Some properties of antistochastic strings

Alexey Milovanov
Moscow State University
almas239@gmail.com

June 27, 2016

Abstract

Antistochastic strings are those strings that do not have any reasonable statistical explanation. We establish the follow property of such strings: every antistochastic string x is "holographic" in the sense that it can be restored by a short program from any of its part whose length equals the Kolmogorov complexity of x. Further we will show how it can be used for list decoding from erasing and prove that Symmetry of Information fails for total conditional complexity.

Keywords: Kolmogorov complexity, algorithmic statistics, stochastic strings, total conditional complexity, Symmetry of Information.

1 Introduction

Algorithmic statistics studies explanations of observed data that are good in the algorithmic sense: an explanation should capture all the algorthmically discoverable regularities in the data. The data is encoded, say, by a string x over a binary alphabet $\{0, 1\}$. In this paper we consider explanations that are statistical hypotheses of the form "x was drawn at random from a finite set A with uniform distribution". (As argued in [15] the class of general probability distributions reduces to the class of uniform distributions over finite sets.)

As an option, Kolmogorov suggested in 1974 [5] to measure the quality of an explanation $A \ni x$ by two parameters, Kolmogorov complexity $C(A)$

of A (the explanation should be simple) and the cardinality $|A|$ of A (the smaller $|A|$ is the more "exact" explanation is). Both parameters cannot be very small simultaneously unless the string x has very small Kolmogorov complexity. Indeed, $C(A) + \log_2 |A| \geqslant C(x)$ (up to $O(\log(l(x)))$), since x can be specified by A and its index in A. Kolmogorov called an explanation $A \ni x$ good if $C(A) \approx 0$ and $\log_2 |A| \approx C(x)$, that is, $\log_2 |A|$ is as small as the inequality $C(A) + \log_2 |A| \geqslant C(x)$ permits given that $C(A) \approx 0$. He called a string *stochastic* if it has such an explanation.

Every string x of length n has two trivial explanations: $A_1 = \{x\}$ and $A_2 = \{0,1\}^n$. The first explanation is good when the complexity of x is small. The second one is good when the string x is random, that is, its complexity $C(x)$ is close to n. Otherwise, when $C(x)$ is far from both 0 and n, both explanations are bad.

Informally, non-stochastic strings are those having no good explanation and antistochastic strings are extreme case of non-stochastic strings (a strict definition will be done in the third section). They were studied in [3, 15]. To define non-stochasticity rigorously we have to introduce the notion of the profile of x, which represents the parameters of possible explanations for x.

Definition 1. The *profile* of a string x is the set P_x consisting of all pairs (m, l) of natural numbers such that there is a finite set $A \ni x$ with $C(A) \leqslant m$ and $\log_2 |A| \leqslant l$.

On the Fig. 1, it is shown how the profile of a string x of length n and complexity k may look like.

The profile of every string x of length n and complexity k has the following three properties. First, P_x is upward closed: if P_x has a pair (m, l) then P_x contains all the pairs (m', l') with $m' \geqslant m$ and $l' \geqslant l$. Second, P_x contains the set

$$P_{\min} = \{(m, l) \mid m + l \geqslant n \text{ or } m \geqslant k\} \quad (1)$$

(the set consisting of all pairs above and to the right of the dashed line on Fig. 1) and is included into the set

$$P_{\max} = \{(m, l) \mid m + l \geqslant k\} \quad (2)$$

(the set consisting of all pairs above and to the right of the dotted line on Fig. 1). In other words, the border line of P_x, called by Kolmogorov the *structure function* of x, lies between the dotted line and the dashed line.

This was a rough formulation of the second property. The accurate statement is the following. For some function $\varepsilon = O(\log n)$ the set $P_{\min}$ is included

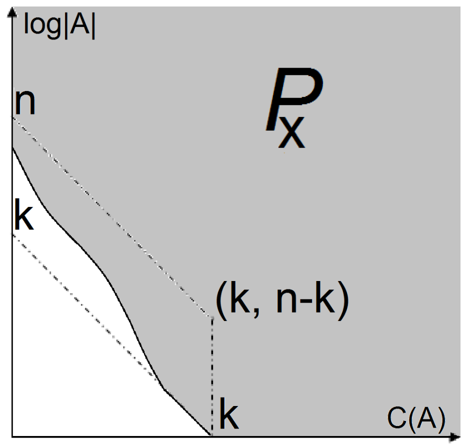

Figure 1: The profile P_x of a string x of length n and complexity k

in the ε-neighborhood of the set P_x, which is included in the ε-neighborhood of the set $P_{\max}$. Speaking about neighborhoods we refer to l_1-metrics on the plane.

And finally, P_x has the following property:

$$\begin{array}{c}\text{if a pair } (m,l) \text{ is in } P_x \text{ then for all } i \leqslant l \\ \text{the pair } (m+i+O(\log l(x)), l-i) \text{ is in } P_x.\end{array} \tag{3}$$

The notion of the profile was introduced by Kolmogorov in [5] and he established these properties.

If for some strings x and y $P_x \subset P_y$ then y is more stochastic then x. The largest possible profile is close to the set $P_{\max}$. Such a profile is possessed, for instance, by a random string of length k appended by $n-k$ zeros. The smaller the set P_x is, the more non-stochastic the sting x is.

The paper [15] shows that every profile that has the above three properties is realizable by a string of length n and complexity $k+O(\log n)$, with certain accuracy:

Theorem 1 ([15])**.** *Assume that we are given an upward closed set P of pairs of natural numbers which includes P_{min} and is included into P_{max} and for all $(m,l) \in P$ and all $i \leqslant l$ we have $(m+i, l-i) \in P$. Then there is a string x of length n and complexity $k+O(\log n)$ whose profile is at most $C(P)+O(\log n)$-close to P.*

In this theorem, we call subsets of $\mathbb{N}^2$ *ε-close* if each of them is in the ε-neighborhood of the other.

Kolmogorov complexity $C(P)$ of the set P is defined as follows. Any set P of pairs of naturals as in Theorem 1 is completely determined by the function $h(l) = \min\{m \mid (m, l) \in P\}$. This function has only finitely many non-zero values, as $h(k) = h(k+1) = \cdots = 0$. Hence h is a finite object and we let $C(P)$ be equal to the Kolmogorov complexity of h.

For the set $P_{\min}$ the function h satisfies $h(m) = n - m$ for $m < k$ and $h(k) = h(k+1) = \cdots = 0$. Thus the Kolmogorov complexity of this set is $O(\log n)$. Hence there is a string x of length about n and complexity about k whose profile P_x is close to the set $P_{\min}$. We call such strings *antistochastic*.

In this paper we show that antistochastic strings have the following property:

Assume that we replace in an antistochastic string of length n and complexity k an arbitrary set of $n - k$ bits by the "blank" symbol. Then the original string can be restored from the resulting string by a short program (Theorem 6). We call this property the "holographic property" of antistochastic strings.

We will use this property to prove the following propositions:

- There are about 2^k "holographic" strings of length n and complexity k and thus they form a binary code of dimension k which is list decodable from $n - k$ erasures with a list of size poly(n) (Theorem 8).

- If y is an anti-stochastic string of length $2k$ and complexity k and x is its first half, then the total complexity of y conditional to x is about k while the plain complexity of y conditional to x is negligible (Theorem **??**). Thus non-stochastic strings provide a new natural example of a pair of strings when total conditional complexity is much less than plain conditional complexity. As the total and plain complexities of x conditional to y coincide (both are negligible), we get a new natural example of asymmetry of information for total conditional complexity.

2 Preliminaries

We consider strings over the binary alphabet $\{0, 1\}$. The set of all strings is denoted by $\{0, 1\}^*$ and the length of a string x by $l(x)$. The empty string is denoted by Λ.

Let D be a partial computable function mapping pairs of strings to strings. Conditional Kolmogorov complexity with respect to D is defined as

$$C_D(x|y) = \min\{l(p) \mid D(p, y) = x\}.$$

In this context the function D is called a *description mode* or a *decompressor*. If $D(p, y) = x$ then p is called a *description* of x conditional to y or a *program* mapping y to x.

A decompressor D is called *universal* if for any other decompressor D' there is a string c such that $D'(cp, y) = D(p, y)$ for all p, y. By Solomonoff—Kolmogorov theorem universal decompressors exist. We pick any universal decompressor D and call $C_D(x|y)$ *the Kolmogorov complexity* of x conditional to y and denote it by $C(x|y)$. Then we define the plain Kolmogorov complexity $C(x)$ of x as $C(x|\Lambda)$

Total conditional complexity is defined as the shortest length of a total program p mapping b to a: $CT(a|b) = \min\{l(p) \mid D(p, b) = a$ and $D(p, y)$ is defined for all $y\}$. Obviously $CT(a|\Lambda) = C(a)+O(1)$ while in general $CT(a|b)$ may be much greater than $C(a|b)$ (such examples are presented below).

Kolmogorov complexity of other finite objects is defined using a computable 1-1 correspondence between those objects and strings. For instance, fix any computable 1-1 correspondence between $\{0, 1\}^*$ and the family of finite subsets of $\{0, 1\}^*$. The string that corresponds to a finite $A \subset \{0, 1\}^*$ is denoted by $[A]$ and is called the *code* of A. Its complexity $C([A])$ is abbreviated to $C(A)$. In the same way we understand the notations $C(x|A)$ and $C(A|x)$.

For properties of Kolmogorov complexity we refer to textbooks [7] or [13]. Here we present only one property established by Kolmogorov and Levin:

Theorem 2 (Symmetry of Information). *For all strings x, y of complexity at most k it holds $C(x) - C(x|y) = C(y) - C(y|x) + O(\log k)$.*

3 Antistochastic strings and their properties

Definition 2. A string x of length n and complexity k is called *ε-antistochastic* if for all $(m, l) \in P_x$ either $m > k - \varepsilon$, or $m + l > n - \varepsilon$.

Notice that ε-antistochasticity implies that P_x is in an ε-neighborhood of the set $P_{\min}$ from Equation (**??**) and the latter implies that x is 2ε-antistochastic.

By Theorem 1 there are ε-antistochastic strings of each length n and complexity $k \leqslant n$ where $\varepsilon = O(\log n)$. More specifically, Theorem 1 has the following consequence

Corollary 3. *For all n and all $k \leqslant n$ there is an $O(\log n)$-antistochastic string x of length n and complexity $k + O(\log n)$.*

This corollary can be proved more easily than the more general Theorem 1. For the sake of completeness we present the proof.

Proof. We first formulate a sufficient condition for antistochasticity.

Lemma 4. *If the profile of a string x of length n and complexity k does not contain the pair $(k - \varepsilon, n - k)$ then x is $\varepsilon + O(\log n)$-antistochastic.*

Notice that the condition of this lemma is implied by the definition of ε-antistochasticity. So, basically Lemma 3 provides a re-formulation of ε-antistochasticity.

Proof. Assume that a pair (m, l) is in the profile of x. We will show that either $m > k - \varepsilon$ or $m + l > n - \varepsilon - O(\log n)$. Assume that $m \leqslant k - \varepsilon$ and hence $l > n - k$. By the third property of profiles we see that the pair

$$(m + (l - (n - k)) + O(\log n), n - k)$$

is in its profile as well. Hence we have

$$m + l - (n - k) + O(\log n) > k - \varepsilon$$

and

$$m + l > n - \varepsilon - O(\log n). \quad \square$$

Consider the family $\mathcal{A}$ consisting of all finite sets A of complexity less than k and log-cardinality at most $n - k$. The number of such sets is less than 2^k and thus the total number of strings in all such sets is less than $2^k 2^{n-k} = 2^n$. Hence there is a string of length n that does not belong to any of those sets. Let x be the lexicographically least such string.

Let us show that the complexity of x is $k + O(\log n)$. It is at least $k - O(1)$, as by construction the singleton $\{x\}$ has complexity at least k. On the other hand, the complexity of x is at most $\log |\mathcal{A}| + O(\log n) \leqslant k + O(\log n)$. Indeed, the list of $\mathcal{A}$ can be found from k, n and $|\mathcal{A}|$, as we can enumerate $\mathcal{A}$ until we get $|\mathcal{A}|$ sets.

By construction x satisfies the condition of the Lemma 3 with $\varepsilon = O(\log n)$. Hence x is $O(\log n)$-antistochastic. $\square$

For any integer i let Ω_i denote the number of strings of complexity at most i. As we can compute from Ω_k and k a string of Kolmogorov complexity more than k, we have $C(\Omega_k) = k + O(\log k)$. If $l \leqslant m$ then the leading l bits of Ω_m contain the same information as Ω_l [15, Theorem VIII.2] and [13, Problem 367]:

Lemma 5. *Assume that $l \leqslant m$ and let $(\Omega_m)_{1:l}$ denote the leading l bits of Ω_m. Then both $C((\Omega_m)_{1:l}|\Omega_l)$ and $C(\Omega_l|(\Omega_m)_{1:l})$ are of order $O(\log m)$.*

Every antistochastic string of x complexity $k < l(x) - O(\log l(x))$ contains the same information as Ω_k:

Lemma 6. *There exists a function $f(n) = O(\log n)$ such that the following holds. Let x be an ε-antistochastic string of length n and complexity $k < n - \varepsilon - f(n)$. Then both $C(\Omega_k|x)$ and $C(x|\Omega_k)$ are less than $\varepsilon + f(n)$.*

Actually this lemma is true for all strings whose profile P_x does not contain the pair $(k - \varepsilon + O(\log k), \varepsilon)$, in which form it was essentially proved in [3]. The lemma goes back to L. Levin (personal communication, see [15] for details).

Proof. Fix an algorithm that given any k enumerates all strings of complexity at most k. Let N denote the number of strings that appear after x in the enumeration of all strings of complexity at most k (N can be equal 0).

Given x, k and N we can find Ω_k just by waiting until N strings have been enumerated after x. Let $l = \lceil \log N \rceil$. We claim that $l \leqslant \varepsilon + O(\log n)$. Indeed, chop the set of all strings enumerated into portions of size 2^l. The last portion might be incomplete, however x does not fall in that portion. Every complete portion can be described by its number and k. The total number of complete portions is less than $2^k/2^l$. Thus the profile P_x contains the pair $(k - l + O(\log k), l)$. By antistochasticity of x, we have $k - l + O(\log k) \geqslant k - \varepsilon$ or $k - l + O(\log k) + l \geqslant n - \varepsilon$. The former inequality implies that $l \leqslant \varepsilon + O(\log k)$. The latter inequality cannot happen provided the function $f(n)$ in the condition of the theorem is large enough.

We have shown that $C(\Omega_k|x) < \varepsilon + O(\log k)$. By Symmetry of Information this implies that $C(x|\Omega_k) < \varepsilon + O(\log n)$ as well. Indeed,

$$C(x) + C(\Omega_k|x) = C(x|\Omega_k) + C(\Omega_k) + O(\log k).$$

The strings x and Ω_k have the same complexity with logarithmic accuracy hence $C(\Omega_k|x) = C(x|\Omega_k)$, also with logarithmic accuracy. □

3.1 A "holographic" property of antistochastic strings

Every antistochastic string x of length n and complexity k can be restored from its first k bits using an auxiliary logarithmic amount of information. Indeed, let A consist of all strings of the same length as x and having the same k first bits as x. The complexity of A is at most $k + O(\log n)$. On the other hand, its complexity is at least $k - O(\log n)$ as the profile of x contains the pair $(C(A), n - k)$. Since $C(A|x) = O(\log n)$, by Symmetry of Information we have $C(x|A) = O(\log n)$ as well.

The same arguments work for every simple k-element subset of indices: if I be a k-element subset of $\{1, \dots, n\}$ and $C(I) = O(\log n)$ then x can be restored from x_I and some auxiliary logarithmic amount of information. Here x_I denotes the string obtained from x by replacing all the symbols with indices outside I by the blank symbol (a fixed symbol, different from 0,1). Surprisingly, this is true for *every* k-element subset of indices, even if that subset be complex: $C(x|x_I) = O(\log n)$. The following theorem provides an even more general formulation of this property.

Theorem 7. *Let x be an ε-antistochastic string of length n and complexity k. Assume that $x \in A$ and $|A| \leqslant 2^{n-k}$. Then $C(x|A) \leqslant 2\varepsilon + O(\log C(A) + \log n)$.*

For instance, let I is a k-element set of indexes and A be the set of all strings of length n that coincide with x on I. Then A can be described in $2n$ bits and hence $C(x|A) \leqslant 2\varepsilon + O(\log n)$.

Proof. W.l.o.g. we may assume that $k < n - \varepsilon - f(n)$ where $f(n) = O(\log n)$ is the function from Lemma 5. Indeed, otherwise A is so small that x can be just identified by its index in A in $\varepsilon + f(n)$ bits. Thus by Lemma 5 both $C(\Omega_k|x)$ and $C(\Omega_k|x)$ are less than $\varepsilon + O(\log n)$.

In all the inequalities below we will ignore additive terms of order $O(\log C(A) + \log n)$. However, we will not ignore additive terms ε. We hope that the exact meaning of the inequalities be clear.

Run the algorithm that enumerates all finite sets of complexity at most $C(A)$. Let N denote the index of the code of A in that enumeration. Let m denote the number of common leading bits of the binary notations of N and $\Omega_{C(A)}$ and l the number of remaining bits. That is, $N = a2^l + b$ and $\Omega_{C(A)} = a2^l + c$ for some integer $a < 2^m$ and $b, c < 2^l$. Thus $l + m$ is equal to the length of the binary notation of $\Omega_{C(A)}$, which is $C(A) + O(1)$. Let us distinguish two cases.

Case 1: $m \geqslant k$. In this case we will use the inequality $C(x|\Omega_k) \leqslant \varepsilon$. The number Ω_k can be retrieved from Ω_m and the latter can be found from m leading bits of $\Omega_{C(A)}$. Finally m leading bits of $\Omega_{C(A)}$ can be found from A as m leading bits of the index N of the code of A in the enumeration of all strings of complexity at most $C(A)$.

Case 2: $m < k$. This case is more elaborated and we need an additional construction.

Lemma 8. *The pair* $(m, l + n - k - C(A|x) - \varepsilon)$ *belongs to* P_x.

Proof. We have to construct a set $B \ni x$ of complexity m and log-size $l + n - k - C(A|x) - \varepsilon$. It is constructed in two steps.

First step. On this step we construct a family $\mathcal{A}$ of sets such that $A \in \mathcal{A}$ and $C(\mathcal{A}) \leqslant m$, $C(\mathcal{A}|x) \leqslant \varepsilon$ and $|\mathcal{A}| \leqslant 2^l$. To this end chop all strings of complexity at most $C(A)$ in chunks of size 2^l in the order they were enumerated. The last chunk may be incomplete, however the code of A does not fall into the last chunk: it belongs to the last complete chunk.

Let $\mathcal{A}$ stand for the family of those finite sets whose code belongs the chunk containing the code of A and log-cardinality at most $n - k$. By construction $|\mathcal{A}| \leqslant 2^l$. Since $\mathcal{A}$ can be found from a as the ath chunk, we have $C(\mathcal{A}) \leqslant m$. To prove that $C(\mathcal{A}|x) \leqslant \varepsilon$ it suffices to show that $C(a|x) \leqslant \varepsilon$. We have $C(\Omega_k|x) \leqslant \varepsilon$ and from Ω_k we can find Ω_m and hence the number a as the m leading bits of $\Omega_{C(A)}$ (Lemma 4).

Second step. We claim that x appears in at least $2^{C(A|x)}$ sets from $\mathcal{A}$. Indeed, assume that x falls in K of them. Given x, we can describe A by its index in $\mathcal{A}$ and about ε bits of additional information to describe $\mathcal{A}$. This implies $C(A|x) \leqslant \log K + \varepsilon$.

Let B be the set of x' that appear in at least $2^{C(A|x)-\varepsilon}$ of sets from $\mathcal{A}$. As shown, x belongs to B. As B can be found from $\mathcal{A}$ we have $C(B) \leqslant m$. It remains to estimate the cardinality of B. The total number of strings in all sets from $\mathcal{A}$ is at most 2^{l+n-k}, counting multiplicities. Thus B has at most $2^{l+n-k-C(A|x)+\varepsilon}$ strings. □

By the lemma either $m \geqslant k - \varepsilon$, or $m + l + n - k - C(A|x) + \varepsilon \geqslant n - \varepsilon$. In the case $m \geqslant k - \varepsilon$ we can just repeat the arguments from Case 1 and show that $C(x|A) \leqslant 2\varepsilon$.

In the case $m+l+n-k-C(A|x)+\varepsilon \geqslant n-\varepsilon$ we recall that $m+l = C(A)$ and by Symmetry of Information $C(A)-C(A|x) = C(x)-C(x|A) = k-C(x|A)$.

Thus we have

$$n - C(x|A) + \varepsilon \geqslant n - \varepsilon. \quad \square$$

Remark 1. Notice that every string with property of Theorem 6 is antistochastic. Indeed, if x is not ε-antistochastic for a large ε, then it belongs to a set A that has 2^{n-k} elements and whose complexity is less than $k-\varepsilon+O(\log n)$ (Lemma 3). Then $C(x|A)$ is large, since

$$k = C(x) \leqslant C(x|A) + C(A) + O(\log n) \leqslant C(x|A) + k - \varepsilon + O(\log n)$$

and hence $C(x|A) \geqslant \varepsilon - O(\log k)$.

3.2 Antistochastic strings and list decoding from erasures

Definition 3. A string x of length n is called ε, k-holographic if for all k-element set of indexes $I \subset \{1, \dots, n\}$ we have $C(x|x_I) < \varepsilon$.

Theorem 9. *For all n and all $k \leqslant n$ there are at least $2^{k-O(\log n)}$ $O(\log n), k$-holographic strings of length n.*

Proof. By Corollary 2 and Theorem 6 for all n and $k \leqslant n$ there is a ε, k-holographic string x of length n and complexity $k+O(\log n)$, where ε denotes a function of n of order $O(\log n)$. This implies that there are many of them. Indeed, the set of all ε, k-holographic strings of length n can be identified by n and k. More specifically, given n and k we can enumerate all ε, k-holographic strings and hence x can be identified by k, n and its ordinal number in that enumeration. As the complexity of x is at least $k-O(\log n)$, we can conclude the logarithm of that number is at least $k - O(\log n)$. $\square$

Theorem 10. *For every m, n with $n \geqslant m$ and for every string x of length m there is a string y of length n such that $C(x|y_I) = O(\log n)$ for every m-element sets of indexes I.*

Proof. Set $k = m + O(\log n)$ and $\varepsilon = O(\log n)$ so that the number of ε, k-holographic strings of length n be 2^m or more. Then start an enumeration of ε, k-holographic strings of length n and number them by strings of length m until we enumerate 2^m holographic strings. Let y_x stand for the ε, k-holographic strings corresponding to the string x of length m. Then $C(x|y) = O(\log n)$ and hence $C(x|y_J) = O(\log n)$ for any k-element set of indexes J.

It remains to notice that every m-element set of index I can be enlarged in a standard way to a k-element set of indexes J so that $C(y_J|y_I) = O(\log n)$. Hence $C(x|y_I) \leqslant C(x|y_J) + C(y_J|y_I) + O(\log n) = O(\log n)$. $\square$

Theorem **??** provides a way to define codes that are list decodable from erasures. Indeed, consider the string y existing by Theorem **??** as a n-bit code for the string x. In this way we obtain a binary code with dimension k and code-length n. This code is list decodable from at most $n - k$ erasures with list size $2^{O(\log n)} = \text{poly}(n)$. Indeed, if an adversary erases at most $n - k$ bits of a code-word y then x can be reconstructed from the resulting strings $\tilde{y}$ (containing zeros, ones and blanks) by a program of length $O(\log n)$. Applying all programs of that size to $\tilde{y}$, we obtain a list of size $\text{poly}(n)$ which contains x.

Although the existence of list decodable codes with such parameters can be established by the probabilistic method [4, Theorem 10.9 on p. 258], we find it interesting that a seemingly unrelated notion of antistochasticity provides such codes.

3.3 Antistochastic strings and total conditional complexity

Total conditional complexity is defined as the shortest length of a total program p mapping b to a: $CT(a|b) = \min\{l(p) \mid D(p,b) = a \text{ and } D(p,y) \text{ is defined for all } y\}$.

The existence of strings where total conditional complexity differs, is attributed in [11] to other places.

The paper [14] shows that there is a string x and its shortest program x^* such that $CT(x|x^*)$ is large (linear in the length of x) while $CT(x^*|x)$ is negligible (of order $O(\log C(x))$. Notice that both plain conditional complexities $C(x|x^*)$ and $C(x^*|x)$ are negligible as well.

Here we show that absolutely antistochastic string provide another example of strings x and y such that all $CT(x|y)$, $C(x|y)$ and $C(y|x)$ are negligible while $CT(y|x)$ is large.

Theorem 11. *For all k there is a string x of length k and a string y of length $2k$ with $C(x) = C(y) + O(\log k) = k + O(\log k)$, $CT(x|y) = O(1)$ (and hence $C(x|y) = O(1)$), $C(y|x) = O(\log k)$ while $CT(y|x) = k + O(\log k)$.*

Proof. Let y be an $O(\log k)$-antistochastic string for length $2k$ and complexity $k+O(\log k)$ existing by Lemma 5. Let x consist of the first k bits of y. Then $C(x) = k + O(\log k)$ and $CT(x|y) = O(1)$.

It suffices to show that $CT(y|x) \geqslant k - O(\log k)$. Let p witness $CT(y|x)$. Consider the set $A = \{D(p, b) \mid b \in \{0,1\}^k\}$. This set witnesses that the profile of y contains the pair $(l(p) + O(\log k), k)$. Therefore either $l(p) + O(\log k) \geqslant k - O(\log k)$ or $l(p) + O(\log k) + k \geqslant 2k - O(\log k)$. In both cases we are done. □

Remark 2. This example, as well as the example from [14], shows that for total conditional complexity the Symmetry of Information (Theorem **??**) does not hold. Indeed, let $CT(a) = CT(a|\Lambda) = C(a) + O(1)$. Then $CT(x) - CT(x|y) > CT(y) - CT(y|x) + k - O(\log k)$ for strings x, y from Theorem **??**.

A big question in time-bounded Kolmogorov complexity is whether the Symmetry of information (Theorem **??**) holds for time-bounded Kolmogorov complexity. Partial answers to this question were obtained in [8, 9, 6].

Total conditional complexity $CT(b|a)$ is defined as the shortest length of a total program p mapping b to a. Being total that program halts on all inputs in time bounded by a total computable function f_p of its input. Thus total conditional complexity may be viewed as a variant of time bounded conditional complexity. Let us stress that the upper bound f_p for time may depend (and does depend) on p in a non-computable way. Thus $CT(b|a)$ is a rather far approximation to time bounded Kolmogorov complexity.

Acknowledgments

I would like to thank Alexander Shen and Nikolay Vereshchagin for useful discussions, advises and remarks.

References

[1] P. Gács, J. Tromp, P.M.B. Vitányi. Algorithmic statistics, *IEEE Trans. Inform. Th.*, 47:6 (2001), 2443–2463.

[2] A.N. Kolmogorov, Talk at the Information Theory Symposium in Tallinn, Estonia, 1974.

[3] Li M., Vitányi P., *An Introduction to Kolmogorov complexity and its applications*, 3rd ed., Springer, 2008 (1 ed., 1993; 2 ed., 1997), xxiii+790 pp. ISBN 978-0-387-49820-1.

[4] A. Shen, V. Uspensky, N. Vereshchagin *Kolmogorov complexity and algorithmic randomness*. MCCME, 2013 (Russian). English translation: http://www.lirmm.fr/~ashen/kolmbook-eng.pdf

[5] Vekatesan Guruswami *List decoding of error-correcting codes: winning thesis of the 2002 ACM doctoral dissertation competition, Springer, 2004*

[6] A. Shen *The concept of* (α, β)*-stochasticity in the Kolmogorov sense, and its properties. Soviet Mathematics Doklady, 271(1):295–299, 1983*

[7] N. Vereshchagin and P. Vitányi *"Kolmogorov's Structure Functions with an Application to the Foundations of Model Selection" . IEEE Transactions on Information Theory 50:12 (2004) 3265-3290. Preliminary version: Proc. 47th IEEE Symp. Found. Comput. Sci., 2002, 751–760.*

[8] L. Longpré and S. Mocas *Symmetry of information and one-way functions. Information Processing Letters, 46(2):95–100, 1993.*

[9] L. Longpré and O. Watanabe *On symmetry of information and polynomial time invertibility. Information and Computation, 121(1):1–22, 1995.*

[10] A. Shen, Game Arguments in Computability Theory and Algorithmic Information Theory. Proceedings of CiE 2012, 655–666.

[11] Troy Lee and Andrei Romashchenko *Resource bounded symmetry of information revisited. Theoretical Computer Science, 345(2-3): 386-405 (2005)*

[12] Nikolay Vereshchagin. On Algorithmic Strong Sufficient Statistics.. In: 9th Conference on Computability in Europe, CiE 2013, Milan, Italy, July 1-5, 2013. Proceedings, LNCS 7921, P. 424-433.