A Random Oracle Does not Help Extract the Mutual Information

Andrei Muchnik[1] and Andrei Romashchenko[2]

[1] Andrei Muchnik (24.02.1958 – 18.03.2007) worked at the Institute of New Technologies in Education

[2] Laboratoire de l'Informatique du Parallélisme (Lyon) and Institute for Information Transmission Problems (Moscow), email: Andrei.Romashchenko@ens-lyon.fr

Abstract. Assume a tuple of words $\bar{x} = \langle x_1, \ldots, x_n \rangle$ has negligible mutual information with another word y. Does this mean that properties of Kolmogorov complexity for $\bar{x}$ do not change significantly if we relativize them conditional to y ? This question becomes very nontrivial when we try to formalize it. We investigate this question for a very particular kind of properties : we show that a random (conditional to $\bar{x}$) oracle y cannot help extract the mutual information from x_i's.

1 Introduction

Kolmogorov complexity $K(x)$ of a word x is the length of a minimal description of this word for an optimal algorithmic description method (see [1, 4]). Respectively, conditional Kolmogorov complexity $K(x|y)$ is the length of a minimal description of x when y is known. In other words, $K(x|y)$ is Kolmogorov complexity of x with the oracle y.

The difference between plain and conditional complexities

$$I(x : y) = K(y) - K(y|x)$$

is called *information in x on y*. The basic result of the algorithmic information theory is the fact that $I(x : y)$ is symmetric up to a small additive term:

Theorem 1 (Kolmogorov–Levin, [1]).

$$I(x : y) = I(y : x) + \mathcal{O}(\log K(x, y)) = K(x) + K(y) - K(x, y) + \mathcal{O}(\log N)$$

If the value $I(x : y)$ is negligible (logarithmic in $K(x, y)$), the words x and y are often called *independent*.

Intuitively it seems that if x and y are 'independent' then 'reasonable' algorithmic properties of x (expressible in terms of Kolmogorov complexity) should not change significantly when we relativize them conditional to y.

Let us find a formal statement corresponding to this intuition. Let us take a tuple $\bar{x} = \langle x_1, x_2, \ldots, x_n \rangle$ instead of a single word[3] x. Suppose that the mutual

[3] More formally, we fix a computable bijection between the set of binary words and the set of all finite tuples of binary words. Now every tuple has a code. When we talk

information between $\bar{x}$ and some y is negligible. Then it is easy to see that *the basic* properties of Kolmogorov complexity for $\bar{x}$ do not really change when we relativize them conditional to y:

$$K(x_i) \approx K(x_i|y),\ K(x_i, x_j) \approx K(x_i, x_j|y), \ldots,$$

for all i, j, etc. (the approximative equations hold up to $I(y : \bar{x}) + \mathcal{O}(\log K(\bar{x}))$, which is negligible by the assumption).

Further we deal with less trivial properties of Kolmogorov complexity. Probably the simplest appropriate example is the property of extractability of common information. Let $\bar{x} = \langle x_1, x_2 \rangle$ be a pair of binary words. We say that α bits of the common information can be extracted from this pair for a precision threshold k if

$$\exists z \text{ such that for } i = 1, 2\ K(z|x_i) < k \text{ and } K(z) \geq \alpha$$

Straightforward arguments imply that for such a word z

$$K(z) \leq I(x_1 : x_2) + \mathcal{O}(k + \log K(x_1, x_2))$$

This is a very natural fact: it means that for a small threshold k we cannot extract from x_1, x_2 much more than $I(x_1 : x_2)$ bits of information.

The question on extracting common information cannot be reduced to the values of complexities $K(x_1)$, $K(x_2)$, $K(x_1, x_2)$. For example, given that $K(x_1) = K(x_2) = 2n$ and $K(x_1, x_2) = 3n$ we cannot say anything nontrivial about extracting common information. On one hand, there exist pairs $\langle x_1, x_2 \rangle$ with the given complexities, such that n bits of common information can be extracted from these words for a very small threshold $k = \mathcal{O}(1)$. On the other hand, there exist pairs with the same complexities such that only negligible amount of information can be extracted for pretty large k. See detailed discussions on this topic in [2, 3, 6, 11]. A similar property of extracting common information can be investigated not only for pairs but also for all finite tuples $\langle x_1, \ldots, x_n \rangle$. For the sake of simplicity in the sequel we restrict ourselves to the case $n = 2$ (though our technique is suitable for all n).

Once again, our intuition says that *negligible mutual information* between $\langle x_1, \ldots, x_n \rangle$ and y actually means that *the relativization conditional to y should not change properties of* $x_1, \ldots, x_n$. Let us formalize this intuitive idea for the problem of extracting common information:

Assume the mutual information between $\bar{x} = \langle x_1, x_2 \rangle$ and y is negligible. Then α bits of common information between x_1 and x_2 can be extracted for a precision threshold k iff the same is true given y as an oracle (for possibly a little different precision threshold).

The 'if' part of the equivalence above is trivial (if some information can be extracted without any oracle, the same can be done also given an oracle). The

about Kolmogorov complexity of pairs, triples, etc., we mean Kolmogorov complexity of codes of these tuples.There is no natural canonical encoding of all tuples. However the choice of a particular code is not essential. Changing this encoding we change Kolmogorov complexity of tuples by only $\mathcal{O}(1)$ additive term.

interesting part is the 'only if' statement. Let us formulate it in the most natural way, with logarithmic thresholds:

Conjecture 1. For every integer $C_1 > 0$ there exists an integer $C_2 > 0$ such that for all $\bar{x} = \langle x_1, x_2 \rangle$ and y, if $I(y : \bar{x}) \le C_1 \log N$ and

$$\exists w : \ K(w|y) \ge \alpha, \ K(w|x_i, y) \le C_1 \log N \ (i = 1, 2),$$

where $N = K(\bar{x}, y)$, (i.e., α bits of information can be extracted from x_1, x_2 for the precision threshold $C_1 \log N$, assuming y is given as an oracle) then

$$\exists z : \ K(z) \ge \alpha, \ K(z|x_i) \le C_2 \log N \ (i = 1, 2),$$

i.e., the same α bits of common information can be extracted without oracles (for another threshold $C_2 \log N$).

This natural statement is surprisingly hard to prove. In [7] this conjecture was proven for $\alpha = I(x_1 : x_2)$. The general case is still an open problem.

In this paper we prove a version of this conjecture for $o(N)$ thresholds instead of logarithmic ones.

Theorem 2. *For every function $f(N)$, $f(N) = o(N)$ there exists a function $g(N)$ (also $g(N) = o(N)$) such that for every $\bar{x} = \langle x_1, x_2 \rangle$ and y if $I(y : \bar{x}) \le f(N)$ and*

$$\exists w : \ K(w|y) \ge \alpha, \ K(w|x_i, y) \le f(N) \ (i = 1, 2),$$

where $N = K(\bar{x}, y)$, (i.e., α bits of information can be extracted from x_1, x_2 for the precision threshold $f(N)$, assuming y is given as an oracle) then

$$\exists z : \ K(z) \ge \alpha, \ K(z|x_i) \le g(N) \ (i = 1, 2),$$

i.e., the same α bits of common information can be extracted without oracles (for another threshold $g(N)$).

It is rather uncommon for algorithmic information theory that a natural statement is proven with $o(\cdot)$-precision but not up to logarithmic terms. Thus, the challenge is to prove Theorem 2 for $g(N) = \mathcal{O}(f(N))$, or at least to show that Conjecture 1 is true.

In the rest of the paper we prove Theorem 2, and in Conclusion discuss some variant of Conjecture 1 that is known to be true.

2 Preliminaries and technical tools

The main proof of this article is based on two technical tools: typization of words with a given profile, and extracting the common information from bunches of words.

2.1 Complexity profiles

For an n-tuple of words $\bar{x} = \langle x_1, \ldots, x_n \rangle$ and a set of indexes $V = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ $(i_1 < i_2 < \ldots < i_k)$ we denote by $\bar{x}_V$ the tuple of words x_j for $j \in V$:

$$\bar{x}_V = \langle x_{i_1}, \ldots, x_{i_k} \rangle.$$

Thus, $K(\bar{x}_V) := K(x_{i_1}, \ldots, x_{i_k})$. We let $K(\bar{x}_\emptyset) := K(\lambda)$ (where λ is the empty word). We use similar notations for conditional complexities: if $V = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ and $W = \{j_1, \ldots, j_l\} \subseteq\{ 1, \ldots, n\}$ we denote

$$K(\bar{x}_V | \bar{x}_W) := K(x_{i_1}, \ldots, x_{i_k} | x_{j_1}, \ldots, x_{j_l}).$$

We also let $K(\bar{x}_V | \bar{x}_\emptyset) := K(\bar{x}_V | \lambda)$ (which is equal to $K(\bar{x})$ up to an additive constant).

Definition 1. *We call by* complexity profile $\boldsymbol{K}$ *of an n-tuple $x_1, \ldots, x_n$ the vector of integers that consists of all complexity quantities $K(\bar{x}_V | \bar{x}_W)$, where $V, W \subseteq \{1, \ldots, n\}$, $V \cap W = \emptyset$ and $V \neq \emptyset$. Note that complexity profile implicitly contains unconditional complexity quantities: if $W = \emptyset$ we have $K(\bar{x}_V | \bar{x}_\emptyset) = K(\bar{x}_V) + \mathcal{O}(1)$. We need to fix somehow the order of components in the complexity profile. Let us suppose that all pairs (V, W) are arranged in the lexicographical order, i.e.,*

$$\boldsymbol{K}(x_1, \ldots, x_n) = (K(x_1), K(x_1|x_2), \ldots, K(x_2|x_1), K(x_2|x_3), \ldots).$$

Similarly we define the conditional complexity profile *of $x_1, \ldots, x_n$ conditional to some y. It is the vector of all complexity quantities $K(\bar{x}_V | \bar{x}_W, y)$:*

$$\boldsymbol{K}(x_1, \ldots, x_n | y) = (K(x_1|y), K(x_1|x_2, y), \ldots, K(x_2|x_1, y), K(x_2|x_3, y), \ldots).$$

We say that a profile $\bar{\alpha}$ is not greater than another profile $\bar{\beta}$ (notation: $\bar{\alpha} \leq \bar{\beta}$) if *every* component of the first vector is not greater than the corresponding component of the second vector.

Denote by $\rho(\alpha, \beta)$ the l_∞-norm of the difference between the vectors α and β.

2.2 Typization

The method if *typization* was proposed in [8, 10, 9].

Definition 2. *Let $\bar{x} = \langle x_1, \ldots, x_n \rangle$ and $\bar{y} = \langle y_1, \ldots, y_m \rangle$ be tuples of words.* The typization *of $\bar{x}$ conditional to $\bar{y}$ is the following set of n-tuples:*

$$T(\bar{x}|\bar{y}) := \{\bar{x}' = \langle x'_1, \ldots, x'_n \rangle \mid \boldsymbol{K}(\bar{x}', \bar{y}) \leq \boldsymbol{K}(\bar{x}, \bar{y})\}.$$

*Further, the k-*strong typization *of $\bar{x}$ conditional to $\bar{y}$ is the following set:*

$$ST_k(\bar{x}|\bar{y}) := T(\bar{x}|\bar{y}) \cap \{\bar{x}' = \langle x'_1, \ldots, x'_n \rangle \mid \rho(\boldsymbol{K}(\bar{x}', \bar{y}), \boldsymbol{K}(\bar{x}, \bar{y})) \leq k\}.$$

Obviously there exists an algorithm that enumerates the list of all elements of $T(\bar{x}|\bar{y})$ given as an input the tuple $\bar{y}$ and the profile $\boldsymbol{K}(\bar{x},\bar{y})$.

The following Lemmas are proven in [8, 9]:

Lemma 1. *For every $\bar{x} = (x_1, \ldots, x_n)$ and $\bar{y} = (y_1, \ldots, y_m)$*

$$\log |T(\bar{x}|\bar{y})| = K(\bar{x}|\bar{y}) + \mathcal{O}(\log N),$$

where $N = K(\bar{x},\bar{y})$. The constant in $\mathcal{O}(\cdot)$-notation depends on n and m.

Lemma 2. *There exists a computable function $C = C(n,m)$ such that for every n-tuple $\bar{x} = \langle x_1, \ldots, x_n\rangle$ and for every m-tuple $\bar{y} = \langle y_1, \ldots, y_m\rangle$ it holds*

$$|ST_{C(n,m)\log N}(\bar{x}|\bar{y})| > \frac{1}{2}|T(\bar{x}|\bar{y})|,$$

where $N = K(\bar{x},\bar{y})$.

For brevity we denote by $ST(\bar{x}|\bar{y})$ the set $ST_{C\log N}(\bar{x}|\bar{y})$, where C is the value from Lemma 2.

2.3 Bunches

The following definition of a *bunch* was given in [12]:

Definition 3. *A set $X \subset \{0,1\}^*$ is called an (α,β,γ)-bunch if*

1. *$|X| = 2^\alpha$,*
2. *$K(x_1|x_2) < \beta$ for every $x_1, x_2 \in X$,*
3. *$K(x) < \gamma$ for all $x \in X$.*

The usage of this definition is based on the following combinatorial lemma:

Lemma 3 ([12]). *There exists an algorithm that takes (α,β,γ) as an input and prints a list of* standard *(α,β,γ)-branches $U_0, \ldots, U_q$ such that:*

- *for every (α,β,γ)-bunch U there exists a number $i \le q$ such that $|U \cap U_i| \ge 2^{\beta-\epsilon}$, $\epsilon = 2(\beta-\alpha) + \mathcal{O}(1)$,*
- *$q < 2^{\beta+\gamma-2\alpha+\mathcal{O}(1)}$.*

Here is a typical usage of Lemma 3: Assume we are given 2^n words a_i of complexity $2n$, and for every pair a_i, a_j it holds $K(a_i|a_j) \le n$. Then the given family of words is an $(n,n,2n)$-bunch. From the lemma it follows that some U_s from the list of 'standard bunches' (here $s < 2^n$) contains at least $\Omega(2^n)$ of the words a_i. It is not hard to show that for *all* given a_i

$$K(a_i|s) \le n + \mathcal{O}(\log n) \text{ and } K(s|a_i) = \mathcal{O}(\log n).$$

Thus, the ordinal number s of a standard bunch U_s is an n-bit 'kernel' of the given family of a_i's; it is a materialization of the mutual information of all these words. See a more detailed discussion and corollaries of these arguments in [12].

We need to modify the definition of a bunch:

Definition 4. *A set $X \subset \{0,1\}^*$ is called an (α, β, γ)-semi-bunch if*

1. $|X| = 2^\alpha$,
2. *for every $x_1 \in X$, for the majority of all words $x_2 \in X$ it holds $K(x_1|x_2) < \beta$*
3. $K(x) < \gamma$ *for all $x \in X$.*

The following statement generalizes Lemma 3:

Lemma 4. *There exists an algorithm that takes (α, β, γ) as an input and prints a list of (α, β, γ)-semi-bunches $U_0, \dots, U_q$ such that:*

- *for every (α, β, γ)-semi-bunch U there exists a number $i \le q$ such that $|U \cap U_i| \ge 2^{\beta - \epsilon}$, where $\epsilon = 2(\beta - \alpha) + \mathcal{O}(1)$,*
- $q < 2^{\beta + \gamma - 2\alpha + \mathcal{O}(1)}$.

The proof of Lemma 4 is almost the same as the proof of Lemma 3 in [12]. We prove this lemma in Appendix. Let us call the semi-bunches $U_0, \dots, U_q$ from Lemma 4 *standard semi-bunches* (i.e., for each triple of parameters α, β, γ we fix a canonical list of standard semi-bunches).

3 Proof of Theorem 2

Let us define some notations and make several assumptions. W.l.o.g. we may suppose that $f(N) > \log N$, and $f(N)$ does not decrease ($f(N+1) \ge f(N)$ for all N).

We chose $g(N)$ and $\delta(N)$ that are not 'too large' and not 'too small', so that the construction of the proof works. Let $\delta(N) = N/\sqrt{\log \frac{N}{f(N)}}$ and

$$g(N) = C(3^{D\sqrt{\log \frac{N}{f(N)}}} \cdot f(N) + \delta(N))$$

(we will fix the constants C and D later). For brevity we will write just δ if the value of N is clear from the context.

The main construction.

Informal idea:

The main trick of the proof is typization of y and w conditional to $\bar{x}$. We take the set of all 'clones' of the pair $\langle y, w \rangle$, which have approximately the same complexity profile (conditional to $\bar{x}$). The two cases are possible:

The good case: Assume this set of 'clones' is well consolidated in the sense that most clones have large enough mutual information. Then we apply Lemma 4 and extract from the class of clones some common kernel z. This word z contains about α bits of information, and it is rather simple conditional to each of x_i. Thus we extract from the words x_i about α bits of common information without any oracle, and we are done.

The bad case: Assume the set of 'clones' is not well consolidated. Then there exist pairs of different clones that have rather small mutual information. At this

stage we cannot extract from x_i's their common information. Instead we change the word y to some y_1 such that conditional to y_1 at least α_1 bits of common information (where α_1 is greater than α) can be extracted from the words x_1, x_2. Thus, we come back to the assumption of the theorem, but with a greater value α_1 instead of α and a new oracle y_1 instead of y. The price for this modification is some loss of precision: instead of the term $f(N)$ we get some greater threshold $f_1(N)$.

The technical question is how to get such a word y_1. The answer is based on the fact that the set of 'clones' is not well consolidated. If we take two of them at random (denote them $\langle y', w' \rangle$ and $\langle y'', w'' \rangle$) then the pair $\langle y', y'' \rangle$ can play the role of y_1. Indeed, with the new oracle we can extract from x_i's both w' and w'', which make up α_1 bits of common information ($\alpha_1 > \alpha$; technically, we will get $\alpha_1 \geq \alpha + \delta/2$).

Then we iterate the trick above again and again, until at some stage we get a well consolidated set of clones...

The formal arguments:

We are given a w such that $K(w|x_i, y) \leq f(N)$ (for $i = 1, 2$). W.l.o.g. we assume that $\alpha = K(w|y)$ (if $K(w|y) > \alpha$, we increase the value of α; this makes the statement only stronger). Denote $m = K(y)$. The aim is to construct z such that $K(z|x_i) \leq g(n)$ and $K(z) \geq \alpha - g(N)$.

We take the strong typization of $\langle y, w \rangle$ conditional to x: $A = ST(y, w|\bar{x})$. From Lemma 1 it follows $|A| = 2^{K(y,w|\bar{x}) - \mathcal{O}(f(N))}$. We have

$$K(y, w|\bar{x}) = K(y|\bar{x}) + K(w|y, \bar{x}) + O(\log N),$$

$K(y|\bar{x}) \geq K(y) - f(N)$ (the mutual information between y and $\bar{x}$ is negligible) and $K(w|y, \bar{x}) \leq f(N)$ (w can be easily extracted from any x_i given y as an oracle). Hence, $|A| = 2^{m - \mathcal{O}(f(N))}$. Note that for every $\langle y', w' \rangle \in A$ it holds

$$K(y', w') = K(y') + K(w'|y) + \mathcal{O}(\log N) = m + \alpha + \mathcal{O}(f(N)).$$

Two cases are possible:

Case 1^0: For every $\langle y', w' \rangle \in A$ for the majority of $\langle y'', w'' \rangle \in A$

$$I(y'w' : y''w'') \geq \alpha - \delta.$$

This inequality implies that

$$K(y'w'|y''w'') = K(y', w') - I(y'w' : y''w'') \leq m + \delta + \mathcal{O}(f(N)).$$

In this case the set A is a semi-bunch with the parameters

$$(m - \mathcal{O}(f(N)), m + \delta + \mathcal{O}(f(N)), m + \alpha + \mathcal{O}(f(N)).$$

We apply Lemma 4: it follows that there exists a *standard semi-bunch* U_j (with the same parameters) such that

$$|A \cap U_j| \geq 2^{m - \delta + \mathcal{O}(f(N))},$$

and j is an integer less than $2^{\alpha+\delta+\mathcal{O}(f(N))}$. So Kolmogorov complexity of j is not greater than $\alpha + \delta + \mathcal{O}(f(N))$.

Further, the words x_i $(i = 1, 2)$ have two properties:

- for every pairs $\bar{v} \in A \cap U_j$ it holds $K(x_i|\bar{v}) \leq K(x_i|y, w)$ (by the definition of $A = ST(y, w|\bar{x})$);
- for every pair $\bar{v} \in A \cap U_j$ it holds $K(\bar{v}|j) \leq \log |U_j| + \mathcal{O}(\log N) \leq m$ (given the number j, the elements of a standard semi-bunch U_j can be enumerated algorithmically).

This means that x_i belong to the set

$$X(i) = \{\hat{x} \mid \text{ there exists at least } 2^{m-\delta+\mathcal{O}(f(N))} \text{ words } \bar{v} \\ \text{s.t. } K(\hat{x}|\bar{v}) \leq K(x_i|y, w) \leq K(x_i) - \alpha + f(N) \text{ and } K(\bar{v}|j) \leq m\}.$$

The set $X(i)$ is enumerable given j and additional $\mathcal{O}(\log N)$ bits of information (we need these additional bits to specify the parameters of the semi-bunch). Also we can bound the size of $X(i)$. Indeed, for each fixed j there exist at most 2^{m+1} different tuple $\bar{v}$ such that $K(\bar{v}|j) \leq m$; for every $\bar{v}$ there exist at most $2^{K(x_i)-\alpha+f(N)}$ different $\hat{x}$ such that $K(\hat{x}|\bar{v}) \leq K(x_i) - \alpha + f(N)$. Since for every $\hat{x} \in X(i)$ there is *at least* $2^{m-\delta+\mathcal{O}(f(N))}$ different $\bar{v}$, we get

$$\log |X(i)| \leq \log \frac{2^m \cdot 2^{K(x_i)-\alpha+f(N)}}{2^{m-\delta+O(f(N))}} \leq K(x_i) - \alpha + \delta + \mathcal{O}(f(N)).$$

It follows that $K(x_i|j) \leq K(x_i) - \alpha + \delta + \mathcal{O}(f(N))$ (in a word, the mutual information between j and x_i is at least $\alpha - \delta - \mathcal{O}(f(N))$). From symmetry of the mutual information we have

$$K(j|x_i) = K(x_i|j) + K(j) - K(x_i) + \mathcal{O}(\log N) \leq 2\delta + \mathcal{O}(f(N)).$$

We set $z = j$. Since $K(z) \geq I(z : x_i) \geq \alpha - \delta - \mathcal{O}(f(N))$, we get $K(z) \geq \alpha - g(N)$.

Thus for the function $g(n)$ defined above it holds $K(z) \geq \alpha - g(N)$ and $K(z|x_i) \leq g(N)$, and we are done.

Case 2^0. For some pair $\langle y', w' \rangle \in A$ and for the majority of $\langle y'', w'' \rangle \in A$ it holds

$$I(y'w' : y''w'') < \alpha - \delta.$$

This means that

$$K(y'y''w'w'') \geq 2m + \alpha + \delta - \mathcal{O}(\log N) \tag{1}$$

Since this inequality holds for the majority of pairs $\langle y'', w'' \rangle \in A$, we can choose one of them such that $\langle y', w' \rangle$ and $\langle y'', w'' \rangle$ are independent conditional to $\bar{x}$. In particular, the words y' and y'' are also independent conditional to $\bar{x}$ (i.e., $I(y' : y''|\bar{x}) = \mathcal{O}(\log N)$). Further, for all $\bar{x}, y', y''$ the following inequality holds:

$$I(y'y'' : \bar{x}) \leq I(y' : \bar{x}) + I(y'' : \bar{x}) + I(y' : y''|\bar{x}) + \mathcal{O}(\log N)$$

(in fact this inequality is equivalent to the sum of two trivial ones:

$$K(y'y'') \le K(y') + K(y'') + \mathcal{O}(\log N),$$
$$K(y'|x) + K(y''|x) = K(y'y''|x) + I(y' : y''|x) + \mathcal{O}(\log N),$$

which follow immediately from the Kolmogorov–Levin theorem [1]). For the given words, the quantities $I(y' : \bar{x})$ and $I(y'' : \bar{x})$ are bounded by $f(N)$ ($\bar{x}$ and y are independent), and $I(y' : y''|\bar{x}) = \mathcal{O}(\log N) \ll f(N)$. Thus, we have

$$I(y'y'' : \bar{x}) \le 3f(N) \tag{2}$$

Also we have $K(y'y'') \le 2K(y) + 3f(N) \le 3N$ (a very rough bound).

From (1) and (2) it follows that for $y^1 = \langle y', y'' \rangle$ and $w^1 = \langle w', w'' \rangle$ it holds

$$K(w^1|y^1) \ge \alpha + \delta - 3f(N) - \mathcal{O}(\log N) \ge \alpha + \delta/2.$$

Thus, we have constructed a word y^1 such that $I(y^1 : \bar{x}) \le 3f(N)$ and

$$\exists w^1 : \ K(w^1|y^1) \ge \alpha + \delta/2, \ K(w^1|x_i, y^1) \le 3f(N) \ (i = 1, 2).$$

We have got a new pair $\langle y^1, w^1 \rangle$ instead of the original one $\langle y, w \rangle$. By the construction, the word y^1 is independent from $\bar{x}$ (though the precision of 'independence' becomes three times worse: $I(y^1 : \bar{x}) \le 3f(N)$). Given y^1 as an oracle, the word w^1 is simple conditional to each x_i (the precision of 'simplicity' also becomes $3f(N)$). Complexity of w^1 conditional to y^1 is not less than $\alpha + \delta/2$. Thus, $\alpha + \delta/2$ bits of common information can be extracted from the words x_1, x_2 with the precision threshold $3f(N)$ given y^1 as an oracle. Note that complexities of the words w^1, y^1 are not greater than $3N$.

Further we iterate the arguments above. We repeat the same procedure with the pair w^1, y^1. Denote $\alpha^1 = \alpha + \delta/2$, $m_1 = K(y^1)$, and $f_1(N) = 3f(N)$. We take the strong typization of the pair $\langle y^1, w^1 \rangle$ conditional to $\bar{x}$:

$$A^1 = ST(y^1, w^1|\bar{x}).$$

Once again, we consider two cases.

Case 1^1. For every $\langle y', w' \rangle \in A^1$ for the majority $\langle y'', w'' \rangle \in A^1$

$$I(y'w' : y''w'') \ge \alpha_1 - \delta.$$

In this case A^1 is a semi-bunch with the following parameters:

$$(m_1 - \mathcal{O}(f_1(N)), m_1 + \delta + \mathcal{O}(f_1(N)), m_1 + \alpha_1 + \mathcal{O}(f_1(N)).$$

From Lemma 4 we get a number j such that for $i = 1, 2$

$$K(j|x_i) \le 2\delta + \mathcal{O}(f_1(N)), \quad I(j : x_i) \ge \alpha_1 - \delta + \mathcal{O}(f_1(N)).$$

Similarly to Case 1^0, we define $z := j$, and we are done.

Case 2^1. Assume that for some $\langle y', w' \rangle \in A^1$ and for the majority of $\langle y'', w' \rangle \in A^1$ it holds $I(y'w' : y''w'') < \alpha_1 - \delta$. Then there exists a pair $\langle y^2, w^2 \rangle$ such that

1. $K(y^2) = m_2 < 3m_1$,
2. $I(y^2 : \bar{x}) \le f_2(N) := 3f_1(N)$,
3. $K(w^2|y^2, x_i) \le f_2(N)$,
4. $K(w^2|y^2) = \alpha_2 \ge \alpha_1 + \delta/2$.

Iterating these arguments again and again, at stage s we get some words w^s, y^s such that

1. $K(y^s) = m_s = 3m_{s-1}$,
2. $I(y^s : \bar{x}) \le f_s(N) := 3f_{s-1}(N) = 3^s f(N)$,
3. $K(w^s|y^s, x_i) \le f_s(N)$,
4. $K(w^s|y^s) = \alpha_s > \alpha_{s-1} + \delta/2 = \alpha + s\delta/2$.

We are iterating the same construction for the 'bad' cases $2^1, 2^2, 2^3 \ldots, 2^j, \ldots$ until at some step s_{max} we come to the 'good' case $1^{j_{max}}$.

This iteration process cannot be too long. Indeed, after $s = D\sqrt{\log \frac{N}{f(N)}}$ steps of the iteration (for large enough D) we get a contradiction with the inequality

$$K(w^s|y^s) \le K(w^s|x_1, y^s) + K(w^s|x_2, y^s) + I(x_1 : x_2|y^s) + \mathcal{O}(\log N)$$

(it is easy to check that this inequality holds for all words, see e.g., the proof of inequality (6) in [8]): the value on the left-hand side of the inequality is at least $DN/2$, and the right-hand side is only

$$2f_s(N) + I(x_1 : x_2|y_s) + \mathcal{O}(\log N) \ll N.$$

Remark: In all the arguments above we ignore additive terms of the order $\mathcal{O}(\log K(y^s, w^s))$ *because* $\log K(y^s, w^s) \ll f(N)$*. This bound is valid since* $K(y^s), K(w^s) < N^2$ *for* $s \ll \log N$*.*

Thus, after several iterations of **Case** 2^s, for some $s_{max} < D\sqrt{\log \frac{N}{f(N)}}$ we get **Case** $1^{s_{max}}$. We obtain some word z such that

$$K(z) \ge \alpha + s_{max}\delta/2 - \mathcal{O}(f_{s_{max}}(N)) > \alpha - g(N)$$

and

$$K(z|x_i) \le 2\delta + f_{s_{max}} < 2\delta + 3^{D\sqrt{\log \frac{N}{f(N)}}} f(N) < g(N) \ (i = 1, 2).$$

In other words, at least α bits of common information can be extracted from the words x_i for the precision threshold $g(N)$.

4 Conclusion

We cannot prove Conjecture 1 in the general case. However we know that it is true for *stochastic* pairs $\langle x_1, x_2 \rangle$.

Definition 5. *A tuple $\bar{x}$ is called (α, β)-stochastic if there exists a finite set $A \ni \bar{x}$ such that (a) complexity of the list of all elements of A (in lexicographical order) is at most α, and (b) $K(\bar{x}|[\text{list of all elements of } A]) \geq \log |A| - \beta$ (c.f. the definition of (α, β)-stochastic sequences [4]).*

In most applications of Kolmogorov complexity all tuples under consideration are (α, β)-stochastic with logarithmic α, β. For stochastic tuples Conjecture 1 is true:

Theorem 3. *For every integer $C_1 > 0$ there exists an integer $C_2 > 0$ such that for all y and all $(C_1 \log N, C_1 \log N)$-stochastic $\bar{x} = \langle x_1, x_2 \rangle$ if $I(y : \bar{x}) \leq C_1 \log N$ and*

$$\exists w : \ K(w|y) \geq \alpha, \ K(w|x_i, y) \leq C_1 \log N \ (i = 1, 2), \ \textit{where } N = K(\bar{x}, y),$$

then $\exists z : \ K(z) \geq \alpha, \ K(z|x_i) \leq C_2 \log N \ (i = 1, 2).$

(We skip the proof due to the lack of space).

Thus, Conjecture 1 is still an open problem. Also there is another interesting question: Does any counterpart of the results above hold for infinite oracles ?

References

1. Zvonkin, A.K., Levin, L.A.: The Complexity of Finite Objects and the Algorithmic Concepts of Information and Randomness. Russian Math. Surveys 25(6), 83-124 (1970)
2. Gács, P., Körner, J.: Common Information Is far Less Than Mutual Information. Problems of Control and Information Theory, 2, 49-62 (1973)
3. Ahlswede, R., Körner, J.: On Common Information and Related Characteristics of Correlated Information Sources. Presented at the 7th Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand. Proc. (1974)
4. Li, M., Vitányi, P.: An introduction to Kolmogorov complexity and its applications. 2nd ed., Springer-Verlag, New York (1997)
5. Zhang, Z., Yeung, R.W.: On Characterization of Entropy Functions via Information Inequalities. IEEE Trans. on Information Theory, 44 1440-1452 (1998)
6. Muchnik, An.A: On Common Information. Theoretical Computer Science, 207, 319–328 (1998)
7. Romashchenko A.: Pairs of Words with Nonmaterializable Mutual Information. Problems of Information Transmission. 36:1, 1-18 (2000)
8. Hammer, D., Romashchenko, A., Shen, A., Vereshchagin, N.: Inequalities for Shannon Entropy and Kolmogorov Complexity. Journal of Computer and System Sciences. 60, 442–464 (2000)
9. Makarychev, K., Makarychev, Yu., Romashchenko, A., Vereshchagin, N., A New Class of non Shannon Type Inequalities for Entropies. Communications in Information and Systems. 2:2, 147-166 (2002)
10. Romashchenko, A. Shen, A. Vereshchagin, N.: Combinatorial Interpretation of Kolmogorov Complexity. Theoretical Computer Science. 271, 111–123 (2002)
11. Chernov, A., Muchnik, An.A., Shen, A., Romashchenko, A., Vereshchagin, N.K.: Upper Semi-Lattice of Binary Strings with the Relation "x Is Simple Conditional to y" Theoretical Computer Science. 271, 69–95 (2002)
12. Romashchenko, A.: Extracting the Mutual Information for a Triple of Binary Strings. Proc. 18th Annual IEEE Conference on Computational Complexity (2003).

5 Appendix

Proof of Lemma 4: First of all, let us fix an algorithm that gets integers α, β, γ as an input, and enumerates the list of *all* (α, β, γ)-semi-bunches. We call this algorithm the *complete enumerator*. Though the number of semi-bunches (for given parameters) is finite, the complete enumerator never stops. We cannot decide effectively if it has already found *all* semi-bunches. We only guarantee that each semi-bunch must be enumerated in the list, soon or late.

Now we describe another enumerator, which chooses some subsequence from the complete enumeration of all semi-bunches as follows. The complete enumerator prints semi-bunches one by one, and we need to select some of them. Assume some semi-bunches $U_0, \ldots, U_s$ are already selected, and the complete enumerator finds a new semi-bunch V. If $|V \cap U_i| < 2^{\beta-\epsilon}$ for all $i = 0, \ldots, s$, where $\epsilon = 2(\beta - \alpha + 2)$, then we *select* this semi-bunch and let $U_{s+1} = V$. Otherwise we skip V and wait for the next item from the complete enumeration.

Let $U_0, \ldots, U_q$ be the list of all selected semi-bunches for given α, β, γ. From the construction it is evident that for every semi-bunch V either $V = U_i$ or at least $|V \cap U_i| \geq 2^{\beta-\epsilon}$ for some $i \leq q$. Also it follows from the construction that $|U_i \cap U_j| < 2^{\beta-\varepsilon}$ for every two different *selected* semi-bunches U_i, U_j. It remains to prove that q is not too large.

In fact it is enough to prove that every x belongs to less than $2^{\beta-\alpha+2}$ selected semi-bunches. Indeed, there are less than 2^γ words x such that $K(x) < \gamma$. If every x belongs to at most $2^{\beta-\alpha+2}$ selected semi-bunches, and every semi-bunch U_i contains 2^α words then the number of all selected semi-bunches is bounded by

$$\frac{2^\gamma \cdot 2^{\beta-\alpha+2}}{2^\alpha} = 2^{\beta+\gamma-2\alpha+2}$$

Thus, it remains to bound the number of selected semi-bunches that contain one fixed word x.

Assume that there exist $N = 2^{\beta-\alpha+2}$ different selected semi-bunches U_i that contain the same word x. Denote

$$U_i' = U_i \cap \{y \mid K(y|x) < \beta\}$$

for all these semi-bunches U_i. From the definition of a semi-bunch it follows that U_i' contains at least $2^{\alpha-1}$ elements.

On one hand, we have

$$\|\bigcup U_i'\| \leq \|\{y \mid K(y|x) < \beta\}\| < 2^\beta$$

On the other hand,

$$\|\bigcup U_i'\| \geq \sum_i \|U_i'\| - \sum_{i<j} \|U_i' \cap U_j'\|$$

As $\|U_i'\| \geq 2^{\alpha-1}$ and $\|U_i' \cap U_j'\| \leq \|U_i \cap U_j\| \leq 2^{\beta-\varepsilon}$, it follows that

$$\|\bigcup U_i'\| \geq N \cdot 2^{\alpha-1} - N^2 \cdot 2^{\beta-\epsilon} = 2^\beta$$

and we get a contradiction. The lemma is proven.