Inequalities for Shannon entropy and Kolmogorov complexity

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

Andrei Romashchenko
Moscow State University
an@romash.mccme.ru

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

Nikolai Vereshchagin
Moscow State University
ver@mech.math.msu.su

Abstract

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 entropy and for Kolmogorov complexity.

It turns out that

(1) all linear inequalities that are valid for Kolmogorov complexity, are also valid for Shannon entropy and vice versa;

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

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

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

We present an inequality for Kolmogorov complexity 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].

1 Shannon entropy and Kolmogorov complexity

Since 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. Let us recall the main definitions.

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

$$H(\alpha) = -\sum_i p_i \log p_i$$

(All logarithms in the paper are base 2 logarithms.) Using the convexity of the function $p \mapsto -p \log p$, one can prove that the Shannon entropy of a random variable does not exceed the logarithm of the cardinality of its range (and is equal to it only for uniformly distributed variables).

Let β be another variable with a finite range $b_1, \ldots, b_k$ 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 logarithm function, one can prove that

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

and that $H(\alpha|\beta) = H(\alpha)$ if and only if α and β are independent. This inequality may be rewritten as

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

The mutual information in α and β is defined as

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

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. Indeed, for any possible value c_i of γ we have

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

Multiplying this inequality by $\Pr[\gamma = c_i]$ and summing over i, we get 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 [9]). 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 $O(\log m)$). As always, $O(\log m)$ additive term is omitted; the precise meaning of this equality is: there exist constants p,q such that

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

for all binary strings 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

$$I(a:b|c) \;\geq\; 0$$

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

2 Inequalities

We have already mentioned several inequalities for Shannon entropy and Kolmogorov complexity. Some others are known. For example, consider the inequality

$$2K(\langle a,b,c\rangle)\leq K(\langle a,b\rangle)+K(\langle a,c\rangle)+K(\langle b,c\rangle) \tag{1}$$

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 (to the coordinate planes Oxy, Oxz and Oyz), then

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

(see [2]).

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

$$I(P:Q|R)\geq 0 \tag{2}$$

used together with equalities

$$K(Q|P)=K(\langle P,Q\rangle)-K(P), \tag{3}$$
$$I(P:Q|R)=K(P|R)+K(Q|R)-K(\langle P,Q\rangle|R), \tag{4}$$

that express mutual information and conditional complexity in terms of unconditional complexity. (Here P,Q,R are some tuples (possibly empty) of binary strings.)

Indeed, (1) is a consequence of the equality

$$\begin{aligned}2K(\langle a,b,c\rangle) \;=\; & K(\langle a,b\rangle)+K(\langle a,c\rangle)+K(\langle b,c\rangle)-\\ & -I(a:b|c)-I(\langle a,b\rangle:c)\end{aligned} \tag{5}$$

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

$$\begin{aligned}
K(a|c)+&K(b|c) - K(\langle a,b\rangle|c) \\
&= K(\langle a,c\rangle) - K(c) + K(\langle b,c\rangle) - K(c) - K(\langle a,b,c\rangle) + K(c) \\
&= K(\langle a,c\rangle) + K(\langle b,c\rangle) - K(\langle a,b,c\rangle) - K(c),
\end{aligned}$$

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 complexity $K(x|a)$ and $K(x|b)$. Indeed, the following inequality holds for any three strings a,b,x:

$$K(x) \leq K(x|a) + K(x|b) + I(a:b). \tag{6}$$

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

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

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

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 complexity. In general, we may always replace conditional complexity and mutual information by linear combinations of unconditional complexity using (3) and (4). Therefore, in the sequel we consider inequalities containing only unconditional complexity. The same applies to inequalities for Shannon entropy.

We call the inequalities

$$K(\langle P,R\rangle) + K(\langle Q,R\rangle) \geq K(\langle P,Q,R\rangle) + K(R) \tag{7}$$

(for any tuples P,Q,R) *basic* inequalities. Let us mention two special cases of inequalities (7). If $P=Q$, we get an inequality

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

or

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

(bigger tuple has bigger complexity) or

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

(conditional complexity is non-negative).
Now we see that the inequality $K(x|\langle a,b\rangle)\geq 0$ in our second example is also a corollary of basic inequalities (7). Another special case: if R is empty, we get the inequality

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

or

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

This inequalities imply that all unconditional complexities are non-negative, too. (Strictly speaking, inequalities imply that complexities are non-negative up to a logarithmic term, i.e., $K(P)\geq -O(\log n)$ where n is the length of P.)

All inequalities mentioned in this section have counterparts that involve Shannon entropy instead of Kolmogorov complexity. We would like to know if (1) the same linear inequalities are true for Shannon entropy and Kolmogorov complexity and if (2) all linear inequalities valid for Shannon entropy (or Kolmogorov complexity) are consequences of basic inequalities. In next section, we obtain a positive answer to the first question (for the general case) and a positive answer to the second question in the case when at most three random variables (binary strings) are involved.

3 Linear inequalities

Consider n variables $a_1,\dots,a_n$ whose values are binary strings (if we consider Kolmogorov complexity) or random variables with finite range (for Shannon entropy). 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_W indexed by non-empty subsets W of the set $\{1, 2, \ldots, n\}$; for example, for $n = 3$ the general form is

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

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

$$\sum_W \lambda_W K(a^W) \geq 0, \tag{8}$$

where the sum is over all nonempty subsets $W \subset \{1, 2, \ldots, n\}$, and a^W stands for the tuple formed by 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 (of tuples of random variables with finite range).

Theorem 1. *Any linear inequality that is true for Kolmogorov complexity is also true for Shannon entropy, and vice versa.*

Proof [Kolmogorov $\to$ Shannon] Let an inequality of the form (8) be true for Kolmogorov complexity (up to $O(\log m)$ term).

Let $\alpha = \langle \alpha_1, \ldots, \alpha_n \rangle$ be a tuple of random variables. We have to prove that

$$\sum_W \lambda_W H(\alpha^W) \geq 0,$$

where the sum is taken over all nonempty subsets $W \subset \{1, 2, \ldots, n\}$ and α^W stands for the tuple formed by all α_i for $i \in W$.

Consider a sequence of random variables

$$\begin{aligned} \alpha^1 &= \langle \alpha_1^1, \ldots, \alpha_n^1 \rangle, \\ \alpha^2 &= \langle \alpha_1^2, \ldots, \alpha_n^2 \rangle, \\ &\ldots \\ \alpha^N &= \langle \alpha_1^N, \ldots, \alpha_n^N \rangle \end{aligned}$$

that form the rows of a $N \times n$ random matrix. We assume that $\alpha^1, \alpha^2, \ldots$ are independent and have the same distribution as α.

Now consider the columns of this matrix. We may assume without loss of generality that all values of $\alpha_1, \alpha_2, \ldots$ are binary strings of some fixed length (the same for all variables). Then the columns of this matrix may be considered as strings whose length is N times bigger. We denote them by

$$\alpha_1^{(N)} = \alpha_1^1\alpha_1^2\ldots\alpha_1^N$$
$$\ldots$$
$$\alpha_n^{(N)} = \alpha_n^1\alpha_n^2\ldots\alpha_n^N$$

It turns out that the complexities of the columns $\alpha_1^{(N)}, \alpha_2^{(N)}, \ldots$ are proportional to the entropies $H(\alpha_1), H(\alpha_2), \ldots$. More precisely, with a probability close to 1, these complexities are close to $NH(\alpha_1), NH(\alpha_2), \ldots$ The same is true for the pairs, triples, etc. So we can apply inequality (8) to get its analogue for Shannon entropy.

More formally, for all possible values of the random tuple

$$\alpha^{(N)} = \langle \alpha_1^{(N)}, \ldots, \alpha_n^{(N)} \rangle$$

we have

$$\sum_W \lambda_W K((\alpha^{(N)})^W) \geq -c\log(N) - c,$$

for some c that does not depend on N. Now we divide this inequality by N and get

$$\sum_W \lambda_W \frac{K((\alpha^{(N)})^W)}{N} \geq -O(\log N/N)$$

The right-hand side has limit 0 as $N \to \infty$. It remains to use the following connection between Shannon entropy and Kolmogorov complexity.

Lemma 1. ([9], *equation* (5.18)) *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 *(i.e., for almost all elements of the sample space where all τ_i are defined).*

We fix W and apply this lemma to $\tau = \alpha^W$. It is easy to see that $K((\alpha^{(N)})^W)$ is equal (up to a $O(1)$ term) to $K(\tau_1 \ldots \tau_N)$. 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 $\sum_W \lambda_W H(\alpha^W) \geq 0$ is true.

[Shannon $\to$ Kolmogorov] Now we have to prove the converse: if the inequality

$$\sum_W \lambda_W H(\alpha^W) \geq 0,$$

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

$$\sum_W \lambda_W K(A^W) \geq -O(\log |A|)$$

is true for any tuple of binary strings $A = \langle a_1, a_2, \ldots, a_n \rangle$, where A^W is a tuple formed by all a_i such that $i \in W$ and $|A| = |a_1| + |a_2| + \cdots + |a_n|$ is the total length of A. (Please note that the constant hidden in $O(\log |A|)$ may depend on n.)

To prove this inequality, for a given A we want to construct random variables $\alpha_1, \ldots \alpha_n$ whose entropies are close to the complexities of $a_1, a_2, \ldots, a_n$. We also want the entropies of all pairs, triples, etc. to be close to the complexities of the corresponding pairs, triples, etc. of binary strings.

The following construction achieves this goal. Assume that a tuple $A = \langle a_1, a_2, \ldots, a_n \rangle$ is fixed. Consider the set of all tuples $B = \langle b_1, b_2, \ldots, b_n \rangle$ that satisfy the following conditions: First, the complexity of each b_i does not exceed the complexity of the corresponding a_i. Moreover, the same is true for all pairs, triples, etc. Finally, the same should be true for all conditional complexities. Formally, we consider the set $\mathcal{B}$ formed by all tuples $B = \langle b_1, b_2, \ldots, b_n \rangle$ such that $K(B^V|B^W) \leq K(A^V|A^W)$ for any two subsets $V, W \subset \{1, 2, \ldots, n\}$. (Here the inequality is understood literally, without any hidden constants or log-terms.)

The set $\mathcal{B}$ contains at least one point, A. This set has a simple description as an enumerable set: to generate its elements it is enough to know all conditional complexities $K(A^V|A^W)$, i.e., several integers not exceeding $|A|$. So the complexity of the program that enumerates $\mathcal{B}$ is $O(\log |A|)$. (The constant hidden in the O-notation depends on n and grows exponentially, but we assume n to be fixed.) And any set X having a simple description (as an enumerable set) and having a point x with high complexity should have many elements. Indeed, any point in X may be identified by its number (in the enumeration order) and the enumeration program, so $K(x)$ cannot be high if $|X|$ is small.

More formally, the following lemma gives the lower bound for the cardinality of $\mathcal{B}$ (denoted by $|\mathcal{B}|$):

Lemma 2. $\log|\mathcal{B}| \geq K(\langle a_1,\dots,a_n\rangle) - O(\log|A|)$.

Proof. Consider the program that prints $\langle a_1,\dots,a_n\rangle$ and works as follows. It enumerates M; a tuple $B = \langle b_1,\dots,b_n\rangle$ is included in the enumeration after we have found that its complexity is in the required range (looking for all programs that print B and finding a short one); and, moreover, the conditional complexities are in the required ranges. The program counts the elements of $\mathcal{B}$ that were already generated; when z elements are found, it prints z-th element and terminates. Here z is the number of A (it is a compiled-in constant in the program). The length of this program does not exceed $\log z + O(\log|A|) + O(1)$, since the program uses z, conditional complexities $\big($the total amount of information is $O(\log|A|)\big)$ and a finite amount of other information. It remains to use that $z \leq |\mathcal{B}|$. □

Now let $\alpha = \langle \alpha_1, \alpha_2, \dots, \alpha_n\rangle$ be a random variable uniformly distributed in $\mathcal{B}$. We know that

$$\sum_W \lambda_W H(\alpha^W) \geq 0,$$

so we get the desired inequality for the complexities of $a_1,\dots,a_n$, their pairs, triples, etc., if we show that their complexities are close to the corresponding entropies of $\alpha_1,\dots,\alpha_n$, their pairs, triples etc. Thus, we have to prove that $H(\alpha^W)$ is close to $K(A^W)$ for any nonempty $W \subset \{1,2,\dots,n\}$.

Let us fix a set $W \subset \{1,2,\dots,n\}$. The random variable α^W is close to the random variable that is uniformly distributed in the set having $2^{K(A^W)}$ elements. Indeed, the cardinality of the set

$$\{b^W \mid b \in \mathcal{B}\}$$

that is the projection of $\mathcal{B}$ onto W-coordinates, is at most $2^{K(A^W)+O(1)}$, since all the elements of this projection have a complexity not exceeding $K(A^W)$. Therefore α^W has no more than $2^{K(A^W)+O(1)}$ values and $H(\alpha^W) \leq K(A^W) + O(1)$.

To prove the converse inequality, let us note that if $\Pr[\xi = x] \leq p$ for all possible values x of a random variable ξ, then $H(\xi) \geq -\log p$. So it suffices to show that $\Pr[\alpha^W = B^W] \leq 2^{-K(A^W)+O(\log|A|)}$ for any $B = \langle b_1,\dots,b_n\rangle$ in $\mathcal{B}$. We have

$$\Pr[\alpha^W = B^W] = |\{C \in \mathcal{B} \mid C^W = B^W\}|/|\mathcal{B}|.$$

The lower bound for the denominator is provided by the lemma we proved. The upper bound for the numerator can be obtained as follows. Let $\neg W$ be the complement of W:

$$\neg W = \{1,2,\dots,n\} \setminus W.$$

All the points C counted in the numerator have the same W-projection C^W and differ only by $\neg W$-projections $C^{\neg W}$. By definition, the complexity $K(C^{\neg W}|C^W)$ for $C \in \mathcal{B}$ does not exceed $K(A^{\neg W}|A^W)$. Therefore, the number of those C is limited; the logarithm of this number does not exceed

$$K(A^{\neg W}|A^W) = K(A) - K(A^W) + O(\log|A|)$$

Combining this bound with the lower bound for $\mathcal{B}$, we get the desired inequality for probabilities. ■

Our next result is about the inequalities for ranks of finite subsets of linear spaces.

Assume that a linear space L over a finite field or over $\mathbb{R}$ is given. Let $\alpha_1, \ldots,$ α_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

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

This inequality can be rewritten in terms of dimensions of subspaces: any set X of vectors generates a subspace, and the dimension of this subspace is $\mathrm{rk}(X)$. Replacing each α_i by a linear subspace A_i generated by α_i, we get

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

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 linear inequality valid for Shannon entropy is valid for ranks (dimensions) in any linear space over any finite field or over* $\mathbb{R}$.

Proof 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 two 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 an 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 an 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

$$k_{A,\varepsilon} \le C(1/\varepsilon)^{\dim A}$$

where C is a constant slightly bigger than the volume of unit disk in A. Therefore,

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

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

$$c\varepsilon^{\dim A}$$

where c is a constant equal to the ratio of volumes of unit disks in Euclidean spaces of dimensions $\dim L - \dim A$ and $\dim L$.

Hence

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

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 entropy is valid for ranks. ■

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

$$\begin{aligned}
\dim(B+C) &\le \dim(A+B+C)\\
\dim(A+C) &\le \dim(A+B+C)\\
\dim(A+B) &\le \dim(A+B+C)\\
\dim(C)+\dim(A+B+C) &\le \dim(A+C)+\dim(B+C)\\
\dim(B)+\dim(A+B+C) &\le \dim(A+B)+\dim(B+C)\\
\dim(A)+\dim(A+B+C) &\le \dim(A+B)+\dim(A+C)\\
\dim(A+B) &\le \dim(A)+\dim(B)\\
\dim(A+C) &\le \dim(A)+\dim(C)\\
\dim(B+C) &\le \dim(B)+\dim(C)
\end{aligned} \tag{9}$$

We claim that any valid 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

$$\dim(A),\ \dim(B),\ \dim(C),\ \dim(A+B),\ \dim(B+C),\ \dim(A+C),\ \dim(A+B+C)$$

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 $\mathfrak{C}'$. It is enough to consider 8 points in $\mathfrak{C}'$ shown on Figure 1.

Here e_1,e_2,e_3 are three pairwise independent vectors in 2-dimensional space; $\{u\}$ stands for the linear subspace generated by u. By 0 we denote the 0-dimensional subspace.

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

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

subspaces			(old) coordinates						
A	B	C	$\dim(A)$	$\dim(B)$	$\dim(C)$	$\dim(A+B)$	$\dim(A+C)$	$\dim(B+C)$	$\dim(A+B+C)$
$\{e_1\}$	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
$\{e_1\}$	$\{e_2\}$	$\{e_3\}$	1	1	1	2	2	2	2

Figure 1. 8 points in $\mathfrak{C}'$.

The relations between new and old variables are:

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

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

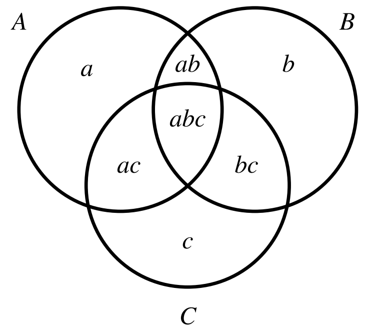

Figure 2. Old and new variables

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

The inequalities (9) rewritten in new variables are

$$
\begin{aligned}
&[a] \geq 0,\ [b] \geq 0,\ [c] \geq 0,\\
&[ab] \geq 0,\ [ac] \geq 0,\ [bc] \geq 0,\\
&[ab] + [abc] \geq 0,\ [ac] + [abc] \geq 0,\ [bc] + [abc] \geq 0.
\end{aligned}
\tag{10}
$$

(Please note that $[abc]$ may be negative.) In new variables, the 8 specified points

vector	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

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

in $\mathfrak{C}'$ are written as shown on Figure 3. Thus we have to show that any vector satisfying the inequalities (10) is a non-negative linear combination of 8 vectors represented on Figure 3 (we denote them by v_1–v_8).

Let $v = ([a],[b],[c],[ab],\dots,[abc])$ be a vector in $\mathfrak{C}$. If $[abc]$ is non-negative, we can represent v as non-negative linear combination of v_1–v_7. Otherwise (when $[abc]$ is negative) we can represent v as non-negative linear combination of v_1–v_8 as follows:

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

Theorem 3 is proven. ■

4 Ingleton's inequality

As we have seen in the preceding section, for $n = 3$ the same inequalities are true for Shannon entropy, Kolmogorov complexity 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 entropy.

A.W. Ingleton [3] established the following necessary condition for a matroid with ground set S and rank function r to be representable over a field F: for any subsets A, B, C, D of S there must hold

$$\begin{aligned} r(A)+r(B)+r(C\cup D)+r(A\cup B\cup C)+r(A\cup B\cup D) \\ \leq\ r(A\cup B)+r(A\cup C)+r(A\cup D)+r(B\cup C)+r(B\cup D). \end{aligned} \tag{11}$$

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

$$\begin{aligned} &\dim A+\dim B+ \\ &\qquad +\dim(C+D)+\dim(A+B+C)+\dim(A+B+D)\leq \\ &\qquad\quad \leq \dim(A+B)+\dim(A+C)+\dim(B+C)+ \\ &\qquad\qquad +\dim(A+D)+\dim(B+D); \end{aligned} \tag{12}$$

It can be rewritten as follows:

$$I(A:B)\leq I(A:B|C)+I(A:B|D)+I(C:D) \tag{13}$$

where $I(A:B)$ stands for $\dim(A)+\dim(B)-\dim(A+B)$, $I(A:B|C)$ stands for $\dim(B+C)+\dim(A+C)-\dim(A+B+C)-\dim(C)$, etc.

To prove inequality (13) 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 factorized over C). See also Section 5 where Ingleton's inequality is proved as a a consequence of Theorem 8.

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

Theorem 4. *There exist four random variables* α, β, γ, 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 *for any fixed value of* δ; however, α and β are dependent.

Proof of Theorem 4. Let the range of all four 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 $\langle\gamma, \delta\rangle$ 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 Q be the subset of S that corresponds to independent random variables; Q 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 Q. The inequality $I(\alpha : \beta) > 0$ means that the point $(P_{00} + P_{01} + P_{10} + P_{11})/4$ does not belong to Q. In other terms, we are looking for a parallelogram (formed by midpoints) whose vertices lie on a quadratic curve but whose center does not, so almost any example will work. Figure 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

(14)

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. ■

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 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).*

Proof The system $\mathcal{F}_4^{(+)}$ of all basic inequalities that involve at most four variables, together with the six Ingleton-type inequalities, determines a convex polyhedral cone $\mathfrak{C}_4^{(+)} \subset \mathbb{R}^{15}$.

It is not hard to show that this cone is generated by (i.e., is the convex hull of) the 35 points shown in Figure 5. Since it requires a great deal of (not very interesting) computation, we refrain from demonstrating this proof here. However, it can be performed by hand (using, for instance, Fourier-Motzkin elimination) or with the help of an appropriate software.

To prove the theorem, it remains to show that for each generator ρ_i there exists a quadruple of subspaces A, B, C, D that represents this point.

subspace	represented generator of $\mathfrak{C}_4^{(+)}$									
	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}
A	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	0	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	0	0
B	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	0	$\{e_1\}$	$\{e_1\}$	0	0	$\{e_1\}$	$\{e_1\}$
C	$\{e_1\}$	$\{e_1\}$	0	$\{e_1\}$	$\{e_1\}$	0	$\{e_1\}$	0	$\{e_1\}$	0
D	$\{e_1\}$	0	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	0	0	$\{e_1\}$	0	$\{e_1\}$

subspace	represented generator of $\mathfrak{C}_4^{(+)}$									
	ρ_{11}	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{15}	ρ_{16}	ρ_{17}	ρ_{18}	ρ_{19}	ρ_{20}
A	0	$\{e_1\}$	0	0	0	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	0	$\{e_1\}$
B	0	0	$\{e_1\}$	0	0	$\{e_2\}$	$\{e_2\}$	0	$\{e_1\}$	$\{e_2\}$
C	$\{e_1\}$	0	0	$\{e_1\}$	0	$\{e_3\}$	0	$\{e_2\}$	$\{e_2\}$	$\{e_3\}$
D	$\{e_1\}$	0	0	0	$\{e_1\}$	0	$\{e_3\}$	$\{e_3\}$	$\{e_3\}$	$\{e_3\}$

ρ_1	$(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)$	ρ_{19}	$(0,1,1,1,1,1,1,2,2,2,2,2,2,2,2)$
ρ_2	$(1,1,1,0,1,1,1,1,1,1,1,1,1,1,1)$	ρ_{20}	$(1,1,1,1,2,2,2,2,2,1,2,2,2,2,2)$
ρ_3	$(1,1,0,1,1,1,1,1,1,1,1,1,1,1,1)$	ρ_{21}	$(1,1,1,1,2,2,2,2,1,2,2,2,2,2,2)$
ρ_4	$(1,0,1,1,1,1,1,1,1,1,1,1,1,1,1)$	ρ_{22}	$(1,1,1,1,2,2,2,1,2,2,2,2,2,2,2)$
ρ_5	$(0,1,1,1,1,1,1,1,1,1,1,1,1,1,1)$	ρ_{23}	$(1,1,1,1,2,2,1,2,2,2,2,2,2,2,2)$
ρ_6	$(1,1,0,0,1,1,1,1,1,0,1,1,1,1,1)$	ρ_{24}	$(1,1,1,1,2,1,2,2,2,2,2,2,2,2,2)$
ρ_7	$(1,0,1,0,1,1,1,1,0,1,1,1,1,1,1)$	ρ_{25}	$(1,1,1,1,1,2,2,2,2,2,2,2,2,2,2)$
ρ_8	$(1,0,0,1,1,1,1,0,1,1,1,1,1,1,1)$	ρ_{26}	$(1,1,1,1,2,2,2,2,2,2,2,2,2,2,2)$
ρ_9	$(0,1,1,0,1,1,0,1,1,1,1,1,1,1,1)$	ρ_{27}	$(1,1,1,1,2,2,2,2,2,2,3,3,3,3,3)$
ρ_{10}	$(0,1,0,1,1,0,1,1,1,1,1,1,1,1,1)$	ρ_{28}	$(2,1,1,1,2,2,2,2,2,2,2,2,2,2,2)$
ρ_{11}	$(0,0,1,1,0,1,1,1,1,1,1,1,1,1,1)$	ρ_{29}	$(1,2,1,1,2,2,2,2,2,2,2,2,2,2,2)$
ρ_{12}	$(1,0,0,0,1,1,1,0,0,0,1,1,1,0,1)$	ρ_{30}	$(1,1,2,1,2,2,2,2,2,2,2,2,2,2,2)$
ρ_{13}	$(0,1,0,0,1,0,0,1,1,0,1,1,0,1,1)$	ρ_{31}	$(1,1,1,2,2,2,2,2,2,2,2,2,2,2,2)$
ρ_{14}	$(0,0,1,0,0,1,0,1,0,1,1,0,1,1,1)$	ρ_{32}	$(2,1,1,1,3,3,3,2,2,2,3,3,3,3,3)$
ρ_{15}	$(0,0,0,1,0,0,1,0,1,1,0,1,1,1,1)$	ρ_{33}	$(1,2,1,1,3,2,2,3,3,2,3,3,3,3,3)$
ρ_{16}	$(1,1,1,0,2,2,1,2,1,1,2,2,2,2,2)$	ρ_{34}	$(1,1,2,1,2,3,2,3,2,3,3,3,3,3,3)$
ρ_{17}	$(1,1,0,1,2,1,2,1,2,1,2,2,2,2,2)$	ρ_{35}	$(1,1,1,2,2,2,3,2,3,3,3,3,3,3,3)$
ρ_{18}	$(1,0,1,1,1,2,2,1,1,2,2,2,2,2,2)$		

Figure 5. The generators of $\mathfrak{C}_4^{(+)}$.

subspace	represented generator of $\mathfrak{C}_4^{(+)}$								
	ρ_{21}	ρ_{22}	ρ_{23}	ρ_{24}	ρ_{25}	ρ_{26}	ρ_{27}	ρ_{28}	ρ_{29}
A	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	$\{e_1\}$	$\{j_1\}$	$\{e_1,e_2\}$	$\{e_1\}$
B	$\{e_2\}$	$\{e_2\}$	$\{e_2\}$	$\{e_2\}$	$\{e_1\}$	$\{e_2\}$	$\{j_2\}$	$\{e_2\}$	$\{e_1,e_2\}$
C	$\{e_3\}$	$\{e_2\}$	$\{e_3\}$	$\{e_1\}$	$\{e_2\}$	$\{e_3\}$	$\{j_3\}$	$\{e_3\}$	$\{e_3\}$
D	$\{e_2\}$	$\{e_3\}$	$\{e_1\}$	$\{e_3\}$	$\{e_3\}$	$\{e_4\}$	$\{j_4\}$	$\{e_4\}$	$\{e_4\}$

subspace	represented generator of $\mathfrak{C}_4^{(+)}$					
	ρ_{30}	ρ_{31}	ρ_{32}	ρ_{33}	ρ_{34}	ρ_{35}
A	$\{e_1\}$	$\{e_1\}$	$\{j_1,j_2\}$	$\{j_1\}$	$\{j_1\}$	$\{j_1\}$
B	$\{e_2\}$	$\{e_2\}$	$\{j_3\}$	$\{j_2,j_3\}$	$\{j_2\}$	$\{j_2\}$
C	$\{e_1,e_3\}$	$\{e_3\}$	$\{j_4\}$	$\{j_4\}$	$\{j_3,j_4\}$	$\{j_3\}$
D	$\{e_4\}$	$\{e_1,e_4\}$	$\{j_5\}$	$\{j_5\}$	$\{j_5\}$	$\{j_4,j_5\}$

It is easy to check that the above indicated quadruples of subspaces meet all necessary requirements. Here e_1, e_2, e_3, e_4 are four pairwise independent vectors in 2-dimensional space; j_1, j_2, j_3, j_4, j_5 are five vectors in 3-dimensional space such that any three of them are independent; $\{u,\ldots\}$ stands for the linear subspace generated by $u,\ldots$. By 0 we denote the 0-dimensional subspace. ■

For Shannon entropy (Kolmogorov complexity) we do not know the complete answer. The only thing we know is the following conditional result.

Theorem 6. *For $n=4$: if for any $\varepsilon>0$ there exist random variables α, β, γ, δ and a real k such that*

$$
\begin{aligned}
&H(\alpha)\approx H(\beta)\approx H(\gamma)\approx H(\delta)\approx 2k,\\
&H(\langle\alpha,\beta\rangle)\approx H(\langle\alpha,\gamma\rangle)\approx H(\langle\alpha,\delta\rangle)\approx H(\langle\beta,\gamma\rangle)\approx H(\langle\beta,\delta\rangle)\approx 3k,\\
&H(\langle\gamma,\delta\rangle)\approx 4k,\\
&H(\langle\beta,\gamma,\delta\rangle)\approx H(\langle\alpha,\gamma,\delta\rangle)\approx H(\langle\alpha,\beta,\delta\rangle)\approx H(\langle\alpha,\beta,\gamma\rangle)\approx 4k,\\
&H(\langle\alpha,\beta,\gamma,\delta\rangle)\approx 4k,
\end{aligned}
$$

where $x\approx y$ means that $|x-y|\le k\varepsilon$, then all the linear inequalities that are valid for Shannon entropy are consequences (positive linear combinations) of the basic inequalities.

Proof The system $\mathcal{F}_4$ of all basic inequalities that involve at most four variables determines a convex polyhedral cone $\mathfrak{C}_4\subset\mathbb{R}^{15}$, and we have $\mathfrak{C}_4^{(+)}\subset\mathfrak{C}_4$. Moreover, it can be shown that the extreme points $\rho_1,\ldots,\rho_{35}$ of $\mathfrak{C}_4^{(+)}$ together with the six points:

$$
\begin{aligned}
\rho_{36} &= (2,2,2,2,4,3,3,3,3,3,4,4,4,4,4)\\
\rho_{37} &= (2,2,2,2,3,4,3,3,3,3,4,4,4,4,4)\\
\rho_{38} &= (2,2,2,2,3,3,4,3,3,3,4,4,4,4,4)\\
\rho_{39} &= (2,2,2,2,3,3,3,4,3,3,4,4,4,4,4)\\
\rho_{40} &= (2,2,2,2,3,3,3,3,4,3,4,4,4,4,4)\\
\rho_{41} &= (2,2,2,2,3,3,3,3,3,4,4,4,4,4,4)
\end{aligned}
$$

generate the cone $\mathfrak{C}_4$. As in the case of $\mathfrak{C}_4^{(+)}$, it requires a rather long computation that can be performed using standard analytic/geometric methods or with the help of an appropriate software.

Now, for every extreme vector ρ_i we would like to find a quadruple of random variables whose entropies' vector is proportional to ρ_i. For the generators ρ_1, $\ldots$, ρ_{35} this can be easily done: Using the method developed in the proof of

Theorem 2, construct for each $i = 1, \dots, 35$ the required quadruple of random variables from the respective quadruple of subspaces A, B, C, D that represents ρ_i.

However, one can prove that for $\rho_{36}, \dots, \rho_{41}$ such a quadruple does not exist. Our assumption says that for every ε there is a quadruple of random variables that gives ε-approximation to the required point.

From here follows the assertion of the theorem. Indeed, assume that there exists a linear inequality that is valid for random variables but is not a positive linear combination of the basic inequalities. Then, at least one of the extreme points ρ_i of $\mathfrak{C}_4$ does not satisfy this inequality. Consequently, for some ε, no point in the ε-neighbourhood of ρ_i is represented by a quadruple of random variables — contradiction. ∎

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 Vámos 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.*

5 One more inequality for Shannon entropy

In this section we present one more inequality for entropy 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*

$$H(\xi) \leq 2H(\xi|\alpha) + 2H(\xi|\beta) + I(\alpha:\beta|\gamma) + I(\alpha:\beta|\delta) + I(\gamma:\delta) \qquad (15)$$

Proof This inequality is a non-negative linear combination of basic inequalities. However, we present a proof that reflects the intuitive meaning of the inequality.

The intuitive meaning of (15) can be explained as follows. As we have seen, Ingleton's inequality

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

is not always true for entropies. However, (15) implies that if a random variable ξ has zero complexities $H(\xi|\alpha)$ and $H(\xi|\beta)$, then

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

The inequality (15) can be proved as follows. As we know from Section 1, inequality (6),

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

Now we use the conditional versions of this inequality,

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

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

We present two corollaries of inequality (15). The first one is the generalization of Ingleton's inequality. We formulate this corollary for Shannon entropy; the similar result is true for Kolmogorov complexity.

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

$$\begin{aligned} H(\xi|\alpha) &= 0 \\ H(\xi|\beta) &= 0 \\ H(\xi) &= I(\alpha:\beta) \end{aligned}$$

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:*

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

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

The inequality of Theorem 8 is valid for dimensions (as a consequence of basic inequalities). That is, for any linear spaces X, A, B, C, D, we have

$$\dim(X) \le 2\dim(X|A) + 2\dim(X|B) + I(A:B|C) + I(A:B|D) + I(C:D).$$

Let $X = A \cap B$. Since $\dim(X) = I(A:B)$, $\dim(X|A) = \dim(X|B) = 0$, we obtain Ingleton's inequality.

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 entropy above): $K(x|a)$ and $K(x|b)$ are small and $K(x)$ is close to $I(a:b)$. (As we know from Section 1, 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 $I(a:b|c)$, $I(a:b|d)$ 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. This can be done as follows.

Consider two coins (random variables) α and β used in the proof of Theorem 4, see (14). 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 with probability 1 : 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: $K(X_N) = o(N)$.

Indeed, the first statement ($I(A_N : B_N) = cN + o(N)$) follows from Lemma 1 above. To validate the second claim, consider four random variables $\alpha, \beta, \gamma, \delta$ used in the proof of theorem 4 and the initial segments $\Gamma_N = \gamma_1 \ldots \gamma_N$ and $\Delta_N = \delta_1 \ldots \delta_N$ of independent trials (each trial involves all four variables). Lemma 1 implies that $I(A_N : B_N|\Gamma_N) = o(N)$ and $I(A_N : B_N|\Delta_N) = o(N)$ as well as $I(\Gamma_N : \Delta_N) = o(N)$. The inequality 15 (for Kolmogorov complexities) guarantees now that

$$\begin{aligned} H(X_N) &\leq 2H(X_N|A_N) + 2H(X_N|B_N) \\ &\quad + I(A_N : B_N|\Gamma_N) + I(A_N : B_N|\Delta_N) + I(\Gamma_N : \Delta_N) = \\ &= 2o(N) + 2o(N) + o(N) = o(N). \end{aligned}$$

Theorem 10 is proved. ∎

This theorem is a very special case of Gács–Körner results [1]; they prove the claim of theorem 10 for any two random variables α and β such that there is no random variable γ such that $H(\gamma) > 0$ while $H(\gamma|\alpha) = H(\gamma|\beta) = 0$. However, their proof seems to be more technical.

Conclusions and open questions

The obtained results are summarized in the following picture: In the general case the class of the basic inequalities (7) and their non-negative linear combinations is a subclass of the class of all inequalities valid for Kolmogorov complexity or Shannon entropy which is a subclass of the class of all inequalities valid for ranks (left side of Figure 6). The latter two classes are separated by Ingleton's inequality, and, hence, the inclusion is strict. Therefore, the area marked by * is nonempty.

For $n = 1, 2, 3$ all these classes coincide (right side of Figure 6).

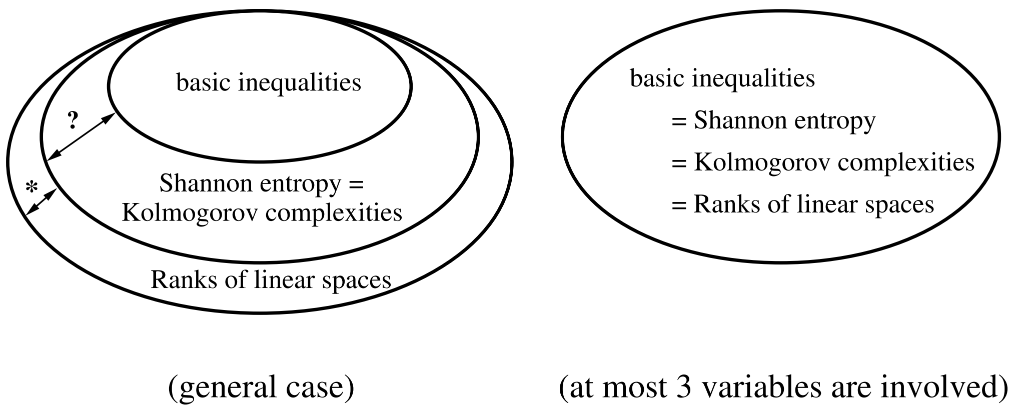

Figure 6. True linear inequalities

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

- Is it true that all inequalities valid for Shannon entropy or Kolmogorov complexity are consequences of basic inequalities? (See the right part of Figure 6 where the respective area is labelled by a question-mark.)

- 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 for a very special α and β; we cannot use it directly even if $3/8$ and $5/8$ are replaced, say, by $1/8$ and $7/8$. (Some extension of our technique allows to cover this case, however.) It is possible to get a simple proof of Gács–Körner's result for general case?

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.

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. *Theory of Computing Systems*, 31(1):1–4, 1998.

[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, second edition, 1997.

[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