COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 2, No. 2, pp. 147-166, December 2002

003

A NEW CLASS OF NON-SHANNON-TYPE INEQUALITIES FOR ENTROPIES*

KONSTANTIN MAKARYCHEV†, YURY MAKARYCHEV†, ANDREI ROMASHCHENKO‡, AND NIKOLAI VERESHCHAGIN§

Abstract. In this paper we prove a countable set of non-Shannon-type linear information inequalities for entropies of discrete random variables, i.e., information inequalities which cannot be reduced to the "basic" inequality $I(X:Y|Z) \geq 0$. Our results generalize the inequalities of Z. Zhang and R. Yeung (1998) who found the first examples of non-Shannon-type information inequalities.

1. Introduction. A central notion of information theory is Shannon's entropy[1]. Given a set of jointly distributed random variables $x_1, \ldots, x_n$, 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 nonempty subsets of $\{x_1, \ldots, x_n\}$). For every n-tuple of random variables we get a point in $\mathbb{R}^{2^n-1}$, representing entropies of the given distribution. Following [10] we call a point in $\mathbb{R}^{2^n-1}$ *constructible* if it represents entropy values of some collection of n random variables. The set of all constructible points is denoted by $\Gamma_n^\star$.

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

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

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 interesting applications of information inequalities beyond information theory were found [11, 13, 14]. So investigation of the class of all valid information inequalities is an interesting problem. We refer the reader to [15] for a comprehensive treatment of the subject.

*Invited paper. Received on December 21, 2001; accepted for publication on August 21, 2002. Supported by RFBR grant #01-01-01028 and grant #02-01-2201.

†Dept. of Mathematical Logic and Theory of Algorithms, Moscow State University, Vorobjevy Gory, Moscow 119899, Russia, E-mail: makarychev@imail.ru

‡Institute of Problems of Information Transmission (Moscow), Bolshoy Karetny, 19, Moscow 101447, Russia, E-mail: anromash@mccme.ru

§Dept. of Mathematical Logic and Theory of Algorithms, Moscow State University, Vorobjevy Gory, Moscow 119899, Russia, E-mail: ver@mccme.ru

[1]An original paper of C. E. Shannon [1] is an excellent introduction in this field. We also recommend to the reader who is not familiar with information theory the books [5] and [3].

In this paper we consider only discrete random variables. We restrict ourselves to random variables with a finite range. This restriction is not very essential: if an inequality is true for all random variables with a finite range, it is also true for random variables with a countable range[2].

Let us give a brief review of the background. For many years only trivial inequalities for entropies were known. Namely, all known inequalities were non-negative linear combinations of *Shannon's basic inequalities*, i.e., inequalities of the form

$$H(A \cup C) + H(B \cup C) - H(A \cup B \cup C) - H(C) \geq 0, \tag{2}$$

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

It can be shown that for any $n \leq 3$ all information inequalities that are valid for n discrete random variables are linear combinations of the basic inequalities [4] (see also [6][7]). In 1998 Zhang and Yeung [10] came up with a linear inequality for entropies of 4 random variables which cannot be reduced to the basic Shannon 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 notation, 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 $n \geq 1$):

$$H(x_1 x_2 \dots x_n) + nI(u : v : x_1) \leq \sum_{j=1}^{n} H(x_j) + \sum_{j=1}^{n} I(u : v|x_j) + I(uv : x_1). \tag{4}$$

Here we use the notation $I(u : v : x_1)$. Let us remind the reader that the mutual information of three random variables $I(a : b : c)$ is defined as

$$I(a : b : c) = H(a) + H(b) + H(c) - H(ab) - H(ac) - H(bc) + H(abc).$$

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

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

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

[2]This fact can be easily proved by approximation of a distribution with a countable range by distributions with a finite range. We do not know if the same is true for *constrained* inequalities (that is, for inequalities which are true assuming another inequality).

(This result gives a new bound for the set $\Gamma_4^\star$, but not for $\overline{\Gamma_4^\star}$.)

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

In the present paper we exhibit a new countable family of unconstrained inequalities for Shannon entropy.

THEOREM 1. *For any random variables* $u, v, z, x_1, \ldots, x_n$

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

Note that the inequality in Theorem 1 implies that of Zhang and Yeung. In fact, by letting $z = x_1$, we get (4).

The rest of the paper is organized as follows. In Section 2 we present a pure syntactical inference rule producing a new inequality given a constrained inequality of a special type. The proof of this inference rule is a refinement of Zhang-Yeung's proof of (4). Theorem 1 will be proved by this method. In Sections 3 and 4 we prove Theorem 1 using another method. It appeals to the idea of *rate region* of two random variables introduced by R. Ahlswede and J. Körner [2, 3]. 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 of the fact that the inequality of Theorem 1 for $n = 2$ does not follow from the basic inequalities and (3), and that the inequalities of Theorem 1 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. A Syntactical Inference Rule for Information Inequalities and First Proof of Theorem 1. We first give a proof for Theorem 1 and then extend the argument to get a general inference rule.

Proof. [Proof of the theorem] First rewrite the inequality of Theorem 1 in the following form:

$$H(x_1 \ldots x_n) + nI(u : v) \le \sum_{i=1}^{n} I(u : v|x_i) + nI(u : v|z) + \sum_{i=1}^{n} H(x_i) + I(uv : z). \tag{6}$$

We first note that this inequality is true under the condition $\langle x_1, \ldots, x_n\rangle$ 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 $x_1 \ldots x_n$ and z are independent given u, v.

Let us prove that (6) holds for any $x_1, \ldots, x_n$, u, v, z. The crucial point is that no term in (6) has any of x_i's together with z. Given $x_1, \ldots, x_n, u, v, z$ we 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 $\tilde{z}$ 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

$$\begin{aligned}&\text{Prob}[\tilde{z} = \mathbf{z} \mid \langle x_1, \dots, x_n, u, v\rangle = \langle \mathbf{x_1}, \dots, \mathbf{x_n}, \mathbf{u}, \mathbf{v}\rangle]\\ &= \text{Prob}[z = \mathbf{z} \mid u = \mathbf{u}, v = \mathbf{v}].\end{aligned}$$

By definition $\tilde{z}$ is independent of $x_1, \dots, x_n$ given u, v. Hence, by Lemma 8 the inequality (6) holds for z replaced by $\tilde{z}$. But this replacement does not change any term in the inequality (6) because for any of its terms all its variables are included either in the set $\{x_1, \dots, x_n, u, v\}$ or 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 $\langle z, u, v\rangle$ have the same distribution. □

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

THEOREM 2 (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 in 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 $V_{t_1}, \dots, V_{t_m}$ are independent given W_s, where $t_1, \dots, t_m$ are all the sons of s, W_s stands for the tuple of random variables assigned to the predecessors of s (including s) and V_t for the tuple of random variables assigned to the 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 $x_1, \dots, x_n$ 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 a tree consisting of 5 vertices: the root, its two sons, and two sons of the left son. The following variables are assigned to them: u (the root), y (the left son), v (the right son), x, z (the sons of the left son).

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

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

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

We construct such $\tilde{x}_1, \ldots, \tilde{x}_n$ by induction. Without loss of generality, 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 $\tilde{x}_1, \ldots, \tilde{x}_{i-1}$ are defined. Let s denote the node x_i is assigned to. Recall that W_s denotes the tuple of all the random variables assigned to predecessors of s (including s). Let W_{si} be the tuple of those random variables from $x_1, \ldots, x_{i-1}$ 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

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

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

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

3. A Collection of Constrained Non-Shannon-Type Inequalities. In this section we give a short proof of one collection of *constrained* non-Shannon-type inequalities.

COROLLARY 3. *For any random variables $u, v, z, x_1, \ldots, x_n$ if*

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

then

$$H(x_1, \ldots, x_n) + (n-1) \cdot I(uv : z) \le \sum_{i=1}^{n} I(u : v|x_i) + \sum_{i=1}^{n} H(x_i). \tag{7}$$

Corollary 3 follows immediately from Theorem 1. Indeed, using the equality $I(u : v : z) = I(uv : z) - I(u : z|v) - I(v : z|u)$, we can rewrite the inequality in Theorem 1 as

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

and we are done. However, we give a direct proof of Corollary 3 as an illustration of our method. In the next section we explain how to modify this reasoning to get another proof of Theorem 1.

LEMMA 4 (Double Markov Property). *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., $I(uv : z|w) = 0$),*
- $H(w) \geq I(uv : z)$.

This lemma was given as an exercise in [3], so we only give a sketch of the proof here.

Sketch of proof: Let w be the conditional distribution of the variable z given fixed values u, v. There is a finite set of different values $\langle u, v\rangle$, so w has a finite range. Since $I(u : z|v) = I(v : z|u) = 0$, the conditional distributions of z given u, given v, and given uv are the same. Hence w is a deterministic function of u and a deterministic function of v (i.e., $H(w|u) = 0$ and $H(w|v) = 0$). The condition $I(uv : z|w) = 0$ is obvious. And $H(w) \geq I(uv : z)$ because $H(w) \geq I(w : z) = I(uv : z)$. □

Let us apply the lemma above to the variables u, v, z given in the condition in Corollary 3, and consider the joint distribution of the tuple $\langle u, v, z, w\rangle$ where w is the variable constructed in the lemma. From [7], we have the following inequality (we present its proof in Appendix A):

$$H(c|d) \leq H(c|ad) + H(c|bd) + I(a : b|d) \tag{8}$$

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

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

Recall that $H(w) \geq I(uv : z)$, and we get the required inequality.

4. The second proof of Theorem 1. We want to generalize the proof from the last section to get the inequality of Theorem 1. To this end we need an analog of the lemma on the Double Markov Property. Such an analog is implied by the following lemma (it follows immediately from Theorem 2 in [2]):

LEMMA 5 (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*

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

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 $i = 1, 2, \ldots, N$ are independent. Let

$$\begin{aligned} U &= u_1, \ldots, u_N, \\ V &= v_1, \ldots, v_N, \\ Z &= z_1, \ldots, z_N. \end{aligned}$$

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

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

The original proof of Ahlswede and Körner refers to a rather non-trivial approximation technique. For the sake of completeness we give a self-contained proof of this lemma in Appendix B.

Remark 1. *A reader familiar with the notion of rate region for two random variables [2, 3] can easily note that both Lemmas 4 and 5 state that some point belongs to the rate region of u and v. Below we show how to use this point to get a non-trivial information inequality for the variables involved. Formally, we do not need here the definition of rate region, so we omit it.*

Remark 2. *The inequality $H(UV|W) \le H(UV|Z) + o(N)$ implies the condition $H(W) \ge I(UV : Z) + o(N)$, so the last equality in the lemma above could be substituted by a weaker condition $H(W) \le I(UV : Z) + o(N)$.*

Let us prove that the random variable W from Lemma 5 satisfies the inequalities

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

The first inequality above is implied by the lemma. To prove the second inequality, note that

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

The third inequality is proved in a similar way.

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

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

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

Let

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

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

Repeat the proof of Corollary 3 for $\langle U, V, Z, X_1, \ldots, X_n\rangle$ in place of the tuple $\langle u, v, z, x_1, \ldots, x_n\rangle$ and use Lemma 5 instead of the lemma on the Double Markov property. Instead of the equalities $H(w|u) = H(w|v) = 0$, we have the inequalities $H(W|U) \le N \cdot I(v : z|u) + o(N)$ and $H(W|V) \le N \cdot I(u : z|v) + o(N)$. Therefore, we obtain the inequality

$$\begin{aligned} &N \cdot (H(x_1, \ldots, x_n) + (n-1) \cdot I(uv : z)) + o(N) \le \\ &N \cdot (\sum_{i=1}^{n} I(u : v|x_i) + \sum_{i=1}^{n} H(x_i) + n(I(v : z|u) + I(u : z|v))) + o(N) \end{aligned}$$

instead of (7). As this inequality holds for any N, we have

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

so the proof is completed. □

5. An Inference Rule for Deducing New Linear Inequalities for Kolmogorov Complexity. In this section, we assume that the reader is familiar with plain Kolmogorov complexity $K(x)$; unfamiliar readers can consult [8]. We will present an 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 $S(x_1, \ldots, x_n) \le 0$ in variables $x_1, \ldots, x_n$ ranging over binary strings whose left hand side is a sum of terms of the form $K(\langle x_{i_1}, \ldots, x_{i_m}\rangle)$. The inference rule allows one to prove that such an inequality is valid up to an additive term logarithmic in the complexity of $x_1, \ldots, x_n$. That is, for some constant c and for any strings $x_1, \ldots, x_n$, $S(x_1, \ldots, x_n) \le c \log k$ holds, where $k = K(x_1) + \cdots + K(x_n)$.

The rule will be defined in terms of infinite sequences of strings. It is easy to see that an inequality $S \le 0$ holds up to an additive $O(\log k)$ term if and only if for any sequences $x_1^i, \ldots, x_n^i$ of strings, $S(x_1^i, \ldots, x_n^i) \le O(\log k^i)$ holds, where $k^i = K(x_1^i) + \cdots + K(x_n^i)$. Indeed, the "only if" part is evident. To prove the "if" part, assume that for any c there are $x_1^c, \ldots, x_n^c$ 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 $S(x_1^c, \ldots, x_n^c) \le O(\log k^c)$ 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 $K(u_1^i|v^i)+\cdots+K(u_m^i|v^i) \le K(u_1^i, \ldots, u_m^i|v^i)+O(\log k^i)$, where k^i denotes the sum of the 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 to 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 $S(x_1^i, \ldots, x_n^i) \le O(\log k^i)$ 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 the variables assigned to all the successors of the jth son of s (including the son itself), and w^i denotes the tuple consisting of all the variables assigned to all the predecessors of s (including s). Then for some c and for any string S, the inequality $S \le c \log k$ holds.*

REMARK 3. *A theorem from [7] states that an unconstrained Kolmogorov complexity inequality $S \le 0$ is true up to an additive logarithmic term if and only if the inequality $S' \le 0$ which is obtained from it by replacing all the strings by random variables and all the Kolmogorov complexities by Shannon entropies is true. So one can 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 an analogy of the theorem from [7] is true for constrained inequalities.*

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

Without loss of generality, we may assume that the number of nodes in the tree is equal to the number of variables and the assignment of 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 by nodes.

It suffices to show that for some c and 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 $x'_{i_1}, \ldots, x'_{i_m}$ differ by $c \log k$ in each component.
- For any internal node s,

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

(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 proved by induction on the height of the tree. To prove the induction step we need to strengthen the statement: we will assume that all the complexities are conditional on some string y. So the conditional version proved by induction is as follows: for some c and for any strings $x_1, \ldots, x_n, y$, there are strings $x'_1, \ldots, x'_n$ such that for any path $x_{i_1}, \ldots, x_{i_m}$ in the tree, the complexity vectors of $x_{i_1}, \ldots, x_{i_m}$ and $x'_{i_1}, \ldots, x'_{i_m}$ conditional on y differ by $O(\log k)$ in each component,

and for any internal node s, $K(u'_1|w',y)+\dots+K(u'_m|w',y)-K(u'_1\dots u'_m|w',y)\le c\log k$.

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

Induction step. To prove the induction step we need to invoke the main tool of [7], the "typization" argument. For some c depending only on n given variables $x_1,\dots,x_n$, 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 $m\ge 1, l\ge 0$),

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

This set is large: for some constant c_1, $\log|M'|\ge K(x_1,\dots,x_n)-c_1\log k$ (because the tuple $x_1,\dots,x_n$ can be described by its index in the enumeration of this set and by the extra $O(\log k)$ bits describing the set itself). By the 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 $j_1,\dots,j_l$,

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

Now we are in the position to prove the induction step. Regard our tree as a collection of subtrees rooted at the sons of the root. Without loss of generality, assume that the string x_1 is assigned to the root, the strings $x_2,\dots,x_l$ are assigned to the nodes of the subtree rooted at the leftmost son of the root, the strings $x_{l+1},\dots,x_m$ are assigned to the 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 has the desired property that the complexity vectors of the tuples $x_1,x'_2,\dots,x'_l$ and $x_1,x_2,\dots,x_l$ conditional on y are the same (up to a $O(\log k)$ term). Repeat the same steps for the other subtrees and gather together all the strings obtained. We get strings $x'_2,\dots,x'_n$ such that the sequence $x_1,x'_2,\dots,x'_n$ satisfies all but one of the the requirements: $K(x'_2,\dots,x'_n|x_1,y)$ 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 on $\langle x_1, y \rangle$ and repeat the same steps for all the subtrees. We thus get sets $M_1, M_2, \dots$. For at least one tuple in their Cartesian product, $K(x'_2, \dots, x'_n | x_1, y) \geq \log |M_1| + \log |M_2| + \dots$ holds, and these sets are big enough: $\log |M_1| + \dots + \log |M_m| \geq K(x'_2, \dots, x'_l | x_1, y) + K(x'_{l+1}, \dots, x'_m | x_1, y) + \dots$ □

6. Conclusion. In the present paper we have proved a family of new non-Shannon-type inequalities. We have used two different methods to prove the main result. The first method (Section 2) uses an inference rule which allows us to deduce new unconstrained inequalities given constrained inequalities of a certain type. The second method (Sections 3 and 4) appeals to the idea of rate region of Ahlswede and Körner. In Section 5 we prove that an analog of our inference rule can be applied to inequalities for Kolmogorov complexity.

The method in Section 2 is based on the ideas of Zhang and Yeung [10]. So it is not surprising that our inequalities are very similar to (4). It is perhaps more surprising that the method in Section 4 gives the same results. Probably there is a relation between these two methods, and an interesting problem is to reveal this relation. Another question is whether any other new inequality can be proved by using our methods.

Here are a few other open questions.

- Are all the inequalities in Theorem 1 independent? We can only prove that inequality (3) together with the basic inequalities do not imply the inequality of Theorem 1 for $n = 2$, and that our inequalities for $n = 2$ and $n = 3$ are not equivalent modulo the basic inequalities (see Appendix C).
- Obviously, inequalities of Zhang and Yeung (4) are implied by Theorem 1. Is the converse true? Namely does all the family of inequalities of type (4) ($n = 1, 2, \dots$) imply Theorem 1?
- Inequalities of Corollary 3 are implied by corresponding non-constraint inequalities of Theorem 1. Is there any true constraint inequality

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

for entropies for which its non-constraint analog

$$S \leq C \cdot (S_1 + \dots + S_n)$$

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 any constraint inequality which is valid for random variables with a finite range but not valid for random variables with a countable range exist?

Acknowledgment

The authors would like to thank Raymond Yeung, who helped us to make the paper much more readable. The authors also thank the reviewer for the detailed comments and suggestions.

7. Appendix.

A. The proof of some useful Shannon-type inequalities

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

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

holds.

Proof. This inequality is an easy consequence of the basic inequality

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

Subtracting $2H(d)$ 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 to be proved. □

LEMMA 8. *The inequality* (6) *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, we first show that

$$I(a:b) \le I(a:b|c) + I(a:b|d) + I(c:d) + I(c:d|ab). \tag{9}$$

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 the entropies of a, b, c, d, their pairs, their triples, etc, to get the symmetric expression

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

To prove (9), we first consider

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

On the other hand, we have

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

Combining these two inequalities, we get (9). Let $a = u$, $b = v$, $c = x_i$, and $d = z$ in this inequality. Then we get

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

Note that the last equality holds as x_i and z are independent given u and v. Summing these inequalities over all i, we obtain

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

So to get (6) it remains to show that

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

or equivalently

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

This can be proven as follows:

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

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

B. The proof of Lemma 5

In this section we prove Lemma 5. We use in our proof the method of *typical sequences*. We omit the tiresome technical details and refer the reader to [3] for an introduction to the standard technique of typical sequences.

We will use $\mathbf{U}, \mathbf{V}, \mathbf{Z}$ to denote n-vectors of i.i.d. copies of the random variables U, V, 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 the values in the triple is close to 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

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

(for every triple $\langle \alpha, \beta, \gamma \rangle$).

REMARK 4. *The choice of the bound for the difference between probability and frequency is not very important here. We can get here any function $\theta(N)$ such that*

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

instead of $\sqrt{N} \log N$. 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 $2^{N \cdot H(u,v,w)+o(N)}$*
- *$\mathrm{Prob}[\langle U, V, Z \rangle$ 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 the 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 *the set of typical pairs* $\langle \mathbf{U}, \mathbf{V} \rangle$. Analogously, we consider the projection of the set of all typical triples onto the third coordinate and call it *the set of all typical* $\mathbf{Z}$.

We use the following additional properties of typical values which can also be proved by a standard counting technique (see detailed proof in Chapter 1 of [3] or in Section 5.3 of [15]):

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

(**) For every typical $\mathbf{Z}$ such that $\langle \mathbf{U}', \mathbf{V}', \mathbf{Z} \rangle$ is typical for some $\mathbf{U}', \mathbf{V}'$, 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 $\langle \mathbf{U}, \mathbf{V} \rangle$ such that $\langle \mathbf{U}, \mathbf{V}, \mathbf{Z}' \rangle$ is typical for some $\mathbf{Z}'$, 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. the 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 satisfies the second and the third conditions above (see (*) and (**)).

So we have candidates for the covering sets (one set for each typical $\mathbf{Z}$). If Z and $\langle U, V \rangle$ are independent, we have got the required collection of covering sets. But in the general case there are 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 the 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), we will choose a collection of sets which covers all typical $\langle \mathbf{U}, \mathbf{V} \rangle$ with a high probability. Specifically, let us fix a typical pair $\langle \mathbf{U}, \mathbf{V} \rangle$ and consider a 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 $\langle \mathbf{U}, \mathbf{V} \rangle$,

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

Let us note that for different typical pairs $\langle \mathbf{U}, \mathbf{V} \rangle$, the probabilities of the event "$\xi(\mathbf{U}, \mathbf{V}) = 1$" differ from each other only by the factor $2^{o(n)}$.

Now introduce another random variable

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

Obviously,

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

Hence, the 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 ϵ such that

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

and the expectation of the number of typical pairs $\langle \mathbf{U}, \mathbf{V} \rangle$ which are *not covered* is less than 1. This means that a collection of k sets covering *all* the pairs $\langle \mathbf{U}, \mathbf{V} \rangle$ does exist, and that is the collection we are looking 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 one of them. Now we want to make the covering sets disjunctive. This can be achieved by reducing the sizes of the covering sets in the collection so that every typical pair $\langle \mathbf{U}, \mathbf{V} \rangle$ belongs to exactly one covering set in the collection. Then we get the collection of all the reduced covering sets with one additional set: the set of all non-typical pairs $\langle \mathbf{U}, \mathbf{V} \rangle$.

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

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

(the sum is 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 $H(UV|W = \mathbf{w}) \leq H(UV|Z) + o(N)$. 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 $\langle u, v \rangle$). This means that the 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 $H(UV|W) \leq H(UV|Z) + o(N)$. Analogously $H(U|W) \leq H(U|Z) + o(N)$ and $H(V|W) \leq H(V|Z) + o(N)$. □

REMARK 5. *Let $u_1, \ldots, u_n, z$ be $(n+1)$ random variables. Denote by U_i a sequence of N i.i.d. variables, each of them having the same distribution as u_i. Then a straightforward generalization of our proof of the Lemma 5 shows that there exists a random variable W such that*

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

and

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

for any $1 \leq i_1 < \ldots < i_k \leq n$.
Question: Can we use the generalization above to prove a new non Shannon type information inequality (with the method from Section 4)?

TABLE 1

tuple of r.v.	entropy of a tuple
u	2
v	4
z	2
x_1	3
x_2	3
u, v	5
u, z	3
v, z	5
u, x_1	4
u, x_2	4
z, x_1	4
z, x_2	4
v, x_1	5
v, x_2	5
x_1, x_2	6

tuple of r.v.	entropy of a tuple
u, x_1, x_2	6
v, x_1, x_2	6
z, x_1, x_2	6
u, v, x_1	6
u, v, x_2	6
u, z, x_1	4.8
u, z, x_2	4.8
v, z, x_1	6
v, z, x_2	5.2
u, v, z	6
u, v, x_1, x_2	6
u, z, x_1, x_2	6
v, v, x_1, x_2	6
u, v, z, x_1	6
u, v, z, x_2	6
u, v, z, x_1, x_2	6

TABLE 2

tuple of r.v.	entropy of a tuple	tuple of r.v.	entropy of a tuple
		u, z, x_1	5.5
u	3	v, z, x_1	5.5
v	3	u, v, x_2	5.5
z	3	u, z, x_2	5.5
x_1	2	v, z, x_2	5.5
x_2	2	u, v, x_3	5.5
x_3	2	u, z, x_3	5.5
u, v	4.5	v, z, x_3	5.5
u, z	4.5	u, v, z	6
v, z	4.5	x_1, x_2, x_3	5
u, x_1	4	u, x_1, x_2, x_3	6
u, x_2	4	v, x_1, x_2, x_3	6
u, x_3	4	z, x_1, x_2, x_3	6
v, x_1	4	u, v, x_1, x_2	6
v, x_2	4	u, z, x_1, x_2	6
v, x_3	4	v, z, x_1, x_2	6
z, x_1	4	u, v, x_1, x_3	6
z, x_2	4	u, z, x_1, x_3	6
z, x_3	4	v, z, x_1, x_3	6
x_1, x_2	3.5	u, v, x_2, x_3	6
x_1, x_3	3.5	u, z, x_2, x_3	6
x_2, x_3	3.5	v, z, x_2, x_3	6
u, x_1, x_2	5	u, v, z, x_1	6
v, x_1, x_2	5	u, v, z, x_2	6
z, x_1, x_2	5	u, v, z, x_3	6
u, x_1, x_3	5	u, v, x_1, x_2, x_3	6
v, x_1, x_3	5	u, z, x_1, x_2, x_3	6
z, x_1, x_3	5	v, z, x_1, x_2, x_3	6
u, x_2, x_3	5	u, v, z, x_1, x_2	6
v, x_2, x_3	5	u, v, z, x_1, x_3	6
z, x_2, x_3	5	u, v, z, x_2, x_3	6
u, v, x_1	5.5	u, v, z, x_1, x_2, x_3	6

C. Why are the new inequalities non-trivial?

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

Inequalities in Theorem 1 for $n = 2$ are not implied by inequalities of type (3) and the basic inequalities.. Let I_n denote the n-th inequality in Theorem 1 (with $(n + 3)$ random variables). We show that inequalities of type (3) together with the basic inequalities do not imply I_2. We consider specific values of the entropy function: it should satisfy all basic inequalities and all inequalities of type (3) but not inequality I_2. We provide values of entropies in Table 1 (e.g. $H(u) = 2$, $H(v) = 4, \ldots, H(z, x_1) = 4$, etc.)

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

Inequalities in Theorem 1 for $n = 2$ and for $n = 3$ are not equivalent.. As in the cases above, we consider specific values of the entropy function (see Table 2) which satisfy all inequalities of type I_2 and all the basic inequalities, but not the inequality I_3.

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 the basic inequalities and all inequalities of type I_2, we need to check an enormous number of inequalities. We have verified this by a computer program.

REFERENCES

[1] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal, 27(July and October 1948), pp. 379–423 and 623–656.

[2] R. Ahlswede and 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, pp. 13–23.
More recent (unpublished) version "*On Common Information and Related Characteristics of correlated information sources*" is available at
`http://www.mathematik.uni-bielefeld.de/ahlwede/pub/ahlswede/sources.ps`

[3] I. Csiszár and 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, 6(1981), pp. 320–321.

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

[6] F. Matúš, *Probabilistic conditional independence structures and matroid theory: Background*, Int. J. of General Syst., 22(1994), pp. 185–196.

[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), pp. 442-464. Preliminary version appeared in Proc. Twelfth Annual IEEE Conference on Computational Complexity, Ulm, Germany June 1997, pp. 13-23.

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

[9] Z. Zhang and R. W. Yeung, *A non-Shannon-type conditional information inequality.* IEEE Transactions on Information Theory, 43(1997), pp. 1982–1986.

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

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

[12] R. W. Yeung and Z. Zhang, *A class of non-Shannon type inequalities and their applications*. Communications in Information and Systems, 1(2001), pp. 87–100, (`http://www.ims.cuhk.edu.hk/~cis`).

[13] T. H. Chan, *A combinatorial approach to information inequalities*, Comm. Inform. & Syst., 1(2001), pp. 241–253, (`http://www.ims.cuhk.edu.hk/~cis/`).

[14] T. H. Chan and R. W. Yeung, *On a relation between information inequalities and group theory*, to appear in IEEE Trans. Inform. Theory, July 2002.

[15] R. W. Yeung, *A First Course in Information Theory*, Kluwer Academic/Plenum Publishers, 2002.