, some inequalities can also be derived for $\alpha > 1$. Combining (23) and (29) with the inequality

$$H_\alpha(X|Y) \geq \{\min p(y)\}^{\alpha-1} \widetilde{H}_\alpha(X|Y) \ , \tag{32}$$

which holds for $\alpha > 1$, we respectively have

$$H_\alpha(X|Y) \geq \{\min p(y)\}^{\alpha-1} \ln_\alpha\!\left(\frac{1}{1-P_e}\right) \qquad (1 < \alpha \leq 2) \ , \tag{33}$$

$$H_\alpha(X|Y) \geq \{\min p(y)\}^{\alpha-1}\, \frac{m \ln_\alpha m}{m-1}\, P_e \qquad (2 < \alpha < \infty) \ . \tag{34}$$

Note that the bounds (33) and (34) coincide for $\alpha = 2$. However, the scope of the lower bounds (33) and (34) is somewhat restricted, since the value of $\min p(y)$ is needed here. It would be of interest to get those bounds on $H_\alpha(X|Y)$ for $\alpha > 1$ that are expressed purely in terms of P_e (and, probably, $m = \#\Omega_X$ or $n = \#\Omega_Y$). In the following section, such a lower bound will be presented for the case of binary input. Nevertheless, the bounds (33) and (34) may sometimes be useful in a practice.

Finally, we shall present another lower bound on the conditional entropy (10). This bound does use the result (28) and does not (20). When the conditional entropies $p(x|y)$ can be estimated from below, we apply the following statement.

Theorem 3.5. Let $\#\Omega_X = m$, and let M be the positive real number such that

$$\min\bigl\{p(x|y) :\ x \in \Omega_X,\ y \in \Omega_Y\bigr\} = M^{-1}\,. \tag{35}$$

For all $\alpha \in (0; \infty)$, the conditional entropy (10) satisfies

$$\widetilde{H}_\alpha(X|Y) \geq \frac{m \ln_\alpha M}{M-1}\ P_e\ . \tag{36}$$

P r o o f. Applying (28) with the concave function $f(\xi) = \bigl(\xi^\alpha - \xi\bigr)/(1-\alpha)$ and $\xi_0 = 1/M$, one obtains

$$\frac{p(x|y)^\alpha - p(x|y)}{1-\alpha} \geq \frac{M^{-\alpha} - M^{-1}}{1-\alpha}\ \frac{1-p(x|y)}{1-M^{-1}} \geq \frac{\ln_\alpha M}{M-1}\bigl[1 - p(\hat{x}|y)\bigr]\ . \tag{37}$$

Hence we have $H(X|y) \geq m \ln_\alpha(M)(M-1)^{-1}\bigl[1-p(\hat{x}|y)\bigr]$ by summing (37) with respect to $x \in \Omega_X$. Substituting this inequality into the right-hand side of (10) completes the proof in view of (13). $\square$

Since $\sum_x p(x|y) = 1$ and $\#\Omega_X = m$, we have $\min\bigl\{p(x|y) :\ x \in \Omega_X\bigr\} \leq 1/m$ for any y, whence $m \leq M$. The lower bound (36) is similar to (29) in structure and almost coincides with the one, when $M \approx m$. At the same time, the bound (36) is valid in wider parametric range, including the standard case $\alpha = 1$. For very large M, however, we rather prefer the bounds (23) and (29). Note that the lower bound (36) can also be combined with (26) for $0 < \alpha < 1$ and with (32) for $\alpha > 1$.

4. THE CASE OF BINARY INPUT

The two symbols, usually “0” and “1”, are quite sufficient for almost all tasks of storage, transmission and protection of information. So the case of binary input set is of great importance in information theory and practice. Rényi pointed out [23] that the lower bound (14) can somewhat be refined in this case. Namely, for $\#\Omega_X = 2$ there holds (see also [1, 27])

$$H_1(X|Y) \geq (2\ln 2)P_e\ . \tag{38}$$

[When the logarithms in (4) and (5) are taken to the base two, the $\ln 2$ should be left out from (38).] We aim to generalize the above inequality to the conditional entropies (6), (8), and (10). First, we prove one auxiliary statement.

Lemma 4.1. Let us define the function $\eta_\alpha(z) \triangleq z^\alpha + (1-z)^\alpha$. For $z \in [0;1/2]$, there holds

$$\eta_\alpha(z) \geq 1 + 2\,\ln_\alpha(2)\,(1-\alpha)\,z \qquad (0<\alpha<1)\ , \tag{39}$$
$$\eta_\alpha(z) \leq 1 + 2\,\ln_\alpha(2)\,(1-\alpha)\,z \qquad (1<\alpha<\infty)\ . \tag{40}$$

Proof. By calculation, $\eta_\alpha(0) = 1$ and $\eta_\alpha(1/2) = 2^{1-\alpha} = 1 + (1-\alpha)\ln_\alpha 2$. Hence the difference $\bigl\{\eta_\alpha(z) - 1 - 2\,\ln_\alpha(2)\,(1-\alpha)\,z\bigr\}$ vanishes for both the points $z=0$ and $z=1/2$. Further, this difference is concave for $\alpha \in (0;1)$ and convex for $\alpha \in (1;\infty)$. So it is positive in the former and negative in the latter. □

Theorem 4.2. Suppose that $\#\Omega_X = 2$. For all $\alpha \in (0;1)$, the conditional entropy (6) satisfies

$$R_\alpha(X|Y) \geq (2\,\ln 2) P_e\ . \tag{41}$$

For all $\alpha \in (1;\infty)$, the conditional entropy (6) satisfies

$$R_\alpha(X|Y) \geq \frac{1}{1-\alpha}\ \ln\Bigl[1 + 2\,\ln_\alpha(2)\,(1-\alpha)\,P_e\Bigr]\ . \tag{42}$$

Proof. For $\Omega_X = \{x_0, x_1\}$ and given y, we put the minimal probability $q(\hat{x}|y) = 1 - p(\hat{x}|y) = \min\bigl\{p(x_0|y), p(x_1|y)\bigr\}$. It is clear that $0 \leq q(\hat{x}|y) \leq 1/2$ and

$$P_e = \sum\nolimits_y p(y)\,q(\hat{x}|y)\ . \tag{43}$$

The function $(1-\alpha)^{-1}\ln\xi$ increasing for $\alpha \in (0;1)$ and decreasing for $\alpha \in (1;\infty)$. Combining the former with (39) and the latter with (40), the formula (7) gives

$$R_\alpha(X|y) = \frac{1}{1-\alpha}\ \ln\eta_\alpha\bigl(q(\hat{x}|y)\bigr) \geq \frac{1}{1-\alpha}\ \ln\Bigl[1 + 2\,\ln_\alpha(2)\,(1-\alpha)\,q(\hat{x}|y)\Bigr]\ . \tag{44}$$

For $\alpha \in (1;\infty)$, the function $(1-\alpha)^{-1}\ln\xi$ is convex. Substituting (44) into the definition (6) and using the Jensen inequality, we get the claim (42) due to (43). In view of the concavity of $(1-\alpha)^{-1}\ln\xi$ for $\alpha \in (0;1)$, we now take some analog of (28). Namely, if concave function $f(\xi)$ obeys $f(1) = 0$ then

$$f(\xi) \geq f(\xi_1)\ \frac{\xi - 1}{\xi_1 - 1} \tag{45}$$

for each $\xi \in [1;\xi_1]$. By $0 \leq q(\hat{x}|y) \leq 1/2$, the term $\xi = 1 + 2\,\ln_\alpha(2)\,(1-\alpha)\,q(\hat{x}|y)$ certainly lies in the interval $[1;\xi_1]$ with $\xi_1 = 1 + \ln_\alpha(2)\,(1-\alpha) = 2^{1-\alpha}$. By relevant substitutions, we then get

$$R_\alpha(X|Y) \geq \frac{\ln\bigl(2^{1-\alpha}\bigr)}{1-\alpha}\ \sum\nolimits_y p(y)\ \frac{2\,\ln_\alpha(2)\,(1-\alpha)\,q(\hat{x}|y)}{2^{1-\alpha}-1}\ , \tag{46}$$

that is merely reduced just to the claim (41). □

The inequality (41) is actually a corollary of the lower bound (38), since the relation $R_\alpha(X|Y) \geq H_1(X|Y)$ takes place for $\alpha \in (0;1)$. We have given the proof as it requires only a few lines. In view of $H_1(X|Y) \geq R_\alpha(X|Y)$ for $\alpha \in (1;\infty)$, no conclusions follow here from (38). So the lower bound (42) is a new result. Because of $1 - \alpha < 0$, the inequality (42) leads to the bound

$$R_\alpha(X|Y) \geq (2\,\ln_\alpha 2) P_e\ , \tag{47}$$

which is linear in the error probability P_e. In a certain sense, both the lower bounds (42) and (47) are proper one-parametric extensions of the original bound (38). We shall now deal with the THC conditional entropies.

Theorem 4.3. Suppose that $\#\Omega_X = 2$. For all $\alpha \in (0;1)$, the conditional entropy (8) satisfies

$$H_\alpha(X|Y) \geq \bigl\{\max p(y)\bigr\}^{\alpha-1} 2\,\ln_\alpha(2)\, P_e\ . \tag{48}$$

For all $\alpha \in (0;\infty)$, the conditional entropy (10) satisfies

$$\widetilde{H}_\alpha(X|Y) \geq (2\,\ln_\alpha 2) P_e\ . \tag{49}$$

P r o o f. First, we will prove (49). The function $(1-\alpha)^{-1}\xi$ increases with ξ for $\alpha \in (0;1)$ and decreases with ξ for $\alpha \in (1;\infty)$. Combining the former with (39) and the latter with (40), the formula (9) leads to

$$H_\alpha(X|y) = \frac{1}{1-\alpha}\left[\eta_\alpha\bigl(q(\hat{x}|y)\bigr) - 1\right] \geq 2\,\ln_\alpha(2)\, q(\hat{x}|y)\ . \tag{50}$$

Substituting this point into (10) and summing with respect to y, we at once get (49). Like the inequality (22), the claim (48) follows from (26) and (49). □

The lower bound (49) is a proper extension of the inequality (38) to the conditional entropy (10) for all the parameter values $\alpha \in (0;\infty)$. Note also that this bound is a special case of (29) for $\alpha > 2$. Similar to the bound (22), the inequality (48) leads to the lower bound

$$H_\alpha(X|Y) \geq (2\,\ln_\alpha 2) P_e\ , \tag{51}$$

which is expressed purely in terms of the error probability P_e. Its scope, however, is restricted to the values $\alpha \in (0;1)$. Of course, for $\alpha > 1$ we can write

$$H_\alpha(X|Y) \geq \{\min p(y)\}^{\alpha-1}\, 2\,\ln_\alpha(2)\, P_e\ , \tag{52}$$

due to (32) and (49). But the formula (52) assumes that the value of $\min p(y)$ is known. We shall now present a lower bound without such a request.

Theorem 4.4. Suppose that $\#\Omega_X = 2$ and $\#\Omega_Y = n$. For all $\alpha \in (1;\infty)$, the conditional entropy (8) satisfies

$$H_\alpha(X|Y) \geq 2^\alpha \ln_\alpha(2)\, n^{1-\alpha} P_e^\alpha\ . \tag{53}$$

P r o o f. We firstly note that for $z \in [0;1/2]$ and $\alpha > 1$, there holds

$$1 - \eta_\alpha(z) = 1 - (1-z)^\alpha - z^\alpha \geq \gamma z^\alpha \ , \tag{54}$$

where the factor $\gamma = 2^\alpha - 2 > 0$. Indeed, the function $1-(1-z)^\alpha - (1+\gamma)z^\alpha$ is concave for $\alpha > 1$ and vanishes for both the points $z = 0$ and $z = 1/2$, i. e. it is positive in the interval $z \in [0;1/2]$. Then we write

$$H_\alpha(X|y) = \frac{1}{\alpha - 1} \Big[1 - \eta_\alpha\big(q(\hat{x}|y)\big)\Big] \geq 2^\alpha \ln_\alpha(2)\, q(\hat{x}|y)^\alpha \ , \tag{55}$$

since $(\alpha-1)^{-1}\gamma = 2^\alpha \ln_\alpha 2$. Combining (8) and (55), for $\alpha > 1$ we have

$$H_\alpha(X|Y) \geq 2^\alpha \ln_\alpha(2) \sum_{y=1}^{n} \big[p(y)\, q(\hat{x}|y)\big]^\alpha \geq 2^\alpha \ln_\alpha(2)\, n^{1-\alpha} \left[\sum_{y=1}^{n} p(y)\, q(\hat{x}|y)\right]^\alpha \tag{56}$$

that is merely (53) in view of (43). The right-hand side of (56) has been inserted due to the inequality

$$\frac{1}{n} \sum\nolimits_{y=1}^{n} a(y)^\alpha \geq \left[\frac{1}{n} \sum\nolimits_{y=1}^{n} a(y)\right]^\alpha \ , \tag{57}$$

which is valid for all $\alpha > 1$ as a special case of the Jensen inequality. $\square$

The lower bound (53) is expressed in terms of P_e and the number $n = \#\Omega_Y$; the latter is always known from the context. The inequality (57) is close to equality, when all the $a(y)$'s are close to each other. If the quantities $p(y)\, q(\hat{x}|y)$ are almost equal for all y, the lower bound (53) seems to be sufficiently tight. In the completely binary case, when $n = 2$ as well, we have

$$H_\alpha(X|Y) \geq (2 \ln_\alpha 2) P_e^\alpha \ . \tag{58}$$

In the limit $\alpha \to 1^+$, both the inequalities (53) and (58) coincide with the inequality (38). Thus, we have obtained a proper extension of the lower bound (38) to the conditional entropy (8) for parameter values $\alpha \in (1;\infty)$.

(Received April 15, 2011)

REFERENCES

[1] L. Baladová: Minimum of average conditional entropy for given minimum probability of error. Kybernetika *2* (1966), 416–422.

[2] T. Cover and J. Thomas: Elements of Information Theory. John Wiley & Sons, New York 1991.

[3] I. Csiszár: Axiomatic characterizations of information measures. Entropy *10* (2008), 261–273.

[4] Z. Daróczy: Generalized information functions. Inform. and Control *16* (1970), 36–51.

[5] M. H. DeGroot: Optimal Statistical Decisions. McGraw-Hill, New York 1970.

[6] 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.

[7] R. M. Fano: Transmission of Information: A Statistical Theory of Communications. MIT Press and John Wiley & Sons, New York 1961.

[8] M. Feder and N. Merhav: Relations between entropy and error probability. IEEE Trans. Inform. Theory *40* (1994), 259–266.

[9] S. Furuichi: Information theoretical properties of Tsallis entropies. J. Math. Phys. *47* (2006), 023302.

[10] M. Gell-Mann and C. Tsallis, eds.: Nonextensive Entropy – Interdisciplinary Applications. Oxford University Press, Oxford 2004.

[11] G. H. Hardy, J. E. Littlewood, and G. Polya: Inequalities. Cambridge University Press, London 1934.

[12] J. Havrda and F. Charvát: Quantification methods of classification processes: concept of structural α-entropy. Kybernetika *3* (1967), 30–35.

[13] P. Jizba and T. Arimitsu: The world according to Rényi: thermodynamics of multifractal systems. Ann. Phys. *312* (2004), 17–59.

[14] R. Kamimura: Minimizing α-information for generalization and interpretation. Algorithmica *22* (1998), 173–197.

[15] A. Novikov: Optimal sequential procedures with Bayes decision rules. Kybernetika *46* (2010), 754–770.

[16] A. Perez: Information-theoretic risk estimates in statistical decision. Kybernetika *3* (1967), 1–21.

[17] A. E. Rastegin: Rényi formulation of the entropic uncertainty principle for POVMs. J. Phys. A: Math. Theor. *43* (2010), 155302.

[18] A. E. Rastegin: Entropic uncertainty relations for extremal unravelings of superoperators. J. Phys. A: Math. Theor. *44* (2011), 095303.

[19] A. E. Rastegin: Continuity estimates on the Tsallis relative entropy. E-print arXiv:1102.5154v2 [math-ph] (2011).

[20] A. E. Rastegin: Fano type quantum inequalities in terms of q-entropies. Quantum Information Processing (2011), doi 10.1007/s11128-011-0347-6.

[21] A. Rényi: On measures of entropy and information. In: Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley – Los Angeles 1961, pp. 547–561.

[22] A. Rényi: On the amount of missing information in a random variable concerning an event. J. Math. Sci. *1* (1966), 30–33.

[23] A. Rényi: Statistics and information theory. Stud. Sci. Math. Hung. *2* (1967), 249–256.

[24] A. Rényi: On some basic problems of statistics from the point of view of information theory. In: Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley – Los Angeles 1967, pp. 531–543.

[25] B. Schumacher: Sending entanglement through noisy quantum channels. Phys. Rev. A *54* (1996), 2614–2628.

[26] C. Tsallis: Possible generalization of Boltzmann–Gibbs statistics. J. Stat. Phys. *52* (1988), 479–487.

[27] I. Vajda: On the statistical decision problem with discrete paprameter space. Kybernetika *3* (1967), 110–126.

[28] I. Vajda: Bounds of the minimal error probability on checking a finite or countable number of hypotheses. Problemy Peredachii Informacii *4* (1968), 9–19 (in Russian); translated as Problems of Information Transmission *4* (1968), 6–14.

[29] K. Życzkowski: Rényi extrapolation of Shannon entropy. Open Sys. Inform. Dyn. *10* (2003), 297–310; corrigendum in the e-print version arXiv:quant-ph/0305062v2.

Alexey Rastegin, Department of Theoretical Physics, Irkutsk State University, Gagarin Bv. 20, Irkutsk 664003. Russia.

e-mail: rast@api.isu.ru