Inequalities for Shannon entropies and Kolmogorov complexities

D. Hammer
Technische Universität, Berlin
hammer@math.tu-berlin.de

A. E. Romashchenko
Moscow State University

A. Shen
Institute of Problems of
Information Transmission, Moscow
shen@landau.ac.ru, shen@mccme.ru

N.K. Vereshchagin
Moscow State University
ver@mech.math.msu.su

Abstract

The paper investigates connections between linear inequalities that are valid for Shannon entropies and for Kolmogorov complexities.

1 Introduction

From the very beginning the notion of complexity of finite objects was considered as an algorithmic counterpart to the notion of Shannon entropy. Kolmogorov's paper [4] was called "Three approaches to the quantitative definition of information"; Shannon entropy and algorithmic complexity were among these approaches.

It was mentioned by Kolmogorov in [5] that the properties of algorithmic complexity and Shannon entropy are similar. We investigate one aspect of this similarity. Namely, we are interested in linear inequalities that are valid for Shannon entropies and for Kolmogorov complexities.

It turns out that

(1) all inequalities that are valid for Kolmogorov complexities, are also valid for Shannon entropies and vice versa;

(2) all inequalities that are valid for Shannon entropies, are valid for ranks of finite subsets of linear space;

(3) the opposite statement is not true: Ingleton's inequality ([3], see also [8]) is valid for ranks but not for Shannon entropies;

(4) for some special cases all three classes of inequalities coincide and have simple description.

We present an inequality for Kolmogorov complexities that implies Ingleton's inequality for ranks; another application of this inequality is a new simple proof of one of Gács–Körner's results on common information [1].

Acknowledgements. The work of Moscow authors was supported in part by the INTAS project No 93-0893. A. Shen thanks also the Volkswagen Foundation for support and Bonn University and Prof. M. Karpinski for hospitality.

2 Shannon entropy and Kolmogorov complexity

Let α be a random variable with a finite range $a_1, \ldots, a_n$. Let p_i be the probability of the event $\alpha = a_i$. Then the Shannon entropy of α is defined as

$$H(\alpha) = -\sum_i p_i \log p_i.$$ [1]

Using the convexity of the function $-x \log x$, one can prove that the Shannon entropy of random variables does not exceed the logarithm of its range (and is

[1] All logarithms in the paper are base 2

equal to the logarithm of the range for uniformly distributed variables).

Let β be another variable with a finite range $b_1, \dots, b_k$ defined on the same probabilistic space as α is. We define $H(\alpha|\beta = b_j)$ in the same way as $H(\alpha)$; the only difference is that p_i is replaced by the conditional probability $\Pr[\alpha = a_i|\beta = b_j]$. Then we define the conditional entropy as

$$H(\alpha|\beta) = \sum_j \Pr[\beta = b_j] \cdot H(\alpha|\beta = b_j).$$

It is easy to check that

$$H(\langle\alpha, \beta\rangle) = H(\beta) + H(\alpha|\beta).$$

Using the convexity of the function $-x \log x$, one can prove that

$$H(\alpha|\beta) \le H(\alpha),$$

and that

$$H(\alpha|\beta) = H(\alpha)$$

if and only if α and β are independent.

In other terms,

$$H(\langle\alpha, \beta\rangle) \le H(\alpha) + H(\beta).$$

The mutual information in α and β is defined as

$$I(\alpha : \beta) = H(\alpha) - H(\alpha|\beta) = H(\alpha) + H(\beta) - H(\langle\alpha, \beta\rangle).$$

The mutual information $I(\alpha : \beta)$ is always non-negative and is equal to 0 if and only if α and β are independent.

The conditional version of mutual information is defined as

$$I(\alpha : \beta|\gamma) = H(\alpha|\gamma) + H(\beta|\gamma) - H(\langle\alpha, \beta\rangle|\gamma)$$

and is always non-negative, too. This is proved as follows. For any possible value c_i of γ we have

$$H(\alpha|\gamma = c_i) + H(\beta|\gamma = c_i) - H(\langle\alpha, \beta\rangle|\gamma = c_i) \ge 0.$$

Multiplying this inequality by $\Pr[\gamma = c_i]$ and summing over i yields the desired inequality.

All these notions have their counterparts in Kolmogorov complexity theory.

The Kolmogorov complexity of a binary string a is defined as the minimal length of a program that generates a. There are different refinements of this idea (called *simple* Kolmogorov complexity, *monotone* complexity, *prefix* complexity, *decision* complexity, see [6], [7]). However, for our purposes the difference is not important, since all these complexity measures differ only by $O(\log m)$ where m is the length of a. Therefore, in the sequel we denote Kolmogorov complexity of a binary string a by $K(a)$ not specifying which version we use, and *all our equalities and inequalities are valid up to $O(\log m)$ term where m is the total length of all strings involved.*

The conditional complexity $K(a|b)$ is defined as the minimal length of a program that produces a having b as input; one can prove that

$$K(b|a) = K(\langle a, b\rangle) - K(a),$$

(see e.g. [9] for the proof). Here $\langle a, b\rangle$ denotes the encoding of the pair a, b by a binary string (different computable encodings lead to complexities that differ only by $O(1)$). As always, $O(\log m)$ additive term is omitted. So the precise meaning of this equality is as follows: there exists constants p, q such that

$$K(b|a) \le K(\langle a, b\rangle) - K(a) + p \log(|a| + |b|) + q,$$
$$K(\langle a, b\rangle) - K(a) \le K(b|a) + p \log(|a| + |b|) + q$$

for all binary words a, b.

The mutual information is defined as

$$I(a : b) = K(b) - K(b|a).$$

An equivalent (up to $O(\log m)$ term) symmetric definition is

$$I(a : b) = K(a) + K(b) - K(\langle a, b\rangle).$$

As for the Shannon case, the mutual information is always non-negative (up to $O(\log m)$ term).

The conditional version of mutual information is defined as

$$I(a : b|c) = K(a|c) + K(b|c) - K(\langle a, b\rangle|c).$$

The inequality

$$I(a : b|c) \ge 0$$

is valid up to a logarithmic term, that is, $I(a : b|c) \ge -O(\log(|a| + |b| + |c|))$. This inequality plays an important role in the sequel.

3 Inequalities

We have already mentioned several inequalities for Shannon entropies and Kolmogorov complexities. Some others are known; for example, the inequality

$$2K(\langle a,b,c\rangle) \le \le K(\langle a,b\rangle) + K(\langle a,c\rangle) + K(\langle b,c\rangle). \quad (1)$$

This inequality is equivalent in a sense to the following geometric fact: if V is the volume of the set $A \subset \mathbb{R}^3$ and S_{xy}, S_{xz} and S_{yz} are areas of its three projections (on OXY, OXZ and OYZ), then

$$V^2 \le S_{xy} \cdot S_{xz} \cdot S_{yz}$$

(see [2]).

It turns out that the inequality (1), as well as all other known inequalities for Kolmogorov complexities, is a corollary of the inequalities and equalities of the following type

$$I(P:Q|R) \ge 0, \quad (2)$$
$$K(Q|P) = K(\langle P,Q\rangle) - K(P), \quad (3)$$
$$I(P:Q|R) = K(P|R) + K(Q|R) - K(\langle P,Q\rangle|R), \quad (4)$$

where P,Q,R are some tuples (possibly empty) of binary strings. Indeed, (1) is a consequence of the equality

$$2K(\langle a,b,c\rangle) = = K(\langle a,b\rangle) + K(\langle a,c\rangle) + K(\langle b,c\rangle) - - I(a:b|c) - I(\langle a,b\rangle : c) \quad (5)$$

and inequalities $I(a:b|c) \ge 0$ and $I(\langle a,b\rangle : c) \ge 0$. To check the equality (5) we express all the quantities included in terms of unconditional complexities. For example, we replace $I(a:b|c)$ by

$$K(a|c) + K(b|c) - K(\langle a,b\rangle|c) = = K(\langle a,c\rangle) - K(c) + K(\langle b,c\rangle) - K(c) - - K(\langle a,b,c\rangle) + K(c) = = K(\langle a,c\rangle) + K(\langle b,c\rangle) - K(\langle a,b,c\rangle) - K(c),$$

and so on.

Let us consider another example. Assume that a and b are two binary strings. Let us prove that the mutual information $I(a:b)$ is an upper bound for complexity $K(x)$ of any string x which has negligible conditional complexities $K(x|a)$ and $K(x|b)$. Indeed, the following inequality holds for any three strings a,b,x:

$$K(x) \le K(x|a) + K(x|b) + I(a:b). \quad (6)$$

This inequality is a consequence of the equality

$$K(x) = I(a:b) + K(x|a) + K(x|b) - - K(x|\langle a,b\rangle) - I(a:b|x)$$

and inequalities $K(x|\langle a,b\rangle) \ge 0$ and $I(a:b|x) \ge 0$.

The inequalities of type (2) can be written in different equivalent forms:

$$\begin{aligned} I(P:Q|R) &\ge 0 \\ K(P|R) + K(Q|R) &\ge K(\langle P,Q\rangle|R) \\ K(P|R) &\ge K(P|\langle Q,R\rangle) \\ K(\langle P,R\rangle) + K(\langle Q,R\rangle) &\ge K(\langle P,Q,R\rangle) + K(R) \end{aligned}$$

Here P, Q and R are strings or tuples of strings; $\langle P,R\rangle$ denotes the union of tuples P and R (it does not matter whether we list strings that are in $P \cap R$ twice or not, the complexity does not change), etc.

The latter form does not involve conditional complexities. In general, we may always replace conditional complexities and mutual informations by linear combinations of unconditional complexities, using equalities (3) and (4). Therefore, in the sequel we will consider inequalities containing only unconditional complexities. The same applies to inequalities for Shannon entropies.

We call the inequalities

$$K(\langle P,R\rangle) + K(\langle Q,R\rangle) \ge K(\langle P,Q,R\rangle) + K(R) \quad (7)$$

(for any tuples P,Q,R) *basic* inequalities.

Let us mention two special cases of inequalities (7). If $P = Q$, we get the inequality

$$K(\langle P,R\rangle) + K(\langle P,R\rangle) \ge K(\langle P,R\rangle) + K(R),$$

or

$$K(\langle P,R\rangle) \ge K(R),$$

or

$$K(P|R) \geq 0.$$

Therefore, the inequality $K(x|\langle a, b\rangle) \geq 0$ in our second example is also the corollary of basic inequalities (7).

If R is empty, we get the inequality

$$K(P) + K(Q) \geq K(\langle P, Q\rangle)$$

or

$$K(P) \geq K(P|Q)$$

All inequalities mentioned in this section have counterparts that involve Shannon entropy instead of Kolmogorov complexity. The questions we are interested in are: 1) whether the same linear inequalities are true for Shannon entropies and Kolmogorov complexities and 2) whether all linear inequalities valid for Shannon entropies (Kolmogorov complexities) are consequences of basic inequalities. In next section, we obtain positive answer to the first question and positive answer to the second question in the case when at most three random variables (binary strings) are involved.

4 Linear inequalities

Consider n variables $a_1, \ldots, a_n$ whose values are binary strings (if we consider Kolmogorov complexities) or random variables (for Shannon entropies). There are $2^n - 1$ nonempty subsets of the set of variables. Therefore, there are $2^n - 1$ tuples whose complexity (or entropy) may appear in the inequality. We consider only linear inequalities. Each inequality has $2^n - 1$ coefficients λ_P indexed by non-empty subsets P of the set $\{1, 2, \ldots, n\}$; for example, for $n = 3$ the general form is

$$\lambda_1 K(a_1) + \lambda_2 K(a_2) + \lambda_3 K(a_3) + \\ + \lambda_{1,2} K(\langle a_1, a_2\rangle) + \lambda_{1,3} K(\langle a_1, a_3\rangle) + \lambda_{2,3} K(\langle a_2, a_3\rangle) + \\ + \lambda_{1,2,3} K(\langle a_1, a_2, a_3\rangle) \geq 0$$

Here a_1, a_2, a_3 are binary strings; for Shannon entropies they should be replaced by random variables, and K should be replaced by H. For arbitrary n the general form of a linear inequality under consideration is:

$$\sum_w \lambda_w K(a^w) \geq 0, \qquad (8)$$

where the sum is over all nonempty subsets w of $\{1, 2, \ldots, n\}$, and a^w stands for the tuple consisting of all a_i for $i \in w$.

Now we consider the set of inequalities that are valid (up to a $O(\log m)$ term, as usual) for all binary strings. This set is a convex cone in $\mathbb{R}^{2^n - 1}$. We want to compare this cone with the similar cone for Shannon entropies.

Theorem 1 *Any linear inequality that is true for Kolmogorov complexities is also true for Shannon entropies and vice versa.*

Proof. Let an inequality of the form (8) be true for Kolmogorov complexities (up to $O(\log m)$ term). Let $\alpha_1, \ldots, \alpha_n$ be random variables. We have to prove that

$$\sum_w \lambda_w H(\alpha^w) \geq 0,$$

where the sum is over all nonempty subsets w of $\{1, 2, \ldots, n\}$, and α^w stands for the tuple consisting of all α_i for $i \in w$.

Consider a sequence of independent tuples of random variables $\alpha^1 = \langle \alpha_1^1, \ldots, \alpha_n^1\rangle, \ldots, \alpha^N = \langle \alpha_1^N, \ldots, \alpha_n^N\rangle, \ldots$ All $\alpha^1, \alpha^2, \ldots$ are independent and have the same distribution as $\alpha = \langle \alpha_1, \ldots, \alpha_n\rangle$. For a given N, consider the random variables $\alpha_1^{(N)} = \alpha_1^1 \alpha_1^2 \ldots \alpha_1^N, \ldots, \alpha_n^{(N)} = \alpha_n^1 \alpha_n^2 \ldots \alpha_n^N$. Values of α_i are elements of a fixed finite set; using a suitable encoding (where all codes have the same length) we may assume that the values of all α_i^j are binary strings of equal length. In this case the values of $\alpha_1^{(N)}, \ldots, \alpha_n^{(N)}$ are binary strings, and any inequality for Kolmogorov complexities may be applied to any values of $\alpha_1^{(N)}, \ldots, \alpha_n^{(N)}$. Therefore, for some c for all N

$$\sum_w \lambda_w K((\alpha^{(N)})^w) \geq -c \log(N) - c,$$

with probability 1. Dividing this by N we get

$$\frac{\sum_w \lambda_w K((\alpha^{(N)})^w)}{N} \geq \frac{-c \log N - c}{N}$$

with probability 1.

It remains to use the following connection between Shannon entropy and Kolmogorov complexity.

Let τ be a random variable whose values are finite binary strings of a fixed length. Consider the sequence $\tau_1, \tau_2, \ldots$ of independent random variables, where each τ_i has the same distribution as τ. Then

$$\lim_{N\to\infty} \frac{K(\tau_1 \ldots \tau_N)}{N} = H(\tau)$$

with probability 1 (see [9], equation (5.18)). let us fix w and apply this to $\tau = \alpha^w$. It is easy to see that $K((\alpha^{(N)})^w)$ is equal (up to $O(1)$ term) to $K(\tau_1 \cdots \tau_N)$. Therefore,

$$\lim_{N\to\infty} \frac{K((\alpha^{(N)})^w)}{N} = \lim_{N\to\infty} \frac{K(\tau_1 \cdots \tau_N)}{N} = H(\alpha^w).$$

with probability 1. Hence the inequality $\sum_w \lambda_w H(\alpha^w) \ge 0$ is true.

Now, we have to prove the converse: if the inequality

$$\sum_w \lambda_w H(\alpha^w) \ge 0,$$

is true for any random variables $\alpha_1, \ldots, \alpha_n$, then the inequality

$$\sum_w \lambda_w K(a^w) \ge -O(\log m)$$

is true for all binary strings $a_1, a_2, \ldots, a_n$, where $m = |a_1| + |a_2| + \cdots + |a_n|$ (the constant hidden in $O(\log m)$ may depend on n).

Let $a_1, a_2, \ldots, a_n$ be binary strings. Given $u \subset \{1, 2, \ldots, n\}$ denote by $\bar{u}$ the set $\{1, 2, \ldots, n\} \setminus u$. Consider the set M consisting of all tuples $\langle b_1, b_2, \ldots, b_n \rangle$ such that

$$K(b^u) \le K(a^u),$$
$$K(b^{\bar{u}}|b^u) \le K(a^{\bar{u}}|a^u)$$

for all nonempty $u \subset \{1, 2, \ldots, n\}$. (If $u = \{1, 2, \ldots, n\}$ the second inequality should be skipped.) There exists a program P that given numbers $K(a^{\bar{u}}|a^u)$ and $K(a^u)$ for all nonempty $u \subset \{1, 2, \ldots, n\}$ eventually prints all the elements in M in some order. (However at any moment of the run of P we can not be sure that *all* the elements of M have been printed.) The number of elements in M is at least $2^{K(\langle a_1, \ldots, a_n\rangle) - O(\log m)}$. Indeed, the tuple $\langle a_1, \ldots, a_n \rangle$ can be specified by identifying the numbers $K(a^{\bar{u}}|a^u)$ and $K(a^u)$ for all u (total $O(\log m)$ bits) and identifying the ordinal number of $\langle a_1, \ldots, a_n \rangle$ in the order in which the program P prints all the elements in P (at most $\log |M|$ bits). Thus, $K(\langle a_1, \ldots, a_n \rangle) \le \log |M| + O(\log m)$ and the inequality $|M| \ge 2^{K(\langle a_1, \ldots, a_n\rangle) - O(\log m)}$ follows.

Let $\beta = \langle \beta_1, \beta_2, \ldots, \beta_n \rangle$ denote the random variable uniformly distributed in M. We have

$$\sum_u \lambda_u H(\beta^u) \ge 0.$$

Let us derive from this the desired inequality for complexities of $a_1, a_2, \ldots, a_n$ its pairs, triples, etc. To this end let us prove that $H(\beta^u)$ is close to $K(a^u)$ for any nonempty $u \subset \{1, 2, \ldots, n\}$.

Let us fix $u \subset \{1, 2, \ldots, n\}$. We will show that β^u is close to the random variable uniformly distributed in the set having $2^{K(a^u)}$ elements. Indeed, the cardinality of the set

$$\{b^u \mid K(b^u) \le K(a^u)\}$$

is at most $2^{K(a^u)+O(1)}$. Therefore β^u has no more than $2^{K(a^u)+O(1)}$ values. Hence, $H(\beta^u) \le K(a^u) + O(1)$.

To prove the converse inequality, let us note that if $\Pr[\xi = x] \le p$ for all possible values x of a random variable ξ, then $H(\xi) \ge -\log p$. So it suffices to show that $\Pr[\beta^u = b^u] \ge 2^{-K(a^u) - O(\log m)}$ for any $b_1, \ldots, b_n$ in M. We have

$$\Pr[\beta^u = b^u] = \frac{|c \in M \mid c^u = b^u|}{|M|}.$$

If $c \in M$ and $c^u = b^u$, then $K(c^{\bar{u}}|b^u) = K(c^{\bar{u}}|c^u) =\le K(a^{\bar{u}}|a^u)$. Therefore, $|c \in M \mid c^u = b^u| \le 2^{K(a^{\bar{u}}|a^u)+O(1)}$. Hence,

$$\begin{aligned}
\Pr[\beta^u = b^u] &\le \frac{2^{K(a^{\bar{u}}|a^u)+O(1)}}{2^{K(\langle a_1, \ldots, a_n\rangle) - O(\log m)}} \\
&= 2^{K(a^{\bar{u}}|a^u) - K(\langle a_1, \ldots, a_n\rangle) + O(\log m)} \\
&\le 2^{-K(a^u) + O(\log m)}.
\end{aligned}$$

(End of proof.)

Assume that a linear space L over a finite field or over $\mathbb{R}$ is given. Let $\alpha_1, \ldots, \alpha_n$ be finite subsets of L. For any subset $\mathcal{A} \subset \{\alpha_1, \ldots, \alpha_n\}$ consider the rank of the union of all $\alpha \in \mathcal{A}$. Now consider all linear inequalities that are valid for ranks of these subsets for all $\alpha_1, \ldots, \alpha_n \subset L$. For example, inequality of type (7) for ranks says that

$$\mathrm{rk}(\alpha_1 \cup \alpha_3) + \mathrm{rk}(\alpha_2 \cup \alpha_3) \geq \mathrm{rk}(\alpha_1 \cup \alpha_2 \cup \alpha_3) + \mathrm{rk}(\alpha_3)$$

This inequality can be rewritten in terms of dimensions of subspaces: Replacing each α_i by a linear subspace A_i generated by α_i, we get

$$\dim(A_1 + A_3) + \dim(A_2 + A_3) \geq \\ \geq \dim(A_1 + A_2 + A_3) + \dim(A_3)$$

It is easy to verify that this inequality is true for any linear subspaces of any linear space. So, all basic inequalities are true when $K(\cdot)$ is replaced by $\mathrm{rk}(\cdot)$ and strings are replaced by vectors. Moreover, the following is true.

Theorem 2 *Any inequality valid for Shannon entropies is valid for ranks in any linear space over any finite field or over* $\mathbb{R}$.

Proof of Theorem 2. Assume that $A_1, \ldots, A_n$ are subspaces of a finite-dimensional linear space L over a field F. It suffices to construct random variables $\alpha_1, \ldots, \alpha_n$ such that $H(\alpha_i)$ is proportional to $\dim A_i$, $H(\langle \alpha_i, \alpha_j \rangle)$ is proportional to $\dim(A_i + A_j)$, …, $H(\langle \alpha_1, \alpha_2, \ldots, \alpha_n \rangle)$ is proportional to $\dim(A_1 + A_2 + \cdots + A_n)$.

If F is finite, the construction is straightforward. Consider a random linear functional $\alpha : L \to F$. For any subspace $A \subset L$ consider the restriction $\alpha|_A$. This is a random variable with $|F|^{\dim A}$ values (here $|F|$ is the number of elements in F); all values have equal probabilities, so $H(\alpha|_A) = \dim A \cdot \log |F|$. If A_i and A_j are different subspaces, the pair $\langle \alpha|_{A_i}, \alpha|_{A_j} \rangle$ is equivalent to (and has the same distribution as) $\alpha|_{A_i + A_j}$. Therefore, the entropy of the pair $\langle \alpha|_{A_i}, \alpha|_{A_j} \rangle$ is equal to $\dim(A_i + A_j) \cdot \log |F|$; the same is true for triples, etc.

Now consider the case $F = \mathbb{R}$. We may assume that L is a Euclidean space. Let α be a random variable uniformly distributed in the unit disk in L. For any subspace A consider a random variable α_A that is the orthogonal projection of α onto A. This random variable has infinite domain, so we need to digitize it. For any $\varepsilon > 0$ and for any subspace $A \subset L$ we divide A into equal cubes of dimension $\dim A$ and size $\varepsilon \times \cdots \times \varepsilon$. By $\alpha_{A,\varepsilon}$ we denote the variable whose value is the cube that contains α_A. Let us prove that

$$H(\alpha_{A,\varepsilon}) = \log(1/\varepsilon) \cdot \dim A + O(1)$$

(when $\varepsilon \to 0$).

If ε is small enough the number $k_{A,\varepsilon}$ of the cubes which are possible values of α_A satisfies the inequality

$$k_{A,\varepsilon} \leq 2(1/\varepsilon)^{\dim A} V_{\dim A},$$

where $V_{\dim A}$ stands for the volume of $\dim A$-dimensional unit disk. Therefore,

$$H(\alpha_{A,\varepsilon}) \leq \log(1/\varepsilon) \cdot \dim A + 1 + \log V_{\dim A}.$$

On the other hand, for any fixed cube the probability of α_A getting into it is at most

$$\frac{\varepsilon^{\dim A} V_{(\dim L - \dim A)}}{V_{\dim L}}.$$

Hence

$$H(\alpha_{A,\varepsilon}) \geq \log(1/\varepsilon) \cdot \dim A + \log V_{(\dim L - \dim A_i)} - \log V_{\dim L}.$$

The projection $\alpha_{A_1+A_2}$ is equivalent to $\langle \alpha_{A_1}, \alpha_{A_2} \rangle$. This is not true for ε-versions; the random variables $\alpha_{A_1+A_2,\varepsilon}$ and $\langle \alpha_{A_1,\varepsilon}, \alpha_{A_2,\varepsilon} \rangle$ do not determine each other completely. However, for any fixed value of one of these variables there exist only a finite number of possible values of the other one, therefore, the conditional entropies are limited and the entropies differ by $O(1)$.

Now we let $\varepsilon \to 0$ and conclude that any inequality that is valid for Shannon entropies is valid for ranks. (End of proof.)

Therefore, we have a sequence of inclusions: (basic inequalities (7) and their non-negative linear combinations) $\subset$ (inequalities valid for Kolmogorov complexities) = (inequalities valid for Shannon entropies) $\subset$ (inequalities valid for ranks).

For $n = 1, 2, 3$ all these sets are equal, as the following theorem shows:

Theorem 3 *For $n = 1, 2, 3$ any inequality valid for ranks is a consequence (linear combination with non-negative coefficients) of basic inequalities (7).*

Proof. The cases $n = 1, 2$ are trivial. Let us consider the case $n = 3$.

Consider the following 9 basic inequalities:

$$\begin{aligned}
\mathrm{rk}(A+B) &\le \mathrm{rk}(A+B+C)\\
\mathrm{rk}(A+C) &\le \mathrm{rk}(A+B+C)\\
\mathrm{rk}(B+C) &\le \mathrm{rk}(A+B+C)\\
\mathrm{rk}(A+B) &\le \mathrm{rk}\,A+\mathrm{rk}\,B\\
\mathrm{rk}(A+C) &\le \mathrm{rk}\,A+\mathrm{rk}\,C \qquad (9)\\
\mathrm{rk}(B+C) &\le \mathrm{rk}\,B+\mathrm{rk}\,C\\
\mathrm{rk}\,A+\mathrm{rk}(A+B+C) &\le \mathrm{rk}(A+B)+\mathrm{rk}(A+C)\\
\mathrm{rk}\,B+\mathrm{rk}(A+B+C) &\le \mathrm{rk}(A+B)+\mathrm{rk}(B+C)\\
\mathrm{rk}\,C+\mathrm{rk}(A+B+C) &\le \mathrm{rk}(A+C)+\mathrm{rk}(B+C).
\end{aligned}$$

We claim that any linear inequality for $\dim A$, $\dim B$, $\dim C$, $\dim(A+B)$, $\dim(A+C)$, $\dim(B+C)$, $\dim(A+B+C)$ is a non-negative linear combination of these nine ones (for instance, so are all other basic inequalities).

The inequalities (9) determine a convex cone $\mathfrak{C}$ in the space $\mathbb{R}^7$ where variables are

$$\mathrm{rk}\,A,\ \mathrm{rk}\,B,\ \mathrm{rk}\,C,$$
$$\mathrm{rk}(A+B),\ \mathrm{rk}(B+C),\ \mathrm{rk}(A+C),\ \mathrm{rk}(A+B+C)$$

Any three subspaces A, B, C determine a point inside $\mathfrak{C}$, let us denote the set of all points in $\mathfrak{C}$ obtained in this way by $\mathfrak{C}'$. To prove Theorem 3 it is enough to show that any point in $\mathfrak{C}$ can be represented as a non-negative linear combination of points from $\mathfrak{C}'$. To this end consider 8 points in $\mathfrak{C}'$ shown on Fig. 1 Let us show that any point in $\mathfrak{C}$ can be represented as a non-negative linear combination of those 8 points. To prove this it is convenient to consider another coordinate system in $\mathbb{R}^7$. We denote new coordinates by

$$[a],\ [b],\ [c],\ [ab],\ [ac],\ [bc],\ [abc]$$

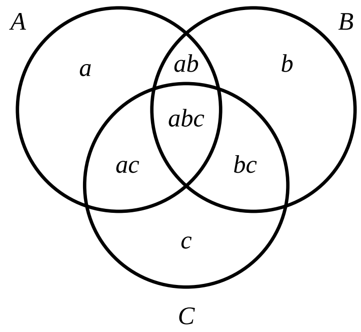

Figure 2: Old and new variables

The relations between new and old variables are:

$$\begin{aligned}
\mathrm{rk}\,A &= [a]+[ab]+[ac]+[abc],\\
\mathrm{rk}(A+B) &= [a]+[b]+[ab]+[ac]+[bc]+[abc],\\
\mathrm{rk}(A+B+C) &= [a]+[b]+[c]+[ab]+[bc]+[ac]+[abc].
\end{aligned}$$

and similar formulae obtained by permutations of letters. (See Fig. 2.) Or, equivalently,

$$\begin{aligned}
[a] &= \mathrm{rk}(A+B+C)-\mathrm{rk}(B+C),\\
[ab] &= \mathrm{rk}(A+C)+\mathrm{rk}(B+C)\\
&\quad -\mathrm{rk}(A+B+C)-\mathrm{rk}\,C,\\
[abc] &= \mathrm{rk}(A+B+C)\\
&\quad -\mathrm{rk}(A+B)-\mathrm{rk}(A+C)-\mathrm{rk}(B+C)\\
&\quad +\mathrm{rk}\,A+\mathrm{rk}\,B+\mathrm{rk}\,C,\\
[b] &= \mathrm{rk}(A+B+C)-\mathrm{rk}(A+C),\\
&\ldots
\end{aligned}$$

The inequalities (9) rewritten in new variables are as follows:

$$\begin{aligned}
&[a]\ge 0,\ [b]\ge 0,\ [c]\ge 0,\\
&[ab]+[abc]\ge 0,\ [ac]+[abc]\ge 0,\ [bc]+[abc]\ge 0,\\
&[ab]\ge 0,\ [ac]\ge 0,\ [bc]\ge 0. \qquad (10)
\end{aligned}$$

(Please note that $[abc]$ may be negative.) In new variables, the 8 specified points in $\mathfrak{C}'$ are written as shown on Fig. 3. Thus we have to show that any vector satisfying the inequalities (10) is a non-negative linear combination of 8 vectors represented on Fig. 3 (we denote them by v_1–v_8).

Figure 1: The 8 points in $\mathfrak{C}'$. By e_1, e_2, e_3 we denote three pairwise independent vectors in 2-dimensional space. $\{u, \ldots\}$ stands for the linear subspace generated by $u, \ldots$. By 0 we denote the 0-dimensional subspace.

A	B	C	$\mathrm{rk}\, A$	$\mathrm{rk}\, B$	$\mathrm{rk}\, C$	$\mathrm{rk}(A+B)$	$\mathrm{rk}(A+C)$	$\mathrm{rk}(B+C)$	$\mathrm{rk}(A+B+C)$
$\{e_1\}$	0	0	1	0	0	1	1	0	1
0	$\{e_1\}$	0	0	1	0	1	0	1	1
0	0	$\{e_1\}$	0	0	1	0	1	1	1
$\{e_1\}$	$\{e_1\}$	0	1	1	0	1	1	1	1
$\{e_1\}$	0	$\{e_1\}$	1	0	1	1	1	1	1
0	$\{e_1\}$	$\{e_1\}$	0	1	1	1	1	1	1
$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	1	1	1	1	1	1	1
$\{e_1\}$	$\{e_2\}$	$\{e_1+e_2\}$	1	1	1	2	2	2	2

Figure 3: The 8 points in $\mathfrak{C}'$ rewritten in new coordinates.

	$[a]$	$[b]$	$[c]$	$[ab]$	$[ac]$	$[bc]$	$[abc]$
v_1	1	0	0	0	0	0	0
v_2	0	1	0	0	0	0	0
v_3	0	0	1	0	0	0	0
v_4	0	0	0	1	0	0	0
v_5	0	0	0	0	1	0	0
v_6	0	0	0	0	0	1	0
v_7	0	0	0	0	0	0	1
v_8	0	0	0	1	1	1	-1

Let $v = ([a], [b], [c], [ab], \ldots, [abc])$ be arbitrary vector in $\mathfrak{C}$. If $[abc]$ is non-negative, we can represent v as non-negative linear combination of v_1–v_8 as follows:

$$v = [a] \cdot v_1 + [b] \cdot v_2 + [c] \cdot v_3 + [ab] \cdot v_4 + \cdots + [abc] \cdot v_7.$$

Otherwise (when $[abc]$ is negative) we can represent v as non-negative linear combination of v_1–v_7 as follows:

$$\begin{aligned} v &= [a]v_1 + [b]v_2 + [c]v_3 + ([ab] + [abc])v_4 \\ &\quad +([ac] + [abc])v_5 + ([bc] + [abc])v_6 - [abc] \cdot v_8. \end{aligned}$$

Theorem 3 is proven.

5 Ingleton's inequality

As we have seen in the preceding section, for $n = 3$ the same inequalities are true for Shannon entropies, Kolmogorov complexities and ranks, namely, the non-negative linear combinations of basic inequalities. However, for $n = 4$ the situation becomes more complicated: there is an inequality that is true for ranks but not for Shannon entropies. This inequality was found by Ingleton [3].

Proposition 1 *The following inequality is true for ranks:*

$$I(A:B) \le I(A:B|C) + I(A:B|D) + I(C:D) \quad (11)$$

In terms of dimensions of subspaces Ingleton's inequality says that

$$\begin{aligned} &\dim A + \dim B + \\ &+ \dim(C+D) + \dim(A+B+C) + \dim(A+B+D) \le \\ &\le \dim(A+B) + \dim(A+C) + \dim(B+C) + \\ &+ \dim(A+D) + \dim(B+D) \quad (12) \end{aligned}$$

To prove Ingleton's inequality one may interpret $I(A:B)$ as the dimension of intersection $A \cap B$, and $I(A:B|C)$ as the dimension of the intersection of A/C and B/C (i.e., A and B factored over C). See also section 6 where Ingleton's inequality is proved as a a consequence of Theorem 9.

The following example shows that Ingleton's inequality is not always true for Shannon entropies.

Theorem 4 *There exist four random variables α, β,*

γ *and* δ *such that*

$$\begin{aligned} I(\alpha:\beta) &> 0 \\ I(\alpha:\beta|\gamma) &= 0 \\ I(\alpha:\beta|\delta) &= 0 \\ I(\gamma:\delta) &= 0 \end{aligned}$$

In other terms, γ and δ are independent, and α and β are independent *for any fixed values of* γ *and* δ; however, α and β are dependent.

Proof of Theorem 4. Let the range of all for variables $\alpha, \beta, \gamma, \delta$ be $\{0, 1\}$. Let γ and δ be independent and uniformly distributed.

Any possible distribution of α, β is determined by four non-negative reals whose sum is 1 (i.e., by the probabilities of all four combinations), so the distribution can be considered as a point in a three-dimensional simplex S in $\mathbb{R}^4$. For any of the four possible values of γ and δ we have a point in S (whose coordinates are conditional probabilities). We denote these points by P_{00}, P_{01}, P_{10} and P_{11}. What are the conditions we need to satisfy? Let $\mathcal{I}$ be the subset of S that corresponds to independent random variables; $\mathcal{I}$ is a quadratic curve (the independence condition means that the determinant of the probabilities matrix is equal to zero). The conditions $I(\alpha : \beta|\gamma) = 0$ and $I(\alpha : \beta|\delta) = 0$ mean that midpoints of segments $P_{00}P_{01}$, $P_{10}P_{11}$, $P_{00}P_{10}$, $P_{01}P_{11}$ belong to $\mathcal{I}$. The inequality $I(\alpha : \beta) > 0$ means that the point $(P_{00} + P_{01} + P_{10} + P_{11})/4$ does not belong to $\mathcal{I}$. In other terms, we are looking for a parallelogram whose vertices lie on a quadratic curve but whose center does not, so almost any example will work. Fig. 4 shows one of them:

	0	1
0	0	0
1	0	1

$\gamma = 0, \delta = 0$

	0	1
0	1/8	3/8
1	3/8	1/8

$\gamma = 1, \delta = 0$

	0	1
0	1/8	3/8
1	3/8	1/8

$\gamma = 0, \delta = 1$

	0	1
0	1	0
1	0	0

$\gamma = 1, \delta = 1$

Figure 4: Conditional probability distributions for $\langle \alpha, \beta \rangle$

It is easy to check that all four conditional distributions (for conditions $\gamma = 0$, $\gamma = 1$, $\delta = 0$, $\delta = 1$) satisfy the independence requirement. However, the unconditional distribution for $\langle \alpha, \beta \rangle$ is

	0	1
0	5/16	3/16
1	3/16	5/16

(13)

so α and β are dependent.

A simpler example, though not so symmetric, can be obtained as follows. Let γ and δ be independent random variables with range $\{0, 1\}$ and uniform distribution, $\alpha = \gamma(1 - \delta)$ and $\beta = \delta(1 - \gamma)$. For any fixed value of γ or δ one of the variables α and β is equal to 0, therefore they are independent. However, α and β are not (unconditionally) independent, since each of them can be equal to 1, but they cannot be equal to 1 simultaneously. (End of proof.)

We see that for $n = 4$ not all the inequalities valid for ranks are valid for entropies, so the rank and entropy cases should be considered separately. For ranks we have the complete answer:

Theorem 5 *For* $n = 4$*, all the inequalities that are valid for ranks, are consequences* (*positive linear combinations*) *of basic inequalities and Ingleton-type inequalities* (*i.e. inequalities obtained from Ingleton's inequality by permutations of variables*).

For entropies we do not know the answer. The only thing we know is the following conditional result. Let $x =_\varepsilon y$ means that $|x - y| < \varepsilon$.

Theorem 6 *If for any* $\varepsilon > 0$ *there exist random* *variables α, β, γ and δ such that*

$$\begin{aligned}
&H(\alpha) =_\varepsilon H(\beta) =_\varepsilon H(\gamma) =_\varepsilon H(\delta) =_\varepsilon 2,\\
&H(\langle\alpha,\beta\rangle) =_\varepsilon H(\langle\alpha,\gamma\rangle) =_\varepsilon H(\langle\alpha,\delta\rangle) =_\varepsilon\\
&\quad =_\varepsilon H(\langle\beta,\gamma\rangle) =_\varepsilon H(\langle\beta,\delta\rangle) =_\varepsilon 3,\\
&H(\langle\gamma,\delta\rangle) =_\varepsilon 4,\\
&H(\langle\beta,\gamma,\delta\rangle) =_\varepsilon H(\langle\alpha,\gamma,\delta\rangle) =_\varepsilon\\
&\quad =_\varepsilon H(\langle\alpha,\beta,\delta\rangle) =_\varepsilon H(\langle\alpha,\beta,\gamma\rangle) =_\varepsilon 4,\\
&H(\langle\alpha,\beta,\gamma,\delta\rangle) =_\varepsilon 4,
\end{aligned}$$

then all the linear inequalities that are valid for Shannon entropies are consequences (positive linear combinations) of basic inequalities.

The proofs of Theorems 5 and 6 require a fairly long computation (it can be performed by hand or using an appropriate software). As before, consider the cone $\mathfrak{C} \subset \mathbb{R}^{15}$ that consists of all the points satisfying basic inequalities (for Theorem 6). Its dual cone $\mathfrak{C}^*$ contains all nonnegative combinations of basic inequalities. Compute all the extreme vectors of the cone $\mathfrak{C}$. If for any extreme vector we can find a quadruple of random variables whose entropies' vector is proportional to the extreme vector, we are done. It turns out that for all the extreme vectors but one it can be done easily; the only exception is the vector given in the statement of the theorem.

Let us note that there are no α, β, γ and δ satisfying the equalities in the above theorem for $\varepsilon = 0$ (the proof will be presented in the whole text).

For the rank case, we have more inequalities, the cone is smaller, and the problematic extreme vectors disappear. (Of course, in this case we need to construct subspaces instead of random variables.)

We may also ask which inequalities are valid for ranks in arbitrary matroids (see [8]). In this case the extreme vector mentioned in Theorem 6 is represented by a Vamos matroid (see [8]), so we get the following

Theorem 7 *For $n = 4$, all the inequalities that are valid for ranks in arbitrary matroids, are consequences (positive linear combinations) of basic inequalities.*

6 One more inequality for entropies

In this section we present one more inequality for entropies and show how it can be used to prove Ingleton's inequality and Gács–Körner result on common information.

Theorem 8 *For any random variables ξ, α, β, γ and δ*

$$\begin{aligned}
H(\xi) \le 2H(\xi|\alpha) + 2H(\xi|\beta) &+\\
+ I(\alpha:\beta|\gamma) + I(\alpha:\beta|\delta) &+ I(\gamma:\delta)
\end{aligned}$$

Proof of Theorem 8. This inequality is a nonnegative linear combination of basic inequalities. However, we present a proof that reflects the intuitive meaning of the inequality.

As we have seen, Ingleton's inequality

$$I(\alpha:\beta) \le I(\alpha:\beta|\gamma) + I(\alpha:\beta|\delta) + I(\gamma:\delta)$$

is not always true for entropies. However, if a binary string ξ has zero complexities $K(\xi|\alpha)$ and $K(\xi|\beta)$, then

$$K(\xi) \le I(\alpha:\beta|\gamma) + I(\alpha:\beta|\delta) + I(\gamma:\delta)$$

Indeed, as we know from section 2, inequality (6),

$$K(\xi) \le K(\xi|\gamma) + K(\xi|\delta) + I(\gamma:\delta)$$

Now we use the conditional versions of the same inequality (6),

$$\begin{aligned}
K(\xi|\gamma) &\le K(\xi|\langle\alpha,\gamma\rangle) + K(\xi|\langle\beta,\gamma\rangle) + I(\alpha:\beta|\gamma)\\
K(\xi|\delta) &\le K(\xi|\langle\alpha,\delta\rangle) + K(\xi|\langle\beta,\delta\rangle) + I(\alpha:\beta|\delta)
\end{aligned}$$

Recalling that $K(\xi|\langle\alpha,\gamma\rangle) \le K(\xi|\alpha)$, $K(\xi|\langle\alpha,\delta\rangle) \le K(\xi|\alpha)$, etc., and combining last three inequalities, we get the inequality of Theorem 8. (End of proof.)

We present two corollaries of this inequality. The first one is the generalization of Ingleton's inequality. We formulate this corollary for Shannon entropies; the similar result is true for Kolmogorov complexities.

Let us call the random variable ξ a *common information* for random variables α and β if

$$H(\xi|\alpha) = 0$$
$$H(\xi|\beta) = 0$$
$$H(\xi) = I(\alpha : \beta)$$

Theorem 9 *Let α, β, γ and δ be random variables. If there exists a random variable that is a common information for α and β, then Ingleton's inequality holds:*

$$I(\alpha : \beta) \leq I(\alpha : \beta|\gamma) + I(\alpha : \beta|\delta) + I(\gamma : \delta)$$

The proof is easy: just apply the Theorem 8 to the random variable ξ that is the common information of α and β.

Ingleton's inequality for ranks is the consequence of Theorem 9. Indeed, recall the proof of Theorem 2. In that proof for each subspace X we considered a random variable α_X that is the restriction of a random linear functional to X. If X and Y are two subspaces, the random variables α_X and α_Y have common information. This common information is α_Z where $Z = X \cap Y$. Therefore, we may apply the inequality of Theorem 9.

Now we understand the reason why Ingleton's inequality is true for ranks in linear spaces (though it is not true for general matroids, Shannon entropy or Kolmogorov complexity): There is an intersection operation on subspaces that extracts the common information!

The second corollary is an easy proof of one of the Gács–Körner [1] results on common information.

Let a and b be two binary strings. We look for the binary string x that represents the common information in a and b in the following sense (cf. the definition for the case of Shannon entropies above): $K(x|a)$ and $K(x|b)$ are small and $K(x)$ is close to $I(a : b)$. (As we know from section 2, equation (6), $K(x)$ cannot exceed $I(a : b)$ significantly if $K(x|a)$ and $K(x|b)$ are small.)

Now we can read the Kolmogorov complexity version of the inequality of Theorem 8 in the following way: *If for given a and b one can find c and d such that $I(a : b|c)$, $I(a : b|c)$ and $I(c : d)$ are small, then any x with small $K(x|a)$ and $K(x|b)$ has small complexity.*

However, $I(a : b)$ may still be significant, and in this case we get an example of two strings with significant mutual information but with no common information. Such an example can be constructed using Theorem 4.

Consider two coins (random variables) α and β used in the proof of Theorem 4, see (13). Each coin has two equiprobable outcomes; α and β are dependent:

$$\Pr[\beta = \alpha] = 5/8, \quad \Pr[\beta \neq \alpha] = 3/8$$

Theorem 10 *Consider the infinite sequence of independent trials $\langle \alpha_i, \beta_i \rangle$ having this distribution. Let A_N be the initial segment $\alpha_1\alpha_2\ldots\alpha_N$; B_N be the initial segment $\beta_1\beta_2\ldots\beta_N$. Then with probability* 1 *we have*

$$I(A_N : B_N) = cN + o(N)$$

where $c = I(\alpha : \beta) > 0$. At the same time the following is true: for any sequence X_N of binary strings of length $O(N)$ such that $K(X_N|A_N) = o(N)$ and $K(X_N|B_N) = o(N)$, the complexity $K(X_N)$ is small: $K(X_N) = o(N)$.

This result was proved (among others) in [1], but the proof is rather technical and long.

7 Questions

Many questions are still unsolved. Here are some of them:

- Is it true that all inequalities valid for Shannon entropies or Kolmogorov complexities are consequences of basic inequalities?
- Is it true that all inequalities valid for ranks are consequences of basic inequalities and Ingleton-type inequalities?
- What inequalities are true for ranks in arbitrary matroids? (For $n = 4$ the answer is given by Theorem 7.)

- The proof of Gács–Körner's result given above works only if the probabilities are close enough to 1/2; we cannot use it directly if 3/8 and 5/8 are replaced, say, by 1/8 and 7/8. Is it possible to modify it and get a simple proof of Gács–Körner's result for general case?

References

[1] P. Gács and J. Körner. Common information is far less than mutual information. *Problems of Control and Inform. Theory*, 2:149–162, 1973.

[2] D. Hammer and A. Shen. A strange application of Kolmogorov complexity. *Mathematical Systems Theory*, accepted for publication.

[3] A.W. Ingleton. Representation of matroids. In: D.J.A. Welsh, editor. *Combinatorial mathematics and its applications*. Academic Press (London), 1971, pp. 149–167.

[4] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, 1(1):1–7, 1965.

[5] A.N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Trans. Inform. Theory*, IT-14(5):662–664, 1968.

[6] M. Li and P. Vitányi. *An introduction to Kolmogorov complexity and its applications*. Springer-Verlag, 1993.

[7] V.A. Uspensky and A. Shen. Relation between varieties of Kolmogorov complexities. *Mathematical Systems Theory*, 29:271–292 (1996)

[8] D.J.A. Welsh. *Matroid theory*. Academic Press, 1976.

[9] A.K. Zvonkin and L.A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, 25(6):83–124, 1970