A New class of non Shannon type inequalities for entropies

Konstantin Makarychev   Yury Makarychev
Andrei Romashchenko   Nikolai Vereshchagin

September 9, 2002

ABSTRACT

Most general laws of information theory can be expressed by linear inequalities for Shannon's entropy. For about 50 years all known inequalities for entropy were just linear combinations of the "basic" Shannon inequalities $I(X:Y|Z) \geq 0$*. Only in 1998 Z. Zhang and R. Yeung proved an inequality for entropies which cannot be reduced to basic inequalities.*

In this paper we prove a countable set of non Shannon type linear inequalities for entropies of discrete random variables. Our results generalize the inequalities of Z. Zhang and R. Yeung.

1 Introduction

A central notion of information theory is Shannon's entropy[1]. Given a set of jointly distributed random variables $x_1, \dots, x_n$, we can consider entropies of all random variables $H(x_i)$, entropies of all pairs $H(x_i, x_j)$, etc. ($2^n - 1$ entropy values for all non empty subsets of $\{x_1, \dots, x_n\}$). For every n-tuple of random variables we get a point in $\mathbb{R}^{2^n-1}$ representing entropies of given distribution. A point in $\mathbb{R}^{2^n-1}$ is called *constructible* if it represents entropy values of some collection of n random variables. Following [10] we denote by $\Gamma_n^\star$ the set of all constructible points.

[1] An original paper of C.Shannon [1] is an excellent introduction in this field. We also recommend to the reader who is not familiar with information theory the books [6] and [3].

It is hard to characterize $\Gamma_n^\star$ for arbitrary n (for $n > 3$ it is not even closed, [9]). A more feasible (but also very non-trivial) problem is to describe the closure $\overline{\Gamma_n^\star}$ of $\Gamma_n^\star$. The set $\overline{\Gamma_n^\star}$ is a convex cone [9], and to characterize it we should describe all linear inequalities of the form

$$\begin{array}{rcl} \lambda_1 H(x_1) + \ldots + \lambda_n H(x_n) + \lambda_{1,2} H(x_1, x_2) + \ldots + & & \\ \lambda_{1,2,3} H(x_1, x_2, x_3) + \ldots + \lambda_{1,2,\ldots,n} H(x_1, \ldots, x_n) & \geq & 0, \end{array} \quad (1)$$

which are true for any random variables $x_1, \ldots, x_n$ (λ_W are real coefficients).

Information inequalities are widely used for proving converse coding theorems in information theory. Recently important application of information inequalities beyond information theory were found [11, 13, 14]. So investigation of the class of linear information inequalities is very important.

In this paper we consider only discrete random variables. We restrict our selves to random variables with finite range. This restriction is not very essential: if an inequality is true for all random variables with finite range, it is also true for random variables with countable range. This fact can be easily proved by approximation of a distribution with countable range by distributions with a finite range. We do not know if the same is true for *constraint* inequalities (that is, for inequalities which are true assuming another inequality). Our main goal is investigation of non-constraint inequalities, i.e. inequalities of type (1), so we can suppose all random variables in the sequel have finite range.

For many years only trivial inequalities for entropies were known. Namely, all known inequalities were non-negative linear combinations of *basic inequalities*, i.e. inequalities of the form

$$H(A \cup C) + H(B \cup C) - H(A \cup B \cup C) - H(C) \geq 0, \quad (2)$$

where A, B, C are arbitrary tuples of random variables (for an empty tuple X we suppose $H(X) = 0$). Note that using standard notations this inequality can be rewritten as $I(A : B|C) \geq 0$.

Let us show a few simple consequences of (2). For example, to obtain well known inequality of sub-convexity, we should let in the basic inequality $A = \{x\}$, $B = \{y\}$, $C = \emptyset$. Then we get

$$H(x) + H(y) - H(x, y) \geq 0.$$

Another example: let $A = \{x, y\}$, $B = \{y\}$, $C = \{x\}$. Then we get

$$H(x, y) + H(x, y) - H(x, y) - H(x) \geq 0,$$

i.e. monotonicity property: $H(x) \leq H(x,y)$.

Basic inequalities were proven in the original papers of C. Shannon. In 1986 N. Pippenger raised the question: are there any “laws of information theory” (inequalities for Shannon entropy), which are not linear combinations of basic inequalities [5]?

It can be shown that for $n \leq 3$ all information inequalities valid for n discrete random variables are linear combinations of basic inequalities [4] (see also [7]). For $n \geq 4$ the question remained open till 1998, when Zhang and Yeung [10] came up with a linear inequality for entropies of 4 random variables which cannot be reduced to the basic (Shannon type) inequalities:

$$\begin{aligned} &H(x,u)+H(x,v)+3(H(u,v)+H(v,y)+H(u,y)) \geq \\ &2H(u)+2H(v)+H(y)+H(x,y)+H(u,v,x)+4H(u,v,y). \end{aligned} \tag{3}$$

Using standard notations, this inequality can be rewritten as

$$2I(u:v) \quad \leq \quad I(x:y)+I(y:uv)+3I(u:v|y)+I(u:v|x).$$

In the same paper (by the same arguments) Zhang and Yeung proved a more general inequality (for any $n \geq 1$):

$$H(x_1x_2\dots x_n)+nI(u:v:x_1) \leq \sum_{j=1}^{n} H(x_j)+\sum_{j=1}^{n} I(u:v|x_j)+I(uv:x_1). \tag{4}$$

The former inequality is obtained from the latter one by letting $n=2$, $x_1=y$, $x_2=x$.

It should be noted that a year earlier the same authors found [9] a *constraint* inequality for entropies which cannot be deduced from the basic inequalities:

$$\begin{aligned} &\text{if } I(x_1:x_2)=I(x_1:x_2|x_3)=0 \text{ then} \\ &I(x_3:x_4) \leq I(x_3:x_4|x_1)+I(x_3:x_4|x_2). \end{aligned} \tag{5}$$

Note: this result gives new bound for the set $\Gamma_4^\star$, but not for $\overline{\Gamma_4^\star}$.

A large collection of non Shannon type constraint inequalities was proved in [12] based on non-constraint inequality (3).

In the present paper we exhibit a new countable family of non-constraint inequalities for Shannon entropy and a new countable family of constraint inequalities.

Theorem 1. *For any random variables* $u, v, z, x_1, \dots, x_n$ *if*

$$I(u:z|v) = I(v:z|u) = 0$$

then

$$H(x_1, \dots, x_n) + (n-1) \cdot I(uv:z) \leq \sum_{i=1}^{n} I(u:v|x_i) + \sum_{i=1}^{n} H(x_i).$$

Theorem 2. *For any random variables* $u, v, z, x_1, \dots, x_n$

$$H(x_1, \dots, x_n) + n \cdot I(u:v:z) \leq \sum_{i=1}^{n} I(u:v|x_i) + \sum_{i=1}^{n} H(x_i) + I(uv:z).$$

Note that the inequality of Theorem 2 implies that of Zhang and Yeung (4). Indeed, letting $z = x_1$ we get (4).

Theorem 1 is an easy corollary of Theorem 2. Indeed, using the equality $I(u:v:z) = I(uv:z) - I(u:z|v) - I(v:z|u)$ we can rewrite the inequality of Theorem 2 as follows:

$$\begin{aligned} H(x_1, \dots, x_n) + (n-1) \cdot I(uv:z) \quad \leq \quad & \sum_{i=1}^{n} I(u:v|x_i) + \sum_{i=1}^{n} H(x_i) + \\ & n(I(u:z|v) + I(v:z|u)). \end{aligned}$$

However, to illustrate our method we will first prove Theorem 1. Actually, we present two different methods to prove new non Shannon type inequalities. The first one appeals to the idea of *rate region* of two random variables introduced by R. Ahlswede and J. Körner [2, 3]. The other method is a pure syntactical inference rule producing a new inequality given a constraint inequality of special type. The proof of this inference rule is a refinement of Zhang-Yeung's proof of (4).

The rest of the paper is organized as follows. In Sections 2 and 3 we prove the main results, Theorem 1 and Theorem 2, by the first method (using properties of rate region of two random variables). In Section 4 we present the syntactical method and give another proof of Theorem 2. In Section 5 we prove that an analog of our syntactical rule is valid also for Kolmogorov complexity.

In the Appendix we give some proofs omitted in the main text. We also sketch the proof that the inequality of Theorem 2 for $n = 2$ does not follow from basic inequalities and (3) and that the inequalities of Theorem 2 for $n = 2$ and $n = 3$ are not equivalent. The full proof includes checking a huge number of conditions and was done using a computer software.

2 Proof of Theorem 1

Lemma 3 (on Double Markov Property). (*[3], chapter 3, exercise 25*) Let random variables u, v, z satisfy the condition

$$I(u : z|v) = I(v : z|u) = 0.$$

Then there exists a random variable w such that

- $H(w|u) = H(w|v) = 0$,
- uv and z are independent given w (i.e. $I(uv : z|w) = 0$),
- $H(w) = I(uv : z)$.

(We omit the proof.) Let as apply the lemma above to the variables u, v, z given in the condition of Theorem 1.

Further we use the following inequality proven in [7] (we present its proof in Appendix A):

$$H(c|d) \leq H(c|ad) + H(c|bd) + I(a : b|d) \tag{6}$$

Letting in this inequality $c = w$, $d = x_i$, $a = u$, $b = v$ we get

$$\begin{aligned} H(x_i|w) + H(w) &= H(x_i) + H(w|x_i) \\ &\leq H(x_i) + H(w|ux_i) + H(w|vx_i) + I(u : v|x_i) \\ &\leq H(x_i) + H(w|u) + H(w|v) + I(u : v|x_i) \\ &= H(x_i) + I(u : v|x_i). \end{aligned}$$

Sum up all such inequalities for $i = 1, \ldots, n$ and add another Shannon type inequality

$$H(x_1, \ldots, x_n) \leq H(w) + \sum_{i=1}^{n} H(x_i|w).$$

Recalling that $H(w) = I(uv : z)$ we get the required inequality.

3 Proof of Theorem 2

We want to generalize the above proof to get the inequality of Theorem 2. To this end we need an analog of the lemma on Double Markov property. Such an analog is implied by the following lemma (it follows immediately from theorem 2 from [2]):

Lemma 4 (Ahlswede-Körner). *Let u, v, z be random variables. For any integer $N > 0$ consider N independent copies of this triple, i.e. N random variables*

$$\langle u_1, v_1, z_1\rangle, \langle u_2, v_2, z_2\rangle, \dots, \langle u_N, v_N, z_N\rangle,$$

such that for any i the triple $\langle u_i, v_i, z_i\rangle$ has the same distribution as $\langle u, v, z\rangle$, and the triples $\langle u_i, v_i, z_i\rangle$ for all $i = 1, 2, \dots, N$ are independent. Let

$$\begin{array}{rcl} U & = & u_1, \dots, u_N, \\ V & = & v_1, \dots, v_N, \\ Z & = & z_1, \dots, z_N. \end{array}$$

(*We omit n in the notations U, V, Z for brevity.*) *Then there is a random variable $W = W(N)$ such that*

$$\begin{array}{l} H(W) \le I(UV : Z) + o(N) = N \cdot I(uv : z) + o(N), \\ H(U|W) \le H(U|Z) + o(N) = N \cdot H(u|z) + o(N), \\ H(V|W) \le H(V|Z) + o(N) = N \cdot H(v|z) + o(N), \\ H(UV|W) \le H(UV|Z) + o(N) = N \cdot H(uv|z) + o(N). \end{array}$$

For the sake of completeness we will prove this lemma in Appendix B.

Remark 1. For the case $I(u : z|v) = I(v : z|u) = 0$ the lemma is a corollary of the lemma on Double Markov Property. Indeed, for any u_i, v_i, z_i we can apply Lemma 3 and get a random variable w_i such that

$$H(w_i) = I(u_i v_i : z_i),\ H(w_i|u_i) = H(w_i|v_i) = 0,\ I(u_i v_i : z_i|w_i) = 0.$$

Since u_i, v_i are independent from z_i given w_i we have $H(u_i|w_i) \le H(u_i|z_i)$, $H(v_i|w_i) \le H(v_i|z_i)$, and $H(u_i, v_i|w_i) \le H(u_i, v_i|z_i)$. Thus we can let $W = \langle w_1, \dots, w_n\rangle$.

Remark 2. A reader familiar with the notion of *rate region for two random variables* [2, 3] can easily note that both lemma 3 and 4 state that some point belongs to rate region of u and v. Then we show how to use this point to get non-trivial information inequality for involved variables. Formally we do not need here the definition of rate region, so we omit it.

Let us prove that random variable W from Lemma 4 satisfies the inequalities

$$\begin{array}{l} H(W) = N \cdot I(uv : z) + o(N), \\ H(W|U) \le N \cdot I(v : z|u) + o(N), \\ H(W|V) \le N \cdot I(u : z|v) + o(N). \end{array}$$

The fourth inequality of Lemma 4 implies $H(W) \geq I(W : UV) = H(UV) - H(UV|W) \geq H(UV) - H(UV|Z) = I(UV : Z) + o(N)$, and we obtain the first equality. To prove the second inequality note that

$$\begin{aligned} H(W|U) &= H(U|W) + H(W) - H(U) \\ &\leq N \cdot (H(u|z) + I(uv : z) - H(u)) + o(N) \\ &= N \cdot I(v : z|u) + o(N). \end{aligned}$$

The third inequality is proven in a similar way.

Let us fix positive integer N and consider N copies of independent tuples $\langle u, v, z, x_1, \dots, x_n \rangle$, i.e. N independent tuples

$$\langle u^1, v^1, z^1, x_1^1, \dots, x_n^1 \rangle, \langle u^2, v^2, z^2, x_1^2, \dots, x_n^2 \rangle, \dots, \langle u^N, v^N, z^N, x_1^N, \dots, x_n^N \rangle,$$

where each $\langle u^i, v^i, z^i, x_1^i, \dots, x_n^i \rangle$ have the same joint distribution as the collection of random variables $\langle u, v, z, x_1, \dots, x_n \rangle$.

Let

$$\begin{aligned} U &= u^1, u^2, \dots, u^N, \\ V &= v^1, v^2, \dots, v^N, \\ Z &= z^1, z^2, \dots, z^N, \\ X_1 &= x_1^1, x_1^2, \dots, x_1^N, \\ \dots & \quad \dots \\ X_n &= x_n^1, x_n^2, \dots, x_n^N. \end{aligned}$$

Note that $H(U) = NH(u)$, $H(V) = NH(v)$, etc.

Repeat the proof of Theorem 1 for $\langle U, V, Z, X_1, \dots, X_n \rangle$ in place of the tuple $\langle u, v, z, x_1, \dots, x_n \rangle$ and using Lemma 4 instead of the lemma on Double Markov property. Instead of equalities $H(w|u) = H(w|v) = 0$ we have the inequalities $H(W|U) \leq N \cdot I(v : z|u) + o(N)$, $H(W|V) \leq N \cdot I(u : z|v) + o(N)$. Therefore we obtain the inequality

$$N \cdot (H(x_1, \dots, x_n) + (n-1) \cdot I(uv : z)) + o(N) \leq$$
$$N \cdot (\sum_{i=1}^{n} I(u : v|x_i) + \sum_{i=1}^{n} H(x_i) + n(I(v : z|u) + I(u : z|v))) + o(N)$$

instead of inequality

$$H(x_1, \dots, x_n) + (n-1) \cdot I(uv : z) \leq \sum_{i=1}^{n} I(u : v|x_i) + \sum_{i=1}^{n} H(x_i).$$

As this inequality holds for any N we have

$$H(x_1,\dots,x_n)+(n-1)\cdot I(uv:z) \leq \sum_{i=1}^n I(u:v|x_i)+\sum_{i=1}^n H(x_i)+ n(I(v:z|u)+I(u:z|v)).$$

4 An inference rule to deduce new linear inequalities for Shannon entropy

We first give another proof for Theorem 2 and then extend the argument to get a general inference rule.

Proof of the theorem. First rewrite the inequality of Theorem 2 in the following form

$$H(x_1\dots x_n)+nI(u:v) \leq \sum_{i=1}^n I(u:v|x_i)+nI(u:v|z)+ \sum_{i=1}^n H(x_i)+I(uv:z). \quad (7)$$

We first note that this inequality is true under the condition $\langle x_1,\dots,x_n\rangle$ and z are independent given $\langle u,v\rangle$. This fact can be easily deduced from Shannon type inequalities (see Appendix A, Lemma 8). It remains to get rid of the assumption that $x_1\dots x_n$ and z are independent given u,v.

Let us prove that (7) holds for any $x_1,\dots,x_n$, u,v,z. The crucial point is that no term of (7) has any of x_i's together with z. Given $x_1,\dots,x_n,u,v,z$ construct a new random variable $\tilde z$ such that $\langle \tilde z,u,v\rangle$ has the same distribution as $\langle z,u,v\rangle$ and such that $x_1\dots x_n$ and z are independent given u,v. This is done as follows: $\tilde z$ has the same range as z and for any $\mathbf{z}$ in its range

$$\mathrm{Prob}[\tilde z=\mathbf{z}\mid \langle x_1,\dots,x_n,u,v\rangle=\langle \mathbf{x_1},\dots,\mathbf{x_n},\mathbf{u},\mathbf{v}\rangle] \\ =\mathrm{Prob}[z=\mathbf{z}\mid u=\mathbf{u},v=\mathbf{v}]$$

By definition $\tilde z$ is independent from $x_1,\dots,x_n$ given u,v. Hence by Lemma 8 the inequality (7) holds for z replaced by $\tilde z$. But this replacement does not change any term of the inequality (7) because for any its term either all its variables are included in the set $x_1,\dots,x_n,u,v$, or in the set u,v,z. Obviously the replacement $z\to\tilde z$ does not change any term of the first type. And it does not change any term of the second type either, as $\langle \tilde{z}, u, v \rangle$ and $\langle \tilde{z}, u, v \rangle$ have the same distribution. $\square$

It is easy to generalize the above argument to the following rule.

Theorem 5 (Inference rule). Assume that each variable from an information inequality $S \le 0$ is assigned to a node of a finite rooted tree so that for any term from S all the variables of that term lie on the same branch of the tree. Assume that the inequality $S \le 0$ is true for any tuple of random variables satisfying the following condition: for any internal node s of the tree the variables $V_{t_1}, \ldots, V_{t_m}$ are independent given W_s, where $t_1, \ldots, t_m$ are all the sons of s, W_s stands for the tuple of random variables assigned to predecessors of s (including s) and V_t for the tuple of random variables assigned to successors of t (including t). Then the inequality $S \le 0$ is true for *any* tuple of random variables.

In the above proof we used this rule for the tree consisting of the root and two its sons, with u, v assigned to the root, z assigned to one son, and $x_1, \ldots, x_n$ assigned to the other one.

Let us give another example of an inequality which might be derived by our rule (unfortunately we have no other real applications of the rule except for the above one). Let an inequality have the form

$$aH(xyu) + bH(yu) + cH(u) + dH(zyu) + eH(vu) \ge 0$$

and assume that it holds for any x, y, z, u, v such that $I(xyz : v|u) = 0$ and $I(x : z|yu) = 0$. Then this inequality holds for *any* x, y, z, u, v. Here we use the tree with 5 vertices consisting of the root with 2 sons, the left son also having 2 sons, with u assigned to the root, y and v assigned to the left and right sons of the root, respectively, and x, z assigned to two sons of the left son of the root.

Proof of the theorem. As above it suffices to show that given any random variables $x_1, \ldots, x_n$ we are able to define new random variables $\tilde{x}_1, \ldots, \tilde{x}_n$ with the following two properties.

- For any path in the tree let $x_{i_1}, \ldots, x_{i_k}$ be variables assigned to all the nodes in this path; then the tuple $\langle \tilde{x}_{i_1}, \ldots, \tilde{x}_{i_k} \rangle$ has the same distribution as $\langle x_{i_1}, \ldots, x_{i_k} \rangle$.

- For an node s in the tree, $\tilde{V}_{t_1}, \dots, \tilde{V}_{t_m}$ are independent given $\tilde{W}_s$, where $t_1, \dots, t_m$ are all the sons of s.

We construct such $\tilde{x}_1, \dots, \tilde{x}_n$ by induction. W.l.o.g. assume the following: if x_i is assigned to a vertex v, x_j is assigned to a vertex w and v precedes w in the tree then $i < j$.

Base of induction: let $\tilde{x}_i = x_i$ for any x_i assigned to the root.

Induction step. Assume that $\tilde{x}_1, \dots, \tilde{x}_{i-1}$ are defined. Let s denote the node x_i is assigned to. Recall that W_s denotes the tuple of random variables assigned to predecessors of s (including s). Let W_{si} be the tuple of those random variables from $x_1, \dots, x_{i-1}$ which belong to W_s. We define $\tilde{x}_i$ so that it has the same range as x_i and for any value $\mathbf{x}_i$ in its range let

$$\mathrm{Prob}[\tilde{x}_i = \mathbf{x}_i \mid \langle \tilde{x}_1, \dots, \tilde{x}_{i-1} \rangle = \langle \mathbf{x}_1, \dots, \mathbf{x}_{i-1} \rangle] = \mathrm{Prob}[x_i = \mathbf{x}_i \mid W_s = \mathbf{W}_{si}].$$

Here $\mathbf{W}_{si}$ denotes the tuple consisting of those values from $\mathbf{x}_1, \dots, \mathbf{x}_{i-1}$ that corresponds to variables from W_{si}.

It is a straightforward consequence of the construction that $\tilde{x}_1, \dots, \tilde{x}_n$ qualify the requirements. □

5 An inference rule to deduce new linear inequality for Kolmogorov complexity

In this section we assume that the reader is familiar with plain Kolmogorov complexity $K(x)$; unfamiliar reader can consult [8]. We will present inference rule to prove inequalities for Kolmogorov complexity of binary strings and their tuples. This rule is an analog of the rule from the previous section.

We consider inequalities of the form $S(x_1, \dots, x_n) \leq 0$ in variables $x_1, \dots, x_n$ ranging over binary strings whose left hand side is a sum of terms of the form $K(\langle x_{i_1}, \dots, x_{i_m} \rangle)$. The inference rule allows to prove that such an inequality is valid up to an additive term logarithmic in complexity of $x_1, \dots, x_n$. That is, for some constant c for any strings $x_1, \dots, x_n$ it holds $S(x_1, \dots, x_n) \leq c \log k$ where $k = K(x_1) + \dots + K(x_n)$.

The rule will be defined in terms of infinite sequences of strings. It is easy to see that an inequality $S \leq 0$ holds up to an additive $O(\log k)$ term if and only if for any sequences $x_1^i, \dots, x_n^i$ of strings it holds $S(x_1^i, \dots, x_n^i) \leq O(\log k^i)$ where $k^i = K(x_1^i) + \dots + K(x_n^i)$. Indeed, "only if" part is evident. To prove the "if" part assume that for any c there are $x_1^c, \dots, x_n^c$ such that

$S(x_1^c, \ldots, x_n^c) > c \log(K(x_1^c) + \cdots + K(x_n^c))$. Then for the sequence $x_1^c, \ldots, x_n^c$, $c = 1, 2, \ldots$, the inequality $S(x_1^c, \ldots, x_n^c) \leq O(\log k^c)$ is not true.

For sequences $u_1^i, \ldots, u_m^i, v^i$ of strings we say that $u_1^i, \ldots, u_m^i$ are independent given v^i if $K(u_1^i|v^i) + \cdots + K(u_m^i|v^i) \leq K(u_1^i, \ldots, u_m^i|v^i) + O(\log k^i)$ where k^i denotes the sum of complexities of all $u_1^i, \ldots, u_m^i, v^i$.

Theorem 6 (Inference rule for Kolmogorov complexity).
Assume that each variable from S is assigned a node of a finite rooted tree so that for any term from S all its variables lie on the same branch of the tree. Assume that the inequality $S(x_1^i, \ldots, x_n^i) \leq O(\log k^i)$ holds for any sequences $x_1^i, \ldots, x_n^i$ with the following property: for any internal node s in the tree the sequences $u_1^i, \ldots, u_m^i$ are independent given w^i, where u_j^i denotes the tuple consisting of all variables assigned to all successors of the jth son of s (including the son itself) and w^i the tuple consisting of all variables assigned to all predecessors of s (including s). Then for some c for any strings the inequality $S \leq c \log k$ holds.

Proof. Given a sequence of strings $x_1, \ldots, x_n$ [and a string y] define its *complexity vector* [conditional to y] to be the sequence of $2^n - 1$ integers consisting of complexities of stings $x_1, \ldots, x_n$, their pairs, triples, etc. [conditional to y].

W.l.o.g. we may assume that the number of nodes in the tree is equal to the number of variables and the assignment variables to nodes is one-to-one: if this is not the case, replace a node to which more than one variable is assigned by a sequence of nodes. Therefore we will identify variables and nodes.

It suffices to show that for some c for any strings $x_1, \ldots, x_n$ there are strings $x'_1, \ldots, x'_n$ such that the following holds.

- For any path $x_{i_1}, \ldots, x_{i_m}$ in the tree the complexity vectors of the tuples $x_{i_1}, \ldots, x_{i_m}$ and $x'_{i_1}, \ldots, x'_{i_m}$ differs by $c \log k$ in each component.
- For any internal node s it holds

$$K(u'_1|w') + \cdots + K(u'_m|w') - K(u'_1 \ldots u'_m|w') \leq c \log k'$$

(here $u_1, \ldots, u_m, w$ are those tuples from the statement of the theorem; u' denotes the tuple obtained from u by replacing x_i by x'_i).

This statement is proven by induction on the height of the tree. To make the induction step we need to strengthen the statement: we will assume that all the complexities are conditional to some string y. So the conditional version proven by induction is as follows: for some c for any strings $x_1,\dots,x_n,y$ there are strings $x'_1,\dots,x'_n$ such that for any path $x_{i_1},\dots,x_{i_m}$ in the tree the complexity vectors of $x_{i_1},\dots,x_{i_m}$ and $x'_{i_1},\dots,x'_{i_m}$ conditional to y differs by $O(\log k)$ in each component and for any internal node s it holds $K(u'_1|w',y)+\dots+K(u'_m|w',y)-K(u'_1\dots u'_m|w',y)\le c\log k$.

Base of induction: for trees of height 0 (consisting of the root only) the statement is trivial.

Induction step. To make the induction step we need to remind the main tool of [7], the "typization" argument. For some c depending only n given $x_1,\dots,x_n$ we can define a set M of at least $2^{K(x_1,\dots,x_n)-c\log k}$ n-tuples of strings such that the complexity vector of any tuple in M differs from that of $x_1,\dots,x_n$ by at most $c\log n$ in every component. This is done as follows. Let M' consist of all n-tuples of strings $\langle x'_1,\dots,x'_n\rangle$ such that for any $i_1,\dots,i_m$ and any $j_1,\dots,j_l$ (where $m\ge 1, l\ge 0$) it holds

$$K(x'_{i_1},\dots,x'_{i_m}|x'_{j_1},\dots,x'_{j_l})\le K(x_{i_1},\dots,x_{i_m}|x_{j_1},\dots,x_{j_l}).$$

This set is large: for some constant c_1 it holds $\log|M'|\ge K(x_1,\dots,x_n)-c_1\log k$ (because the tuple $x_1,\dots,x_n$ can be described by its index in the enumeration of this set and by extra $O(\log k)$ bits describing the set itself). By construction the complexity vector of any point in M' is not larger than that of $x_1,\dots,x_n$. The inverse however is not true: M' does contain points with low complexity vector. But an easy counting argument shows that the fraction of points in M' such that $K(x'_{i_1},\dots,x'_{i_m}|x'_{j_1},\dots,x'_{j_l})<K(x_{i_1},\dots,x_{i_m}|x_{j_1},\dots,x_{j_l})-c_2\log k$ for some $i_1,\dots,i_m,j_1,\dots,j_l$ is small (for appropriate c_2). Let M be the set of all points in M' which do not satisfy this inequality.

We will use a conditional version of this construction, which has a similar proof: for any $x_1,\dots,y_m$ and y there is a set M of cardinality at least $2^{K(x_1,\dots,x_n|y)-O(\log k)}$ such that for any $i_1,\dots,i_m$ and any $j_1,\dots,j_l$ it holds

$$|K(x'_{i_1},\dots,x'_{i_m}|x'_{j_1},\dots,x'_{j_l},y)-K(x_{i_1},\dots,x_{i_m}|x_{j_1},\dots,x_{j_l},y)|=O(\log k).$$

Now we are in position to make the induction step. Regard our tree as a collection of subtrees rooted at sons of the root. W.l.o.g. assume the string x_1 is assigned to the root, strings $x_2,\dots,x_l$ are assigned to nodes of the subtree rooted at the leftmost son of the root, $x_{l+1}, \dots, x_m$ are assigned to nodes of the tree rooted at the next son of the root etc. Apply the induction hypothesis to the leftmost subtree, to the sequence of strings $x_2, \dots, x_l$ and to $\langle x_1, y\rangle$ as condition. We get some strings $x'_2, \dots, x'_l$. The complexity vector of this tuple is OK conditional to $\langle x_1, y\rangle$, that is, the complexity vectors of tuples $x_1, x'_2, \dots, x'_l$ and $x_1, x_2, \dots, x_l$ conditional to y are the same (up to a $O(\log k)$ term). Make the same thing for the other subtrees and gather together all obtained strings. We get strings $x'_2, \dots, x'_n$ such that the sequence $x_1, x'_2, \dots, x'_n$ satisfies all the requirements but one exception: $K(x'_2, \dots, x'_n|x_1, y)$ may be much less than the sum $K(x'_2, \dots, x'_l|x_1, y) + K(x'_{l+1}, \dots, x'_m|x_1, y) + \dots$ (the sum is over all subtrees). To overcome this difficulty we use typization. We apply typization to the tuple $\langle x'_2, \dots, x'_l\rangle$ conditional to $\langle x_1, y\rangle$ and make the same for all subtrees. We thus get sets $M_1, M_2, \dots$. For at least one tuple in their Cartesian product it holds $K(x'_2, \dots, x'_n|x_1, y) \geq \log|M_1| + \log|M_2| + \dots$. And these sets are big enough: $\log|M_1| + \dots + \log|M_m| \geq K(x'_2, \dots, x'_l|x_1, y) + K(x'_{l+1}, \dots, x'_m|x_1, y) + \dots$ □

Remark 3. A theorem from [7] states that a Kolmogorov complexity inequality $S \leq 0$ is true up to an additive logarithmic term iff the inequality $S' \leq 0$ which is obtained from it by replacing all strings by random variables and Kolmogorov complexity by Shannon entropy is true. So one could try to reduce the inference rule for Kolmogorov complexity to that for Shannon entropy. However it is not clear how to do this: we do not know whether the analogous statement is true for constraint inequalities.

6 Conclusion

In the present paper we proved two families of new non-Shannon type inequalities for entropies: a countable family of constraint inequalities and a countable family of non-constraint inequalities. Each constraint inequality follows from the corresponding non-constraint one. However the proof of constraint inequalities is interesting in its own right.

We used two different methods to prove the main results. The first method (Sections 2 and 3) appeals to the idea of rate region of Ahlswede and Körner. The second method (Section 4) uses an inference rule which allows us to deduce new non-constraint inequalities given constraint inequalities of certain type. In Section 5 we prove that the same inference rule can be applied to inequalities for Kolmogorov complexity.

The method of Section 4 is based on ideas of Zhang and Yeung [10]. So it is not surprising that our inequalities are very similar to (4). More surprising is that the method from Section 3 gives the same results. Probably there is a relation between these two methods, and an interesting problem is to reveal this relation. Another intriguing question is whether any other new inequality can be proven using our methods.

Here are other interesting open questions.

- Are all inequalities of Theorem 2 independent? We can prove only that inequality (3) together with basic inequalities do not imply the inequality of Theorem 2 for $n = 2$, and that our inequalities for $n = 2$ and $n = 3$ are not equivalent modulo basic inequalities (see Appendix C).

- Obviously, inequalities of Zhang and Yeung (4) are implied by Theorem 2. Is the converse true? Namely does the family of all inequalities of type (4) ($n = 1, 2, \ldots$) implies Theorem 2?

- Inequalities of Theorem 1 are implied by corresponding non-constraint inequalities of Theorem 2. Is there any true constraint inequality

$$(S_1 \leq 0 \wedge S_2 \leq 0 \wedge \ldots S_n \leq 0) \Rightarrow S \leq 0$$

  for entropies for which its non-constraint analog, $S \leq C \cdot (S_1 + \ldots + S_n)$, is false for any positive constant C? In particular, is the inequality $I(x_3 : x_4) \leq I(x_3 : x_4|x_1) + I(x_3 : x_4|x_2) + C \cdot I(x_1 : x_2) + C \cdot I(x_1 : x_2|x_3)$ true for some C (the non-constraint analog of (5))?

- Does exist any constraint inequality which is valid for random variables with finite range, but is not true for random variables with countable range?

7 Appendix

7.1 A. The proof of useful Shannon type inequalities.

Lemma 7. *For any random variables a, b, c, d the inequality*

$$H(c|d) \leq H(c|ad) + H(c|bd) + I(a : b|d)$$

holds.

Proof. This inequality is an easy consequence of a basic inequality:

$$H(a,b,d)+H(c,d)\le H(a,b,c,d)+H(c,d)\le H(a,c,d)+H(b,c,d).$$

Deducing $2H(d)$ from both sides we obtain

$$H(a,b|d)+H(c|d)\le H(a,c|d)+H(b,c|d).$$

Hence

$$H(c|d)\le H(c|ad)+H(c|bd)+(H(a|d)+H(b|d)-H(a,b|d))$$

and we get the inequality. □

Lemma 8. *The inequality* (7) *is true provided* $\langle x_1,\dots,x_n\rangle$ *and* z *are independent given* $\langle u,v\rangle$*.*

Proof. For any a,b,c,d it holds

$$I(a:b)\le I(a:b|c)+I(a:b|d)+I(c:d)+I(c:d|ab). \qquad (8)$$

Indeed, consider the mutual information in a,b,c,d defined as

$$I(a:b:c:d)=I(a:b:c)-I(a:b:c|d).$$

It is easy to verify that it is symmetric: just rewrite it as an algebraic sum of entropies a,b,c,d, their pairs, triples etc. to get the symmetric expression

$$\begin{aligned} I(a:b:c:d) \;=\;& H(a)+H(b)+H(c)+H(d)\\ &-H(ab)-H(ac)-\dots-H(cd)\\ &+H(abc)+\dots+H(bcd)\\ &-H(abcd). \end{aligned}$$

First we have

$$\begin{aligned} I(a:b:c:d) &= I(a:b:c)-I(a:b:c|d)\\ &= I(a:b)-I(a:b|c)-I(a:b|d)+I(a:b|cd)\\ &\ge I(a:b)-I(a:b|c)-I(a:b|d). \end{aligned}$$

On the other hand we have

$$\begin{aligned} I(a:b:c:d) &= I(c:d)-I(c:d|a)-I(c:d|b)+I(c:d|ab)\\ &\le I(c:d)+I(c:d|ab). \end{aligned}$$

Comparing these two inequalities we get (8).

Let in this inequality $a = u$, $b = v$, $c = x_i$, $d = z$. We get

$$\begin{aligned} I(u:v) &\le I(u:v|x_i) + I(u:v|z) + I(x_i:z) + I(x_i:z|uv) \\ &= I(u:v|x_i) + I(u:v|z) + I(x_i:z). \end{aligned}$$

The last equality holds as x_i and z are independent given u, v.

Summing these inequalities over all i we obtain

$$nI(u:v) \le \sum_{i=1}^{n} I(u:v|x_i) + nI(u:v|z) + \sum_{i=1}^{n} I(x_i:z).$$

So to get (7) it remains to show that

$$H(x_1 \dots x_n) + \sum_{i=1}^{n} I(x_i:z) \le \sum_{i=1}^{n} H(x_i) + I(uv:z),$$

or equivalently

$$H(x_1 \dots x_n) \le \sum_{i=1}^{n} H(x_i|z) + I(uv:z).$$

This can be proven as follows:

$$\begin{aligned} H(x_1 \dots x_n) &= H(x_1 \dots x_n|z) + I(x_1 \dots x_n : z) \\ &\le \sum_{i=1}^{n} H(x_i|z) + I(x_1 \dots x_n : z) \le \sum_{i=1}^{n} H(x_i|z) + I(uv:z) \end{aligned}$$

(the last inequality is true because $x_1, \dots, x_n$ and z are independent given u, v). □

7.2 B. The proof of Lemma 4

In this section we prove Lemma 4. Our proof uses the method of *typical sequences*. We omit tiresome technical details, and refer a reader to [3] for the introduction in a standard technique for typical sequences.

We shall use the denotations $\mathbf{U}, \mathbf{V}, \mathbf{Z}$ for values of random variables U, V, Z (each of them is a string of n values of u, v or z respectively).

Let us consider *typical* values of the triple $\langle \mathbf{U}, \mathbf{V}, \mathbf{Z} \rangle$, i.e. the triples $\langle \mathbf{U}, \mathbf{V}, \mathbf{Z} \rangle$ such that frequencies of all values in the triple is about its probability for the distribution u, v, z. More precisely, we say that $\mathbf{U}, \mathbf{V}, \mathbf{Z}$ is a typical triple if for the number $N(\alpha, \beta, \gamma)$ of places i such that $\langle \mathbf{U}_i, \mathbf{V}_i, \mathbf{Z}_i \rangle = \langle \alpha, \beta, \gamma \rangle$ we have

$$|N(\alpha, \beta, \gamma) - N \cdot \text{Prob}[u = \alpha \wedge v = \beta \wedge z = \gamma]| \leq \sqrt{N} \log N.$$

Remark 4. The choice of the bound for the difference between probability and frequency is not very important here. We could get instead of $\sqrt{N} \log N$ any function $\theta(N)$ such that

$$\frac{\theta(N)}{N} \to 0 \text{ and } \frac{\theta(N)}{\sqrt{N}} \to \infty \text{ as } N \to \infty.$$

Only the following properties of typical sequences are required:

- the number of typical values $\langle \mathbf{U}, \mathbf{V}, \mathbf{Z} \rangle$ is equal to $2^{N \cdot H(u,v,w)+o(N)}$
- $\text{Prob}[\langle U, V, Z \rangle$ is typical $] \to 1$ as $N \to \infty$.

Both properties can be proved by standard counting arguments (see chapter 1 in [3]).

Now let us consider projection of the set of all typical triples $\langle \mathbf{U}, \mathbf{V}, \mathbf{W} \rangle$ onto the first two coordinates. We call this projection by *the set of typical pairs* $\langle \mathbf{U}, \mathbf{V} \rangle$. Analogously we consider projection of the set of all typical triples on the third coordinate and call it by *the set of all typical* $\mathbf{Z}$.

We formulate without proof a few additional properties of typical values. They also can be proves with a standard counting technique:

(*) There $2^{H(U,V)+o(N)}$ typical pairs $\langle \mathbf{U}, \mathbf{V} \rangle$ and $2^{H(\mathbf{Z})+o(N)}$ typical $\mathbf{Z}$.

(**) For every typical $\mathbf{Z}$ there are $2^{H(U,V|Z)+o(N)}$ pairs $\langle \mathbf{U}, \mathbf{V} \rangle$ such that the triple $\langle \mathbf{U}, \mathbf{V}, \mathbf{Z} \rangle$ is typical.

(***) For every typical pair $\langle \mathbf{U}, \mathbf{V} \rangle$ there are $2^{H(Z|U,V)+o(N)}$ different $\mathbf{Z}$ such that the triple $\langle \mathbf{U}, \mathbf{V}, \mathbf{Z} \rangle$ is typical.

Our first goal is to cover the set of all typical pairs $\langle \mathbf{U}, \mathbf{V} \rangle$ by some collection of sets such that

1. the number of covering sets is not larger then $2^{I(UV:Z)+o(N)}$,

2. projections of each covering set onto the first and the second coordinates are not larger then $2^{H(U|Z)+o(N)}$ and $2^{H(V|Z)+o(N)}$ respectively,

3. the size of each covering set is not larger then $2^{H(UV|Z)+o(N)}$.

Let us choose an arbitrary typical $\mathbf{Z}$. Consider the set of all pairs $\mathbf{U}, \mathbf{V}$ which form together with the fixed $\mathbf{Z}$ a typical triple. This set satisfy the second and the third conditions above (see (*) and (**)).

So we have candidates for covering sets (one set for each typical $\mathbf{Z}$). If Z and $\langle U, V\rangle$ are independent, we get required collection of covering sets. But in general case there too many candidates: we have $2^{H(Z)+o(N)}$ different typical $\mathbf{Z}$, and only $2^{I(UV:Z)+o(N)}$ sets should be in a covering family. Now we explain how to choose an appropriate number of candidates satisfying all three conditions above.

let $k = 2^{I(UV:Z)+\epsilon}$. Choose at random k typical $\mathbf{Z}$ and get k corresponding sets of typical pairs $\langle \mathbf{U}, \mathbf{V}\rangle$. For an appropriate ϵ (to be specified later) chosen sets cover all typical $\langle \mathbf{U}, \mathbf{V}\rangle$ with a high probability. Really, let us fix a typical pair $\langle \mathbf{U}, \mathbf{V}\rangle$ and consider random variable

$$\xi(\mathbf{U}, \mathbf{V}) = \begin{cases} 1, & \text{if the triple } \langle \mathbf{U}, \mathbf{V}, \mathbf{Z}\rangle \text{ is typical for randomly chosen } \mathbf{Z} \\ 0, & \text{otherwise.} \end{cases}$$

Then by (*) and (**) for each typical pair $\langle \mathbf{U}, \mathbf{V}\rangle$

$$\text{Prob}[\xi(\mathbf{U}, \mathbf{V}) = 1] = \frac{2^{H(UV|Z)} + o(n)}{2^{H(U,V)+o(n)}} = 2^{-I(UV:Z)+o(n)}.$$

Note: for different typical pairs $\mathbf{U}, \mathbf{V}$ the probabilities of the event $\xi(\mathbf{U}, \mathbf{V}) = 1$ differ one from other only by the factor $2^{o(n)}$.

Now denote by $\tilde{\xi}(\mathbf{U}, \mathbf{V})$ another random variable:

$$\tilde{\xi}(\mathbf{U}, \mathbf{V}) = \begin{cases} 1, & \text{if the triple } \langle \mathbf{U}, \mathbf{V}, \mathbf{Z}_i\rangle \text{ is typical for at least one of } k \\ & \text{randomly chosen } \mathbf{Z}_i \ (i = 1, \dots, k) \\ 0, & \text{otherwise} \end{cases}$$

Obviously,

$$\text{Prob}[\tilde{\xi}(\mathbf{U}, \mathbf{V}) = 1] = 1 - (1 - \text{Prob}[\tilde{\xi}(\mathbf{U}, \mathbf{V}) = 1])^k = 1 - (1 - 2^{-I(UV:Z)+o(n)})^k.$$

Hence expectation of the number of typical pairs $\langle \mathbf{U}, \mathbf{V}\rangle$ covered by at least one of k randomly chosen sets is equal to

$$E(\sum_{\mathbf{U},\mathbf{V}} \tilde{\xi}(\mathbf{U},\mathbf{V})) = \sum_{\mathbf{U},\mathbf{V}} E(\tilde{\xi}(\mathbf{U},\mathbf{V})) = 2^{H(U,V)+o(n)} \cdot (1-(1-2^{-I(UV:Z)+o(n)})^k).$$

Now we can choose k such that

$$k = 2^{I(UV:Z)+o(n)} \cdot H(U,V) \cdot O(1)$$

and expectation of the number of *not covered* typical pairs $\langle \mathbf{U}, \mathbf{V}\rangle$ is less then 1. That means that a collection of k sets covering *all* pairs $\langle \mathbf{U}, \mathbf{V}\rangle$ does exist. And that is the collection we were looked for.

In fact, we have proved that there are lots of covering collections satisfying all our requirements. But we need only one such collection. Let us fix any of them. We can reduce covering sets from this collection so that every typical pair $\langle \mathbf{U}, \mathbf{V}\rangle$ belongs to exactly one covering set from the collection. Then get the collection of all reduced covering sets and add one more set: the set of all non-typical pairs $\mathbf{U}, \mathbf{V}$.

The collection of covering sets is constructed, and we can define random variable W. Let W be a covering set corresponding to the value of random pair $\langle U, V\rangle$. Now it is easy to check that it satisfies all the requirements of the lemma. Really, the number of values of W is k. Hence its entropy is not larger then $\log k = I(UV:Z) + o(N)$. Further,

$$H(UV|W) = \sum H(UV|W=\mathbf{w}) \cdot \text{Prob}[W=\mathbf{w}]$$

(the sum over all values $\mathbf{w}$ of W). Every $\mathbf{w}$ is a covering set from the constructed collection. If $\mathbf{w}$ is a set of typical pairs $\langle \mathbf{U}, \mathbf{V}\rangle$ then it contains at most $2^{H(UV|Z)+o(N)}$ values, and $H(UV|W=\mathbf{w}) \le H(UV|Z) + o(N)$. If $\mathbf{w}$ is the set of all *non-typical* pairs $\langle \mathbf{U}, \mathbf{V}\rangle$, then it contains much more elements (about c^N, where c is the number of all values $\langle u, v\rangle$). That means that entropy $H(UV|W=\mathbf{w})$ may be very large (about $N \log c$). But the probability of the event "W is the set of all non-typical pairs" tends to zero as $N \to \infty$. So after averaging we get $H(UV|W) \le H(UV|Z) + o(N)$. Analogously one can show $H(U|W) \le H(U|Z) + o(N)$ and $H(V|W) \le H(V|Z) + o(N)$. □

Remark 5. Let we have $(n+1)$ random variables $u_1, \dots, u_n, z$. Denote by U_i a sequence of N i.i.d. variables, each of them has the same distribution as

u_i. Then a straightforward generalization of our proof of the lemma 4 shows that there exist random variable W such that

$$H(W) \leq N \cdot I(u_1 \dots u_n : z) + o(N)$$

and

$$H(U_{i_1} \dots U_{i_k} | W) \leq N \cdot H(u_{i_1} \dots u_{i_k} | z) + o(N)$$

for any $1 \leq i_1 < \dots < i_k \leq n$. We do not know if this generalization can be used to prove new non Shannon type information inequality.

7.3 C. Why new inequalities are non-trivial?

We give here an explanation why constraint inequalities of Theorem 1 are not implied by basic inequalities, why inequalities from Theorem 2 are not implied by (3), and why the inequalities of Theorem 2 for $n = 2$ and $n = 3$ are not equivalent.

The constraint inequality of Theorem 1 is not implied by basic inequalities. Let us consider the following values of entropy function for a tuple of random variables $\langle a, b, c, d \rangle$:

$$\begin{array}{l} H(a) = H(b) = H(c) = H(d) = 2m, \\ H(a,b) = H(a,c) = H(a,d) = H(b,c) = H(b,d) = 3m, \\ H(c,d) = H(a,b,c) = H(a,b,d) = H(a,c,d) = \\ H(b,c,d) = H(a,b,c,d) = 4m, \end{array}$$

where m is any positive real (these values of entropy function was used in [10] to prove that (3) is a non Shannon type inequality). It is not hard to check that these entropy values satisfy all basic inequalities and conditions of Theorem 1 but do not satisfy the statement of Theorem 1 (for $n = 2, u = a, v = b, z = c, x_1 = c, x_2 = d$).

Inequalities from Theorem 2 for $n = 2$ is not implied by inequality 3 and basic inequalities. Let I_n denote the n-th inequality from Theorem 2 (with $(n + 3)$ random variables). We show that inequalities of type (3) together with basic inequalities do not imply I_2. Again we consider specific value of entropy function: it should satisfy all basic inequalities and all inequalities of type (3) but not satisfy inequality of I_2.

$$
\begin{aligned}
&H(u,v,z,x_1,x_2)=6,\\
&H(u,v,z)=H(u,v,z,x_1)=H(u,v,z,x_2)=6,\\
&H(u,v,x_1)=H(u,v,x_2)=H(u,v,x_1,x_2)=6,\\
&H(u,x_1,x_2)=H(v,x_1,x_2)=H(z,x_1,x_2)=6,\\
&H(v,z,x_1,x_2)=H(u,z,x_1,x_2)=H(v,z,x_1)=6,\\
&H(u)=2, H(v)=4, H(z)=2, H(x_1)=3, H(x_2)=3,\\
&H(u,v)=5, H(u,z)=3, H(v,z)=5,\\
&H(u,x_1)=H(u,x_2)=H(z,x_1)=H(z,x_2)=4, H(v,x_1)=H(v,x_2)=5,\\
&H(x_1,x_2)=6,\\
&H(u,z,x_1)=H(u,z,x_2)=4.8, H(v,z,x_2)=5.2.
\end{aligned}
$$

It is easy to check that for such entropy values I_2 is not satisfied. To check that the basic inequality and inequality of type (3) are true is not so easy. We used a computer program to verify this.

Inequalities from Theorem 2 for $n=2$ and for $n=3$ are not equivalent. As in the cases above, we consider specific values of entropy function, which satisfy all inequalities of type I_2 and all basic inequalities, but do not satisfy the inequality I_3.

$$
\begin{aligned}
&H(u) = H(v) = H(z) = 3,\\
&H(x_1) = H(x_2) = H(x_3) = 2,\\
&H(u,v) = H(u,z) = H(v,z) = 4.5,\\
&H(u,x_1) = H(u,x_2) = H(u,x_3) = 4,\\
&H(v,x_1) = H(v,x_2) = H(v,x_3) = 4,\\
&H(z,x_1) = H(z,x_2) = H(z,x_3) = 4,\\
&H(x_1,x_2) = H(x_1,x_3) = H(x_2,x_3) = 4,\\
&H(z,x_1,x_2,x_3) = H(u,v,z,x_1) = H(u,v,z,x_2) = 6,\\
&H(u,v,z,x_3) = H(u,v,x_1,x_2) = H(u,v,x_1,x_3) = H(u,v,x_2,x_3) = 6,\\
&H(u,z,x_1,x_2) = H(u,z,x_1,x_3) = H(u,z,x_2,x_3) = 5.5,\\
&H(u,x_1,x_2,x_3) = H(u,v,z,x_1,x_2,x_3) = H(v,z,x_1,x_2,x_3) = 6,\\
&H(u,v,z,x_1,x_2) = H(u,v,z,x_1,x_3) = H(u,v,z,x_2,x_3) = 6,\\
&H(u,v,x_1,x_2,x_3) = H(u,v,z) = 6,\\
&H(u,v,x_1) = H(u,v,x_2) = H(u,v,x_3) = 5.5,\\
&H(u,z,x_1) = H(u,z,x_2) = H(u,z,x_3) = 5,\\
&H(u,x_1,x_2) = H(u,x_1,x_3) = H(u,x_2,x_3) = 5,\\
&H(v,z,x_1) = H(v,z,x_2) = H(v,z,x_3) = 5.25,\\
&H(v,x_1,x_2) = H(v,x_1,x_3) = H(v,x_2,x_3) = 5,\\
&H(z,x_1,x_2) = H(z,x_1,x_3) = H(z,x_2,x_3) = 5,\\
&H(v,z,x_1,x_2) = H(v,z,x_1,x_3) = H(v,z,x_2,x_3) = 6,\\
&H(x_1,x_2,x_3) = 6,\\
&H(v,x_1,x_2,x_3) = 6,\\
&H(u,z,x_1,x_2,x_3) = 6.
\end{aligned}
$$

It is not hard to check that these entropy values do not satisfy the inequality I_3. To prove that these entropy values satisfy all basic inequalities and all inequalities of type I_2 we should check enormous number of inequalities. We verified this using a computer program.

References

[1] C. Shannon "A Mathematical Theory of Communication", Bell System Technical Jouranal, Vol. 27 (July and October 1948), pp. 379-423 and 623-656.

[2] R. Ahlswede, J. Körner, On the connection between the entropies of input and output distributions of discrete memoryless channels, *Proceedings of*

the 5th Brasov Conference on Probability Theory, Brasov, l974. Editura Academiei, Bucuresti, 1977, 13–23

[3] I. Csiszár, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless System.* Akademiai Kiado, Budapest, 1981.

[4] T. S. Han, "A uniqueness of Shannon's information distance and related nonnegativity problems." Journal of Combinatorics, Information and System Sciences, vol. 6, pp. 320-321, 1981.

[5] N. Pippenger, "What are the laws of information theory?", Specific Problems on Communication and Computational Conference, 1986.

[6] T. M. Cover, J. A. Thomas, *Elements of Information Theory.* Wiley, New York, 1991.

[7] D. Hammer, A. Romashchenko, A. Shen, and N. Vereshchagin. Inequalities for Shannon entropy and Kolmogorov complexity. Journal of Computer and Systems Sciences 60 (2000) 442-464. Preliminary version appeared in Proc. Twelfth Annual IEEE Conference on Computational Complexity, Ulm, Germany June 1997, 13-23.

[8] M. Li, P. Vitányi. *An introduction to Kolmogorov complexity and its applications.* Second edition. Springer Verlag, 1997.

[9] Z. Zhang, R. W. Yeung, "A non-Shannon-type conditional information inequality." IEEE Trans. Inform. Theory, vol.43, pp.1982-1986, November 1997.

[10] R. W. Yeung, Z. Zhang, "On Characterization of entropy function via information inequalities." IEEE Trans. Inform. Theory, vol.44, pp.1440-1450, July 1998.

[11] A. Romashchenko, N. Vereshchagin, A. Shen, "Combinatorial Interpretation of Kolmogorov Complexity". Proc. of 15th Annual IEEE Conference on Computational Cimplexity, July 2000, Florence, Italy, pp. 131-137.

[12] R. W. Yeung, Z. Zhang, "A class of non-Shannon type inequalities and their applications." Communications in Information and Systems, vol. 1, pp. 87-100, 2001.

[13] H.-L. Chan, R. W. Yeung, "New approach to information inequalitie, Part I: A combinatorial approach," submitted to IEEE Transactions on Information Theory.

[14] H.-L. Chan, R. W. Yeung, "New approach to information inequalitie, Part II: Algebraic analysis of entropy function," submitted to IEEE Transactions on Information Theory.