HAL
open science

On the Non-robustness of Essentially Conditional Information Inequalities

Tarik Kaced, Andrei Romashchenko

To cite this version:

Tarik Kaced, Andrei Romashchenko. On the Non-robustness of Essentially Conditional Information Inequalities. ITW'12: Information Theory Workshop, Sep 2012, Switzerland. pp.1935-1939. lirmm-00736192

HAL Id: lirmm-00736192

https://hal-lirmm.ccsd.cnrs.fr/lirmm-00736192v1

Submitted on 27 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Non-robustness of Essentially Conditional Information Inequalities

Tarik Kaced
LIRMM & Univ. Montpellier 2
Email: tarik.kaced@lirmm.fr

Andrei Romashchenko
LIRMM, CNRS & Univ. Montpellier 2
On leave from IITP, Moscow
Email: andrei.romashchenko@lirmm.fr

***Abstract*—We show that two essentially conditional linear inequalities for Shannon's entropies (including the Zhang–Yeung'97 conditional inequality) do not hold for asymptotically entropic points. This means that these inequalities are non-robust in a very strong sense. This result raises the question of the meaning of these inequalities and the validity of their use in practice-oriented applications.**

I. Introduction

Following Pippenger [15] we can say that the most basic and general "laws of information theory" can be expressed in the language of information inequalities (inequalities which hold for the Shannon entropies of jointly distributed tuples of random variables for every distribution). The very first examples of information inequalities were proven (and used) in Shannon's seminal papers in the 1940s. Some of these inequalities have a clear intuitive meaning. For instance, the entropy of a pair of jointly distributed random variables a, b is not greater than the sum of the entropies of the marginal distributions, i.e., $H(a, b) \leq H(a) + H(b)$. In standard notations, this inequality means that the mutual information between a and b is non-negative, $I(a{:}b) \geq 0$; this inequality becomes an equality if and only if a and b are independent in the usual sense of probability theory. These properties have a very natural meaning: a pair cannot contain more "uncertainty" than the sum of "uncertainties" in both components. This basic statement can be easily explained, e.g., in term of standard coding theorems: the average length of an optimal code for a distribution (a, b) is not greater than the sum of the average lengths for two separate codes for a and b. Another classic information inequality $I(a{:}b|c) \geq 0$ is slightly more complicated from the mathematical point of view, but is also very natural and intuitive. Inequalities of this type are called basic Shannon's inequality, [19].

We believe that the success of Shannon's information theory in a myriad of applications (in engineering and natural sciences as well as in mathematics and computer science) is due to the intuitive simplicity and natural interpretations of the very basic properties of Shannon's entropy.

Formally, information inequalities are just a dual description of the set of all entropy profiles. That is, for every joint distribution of an n-tuple of random variables we have a vector of $2^n - 1$ ordered entropies (entropies of all random variables involved, entropies of all pairs, triples, of quadruples, etc. in some fixed order). A vector in $\mathbb{R}^{2^n-1}$ is called *entropic* if it represents entropy values of some distribution. The fundamental (and probably very difficult) problem is to describe the set of entropic vectors for all n. It is known, see [20], that for every n the closure of the set of all entropic vectors is a convex cone in $\mathbb{R}^{2^n-1}$. The points that belong to this closure are called *asymptotically entropic* or *asymptotically constructible* vectors, [12], say a.e. vectors for short. The class of all linear information inequalities is exactly the dual cone to the set of a.e. vectors. In [15] and [5] a natural question was raised: *What is the class of all universal information inequalities?* (Equivalently, how to describe the cone of a.e. vectors?) More specifically, does there exist any linear information inequality that cannot be represented as a convex combination of Shannon's basic inequality?

In 1998 Z. Zhang and R.W. Yeung came up with the first example of a *non-Shannon-type* information inequality [21]:

$$I(c{:}d) \leq 2I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(a{:}c|d) + I(a{:}d|c).$$

This unexpected result raised other challenging questions: *What does this inequality mean? How to understand it intuitively?* Although we still do not know a complete and comprehensive answer to the last questions, we have several interpretations and explanations of this inequality. Some information-theoretic interpretations were discussed, e.g., in [17], [22]. This inequality is closely related to *Ingleton's inequality* for ranks of linear spaces, [3], [6], [12]. This connection was explained by F. Matúš in his paper [11], where the connection between information inequalities and polymatroids was established. Matúš proved that a polymatroid with the ground set of cardinality 4 is selfadhesive if and only if it satisfies the Zhang–Yeung inequality formulated above (more precisely, a polymatroid must satisfy all possible instances of this inequality for different permutations of variables).

Thus, the inequality from [21] has some explanations and intuitive interpretations. However, another type of inequalities is still much less understood. We mean other "universal laws of information theory", those that can be expressed as *conditional linear information inequalities* (linear inequalities for entropies which are true for distributions whose entropies satisfy some linear constraints; they are also called in the literature *constrained information inequalities*, see [19]). We do not give a general definition of a "conditional linear information inequality" since the entire list of all known

nontrivial inequalities in this class is very short. Here are three of them:

(1) [20]: if $I(a{:}b|c) = I(a{:}b) = 0$, then

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b),$$

(2) [9]: if $I(a{:}b|c) = I(b{:}d|c) = 0$, then

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b),$$

(3) [7]: if $I(a{:}b|c) = H(c|a,b) = 0$, then

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b).$$

It is known that (1-3) are "essentially conditional", i.e., they cannot be extended to any unconditional inequalities, [7], e.g., for (1) this means that for any values of "Lagrange multipliers" λ_1, λ_2 the corresponding unconditional extension

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + \lambda_1 I(a{:}b) + \lambda_2 I(a{:}b|c)$$

does not hold for some distributions (a, b, c, d). In other words, (1-3) make some very special kind of "information laws": they cannot be represented as "shades" of any unconditional inequalities on the subspace corresponding to their linear constraints.

A few other nontrivial conditional information inequalities can be obtained from the results of F. Matúš in [9]. For example, Matúš proved that for every integer $k > 0$ and for all (a, b, c, d)

$$(*) \quad \begin{aligned} I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \frac{1}{k} I(c{:}d|a) \\ + \frac{k+1}{2}(I(a{:}c|d) + I(a{:}d|c)) \end{aligned}$$

(this is a special case of theorem 2 in [9]). Assume that $I(a{:}c|b) = I(b{:}c|a) = 0$. Then, as $k \to \infty$ we get from (*) another conditional inequality:

(4) if $I(a{:}c|d) = I(a{:}d|c) = 0$, then

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b).$$

It can be proven that (4) is also an essentially conditional inequality, i.e., whatever are the coefficients λ_1, λ_2,

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \lambda_1 I(a{:}c|d) + \lambda_2 I(a{:}d|c)$$

does not hold for some distribution (a, b, c, d).

Since (*) holds for a.e. vectors, (4) is also true for a.e. vectors. Inequality (4) is robust in the following sense. Assume that entropies of all variables involved are bounded by some h. Then for every $\varepsilon > 0$ there exists a $\delta = \delta(h, \varepsilon)$ such that

if $I(a{:}c|d) \le \delta$ and $I(a{:}d|c) \le \delta$, then

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \varepsilon$$

(note that δ is not linear in ε). In this paper we prove that this is not the case for (1) and (3) – these inequalities do not hold for a.e. vectors, and they are not robust. So, these inequalities are, in some sense, similar to the nonlinear (piecewise linear) conditional information inequality from [10].

Together with [7], where (1–3) are proven to be essentially conditional, our result indicates that (1) and (3) are very fragile and non-robust properties of entropies. We cannot hope that similar inequalities hold when the constraints become soft. For instance, assuming that $I(a{:}b)$ and $I(a{:}b|c)$ are "very small" we cannot say that

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b)$$

holds also with only "a small error"; even a negligible deviation from the conditions in (1) can result in a dramatic effect $I(c{:}d) \gg I(c{:}d|a) + I(c{:}d|b)$.

Conditional information inequalities (in particular, inequality (2)) were used in [9] to describe conditional independences among several jointly distributed random variables. Conditional independence is known to have wide applications in statistical theory (including methods of parameter identification, causal inference, data selection mechanisms, etc.), see, e.g., surveys in [2], [16]. We are not aware of any direct or implicit *practical* usage of (1-3), but it would not be surprising to see such usages in the future. However, our results indicate that these inequalities are non-robust and therefore might be misleading in practice-oriented applications.

The rest of the paper is organized as follows. We provide a new proof of why two conditional inequalities (1) and (3) are essentially conditional. This proof uses a simple algebraic example of random variables. Then, we show that (1) and (3) are not valid for a.e. vectors, leaving the question for (2) open.

II. Why "essentially conditional" : an algebraic counterexample

Consider the quadruple $(a, b, c, d)_q$ of geometric objects, resp. $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$, on the affine plane over the finite field $\mathbb{F}_q$ defined as follows :

- First choose a random non-vertical line $\mathcal{C}$ defined by the equation $y = c_0 + c_1 x$ (the coefficients c_0 and c_1 are independent random elements of the field);
- pick points $\mathcal{A}$ and $\mathcal{B}$ on $\mathcal{C}$ independently and uniformly at random (these points coincide with probability $1/q$);
- then pick a parabola $\mathcal{D}$ uniformly at random in the set of all non-degenerate parabolas $y = d_0 + d_1 x + d_2 x^2$ (where $d_0, d_1, d_2 \in \mathbb{F}_q, d_2 \ne 0$) that intersect $\mathcal{C}$ at $\mathcal{A}$ and $\mathcal{B}$; (if $\mathcal{A} = \mathcal{B}$ we require that $\mathcal{C}$ is a tangent line to $\mathcal{D}$). When $\mathcal{C}$ and $\mathcal{A}, \mathcal{B}$ are chosen, there exist $(q-1)$ different parabolas $\mathcal{D}$ meeting these conditions.

A typical quadruple is represented on Figure 1.

Remark 1. *This picture is not strictly accurate, for the plane is discrete, but helps grasping the general idea since the relevant properties used are also valid in the continuous case.*

Let us now describe the entropy profile of this quadruple.

- Every single random variable is uniform over its support.
- The line and the parabola share some mutual information, (the fact that they intersect) which is approximately one bit. Indeed, $\mathcal{C}$ and $\mathcal{D}$ intersect iff the corresponding equation discriminant is a quadratic residue, which happens

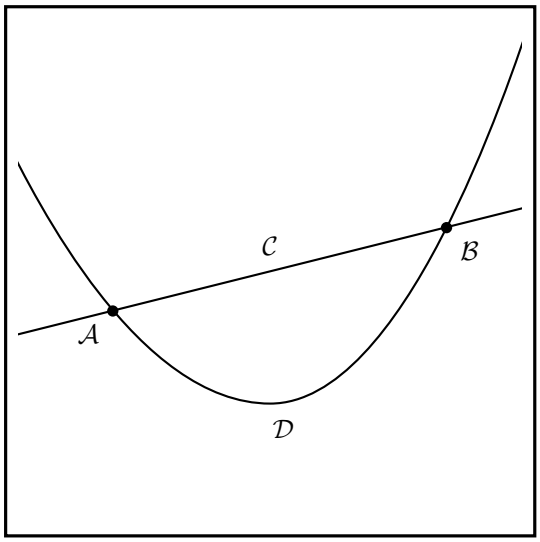

Fig. 1. An algebraic example

almost half of the time.

$$I(c{:}d) = \frac{q-1}{q}$$

- When an intersection point is given, the line does not give more information about the parabola.

$$I(c{:}d|a) = I(c{:}d|b) = 0$$

- When the line is known, an intersection point does not help knowing the other (by construction).

$$I(a{:}b|c) = 0$$

- The probability that there is only one intersection point is $1/q$. In that case, the line can be any line going through this point.

$$I(a{:}b) = H(c|a,b) = \frac{\log_2 q}{q}$$

Now we plug the computations into the following inequalities

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + \lambda_1 I(a{:}b) + \lambda_2 I(a{:}b|c)$$

or

$$I(c{:}d) \le I(c{:}d|a)+I(c{:}d|b)+I(a{:}b)+\lambda_1 I(a{:}b|c)+\lambda_2 H(c|a,b),$$

which are "unconditional" counterparts of (1) and (3) respectively. For every constants λ_1, λ_2 we get

$$1 - \frac{1}{q} \le (\lambda_1 + \lambda_2)\frac{\log_2 q}{q}$$

and conclude they can not hold when q is large. Thus, we get the following theorem (originally proven in [7]):

Theorem 1. *Inequalities (1) and (3) are essentially conditional.*

III. Why (1) and (3) do not hold for a.e. vectors

We are going to use the previous example to show that conditional inequalities (1) and (3) are not valid for asymptotically entropic vectors. We will use the Slepian–Wolf coding theorem (*cf.* [18]) as our main tool.

Lemma 1 (Slepian–Wolf)**.** *Let (x, y) be joint random variables and (X, Y) be N independent copies of this distribution. Then there exists X' such that $H(X'|X) = 0$, $H(X') = H(X|Y) + o(N)$ and $H(X|X', Y) = o(N)$.*

This lemma constructs a hash of a random variable X which is almost independent of Y and has approximately the entropy of X given Y. We will say that X' is the Slepian–Wolf hash of X given Y and write $X' = SW(X|Y)$.

In what follows we call by the *entropy profile* of $(x_1, \dots, x_n)$ the vector of entropies for all non-empty subset of these random variable in the lexicographic order. We denote it

$$\vec{H}(x_1, \dots, x_n) = (H(\mathcal{S}))_{\emptyset \ne \mathcal{S} \subseteq \{x_1, \dots, x_n\}}.$$

This is a vector in $\mathbb{R}^{2^n-1}$ (dimension is equal to the number of nonempty subsets in the set of n elements).

Theorem 2. *(1) and (3) are not valid for a.e. vectors.*

Proof: For each given inequality, we construct an asymptotically entropic vector which excludes it. The main step is to ensure, via Slepian–Wolf lemma, that the constraints are met.

a) An a.e. counterexample for (1):

1. Start with the quadruple $(a, b, c, d)_q$ from the previous section for some fixed q to be defined later. Notice that it does not satisfy the constraints.
2. Serialize it: define a new quadruple (A, B, C, D) such that each entropy is N times greater. (A, B, C, D) is obtained by sampling N times independently (a_i, b_i, c_i, d_i) according to the distribution (a, b, c, d) and letting, e.g., $A = (a_1, a_2, \dots, a_N)$.
3. Apply Slepian–Wolf lemma to get $A' = SW(A|B)$ such that $I(A'{:}B) = o(N)$, and replace A by A' in the quadruple. The entropy profile of (A', B, C, D) cannot vary much from the profile of (A, B, C, D). More precisely, entropies for A', B, C, D differ from the corresponding entropies for A, B, C, D by at most $I(A{:}B) + o(N) = O\left(\frac{\log_2 q}{q} N\right)$. Notice that $I(A'{:}B|C) = 0$ since A' functionally depends on A and $I(a{:}b|c) = 0$.
4. Scale down the entropy profile of (A', B, C, D) by a factor of $1/N$. This operation can be done within a precision of, say, $o(N)$. Basically, this can be done because the set of all a.e. points is convex (see, e.g., [19])
5. Tend N to infinity to define an a.e. vector. This limit vector is **not** an entropic vector. For this a.e. vector, inequality (1) does not hold when q is large. Indeed $I(A{:}B)/N$ and $I(A{:}B|C)/N$ both approaches zero as N tends to infinity. On the other hand, for the resulting limit vector, inequality (1) turns into

$$1 + O\left(\frac{\log_2 q}{q}\right) \le O\left(\frac{\log_2 q}{q}\right),$$

which can not hold if q is bigger than some constant.

b) An a.e. counterexample for (3): We start with another lemma based on the Slepian–Wolf coding theorem.

Lemma 2. *For every distribution (a, b, c, d) and every integer N there exists a distribution (A', B', C', D') such that*

- $H(C'|A', B') = o(N)$,
- *The difference between corresponding components of the entropy profile $\vec{H}(A', B', C', D')$ and $N \cdot \vec{H}(a, b, c, d)$ is at most $N \cdot H(c|a, b) + o(N)$.*

Proof: First we serialize (a, b, c, d), i.e., we take M i.i.d. copies of the initial distribution. The result of this serialization is a distribution (A, B, C, D) whose entropy profile is the exactly the entropy profile of (a, b, c, d) multiplied by M. In particular, we have $I(A{:}B|C) = 0$. Then, we apply Slepian–Wolf encoding (Lemma 1) and get a $Z = SW(C|A, B)$ such that

- $H(Z|C) = 0$,
- $H(Z) = H(C|A, B) + o(M)$,
- $H(C|A, B, Z) = o(M)$.

The entropy profile of the conditional distribution of (A, B, C, D) given Z differs from then entropy profile of (A, B, C, D) by at most $H(Z) = M \cdot H(c|a, b) + o(M)$. Also, if in the original distribution $I(a{:}b|c) = 0$, then $I(A{:}B|C, Z) = I(A{:}B|C) = 0$.

We would like to "relativize" (A, B, C, D) conditional on Z and get a new distribution for a quadruple (A', B', C', D') whose unconditional entropies are equal to the corresponding entropies of (A, B, C, D) conditional on Z. For different values of Z, the corresponding conditional distributions on (A, B, C, D) can be very different. So there is no well-defined "relativization" of (A, B, C, D) conditional on Z. The simplest way to overcome this obstacle is the method of quasi-uniform distributions suggested by T.H. Chan and R.W. Yeung, see [1].

Definition 1 (Quasi-uniform random variables, [1])**.** *A random variable u distributed on a finite set $\mathcal{U}$ is called* quasi-uniform *if the probability distribution function of u is constant over its support (all values of u have the same probability). That is, there exists $c > 0$ such that $\mathrm{Prob}[u = \mathfrak{u}] \in \{0, c\}$ for all $\mathfrak{u} \in \mathcal{U}$. A set of random variables $(x_1, \ldots, x_n)$ is called quasi-uniform if for any non-empty subset $\{i_1, \ldots, i_s\} \subset \{1, \ldots, n\}$ the joint distribution $(x_{i_1}, \ldots, x_{i_s})$ is quasi-uniform.*

In [1][theorem 3.1] it is proven that for every distribution (A, B, C, D, Z) and every $\delta > 0$ there exists a quasi-uniform distribution $(A'', B'', C'', D'', Z'')$ and an integer k such that

$$\|\vec{H}(A, B, C, D, Z) - \frac{1}{k}\vec{H}(A'', B'', C'', D'', Z'')\| < \delta.$$

For a quasi-uniform distribution for all values $\mathfrak{z}$ of Z'' the corresponding conditional distributions (A'', B'', C'', D'') have the same entropies, which are equal to the conditional entropies. That is, entropies of the distribution of A'', B'', (A'', B''), etc. given $Z'' = \mathfrak{z}$ are equal to $H(A''|Z'')$, $H(B''|Z'')$, $H(A'', B''|Z'')$ and so on. Thus, for a quasi-uniform distribution we can do "relativization" as follows.

Fix any value $\mathfrak{z}$ of Z'' and take the conditional distribution on (A'', B'', C'', D'') given $Z'' = \mathfrak{z}$. In this conditional distribution the entropy of C'' given (A'', B'') is not greater than

$$k \cdot (H(C|A, B, Z) + \delta) = k \cdot (\delta + o(M)).$$

Also, by letting δ be small enough (e.g., $\delta = 1/M$), all entropies of (A'', B'', C'', D'') given $Z'' = \mathfrak{z}$ differ from the corresponding entropies of $kM \cdot \vec{H}(a, b, c, d)$ by at most $H(Z'') \le kM \cdot H(c|a, b) + o(kM)$.

Moreover, entropies of (A'', B'') given (C'', Z'') are the same as entropies of (A'', B'') given C'', since Z functionally depends on C. If in the original distribution $I(a{:}b|c) = 0$, then the mutual information between A'' and B'' given (C'', Z'') is $o(kM)$.

Denote $N = kM$ and (A', B', C', D') the above-defined conditional distribution to get the theorem. ■

c) Rest of the proof for (3):

1. Start with the distribution $(a, b, c, d)_q$ for some q, to be fixed later, from the previous section.
2. Apply the "relativization" lemma 2 and get (A', B', C', D') such that $H(C'|A', B') = o(N)$. Lemma 2 guarantees that other entropies are about N times larger than the corresponding entropies for (a, b, c, d), possibly with an overhead of size
$$O(N \cdot H(c|a, b)) = O\left(\frac{\log_2 q}{q} N\right).$$
Moreover, since the quadruple (a, b, c, d) satisfies $I(a{:}b|c) = 0$, we also have $I(A'{:}B'|C') = 0$ by construction of the random variables in Lemma 2.
3. Scale down the entropy profile of (A', B', C', D') by a factor of $1/N$ within a $o(N)$ precision.
4. Tend N to infinity to get an a.e. vector. Indeed, all entropies from the previous profile converge when N goes to infinity. Conditions of inequality (3) are satisfied for $I(A'{:}B'|C')$ and $H(C'|A', B')$ both vanish at the limit. Inequality (3) eventually reduces to
$$1 + O\left(\frac{\log_2 q}{q}\right) \le O\left(\frac{\log_2 q}{q}\right)$$
which can not hold for large enough q.

■

Remark 2. *In both cases of the proof we constructed an a.e. vector such that the corresponding unconditional inequalities with Lagrange multipliers reduces (as $N \to \infty$) to*

$$1 + O\left(\frac{\log_2 q}{q}\right) \le O\left(\frac{\log_2 q}{q}\right) + o(\lambda_1 + \lambda_2),$$

which cannot hold if we choose q appropriately.

Remark 3. *Notice that in our proof even one fixed value of q suffices to prove that (1) and (3) do not hold for a.e. points. The*

choice of the value of q provides some freedom in controlling the gap between the lhs and rhs of both inequalities.

In fact, we may combine the two above constructions into one to get a single a.e. vector to prove the previous result.

Proposition 1. *There exists one a.e. vector which excludes both (1) and (3) simultaneously.*

Proof sketch:

1. Generate (A, B, C, D) from $(a, b, c, d)_q$ with entropies N times greater.
2. Construct $A'' = SW(A|B)$ and $C' = SW(C|A, B)$ simultaneously (with the same serialization (A, B, C, D)).
3. Since A'' is a Slepian–Wolf hash of A given B, we have
   - $H(C|A'', B) = H(C|A, B) + o(N)$ and
   - $H(C|A'', B, C') = H(C|A, B, C') + o(N) = o(N)$.
4. By inspecting the proof of the Slepian–Wolf theorem we conclude that A'' can be plugged into the argument of Lemma 2 instead of A. The entropy profile of the quadruple (A', B', C', D') thusly obtained from Lemma 2 is approximately N times the entropy profile of $(a, b, c, d)_q$ with a possible overhead of

$$O(I(A{:}B) + H(C|A, B)) + o(N) = O\left(\frac{\log_2 q}{q} N\right),$$

   and further :
   - $I(A'{:}B'|C') = 0$,
   - $I(A'{:}B') = o(N)$,
   - $H(C'|A', B') = o(N)$.
5. Scale the corresponding entropy profile by a factor $1/N$ and tend N to infinity to define the desired a.e. vector.

IV. Conclusion & Discussion

In this paper we discussed the known conditional information inequalities. We presented a simple algebraic example which provides a new proof that two conditional information inequalities are essentially conditional (they cannot be obtained as a direct corollary of any unconditional information inequality). Then, we prove a stronger result: two linear conditional information inequalities are not valid for *asymptotically entropic* vectors.

This last result has a counterpart in the Kolmogorov complexity framework. It is known that unconditional linear information inequalities for Shannon's entropy can be directly translated into equivalent linear inequalities for Kolmogorov complexity, [4]. For conditional inequalities the things are more complicated. Inequalities (1) and (3) could be rephrased in the Kolmogorov complexity setting; but the natural counterparts of these inequalities prove to be not valid for Kolmogorov complexity. The proof of this fact is very similar to the argument in Theorem 2 (we need to use Muchik's theorem on conditional descriptions [14] instead of the Slepian–Wolf theorem employed in Shannon's framework). We skip details for the lack of space.

Open problem 1: Does (2) hold for a.e. vectors?

Every essentially conditional linear inequality for a.e. vectors has an interesting geometric interpretation: it provides a proof of Matúš' theorem from [13], which claims that the convex cone of a.e. vectors for 4 variables is not polyhedral.

Open problem 2: Do (1) and (3) (that hold for entropic but not for a.e. vectors) have any geometric or "physical" meaning?

Acknowledgements

The authors are grateful to Alexander Shen for many useful discussions. This work was supported in part by NAFIT ANR-08-EMER-008-01 grant.

References

[1] T.H. Chan, R.W. Yeung, On a relation between information inequalities and group theory , IEEE Trans. on Inform. Theory, 48, pp. 1992–1995, July 2002.

[2] A.P. Dawid, Conditional independence in statistical theory, Journal of the Royal Statistical Society. 41(1), 1979, pp. 1–31.

[3] R Dougherty, C. Freiling, and K. Zeger. Networks, Matroids, and Non-Shannon Information Inequalities. IEEE Transactions on Information Theory 53(6), June 2007, pp. 1949–1969.

[4] D. Hammer, A. Romashchenko, A. Shen, N. Vereshchagin. Inequalities for Shannon Entropy and Kolmogorov Complexity. Journal of Computer and System Sciences. 60 (2000) pp. 442–464.

[5] T.S. Han, A uniqueness of Shannon's information distance and related nonnegativity problems, J. Comb., Inform. Syst. Sci., vol. 6, pp. 320–321, 1981.

[6] A. W. Ingleton, Representation of Matroids in Combinatorial Matheatics and Its Applications, D. J. A. Welsh, Ed. London, U.K.: Academic, 1971, pp. 149–167.

[7] T. Kaced, A. Romashchenko. On essentially conditional information inequalities. IEEE ISIT 2011, pp. 1935–1939.

[8] K. Makarychev, Yu. Makarychev, A. Romashchenko, N. Vereshchagin. A New Class of non-Shannon Type Inequalities for Entropies. Communications in Information and Systems. 2(2), 2002, pp. 147–166.

[9] F. Matúš. Conditional independences among four random variables III: final conclusion. Combinatorics, Probability & Computing, 8 (1999), pp. 269–276.

[10] F. Matúš, Piecewise linear conditional information inequality, IEEE Transactions on Information Theory, 2006, pp. 236–238.

[11] F. Matúš, Adhesivity of polymatroids, Discrete Math., 307, 2007, pp. 2464–2477.

[12] F. Matúš, Two constructions on limits of entropy functions. IEEE Transactions on Information Theory, 53(1), Jan. 2007, pp. 320–330.

[13] F. Matúš, Infinitely many information inequalities. IEEE ISIT 2007, pp. 41–44.

[14] An. Muchnik, Conditional complexity and codes, Theoretical Computer Science, 271(1–2), 2002, pp. 97–109.

[15] N. Pippenger, What are the laws of information theory. 1986 Special Problems on Communication and Computation Conf., Palo Alto, CA.

[16] Judea Pearl, Causal inference in statistics: An overview. Statistics Surveys, 3, 2009, pp. 96–146.

[17] A. Romashchenko, Extracting the Mutual Information for a Triple of Binary Strings. Proc. 18th Annual IEEE Conference on Computational Complexity (2003). Aarhus, Denmark, July 2003, pp. 221–235.

[18] D. Slepian and J. K. Wolf, Noiseless coding of correlated information sources, IEEE Transactions on Information Theory, 19(4), 1973, pp. 471–480.

[19] R.W. Yeung, A first course in information theory. Norwell, MA/New York: Kluwer/Plenum, 2002.

[20] Z. Zhang and R. W. Yeung, A non-Shannon-type conditional information inequality. IEEE Trans. Inform. Theory, vol. 43, pp. 1982–1985, Nov. 1997.

[21] Z. Zhang and R. W. Yeung, On characterization of entropy function via information inequalities. IEEE Transactions on Information Theory, 44(1998), pp. 1440–1450.

[22] Z. Zhang, On a new non-Shannon-type information inequality, Communications in Information and Systems. 3(1), pp. 47–60, June 2003.