

# Combinatorial meaning of Kolmogorov complexity

A. Romashchenko, A. Shen, N. Vereshchagin

## Abstract

The very first Kolmogorov's paper on algorithmic information theory [1] was entitled "Three approaches to the definition of the quantity of information". These three approaches were called *combinatorial*, *probabilistic* and *algorithmic*. Trying to establish formal connections between combinatorial and algorithmic approaches, we prove that every linear inequality including Kolmogorov complexities could be translated to an equivalent combinatorial statement.

Entropy (complexity) proofs of combinatorial inequalities given in [5] and [2] can be considered as a special cases (and a natural starting points) for this translation.

## 1 Introduction and examples

Kolmogorov complexity  $K(x)$  of a binary string  $x$  is defined as the length of shortest program that produces  $x$ . (Programs are binary strings, too.) Complexity depends on the programming system, and we assume that programming system is optimal (complexity in minimal up to  $O(1)$  additive term). Conditional complexity  $K(x|y)$  is defined as the length of shortest program that produces  $x$  given input  $y$ .

This approach was called "algorithmic" in [1]. Combinatorial approach was explained in the same paper as follows:

Consider a variable  $x$  whose range is a finite set  $X$  of cardinality  $N$ . One can say that the "entropy" of variable  $x$  is equal to  $H(x) = \log_2 N$ . When a specific value  $x = a$  is fixed, we "eliminate" this entropy by providing  $I = \log_2 N$  bits of "information". For  $k$  independent variables  $x_1, \dots, x_k$  whose range have cardinalities  $N_1, \dots, N_k$  we have  $H(x_1, x_2, \dots, x_k) = H(x_1) + H(x_2) + \dots + H(x_k)$ .

And later:

Let  $x$  and  $y$  be variables (with ranges  $X$  and  $Y$ ) that are dependent in the following sense: not all pairs  $x, y$  from  $X \times Y$  are allowed as values. Let  $U$  be the set of all allowed pairs. For any  $a \in X$  we consider the set  $Y_a$  of all  $y$  such that  $(a, y) \in U$ . Now the conditional entropy can be naturally defined as follows:  $H(y|a) = \log_2 N(Y_a)$  where  $N(Y_a)$  stands for the cardinality of  $Y_a$ .

There are some evident connections between combinatorial and algorithmic approaches. First, *the set of all strings having complexity less than  $n$  contains at most  $2^n$  elements* (since

different strings correspond to different programs and the number of programs does not exceed  $1 + 2 + \dots + 2^{n-1}$ ). On the other hand, as Kolmogorov says, *if a finite set  $M$  with large cardinality  $N$  can be defined by a program of a negligible length (compared to  $\log_2 N$ ), then almost all elements of  $M$  have complexity close to  $\log_2 N$  [1].*

Therefore the statement  $K(x) < n$  can be informally translated into combinatorial language as  *$x$  belongs to a naturally defined set of cardinality about  $2^n$ .*

In this section we develop parallelism between combinatorial and algorithmic approaches giving more examples. In the next section we formulate three theorems that provide combinatorial translation for linear inequalities involving Kolmogorov complexities.

Our first example is the inequality

$$K(x, y) \leq K(x) + K(y) + O(\log(K(x) + K(y))) \quad (1)$$

Here  $x$  and  $y$  are binary strings;  $K(x, y)$  denotes the complexity of pair  $(x, y)$  defined as complexity of the string  $[x, y]$  for a computable encoding  $x, y \mapsto [x, y]$  (different encodings give different complexities, but the difference is  $O(1)$ ).

The combinatorial counterpart of this inequality is the following statement: Let  $A$  be a subset of the product  $X \times Y$  of two finite sets  $X$  and  $Y$ . Then

$$\#A \leq \#\pi_X(A) \cdot \#\pi_Y(A) \quad (2)$$

where  $\#$  stands for cardinality,  $\pi_X$  and  $\pi_Y$  are projections (e.g.,  $\pi_X(A) = \{x \in X \mid \exists y \langle x, y \rangle \in A\}$ ).

The similarity is straightforward: take logarithms and recall that “combinatorial entropy” is logarithm of cardinality of range. If a pair of variables  $x, y$  ranges over  $A \subset X \times Y$ , then  $x$  ranges over  $\pi_X(A)$  and  $y$  ranges over  $\pi_Y(A)$ .

Now consider a stronger inequality

$$K(x, y) \leq K(x) + K(y|x) + O(\log(K(x) + K(y))) \quad (3)$$

(Let us note that all inequalities with complexities are considered up to a  $O(\log m)$  term where  $m$  is the sum of complexities of all strings involved; we omit  $O(\log m)$  terms (and  $O(1)$  terms) in the sequel.)

The inequality (3) is stronger than (1) since  $K(y|x) \leq K(y)$ .

Recalling Kolmogorov’s explanation of the combinatorial meaning of conditional entropy, we come to the following inequality:

$$\#A \leq \#\pi_X(A) \cdot [\max_{x \in X} \#A_x], \quad (4)$$

where  $A_x$  stands for the set  $\{y \mid \langle x, y \rangle \in A\}$ . Again the inequality (4) is stronger than (2) since  $\#A_x \leq \#\pi_Y(A)$  for any  $x \in X$ .

The next example involves three variables and is considered in detail in [2]. The inequality

$$2K(x, y, z) \leq K(x, y) + K(x, z) + K(y, z) \quad (5)$$

is true (up to logarithmic terms) for any three strings  $x, y, z$ . Its combinatorial counterpart says that

$$(\#A)^2 \leq \#\pi_{XY}(A) \cdot \#\pi_{XZ}(A) \cdot \#\pi_{YZ}(A) \quad (6)$$

for any subset  $A$  of the cartesian product  $X \times Y \times Z$  of three finite sets  $X$ ,  $Y$  and  $Z$ . (Here  $\pi_{XY}$  stands for the projection of  $X \times Y \times Z$  onto  $X \times Y$  etc.)

This inequality also can be strengthened by replacing unconditional complexity by conditional one:

$$2K(x, y, z) \leq K(x, y) + K(x, z) + K(y, z|x) \quad (7)$$

The combinatorial counterpart is

$$(\#A)^2 \leq \#\pi_{XY}(A) \cdot \#\pi_{XZ}(A) \cdot [\max_{x \in X} \#A_x] \quad (8)$$

where  $A_x = \{\langle y, z \rangle \mid \langle x, y, z \rangle \in A\}$ . (Note the we can add condition only to one term in the right-hand side.)

All four examples given above follow the same pattern and are covered by theorem 1 below; it says that combinatorial statement can be derived from corresponding inequality and vice versa.

More subtle example is provided by an inequality

$$K(x) + K(y|x) \leq K(x, y) \quad (9)$$

where, as usual, logarithmic terms are omitted. (This inequality is a reversed form of (3), so in fact inequality (3) is an equality.) What is the corresponding combinatorial statement? One could try

$$\#\pi_X(A) \cdot [\max_{x \in X} \#A_x] \leq \#A$$

but this statement is false for evident reasons (consider  $A$  that has large  $A_x$  for some  $x$  and small  $A_x$  for many other  $x$ 's). However, one can find a true statement which looks parallel to (9). Here it is:

Let  $X$  and  $Y$  be two finite sets and let  $A$  be a subset of  $X \times Y$ . Let  $u$  and  $v$  be two integers such that  $uv \geq \#A$ . Then  $A$  can be partitioned into  $A = U \cup V$  (10) with  $\#\pi_X(U) \leq u$  and  $\max_{x \in X} \#V_x \leq v$ .

To prove (10) consider the set  $T$  of all  $x \in X$  such that  $\#A_x > v$ . This set contains at most  $u$  elements (otherwise  $\#A > uv$ ). Now let  $U$  be the set of all  $\langle x, y \rangle \in A$  such that  $x \in T$  and let  $V$  be the remaining part of  $A$ . Then  $\pi_X(U) = T$  and  $\#\pi_X(U) \leq u$ ; on the other hand,  $\#V_x$  is zero for  $x \in T$  and does not exceed  $v$  for  $x \notin T$ , therefore,  $\max_{x \in X} \#V_x \leq v$ .

In fact, the statement (10) can be used as an intermediate step in the proof of (9).

Our last example is the so-called “basic inequality” from [4], i.e., the inequality

$$K(x) + K(x, y, z) \leq K(x, y) + K(x, z) \quad (11)$$

This inequality follows from the inequality  $K(y, z|x) \leq K(y|x) + K(z|x)$  (which is a “conditional version” of (1)) using the equalities  $K(x, y) = K(y|x) + K(x)$ ,  $K(x, z) = K(z|x) + K(x)$  and  $K(x, y, z) = K(y, z|x) + K(x)$ ; all three equalities mentioned follow from (3) and (9).

Inequality (11) corresponds to the following combinatorial statement:

Let  $X, Y$  and  $Z$  be three finite sets and let  $A$  be a subset of  $X \times Y \times Z$ . Let  $u$  and  $v$  be two integers such that  $uv \geq \#\pi_{XY}(A) \cdot \#\pi_{XZ}(A)$ . Then  $A$  can be partitioned into  $A = U \cup V$  with  $\#\pi_X(U) \leq u$  and  $\#V \leq v$ . (12)

This statement can be proved as follows. For each  $x \in X$  consider the set

$$A_x = \{\langle y, z \rangle \mid \langle x, y, z \rangle \in A\}$$

The set  $X$  can be linearly ordered in such a way that  $\#A_x$  decreases as  $x$  increases. Consider  $l$  first elements of  $X$  in this ordering. Corresponding triples form the set  $U$ ; the remaining part of  $A$  goes to  $V$ . (It is easy to see that this choice is optimal; we want to make  $\#V$  smaller, so we include large  $A_x$  into  $U$ .) The construction guarantees that  $\#\pi_X(U) \leq l$ . It remains to prove that  $\#V \leq v$ .

Let  $S_1$  and  $S_2$  be the cardinalities of  $\pi_{XY}(A)$  and  $\pi_{XZ}(A)$ . Let us prove first that all  $A_x$  outside  $U$  have cardinalities at most  $S_1 S_2 / l^2$ . Let  $p(x)$  and  $q(x)$  be the cardinalities of projections of  $A_x$  onto  $Y$  and  $Z$ . Then  $\sum_x p(x) = S_1$  and  $\sum_x q(x) = S_2$ . Therefore, the average value of  $p(x)$  for  $l$  first values of  $x$  (corresponding to the set  $U$ ) does not exceed  $S_1/l$ ; the average value of  $q(x)$  for  $l$  first values of  $x$  does not exceed  $S_2/l$ . Using Cauchy inequality, we conclude that the geometric mean of  $l$  first values of  $p(x)$  [of  $q(x)$ ] does not exceed  $S_1/l$  [resp.  $S_2/l$ ]. Therefore, the geometric mean of the product  $p(x)q(x)$  does not exceed  $S_1 S_2 / l^2$ , and the minimal value of  $p(x)q(x)$  does not exceed  $S_1 S_2 / l^2$ . Since  $\#A_x \leq p(x)q(x)$ , the minimal value of  $\#A_x$  in  $U$  (and all the values outside  $U$ ) does not exceed  $S_1 S_2 / l^2$ .

Now we know that  $\#V_x \leq S_1 S_2 / l^2$  for all  $x$  (here  $V_x = \emptyset$  for  $l$  first values of  $x$  and  $V_x = A_x$  for remaining  $x$ ). It remains to apply the inequality (8) to get the desired result:

$$\#V \leq \sqrt{S_1 \cdot S_2 \cdot \frac{S_1 S_2}{l^2}} = \frac{S_1 S_2}{l} \leq v.$$

The statement (12) is proved.

## 2 Linear inequalities

We hope that the examples above make clear the correspondence between inequalities about complexities and combinatorial statements. However, let us give the exact definitions for the general case.

We consider linear inequalities involving strings  $x_1, \dots, x_s$ . (The number  $s$  of strings is a constant.) For any set  $I \subset \{1, \dots, s\}$  containing elements  $i_1, \dots, i_m$  we denote by  $x_I$  the tuple  $\langle x_{i_1}, \dots, x_{i_m} \rangle$ . Its complexity (defined in a natural way using encodings) is denoted by  $K(x_I)$ . For example, the basic inequality (11) can be written in this notation as

$$K(x_{\{1\}}) + K(x_{\{1,2,3\}}) \leq K(x_{\{1,2\}}) + K(x_{\{1,3\}})$$

The general form of the linear inequality involving complexities of strings  $x_1, \dots, x_s$  and their combinations, is

$$\sum_I \lambda_I K(x_I) \geq 0.$$

The general form for an inequality involving conditional complexities is

$$\sum_{I \cap J = \emptyset} \lambda_{I,J} K(x_I | x_J) \geq 0. \quad (13)$$

We assume that  $I \cap J = \emptyset$  since  $K(x_I | x_J) = K(x_{I \setminus J} | x_J)$ .

Now we need to introduce the notation for combinatorial statements. Let  $X_1, \dots, X_s$  be sets. For each  $I \subset \{1, \dots, s\}$  we consider a projection function  $\pi_I$  that maps  $X_1 \times \dots \times X_s$  onto  $\prod_{i \in I} X_i$ . For any  $A \subset X_1 \times \dots \times X_s$  by  $\pi_I(A)$  we denote the image of  $A$  under this projection;  $n_I(A) = \#\pi_I(A)$  is its cardinality. (According to Kolmogorov,  $\log_2 n_I(A)$  can be considered as “combinatorial entropy” of projection  $x_I$  if  $x$  ranges over  $A$ .)

Conditional combinatorial entropy can be defined in a similar way. Let  $I$  and  $J$  be disjoint subsets of the index set  $\{1, \dots, s\}$ . For any  $a \in A$  consider a section of  $A$  going through  $a$  obtained when all  $J$ -coordinates are fixed; consider  $I$ -projection of this section. Cardinality of this projection depends on  $a$ ; let  $n_{I|J}(a)$  be the maximal cardinality. Reformulation: fix  $J$ -coordinates of a variable  $a \in A$  and consider the set of all possible values of  $I$ -coordinates. (This set depends on the values of  $J$ -coordinates.) Maximal cardinality of this set is denoted by  $n_{I|J}(A)$ .

The connection between combinatorial entropy and Kolmogorov complexity can be informally described as follows. Let  $A$  be a set whose elements are tuples of strings  $\langle x_1, \dots, x_s \rangle$ . Assume that Kolmogorov complexity of  $A$  is small. Then the maximal value of  $K(x_I | x_J)$  over all  $\langle x_1, \dots, x_n \rangle \in A$  is close to  $\log_2 n_{I|J}(A)$ . Indeed, to specify  $x_J$  when  $x_I$  is known, we need  $\log_2 N$  bits, where  $N$  is the number of possible values of  $x_I$  when  $x_J$  is known. This simple observation (refined in an appropriate way) is the main point of the proofs given below.

Our first theorem considers the case when only one coefficient  $\lambda_{I,J}$  is negative. In other words, we consider inequality of type

$$K(x_{I_0} | x_{J_0}) \leq \sum_{I,J} \lambda_{I,J} K(x_I | x_J) \quad (14)$$

where summation is over pairs of disjoint sets different from  $(I_0, J_0)$  and all  $\lambda_{I,J}$  are non-negative.

**Theorem 1** *The inequality (14) is valid for all binary strings  $x_1, \dots, x_s$  (up to  $O(\log(K(x_1) + \dots + K(x_s)))$  term) if and only if*

$$n_{I_0|J_0}(A) \leq \prod_{I,J} [n_{I|J}(A)]^{\lambda_{I,J}} \quad (15)$$

for any subset  $A \subset X_1 \times \dots \times X_s$  ( $X_1, \dots, X_s$  are arbitrary finite sets).

This theorem can be applied to the examples given above: it says that (1) is equivalent to (2), that (3) is equivalent to (4), that (5) is equivalent to (6), and that (7) is equivalent to (8). A special case of this theorem (inequalities (5) and (6)) was considered in [2]. Other special cases of this theorem and theorem 2 below are considered in [5]; in this paper

Shannon entropy is used instead of Kolmogorov complexity and all  $X_i$  have two elements (this restriction is not essential).

**Proof.** Let us prove (15)  $\Rightarrow$  (14) first. Let  $x_1, \dots, x_s$  be arbitrary strings and  $k_{I|J} = K(x_I|x_J)$ . Consider the set  $A$  of all tuples  $y = \langle y_1, \dots, y_s \rangle$  such that  $K(y_I|y_J) \leq k_{I|J}$  for all  $(I, J) \neq (I_0, J_0)$ . We want to apply (15) to  $A$ . It is easy to see that  $\log_2 n_{I|J}(A) \leq k_{I|J} + O(1)$ . Indeed, if  $y_J$  is fixed, only  $2 \cdot 2^{k_{I|J}}$  values of  $y_I$  are possible, since these values are obtained from  $y_J$  by programs of length at most  $k_{I|J}$ . Applying (15) to  $A$ , we conclude that  $\log_2 n_{I_0|J_0}(A) \leq \sum \lambda_{I,J} k_{I|J} + O(1)$ . Note also that the set  $A$  can be enumerated effectively provided all  $k_{I|J}$  are given (we need  $O(\log(K(x_1) + \dots + K(x_s)))$  bits to specify all  $k_{I|J}$ ). Now we see that  $I_0$ -coordinates of any element  $y$  of  $A$  are determined by  $J_0$ -coordinates of  $y$  and its ordinal number in the enumeration of all  $A$ -elements having given  $J$ -coordinates. This number takes  $\log_2 n_{I_0|J_0}$  bits, so we get (14).

Formally speaking, there is an error in this argument: we cannot apply (15) to  $A$  directly, since it is infinite. However, we can apply (15) to all finite subsets of  $A$ : if  $n_{I_0|J_0}(A') \leq c$  for all finite  $A' \subset A$ , then  $n_{I_0|J_0}(A) \leq c$ .

Now let us prove (14)  $\Rightarrow$  (15). This proof is given in [2] for the special case of inequalities (5) and (6). It uses some trick: to get rid of logarithmic terms, we consider a sequence of elements of  $A$  instead of one element.

We may assume w.l.g. that  $X_1, \dots, X_s$  are sets of binary strings. Let  $\mathbf{y} = y^1, \dots, y^M$  be a sequence of arbitrary elements of  $A$ . Each  $y^i$  is a sequence of strings  $y_1^i, \dots, y_s^i$ , so  $\mathbf{y}$  can be considered as a matrix with  $M$  rows and  $s$  columns. For any set  $I \subset \{1, \dots, s\}$  we denote the sequence  $y_1^1, \dots, y_I^M$  by  $\mathbf{y}_I$ . (To get  $\mathbf{y}_I$  from  $\mathbf{y}$  we consider only columns of the matrix whose numbers belong to  $I$ .)

Now we apply the inequality (14) to the columns of the matrix. For any disjoint sets  $I, J \subset \{1, \dots, s\}$  the complexity  $K(\mathbf{y}_I|\mathbf{y}_J)$  does not exceed  $M \log_2 n_{I|J}(A) + O(\log_2 M)$  where the constant in  $O$ -notation depends on  $A$  but not on  $M$ . Indeed, to specify  $\mathbf{y}_I$  when  $\mathbf{y}_J$  is known we need (for each row  $i$ ) to use  $\log_2 n_{I|J}$  bits for the ordinal number of  $y_I^i$  in the set of all possibilities (for given  $y_J^i$ ).

Therefore, for any  $y^1, \dots, y^M \in A$  we have  $K(\mathbf{y}_{I_0}|\mathbf{y}_{J_0}) \leq M \sum \lambda_{I,J} \log_2 n_{I|J}(A) + O(\log_2 M)$ . Now we want to get an upper bound for  $n_{I_0, J_0}(A)$ . Fix some value of  $J_0$ -coordinates. We want to get an upper bound for the number  $N$  of possible values of  $I_0$ -coordinates compatible with fixed  $J_0$ -coordinates. Consider an arbitrary matrix  $\mathbf{y}$  where all rows have given  $J_0$ -coordinates. Since  $J_0$ -coordinates are fixed,  $K(\mathbf{y}_{J_0}) = O(\log_2 M)$  and  $K(\mathbf{y}_{I_0}) \leq M \sum \lambda_{I,J} \log_2 n_{I|J}(A) + O(\log_2 M)$ . On the other hand, there are still  $N^M$  possible values of  $\mathbf{y}_{I_0}$ , and all of them have bounded complexity, therefore  $\log_2(N^M) = M \log_2 N \leq M \sum \lambda_{I,J} \log_2 n_{I|J}(A) + O(\log_2 M)$ . Since  $\log_2 M/M \rightarrow 0$  as  $M \rightarrow \infty$ , we get the required upper bound for  $N$ .

Theorem 1 is proved.

Let us consider a special case of (14) when no conditional complexities are involved:

$$K(x_1, \dots, x_s) \leq \sum \lambda_I K(x_I) \quad (16)$$

Here  $\lambda_I$  are non-negative reals (for all  $I \subsetneq \{1, \dots, s\}$ ).

**Theorem 2** *The inequality (16) is true for all  $x_1, \dots, x_s$  (up to a logarithmic term) if and only if for any  $j = 1, \dots, s$  the sum of coefficients  $\lambda_I$  for all  $I$  containing  $j$  is at least 1.*

**Proof.** Let  $x_i$  be empty strings for all  $i \neq j$ . Then the inequality (16) can be rewritten as  $K(x_j) \leq \sum \lambda_I K(x_j)$  where the sum is taken over all  $I$  containing  $j$ . Therefore, if (16) is true for all strings, this sum is at least 1.

On the other hands, if all these sums are at least 1, we can prove (16) as follows. Using (3) and (9), we rewrite  $K(x_1, \dots, x_s)$  as

$$K(x_1) + K(x_2|x_1) + K(x_3|x_1, x_2) + \dots + K(x_s|x_1, \dots, x_{s-1})$$

and rewrite complexities in the right-hand side in the same way (in the same order of indices). For example, the term  $K(x_1, x_3)$  in the right-hand side becomes  $K(x_1) + K(x_3|x_1)$ . We then add omitted conditions in the right-hand side (e.g., replace  $K(x_3|x_1)$  by  $K(x_3|x_1, x_2)$ ) and get a stronger inequality; this stronger inequality is valid according to our assumption (sum of coefficients for each  $K(x_i|x_1, \dots, x_{i-1})$  is at least 1).

Theorem 2 is proved.

This argument shows also that any valid inequality of type (16) is a positive linear combination of basic inequalities in the sense of [4].

Now we return to the general case and consider inequalities of type  $\sum \lambda_{I,J} K(x_I|x_J) \geq 0$  where several coefficients may be negative. It is convenient to separate positive and negative coefficients and consider inequalities of type

$$\sum_{(I,J) \in \mathcal{A}} \alpha_{I,J} K(x_I|x_J) \leq \sum_{(I,J) \in \mathcal{B}} \beta_{I,J} K(x_I|x_J) \quad (17)$$

where all  $\alpha_{I,J}$  and  $\beta_{I,J}$  are positive and  $\mathcal{A}, \mathcal{B}$  are disjoint sets of pairs of disjoint subsets of  $\{1, \dots, s\}$ .

The following theorem gives a combinatorial statement that is equivalent to (17). Unfortunately, this condition is more complicated than one could expect looking at the relations between (9) and (10) or between (11) and (12). It include a polynomial factor that corresponds to additive logarithmic term in the inequality about complexities.

Notation:  $\mathbb{B}^n$  is a set of all binary strings of length  $n$ .

**Theorem 3** *The inequality (17) is valid for given coefficients  $\alpha_{I,J}$  and  $\beta_{I,J}$  and for any strings  $x_1, \dots, x_s$  (up to a logarithmic term) if and only if the following combinatorial statement is true:*

*there exists a constant  $c$  such that for any  $n$ , for any set  $A \subset (\mathbb{B}^n)^s$  and for any integers  $a_{I,J}$  such that*

$$\prod_{(I,J) \in \mathcal{B}} [n_{I|J}(A)]^{\beta_{I,J}} \leq \prod_{(I,J) \in \mathcal{A}} a_{I,J}^{\alpha_{I,J}} \quad (18)$$

*the set  $A$  can be covered by sets  $U_{I,J}$  (for  $(I, J) \in \mathcal{A}$ ) such that*

$$n_{I|J}(U_{I,J}) \leq a_{I,J} \cdot n^c$$

Before proving this theorem, let us look at the combinatorial translation for the basic inequality (11): there exists a constant  $c$  such that for all  $n$ , for any set  $A \subset X \times Y \times Z$  (where  $X = Y = Z = \mathbb{B}^n$ ) and for any  $u$  and  $v$  such that  $\#\pi_{XY}(A)\#\pi_{XZ}(A) \leq uv$  there exist  $U$  and  $V$  such that  $A \subset U \cup V$ ,  $\#\pi_X(U) \leq un^c$  and  $\#V \leq vn^c$ . We see that the only difference between this statement and (12) is the factor  $n^c$ . (It seems quite possible that theorem 3 remains true without this factor. However, it is needed in our proof.)

**Proof** of theorem 3. Assume that the inequality (17) is valid up to a logarithmic term  $O(\log(K(x_1) + \dots + K(x_s)))$ . We want to prove (18). For a given  $n$  and given  $A$  there exists some constant  $c(n, A)$  that makes the statement (18) true. This is evident; what we need to prove is that the same constant works for all  $n$  and all  $A$ . For a given  $n$  consider the “worst-case” set  $A_n$ , i.e., the set which needs maximal constant. This set can be effectively found (try all possibilities; it is very long, but finite process). Therefore, complexity of  $A_n$  is  $O(\log n)$ . For any  $x \in A_n$  and for any disjoint  $I, J \subset \{1, \dots, s\}$  we have  $K(x_I|x_J) \leq \log_2 n_{I|J}(A_n) + O(\log n)$  (to specify  $x_I$  when  $x_J$  is fixed we need to specify  $A_n$  and the ordinal number of  $x_I$ ). Therefore, if numbers  $a_{I,J}$  satisfy the inequality

$$\prod_{(I,J) \in \mathcal{B}} [n_{I|J}(A_n)]^{\beta_{I,J}} \leq \prod_{(I,J) \in \mathcal{A}} a_{I,J}^{\alpha_{I,J}}$$

then

$$\sum_{(I,J) \in \mathcal{B}} \beta_{I,J} K(x_I|x_J) \leq \sum_{(I,J) \in \mathcal{B}} \beta_{I,J} \log_2 n_{I|J}(A) + O(\log n) \leq \sum_{(I,J) \in \mathcal{A}} \alpha_{I,J} \log_2 a_{I,J} + O(\log n)$$

Combining this inequality with (17), we conclude that

$$\sum_{(I,J) \in \mathcal{A}} \alpha_{I,J} K(x_I|x_J) \leq \sum_{(I,J) \in \mathcal{A}} \alpha_{I,J} \log_2 a_{I,J} + C \log n$$

for any  $x \in A_n$  and for some fixed  $C$  (not depending on  $n$ ). Therefore, if  $x \in A_n$ , then

$$\alpha_{I,J} K(x_I|x_J) \leq \alpha_{I,J} \log_2 a_{I,J} + \frac{C}{\#\mathcal{A}} \log n$$

for at least one  $(I, J) \in \mathcal{A}$ . In other terms, sets

$$U_{I,J} = \{x \mid K(x_I|x_J) \leq \log_2 a_{I,J} + \frac{C}{\alpha_{I,J}\#\mathcal{A}} \log n\}$$

cover  $A$ . And  $\log_2 n_{I|J}(U_{I,J}) \leq \log_2 a_{I,J} + c \log n$  for some constant  $c$  that does not depend on  $n$ . Since  $A_n$  is the “worst-case” set by our assumption, we conclude that  $c(n, A_n)$  are bounded by a constant not depending on  $n$ , and (18) is true.

To prove the second part of the theorem, assume that the statement (18) is true. We need to prove (17) for arbitrary tuple  $x = \langle x_1, \dots, x_s \rangle$ . To do that, we “generalize”  $x$  and include it in the set  $A$  of tuples of strings that have “similar complexity behavior”. Then we apply the statement (18) to  $A$ .



Formally  $A$  is defined as the set of all tuples  $y = \langle y_1, \dots, y_s \rangle$  such that  $K(y_I|y_J) \leq K(x_I|x_J)$  for any disjoint sets  $I, J \subset \{1, \dots, s\}$ . (This set was already used in the proof of theorem 1.) The set  $A$  is not empty since it contains  $x$ . Moreover,  $\log_2 \#A$  is close to  $K(x_1, \dots, x_s)$ . Indeed,  $\log_2 \#A$  cannot be significantly larger than  $K(x_1, \dots, x_s)$  because all  $y \in A$  have complexity not exceeding  $K(x_1, \dots, x_s)$ . On the other hand,  $A$  can be enumerated by a program that has logarithmic (in  $K(x_1) + \dots + K(x_s)$ ) length (we need to specify all complexity bounds; number of these bounds is exponential in  $s$ , but  $s$  is considered as a constant). Therefore, complexity of any  $y \in A$  (including  $x$ ) does not exceed significantly  $\log_2 \#A$ , so  $\log \#A$  cannot be significantly less than  $K(x_1, \dots, x_s)$ .

The same argument shows that for any  $(I, J)$  the number  $\log_2 n_{I|J}(U_{I,J})$  differs from  $K(x_I|x_J)$  at most by  $O(\log(K(x_1) + \dots + K(x_n)))$ .

To apply the statement (18) to  $A$  we need to choose some value of  $n$ . Let  $n$  be equal to  $K(x_1) + \dots + K(x_n) + 1$ . Using this value, we cannot apply (18) directly: an element  $y = \langle y_1, \dots, y_s \rangle \in A$  can contain very long  $y_i$ . However, the purely combinatorial nature of (18) allows us to rename all  $y_i$ . There is at most  $2^n$  of them (since all  $y_i$  have complexity less than  $n$ , so they can be replaced by strings of length  $n$ ).

Now suppose that (in contradiction with (17))

$$\sum_{(I,J) \in \mathcal{B}} \beta_{I,J} K(x_I|x_J) < \sum_{(I,J) \in \mathcal{A}} \alpha_{I,J} K(x_I|x_J) - C \log_2 n,$$

for some constant  $C$  (to be fixed later).

Choose numbers  $a_{I,J}$  such that  $\log_2 a_{I,J} \leq K(x_I|x_J) - \frac{C \log n}{\alpha_{I,J} \#A}$  and still

$$\sum_{(I,J) \in \mathcal{B}} \beta_{I,J} K(x_I|x_J) < \sum_{(I,J) \in \mathcal{A}} \alpha_{I,J} \log_2 a_{I,J},$$

i.e.,

$$\prod_{(I,J) \in \mathcal{B}} K(x_I|x_J)^{\beta_{I,J}} \leq \sum_{(I,J) \in \mathcal{A}} [a_{I,J}]^{\alpha_{I,J}}.$$

Then by (18) the set  $A$  can be covered by sets  $U_{I,J}$  such that

$$n_{I|J}(U_{I,J}) \leq a_{I,J} \cdot n^c.$$

Let  $(I_0, J_0)$  be a pair such that  $x \in U_{I_0, J_0}$ . Then for some constant  $C_1$

$$\begin{aligned} K(x_{I_0}|x_{J_0}) &\leq \log_2 n_{I_0|J_0}(U_{I_0, J_0}) + C_1 \log_2 n \leq \log(a_{I_0, J_0} \cdot n^c) + C_1 \log n \\ &\leq K(x_{I_0}|x_{J_0}) - \frac{C \log n}{\alpha_{I_0, J_0} \#A} + cn + C_1 \log n \end{aligned}$$

For  $C$  large enough we get a contradiction. Theorem 3 is proved.

The underlying reason for the second part of the proof can be explained as follows.  $A$  is uniform: most of its sections (in a given direction) have approximately the same size. (The same is true for projections.) Therefore, if  $U$  is some part of  $A$  that has small sections in some direction,  $\#U$  is small compared to  $\#A$  and such  $U$ 's cannot cover  $A$ .

### 3 Prefix complexity

All inequalities for Kolmogorov complexities were considered up to  $O(\log n)$  term, where  $n$  is a sum of complexities of strings involved. Therefore we could safely ignore the difference between several existing versions of this notion. We can use plain complexity defined by Kolmogorov in [1]), denoted by  $C(x)$  in [6] and  $KS(x)$  in [7], or prefix complexity, denoted by  $K(x)$  in [6] and  $KP(x)$  in [7].

In this section we are interested in equalities valid up to  $O(1)$ . Therefore we should be careful and specify exactly the version of complexity we use. Most useful here is prefix complexity  $KP(x)$ . For example, the inequality  $KP(x, y) \leq KP(x) + KP(y) + O(1)$  is well known (see [6], example 3.1.2, p. 194). The inequality  $2KP(x, y, z) \leq KP(x, y) + KP(x, z) + KP(y, z)$  was proved (using Cauchy–Schwartz inequality) in [2]. These examples make the following conjecture plausible:

**Conjecture.** Any linear inequality involving unconditional complexities that is valid up to logarithmic term is valid up to  $O(1)$  for prefix complexity.

A partial result in this direction:

**Theorem 4** *Basic inequality (11) is valid up to  $O(1)$ -term for prefix complexity:*

$$KP(x) + KP(x, y, z) \leq KP(x, y) + KP(x, z). \quad (19)$$

**Proof.** This theorem can be easily derived from L.A. Levin’s result saying that  $KP(x, y) = KP(x) + KP(y|x, KP(x))$  (for the proof see, e.g., [6], theorem 3.9.1 p. 232). Using this result, one can rewrite (19) as

$$KP(y, z|x, KP(x)) \leq KP(y|x, KP(x)) + KP(z|x, KP(x)),$$

and this inequality is a “relativized” version of the inequality  $KP(y, z) \leq KP(y) + KP(z)$ .

However, we provide also a direct proof using a priori probabilities. Recall that  $KP(x) = -\log_2 \mathbf{m}(x)$ , where  $\mathbf{m}$  is universal enumerable semimeasure (see [6], p. 247). Therefore, we need to prove that

$$\mathbf{m}(x, y, z) \mathbf{m}(x) \geq \mathbf{m}(x, y) \mathbf{m}(x, z).$$

or

$$\mathbf{m}(x, y, z) \geq \frac{\mathbf{m}(x, y) \mathbf{m}(x, z)}{\mathbf{m}(x)}.$$

Since  $\sum_y \mathbf{m}(x, y) \leq \mathbf{m}(x)$  and  $\sum_z \mathbf{m}(x, z) \leq \mathbf{m}(x)$ , we conclude that

$$\sum_{x,y,z} \frac{\mathbf{m}(x, y) \mathbf{m}(x, z)}{\mathbf{m}(x)} \leq \sum_x \mathbf{m}(x) < 1.$$

If the fraction  $\mathbf{m}(x, y) \mathbf{m}(x, z) / \mathbf{m}(x)$  were enumerable from below, the proof would be complete, since  $\mathbf{m}$  is maximal. However, we have  $\mathbf{m}$  in the denominator, so we need to find an enumerable upper bound for this fraction having finite sum. For each  $n$  by  $\mathbf{m}_n(x, y)$  we denote the enumerable function obtained from  $\mathbf{m}$  by adding an additional requirement

$\sum_y \mathbf{m}_n(x, y) \leq 2^{-n}$ . (We eliminate values of  $\mathbf{m}$  that can violate this requirement.) Now consider the function

$$\sum_{n \geq KP(x)} \frac{\mathbf{m}_n(x, y) \mathbf{m}_n(x, z)}{2^{-n}}$$

It is an enumerable upper bound we asked for, since

$$\begin{aligned} \sum_{x, y, z} \sum_{n \geq KP(x)} \frac{\mathbf{m}_n(x, y) \mathbf{m}_n(x, z)}{2^{-n}} &\leq \sum_x \sum_{n \geq KP(x)} \frac{\sum_y \mathbf{m}_n(x, y) \sum_z \mathbf{m}_n(x, z)}{2^{-n}} \\ &\leq \sum_x \sum_{n \geq KP(x)} 2^{-n} \leq \sum_x 2 \mathbf{m}(x) \leq 2. \end{aligned}$$

Theorem 4 is proved.

**Corollary:** all inequalities involving unconditional complexities, having one term in the left-hand side and being true up to logarithmic term, are true up to  $O(1)$  for prefix complexity.

(Indeed, theorem 2 guarantees that such an inequality is a positive linear combination of basic inequalities, so we can apply theorem 4.)

This corollary can be proved directly using semimeasures and the following version of Jensen's inequality: if  $\alpha_1 + \dots + \alpha_s = 1$ ,  $\alpha_i \geq 0$ , then

$$\int [f_1(x)]^{\alpha_1} \dots [f_s(x)]^{\alpha_s} dx \leq \left[ \int f_1(x) dx \right]^{\alpha_1} \dots \left[ \int f_s(x) dx \right]^{\alpha_s}$$

## Acknowledgements

This work was done while visiting Ecole Normale Supérieure de Lyon, Laboratoire de l'Informatique du Parallélisme.

Authors are grateful to all participants of Kolmogorov seminar (Moscow) and especially to M. Ushakov for his help in writing this paper.

## References

- [1] A.N.Kolmogorov. Tri podkhoda k opredeleniju ponjatija "kolichestvo informatsii", *Problemy peredachi informatsii*, tom 1, no. 1, 1965, pp. 3–11. (Translation: Three approaches to the definition of the quantity of information. *Prob. Inform. Transmission*, **1**(1), 1965, pp. 4–7.)
- [2] D. Hammer, A. Shen. A Strange Application of Kolmogorov Complexity. *Theory of Computing Systems*, **31**, 1998, pp. 1–4. (See also Tech. Report CS-R9328, CWI, Netherlands, 1993.)
- [3] Hammer Ph.D. thesis

- [4] D. Hammer, A. Romashenko, A. Shen, N. Vereshchagin. Inequalities for Kolmogorov complexities and Shannon entropies. *Proceedings of CCC'97*, Ulm. (To be published in JCSS.)
- [5] J. Llewellyn, J. Radhakrishnan, On Shearer's Lemma.
- [6] M. Li, P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Second edition, Springer-Verlag, 1997.
- [7] V.A. Uspensky, A. Shen. Relations between varieties of Kolmogorov complexities. *Math. Systems Theory*, **29**, 1996, pp. 271–292.