Shannon entropy vs. Kolmogorov complexity

An. Muchnik[1] and N. Vereshchagin[2⋆]

[1] Institute of New Technologies, 10 Nizhnyaya Radischewskaya, Moscow, Russia 109004

[2] Department of Mathematical Logic and Theory of Algorithms, Faculty of Mechanics and Mathematics, Moscow State University, Leninskie Gory, Moscow, Russia 119992. E-mail: ver@mccme.ru

Abstract. Most assertions involving Shannon entropy have their Kolmogorov complexity counterparts. A general theorem of Romashchenko [4] states that every information inequality that is valid in Shannon's theory is also valid in Kolmogorov's theory, and vice verse. In this paper we prove that this is no longer true for $\forall\exists$-assertions, exhibiting the first example where the formal analogy between Shannon entropy and Kolmogorov complexity fails.

1 Introduction

Since the very beginning the notion of complexity of finite objects was considered as an algorithmic counterpart to the notion of Shannon entropy [9]. Kolmogorov's paper [6] was called "Three approaches to the quantitative definition of information"; Shannon entropy and algorithmic complexity were among these approaches. Let us recall the main definitions.

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

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

(All logarithms in the paper are base 2.) Using the concavity of the function $p \mapsto -p \log p$, one can prove that the Shannon entropy of every random variable does not exceed its *max-entropy*, $H_0(\alpha)$, defined as the logarithm of the cardinality of the range of α (and is equal to $H_0(\alpha)$ only for uniformly distributed variables).

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

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

⋆ Supported in part by Grants 03-01-00475, 06-01-00122, NSh-358.2003.1 from Russian Federation Basic Research Fund.

It is easy to check that

$$H(\langle\alpha,\beta\rangle) \;=\; H(\beta)+H(\alpha|\beta). \tag{1}$$

Using the concavity of logarithm function, one can prove that

$$H(\alpha|\beta) \;\le\; H(\alpha), \tag{2}$$

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

$$H(\langle\alpha,\beta\rangle) \;\le\; H(\alpha)+H(\beta). \tag{3}$$

All these notions have their counterparts in Kolmogorov complexity theory.

Roughly speaking, the Kolmogorov complexity of a binary string a is defined as the minimal length of a program that generates a; the conditional complexity $K(a|b)$ of a conditional to b is the minimal length of a program that produces a having b as input. There are different refinements of this idea (called *simple* Kolmogorov complexity, *monotone* complexity, *prefix* complexity, *decision* complexity, see [5], [11]). However, for our purposes the difference is not important, since all these complexity measures differ only by $O(\log n)$ where n is the length of a.

Now we define these notions rigorously. A *conditional description method* is a partial computable function F mapping pairs of binary strings to binary strings. A string p is called a *description of a conditional to b* with respect to F if $F(p,b)=a$. The complexity of a conditional to b with respect to F is defined as the minimal length of a description of a conditional to b with respect to F:

$$K_F(a|b)=\min\{l(p)\mid F(p,b)=a\}.$$

A conditional description method F is called *optimal* if for all other conditional description methods G there is a constant C such that

$$K_F(a|b)\le K_G(a|b)+C$$

for all a,b. The Solomonoff–Kolmogorov theorem [6, 10] (see also the textbook [5]) states that optimal methods exist. We fix an optimal F and define conditional Kolmogorov complexity $K(a|b)$ as $K_F(a|b)$. The (unconditional) Kolmogorov complexity $K(a)$ is defined as Kolmogorov complexity of a conditional to the empty string. Comparing the optimal function F with the function $G(p,b)=p$ we see that Kolmogorov complexity does not exceed the length:

$$K(a)\le l(a)+O(1).$$

Fix a computable injective function $a,b\mapsto[a,b]$ encoding pairs of binary strings by binary strings (different computable encodings lead to complexities of

$K([a,b])$ that differ only by $O(1)$). The inequalities (1), (2), and (3) translate to Kolmogorov complexity as follows

$$K([a,b]) = K(b) + K(a|b) + O(\log n), \tag{4}$$

$$K(a|b) \leq K(a) + O(1), \tag{5}$$

$$K([a,b]) \leq K(a) + K(b) + O(\log n). \tag{6}$$

Here $n = l(x) + l(y)$. The inequalities (5) and (6) are easy. The inequality (4) is easy in one direction:

$$K([a,b]) \leq K(b) + K(a|b) + O(\log n).$$

The inverse inequality is the famous theorem of Kolmogorov and Levin, see [5].

Following this analogy between Shannon entropy and Kolmogorov complexity, Romashchenko proved in [4] that the class of linear inequalities for Shannon entropy coincides with the class of inequalities for Kolmogorov complexity. To state this result rigorously, we introduce the following notation. Let $\alpha_1, \alpha_2, \ldots, \alpha_m$ be random variables having a joint distribution. For a set $A \subset \{1,2,\ldots,m\}$ let α_A denote the tuple $\langle \alpha_i \mid i \in A \rangle$. For instance, $\alpha_{\{1,2,4\}} = \langle \alpha_1, \alpha_2, \alpha_4 \rangle$. Similarly, for a sequence $x_1, \ldots, x_n$ of binary strings x_A denotes $[x_i \mid i \in A]$, for example, $x_{\{1,2,4\}} = [[x_1, x_2], x_4]$.

Theorem 1 (Romashchenko). *If an inequality of the form*

$$\sum_{A,B} \lambda_{A,B} H(\alpha_A|\alpha_B) \leq 0 \tag{7}$$

is true for all random variables $\alpha_1, \ldots, \alpha_m$ then for some function $f(n) = O(\log n)$ the inequality

$$\sum_{A,B} \lambda_{A,B} K(x_A|x_B) \leq f(n) \tag{8}$$

holds for all binary strings $x_1, \ldots, x_m$. Here n stands for the sum of lengths of x_i, the summation is over all subsets A, B of $\{1,2,\ldots,m\}$, and $\lambda_{A,B}$ denote arbitrary real numbers. Conversely, if for some function $f(n) = o(n)$ the inequality (8) *is true for all $x_1, \ldots, x_m$ then* (7) *holds for all $\alpha_1, \ldots, \alpha_m$.*

This theorem shows that all "information inequalities" for Shannon entropy of random variables are true for Kolmogorov complexity of binary strings with logarithmic accuracy, and vice versa. Information inequalities can be considered as universal formulas in a language having $\leq$ as the only predicate symbol and terms of the form $\sum_{A,B} \lambda_{A,B} H(\alpha_A|\alpha_B)$. In this paper we compare Shannon's and Kolmogorov's information theories using $\forall\exists$-formulas in this language. We show that there is $\forall\exists$-formula that is valid in Kolmogorov's theory but is false in Shannon's theory. Then we exhibit another $\forall\exists$-formula that is true in Shannon's theory (assuming that all universal quantifiers range over sequences of independent identically distributed random variables) but is false in Kolmogorov's theory.

2 Relating Shannon entropy and Kolmogorov complexity using $\forall\exists$-formulas

Consider $\forall\exists$-formulas with atomic formulas being OR of ANDs of information inequalities:

$$\forall\alpha_1 \dots \forall\alpha_k \exists\alpha_{k+1} \dots \exists\alpha_{k+l} \bigvee_i \bigwedge_j \sum_{A,B} \lambda^{ij}_{A,B} H(\alpha_A|\alpha_B) \leq 0. \tag{9}$$

Here the summation is over all subsets A, B of $\{1, 2, \dots, k+l\}$, and $\lambda_{A,B}$ denote arbitrary real numbers. This formula expresses in a succinct form the following statement: For all finite sets $A_1, \dots, A_k$ and jointly distributed random variables $\tilde{\alpha}_1, \dots, \tilde{\alpha}_k$ in $A_1, \dots, A_k$ there are finite sets $A_{k+1}, \dots, A_{k+l}$ and jointly distributed random variables $\alpha_1, \alpha_2, \dots, \alpha_{k+l}$ in $A_1, A_2, \dots A_{k+l}$ such that the marginal distribution of $\langle\alpha_1, \dots, \alpha_k\rangle$ is the same as that of $\langle\tilde{\alpha}_1, \dots, \tilde{\alpha}_k\rangle$ and $\bigvee_i \bigwedge_j \sum_{A,B} \lambda^{ij}_{A,B} H(\alpha_A|\alpha_B) \leq 0$. For every such formula consider the corresponding formula for Kolmogorov complexity:

$$\forall x_1 \dots \forall x_k \exists x_{k+1} \dots \exists x_{k+l} \bigvee_i \bigwedge_j \sum_{A,B} \lambda^{ij}_{A,B} K(x_A|x_B) \leq o(n). \tag{10}$$

Here n denotes $l(x_1) + \dots + l(x_k)$. Note that we include in the sum only the length of strings under universal quantifiers. Otherwise, if we included also the length of other strings, the assertion could become much weaker. We could choose $x_{j+1}, \dots, x_{j+m}$ of length much greater than that of $x_1, \dots, x_j$, and the accuracy $o(n)$ might become larger than $K(x_i)$ for $i \leq k$.

Is it true that for all m and $\lambda^{ij}_{A,B}$ Equation (9) holds if and only if Equation (9) holds? The following trivial counter-example shows that this is not the case. Consider the formula:

$$\forall\alpha\exists\beta\ H(\beta) = H(\alpha)/2,\ H(\beta|\alpha) = 0.$$

This statement is false: let α be the random variable with 2 equiprobable outcomes, thus $H(\alpha) = 1$. If $H(\beta|\alpha) = 0$ then β is a function of α and thus $H(\beta)$ can take only values $0, 1$. On the other hand, the similar assertion for Kolmogorov complexity is true:

$$\forall x\exists y\ K(y) = K(x)/2 + O(\log n),\ K(y|x) = O(\log n),$$

where $n = l(x)$. Indeed, as y we can take the first half of the shortest description of x. However, we think that this counter-example is not honest. Indeed, the statement holds for Shannon entropy with accuracy $O(1)$:

$$\forall\alpha\exists\beta\ H(\beta) = H(\alpha)/2 + O(1),\ H(\beta|\alpha) = 0.$$

To prove this, define a sequence $\beta_0, \beta_1, \dots, \beta_N$ of random variables, where N is the number of outcomes of α, as follows. Let $\beta_0 = \alpha$ and β_{i+1} is obtained from β_i

by gluing any two outcomes of β_i. Each β_i is a function of α, hence $H(\beta_i|\alpha)=0$. It is easy to verify that gluing any two outcomes can decrease the entropy at most by 1. As $H(\beta_0)=H(\alpha)$ and $H(\beta_n)=0$ there is i with $H(\beta_i)=H(\alpha)/2\pm1$.

We think that it is natural, in the comparison of Shannon and Kolmogorov theories, to consider the information inequalities for Shannon entropy also with accuracy $o(n)$ where n is the sum of "lengths" of the involved random variables. As a "length" of a random variable α it is natural to consider its max-entropy $H_0(\alpha)$. Thus, instead of Equation (9) we will consider the following formula:

$$\forall\alpha_1\dots\forall\alpha_k\exists\alpha_{k+1}\dots\exists\alpha_{k+l}\ \bigvee_i\bigwedge_j\sum_{A,B}\lambda^{ij}_{A,B}H(\alpha_A|\alpha_B)\le o(n) \tag{11}$$

where $n=H_0(\alpha_1)+\dots+H_0(\alpha_k)$. This formula is a succinct representation of the following assertion: there is a function $f(n)=o(n)$ such that the formula

$$\forall\alpha_1\dots\forall\alpha_k\exists\alpha_{k+1}\dots\exists\alpha_{k+l}\ \bigvee_i\bigwedge_j\sum_{A,B}\lambda^{ij}_{A,B}H(\alpha_A|\alpha_B)\le f(n)$$

holds in the same sense, as (9) does. Is it true that Equation (11) holds if and only if Equation (10) holds? This is true for formulas without existential quantifiers ($l=0$), as Romashchenko's theorem holds (with the same proof) if we replace 0 by $o(n)$ in the right hand side of (7).

3 Separating Shannon's and Kolmogorov's information using max-entropy in formulas

It is easy to find a counter-example if we allow to use the max-entropy in formulas (and the length of strings in the corresponding formulas for Kolmogorov complexity). Namely, in Kolmogorov theory, for every string x it is possible to extract about $K(x)$ bits of randomness from x: For every string x there is a string y with

$$K(y|x)=O(\log l(x)),\ K(y)=l(y)+O(1)=K(x)+O(1)$$

(let y to be the shortest description of x). This property of Kolmogorov complexity translates to Shannon theory as follows. For every random variable α there is a random variable β with

$$H(\beta|\alpha)=o(n),\ H(\beta)=H_0(\beta)+o(n)=H(\alpha)+o(n), \tag{12}$$

where $n=H_0(\alpha)$. This statement is false. This is implied by the following Theorem 2. Indeed, the inequalities (12) and the equality (1) imply that $H(\alpha|\beta)=o(n)$. Thus the left hand side of the inequality (13) is equal to $H_0(\beta)+o(n)=H(\alpha)+o(n)$, which is much less than its right hand side.

Theorem 2. *For every n there is a random variable α with 2^n+1 outcomes such that for all random variables β it holds*

$$H_0(\beta)+64H(\alpha|\beta)>H(\alpha)+n/2-2. \tag{13}$$

Proof. Let the outcomes of α be $a_0, a_1, \dots, a_{2^n}$ and have probabilities $p_0 = 1/2$ and $p_i = 2^{-n-1}$ for $i = 1, \dots, 2^n$. Obviously, $H(\alpha) = n/2 + 1$. Thus, given a random variable β in a set B of cardinality 2^d, we have to prove that $d + 64H(\alpha|\beta) > n - 1$.

Let A denote the set $\{a_1, \dots, a_{2^n}\}$ (the element a_0 is not included). Divide all the pairs $\langle a, b\rangle$ in the set $A \times B$ into three groups:
(1) the pairs $\langle a, b\rangle$ with $H(\alpha|\beta = b) \geq 8H(\alpha|\beta)$;
(2) the pairs $\langle a, b\rangle$ outside (1) with with $\Pr[\alpha = a|\beta = b] \leq 2^{-64H(\alpha|\beta)}$;
(3) the pairs $\langle a, b\rangle$ with $\Pr[\alpha = a|\beta = b] > 2^{-64H(\alpha|\beta)}$.

The sum of probabilities of pairs in (1) is at most $1/8$, as the probability that $H(\alpha|\beta = b)$ exceeds its expectation 8-fold is at most $1/8$. The same argument applies for pairs $\langle a, b\rangle$ in (2): for every fixed b the value $-\log \Pr[\alpha = a|\beta = b]$ is more than $64H(\alpha|\beta)$ hence exceeds its expectation $H(\alpha|\beta = b)$ more than 8-fold. And the total probability of pairs in (3) is at most $2^{d+64H(\alpha|\beta)-n-1}$. Indeed, for every $b \in B$ there are less than $2^{64H(\alpha|\beta)}$ pairs $\langle a, b\rangle$ in (3). Thus, the total number of pairs in (3) is less than $2^{d+64H(\alpha|\beta)}$. The probability of each of them is at most 2^{-n-1}. Summing all probabilities we should obtain at least $1/2$, the probability of $A \times B$. Thus we have

$$1/8 + 1/8 + 2^{d+64H(\alpha|\beta)-n-1} > 1/2 \;\Rightarrow\; d + 64H(\alpha|\beta) > n - 1. \square$$

It is worth to mention here that a result on implicit extractors from [1] implies that in Shannon's theory it is possible to extract about $H_\infty(\alpha)$ random bits from every random variable α. Here $H_\infty(\alpha)$ denotes the min-entropy of α defined as $\min\{-\log p_i, \dots, -\log p_N)$ where $p_1, \dots, p_N$ are probabilities of outcomes of α. More specifically, the following is true.

Theorem 3. *For every random variable α there is a random variable β with*

$$H(\beta|\alpha) = O(\log n),\ H(\beta) = H_0(\beta) + O(\log n) = H_\infty(\alpha) + O(\log n),$$

where $n = H_0(\alpha)$.

Proof. We use the following theorem on extractors from [1]. For all integer $n \geq m$ and positive ε there is a set C of cardinality $O(n/\varepsilon^2)$ and a function $f : \{0,1\}^n \times C \to \{0,1\}^m$ with the following property. Let α be a random variable in $\{0,1\}^n$ with min-entropy at least m and let u be uniformly distributed in C and independent of α. Then the distribution of $f(\alpha, u)$ is at most ε apart from the uniform distribution over $\{0,1\}^m$. This means that for every subset B of $\{0,1\}^m$ the probability that $f(\alpha, u)$ gets into B is $|B|2^{-m} \pm \varepsilon$.

Apply this theorem to $n = \lceil H_0(\alpha)\rceil$, $m = \lfloor H_\infty(\alpha)\rfloor$ and $\varepsilon = \log m/m$. Let $\beta = f(\alpha, u)$ (we may assume that α takes values in $\{0,1\}^n$). Then we have $H(\beta|\alpha) \leq \log|C| \leq \log n - 2\log\varepsilon + O(1) = O(\log n)$.

To estimate $H(\beta)$ we need the following

Lemma 1. *If β is at most ε apart from the uniform distribution over $\{0,1\}^m$ then $H(\beta) \geq m(1-\varepsilon) - 1$.*

Proof. Let μ stand for the probability distribution of β, that is, $\mu(b) = \Pr[\beta = b]$. Let B_i denote the set of all $b \in \{0,1\}^m$ with $\mu(b) > 2^{i-m}$. As β is at most ε apart from the uniform distribution, we can conclude that $\mu(B_i) \le |B_i| 2^{-m} + \varepsilon < 2^{-i} + \varepsilon$. For all b outside B_i we have $-\log \mu(b) \ge m - i$. Thus the entropy of β can be lower bounded as

$$H(\beta) \ge m - \sum_{i=1}^{m} i \cdot \mu(B_{i-1} - B_i) = m - \sum_{i=1}^{m} i \cdot (\mu B_{i-1} - \mu B_i)$$
$$= m - \sum_{i=1}^{m-1} \mu B_i \; > m - \sum_{i=1}^{m-1} (2^{-i} + \varepsilon) \ge m - 1 - \varepsilon m.$$

The lemma implies that

$$H(\beta) \ge m - \log m - 1 \ge H_0(\beta) - O(\log n).$$

As $H_\infty(\alpha) = m + O(1)$ and $H(\beta) \le H_0(\beta)$, this inequality implies that the difference between $H(\beta), H_0(\beta), H_\infty(\alpha)$ is $O(\log n)$.

4 Separating Shannon's and Kolmogorov's information theories

Looking for an assertion of the form (11) that distinguishes Shannon entropy and Kolmogorov complexity, it is natural to exploit the following difference between Shannon and Kolmogorov definitions of conditional entropy. In Kolmogorov's approach conditional complexity $K(a|b)$ is defined as the length of a string, namely the shortest description of a conditional to β. In Shannon's approach $H(\alpha|\beta)$ is not defined as the max-entropy or Shannon entropy of any random variable. Thus the following easy statement could distinguish Kolmogorov's and Shannon's theories:

$$\forall x \forall y \exists z \; K(z) \le K(x|y) + O(1), \;\; K(x|[y,z]) = O(1)$$

(let z be the shortest description of x conditional to y). However it happens that its analog holds also in Shannon's theory:

Theorem 4. *For all random variables α, β there is random variable γ such that $H(\gamma) \le H(\beta|\alpha) + O(\log n)$ and $H(\beta|\langle\alpha,\gamma\rangle) = 0$, where $n = H(\beta|\alpha) \le H_0(\alpha) + H_0(\beta)$.*

Proof. Let A, B denote the set of outcomes of α, β, respectively. Fix $a \in A$ and let β_a denote the random variable in B which takes every value $b \in B$ with probability $\Pr[\beta = b|\alpha = a]$. Using Shannon or Fano code we can construct, for each a, an injective mapping f_a from B to the set of binary strings such that the expected length of $f_a(\beta_a)$ is at most $H(\beta_a) + O(1)$. Let $\gamma = f_\alpha(\beta)$. By construction the outcomes of α and γ together determine the outcome of β uniquely. This shows that $H(\beta|\langle\alpha,\gamma\rangle) = 0$.

It remains to show that $H(\gamma) \le H(\beta|\alpha) + O(\log n)$. Let us first upper bound the expectation of $l(\gamma)$. It is less than the expectation of $H(\beta_\alpha) + O(1)$, which is equal to $H(\beta|\alpha) + O(1)$. Thus it suffices to show that Shannon entropy of every random variable γ in the set of binary strings with expected length n is at most $n + O(\log n)$. To this end consider the following "self-delimiting" encoding $\bar{x}$ of a binary string x. Double each bit of binary representation of the length of x then append the string 10 to it, and then append x. Obviously $l(\bar{x}) \le 2\log l(x) + 2 + l(x)$. The set of all strings of the form $\bar{x}$ is a prefix code. Thus the set of all strings $\bar{c}$ where c is an outcome of γ is a prefix code, too. By the Shannon noiseless coding theorem [9] its expected length is at least $H(\gamma)$. Therefore $H(\gamma)$ is less than the expectation of $l(\gamma) + 2\log l(\gamma) + 2$. The expectation of the first term here is equal to n. The expectation of the second term is at most $2\log n$ by concavity of the logarithm function.

The previous discussion shows that it is not so easy to find a counter-example. Looking for a candidate in the literature we find the following:

Theorem 5 ([2]). *For all strings x, y there is a string z such that $K(z) \le \max\{K(x|y), K(y|x)\} + O(\log n)$ and $K(y|z,x) = O(\log n)$, $K(x|z,y) = O(\log n)$ where $n = K(x|y) + K(y|x)$.*

This theorem implies the following statement

$$\forall x \forall y \exists z \; K(z) + K(y|z,x) + K(x|z,y) \le \max\{K(x|y), K(y|x)\} + O(\log n)$$

where $n = l(x) + l(y)$. The inner quantifier free formula here can be expressed as an OR of two inequalities. Thus this formula has the form (10). And the analogous statement for Shannon entropy is false:

Theorem 6. *For every n there are random variables α, β with $2^n + 1$ outcomes each such that for every random variable γ we have*

$$H(\gamma) + H(\alpha|\beta,\gamma) + H(\beta|\alpha,\gamma) \ge \max\{H(\alpha|\beta), H(\beta|\alpha)\} + n/2. \qquad (14)$$

Proof. Let δ be a random variable having two equiprobable outcomes 0,1. The random variables α and β have the range $\{a_0, a_1, \dots, a_{2^n}\}$ and are defined as follows. If $\delta = 0$ then α is equal to a_0 and β is uniformly distributed in $\{a_1, \dots, a_{2^n}\}$. If $\delta = 1$ then β is equal to a_0 and α is uniformly distributed in $\{a_1, \dots, a_{2^n}\}$. Note that $H(\alpha|\beta) = H(\alpha|\beta) = n/2$, thus the right hand side of Equation (14) is equal to n.

Let γ be a random variable. If $\delta = 0$ then α is constant and β is uniformly distributed in a set of cardinality 2^n, therefore

$$n = H(\beta|\alpha,\delta=0) \le H(\beta|\alpha,\gamma,\delta=0) + H(\gamma|\delta=0) \le 2H(\beta|\alpha,\gamma) + H(\gamma|\delta=0).$$

In a similar way we have

$$n \le 2H(\alpha|\beta,\gamma) + H(\gamma|\delta=1).$$

Taking the arithmetical mean of these inequalities we get

$$n \le H(\beta|\alpha,\gamma) + H(\alpha|\beta,\gamma) + H(\gamma|\delta) \le H(\beta|\alpha,\gamma) + H(\alpha|\beta,\gamma) + H(\gamma). \square$$

5 Sequences of identically distributed independent random variables

A large part of the classical information theory is devoted to the study of sequences of independent identically distributed random variables. Following this line, assume that the universal quantifiers in (11) range over sequences of i.i.d. variables. More specifically let ξ_s^n, $s = 1, \ldots, k$, $n = 1, 2, \ldots$ be random variables such that the k-tuples $\langle \xi_1^n, \ldots, \xi_k^n \rangle$ for $n = 1, 2, \ldots$, have the same distribution and are independent. Let α_s^n denote the sequence of n first outcomes of ξ_s:

$$\alpha_s^n = \xi_s^{(n)} = \xi_s^1, \ldots, \xi_s^n.$$

Consider the following formula:

$$(\forall \text{ i.i.d. } \langle \xi_1^n, \ldots, \xi_k^n \rangle)\ \exists \alpha_{k+1}^n \ldots \exists \alpha_{k+l}^n \bigvee_i \bigwedge_j \sum_{A,B} \lambda_{A,B}^{ij} H(\alpha_A^n | \alpha_B^n) \leq o(n). \quad (15)$$

This formula represents the following statement: For all random variables ξ_s^n, $s = 1, \ldots, k$, $n = 1, 2, \ldots$ such that the k-tuples $\langle \xi_1^n, \ldots, \xi_k^n \rangle$ for $n = 1, 2, \ldots$, have the same distribution and are independent there are sequences of random variables $\alpha_1^n, \ldots, \alpha_{k+l}^n$, $n = 1, 2, \ldots$, and a function $f(n) = o(n)$ with $\bigvee_i \bigwedge_j \sum_{A,B} \lambda_{A,B}^{ij} H(\alpha_A^n | \alpha_B^n) \leq f(n)$, and $\langle \alpha_1^n, \ldots, \alpha_k^n \rangle$ having the same distribution as $\langle \xi_1^{(n)}, \ldots, \xi_k^{(n)} \rangle$ has (for all n).

An example of (15) is the Slepian—Wolf theorem [7]: for every sequence of i.i.d. pairs $\langle \xi^n, \eta^n \rangle$, $n = 1, 2, \ldots$ of random variables there is a sequence of random variables $\{\beta^n\}$ such that

$$H(\beta^n) = H(\xi^{(n)} | \eta^{(n)}) + o(n),\ H(\beta^n | \xi^{(n)}) = o(n),\ H(\xi^{(n)} | \langle \eta^{(n)}, \beta^n \rangle) = o(n).$$

(To fit in our framework, we give here a formulation of Slepian—Wolf theorem that differs slightly from that in [7].)

Is it true that for every theorem of the form (15) the analogous statement (10) for Kolmogorov complexity is also true, and vice versa? We will show that this is not the case. As a counter-example, it is natural to try the Slepian-Wolf theorem, since its proof is very Shannon-theory-specific. Surprisingly, it turns out that the analogous theorem holds for Kolmogorov complexity, too:

Theorem 7 ([8]). *Let x and y be arbitrary strings of length less than n. Then there exists a string z of complexity $K(x|y) + O(\log n)$ such that $K(z|x) = O(\log n)$ and $K(x|z, y) = O(\log n)$. (The constants in $O(\log n)$-notation do not depend on n, x, y.)*

The following easy fact about i.i.d. sequences of random variables gives an example of a true statement of the form (15) whose analog is false for Kolmogorov complexity.

Theorem 8. *For every sequence of i.i.d. pairs* $\langle \xi^n, \eta^n \rangle$, $n = 1, 2, \ldots$ *of random variables there is a sequence* $\{\beta^n\}$ *of random variables such that*

$$H(\beta^n) \leq \frac{H(\xi^{(n)}) + H(\eta^{(n)})}{2} + O(1),$$
$$H(\xi^{(n)}|\beta^n) \leq \frac{H(\xi^{(n)}|\eta^{(n)})}{2} + O(1), \quad H(\eta^{(n)}|\beta^n) \leq \frac{H(\eta^{(n)}|\xi^{(n)})}{2} + O(1).$$

Proof. Let $\beta^n = \xi^1, \xi^2, \ldots, \xi^{n/2}, \eta^{n/2+1}, \eta^{n/2+2}, \ldots, \eta^n$.

On the other hand, the similar statement for Kolmogorov complexity is false:

Theorem 9. *There are sequences of strings* $\{x_n\}, \{y_n\}$ *of length* $O(n)$ *such that there is no sequence* $\{z_n\}$ *with*

$$\begin{gathered} K(z_n) \leq \frac{K(x_n) + K(y_n)}{2} + o(n), \\ K(x_n|z_n) \leq \frac{K(x_n|y_n)}{2} + o(n), \quad K(y_n|z_n) \leq \frac{K(y_n|x_n)}{2} + o(n). \end{gathered} \tag{16}$$

Proof. The proof easily follows from a theorem from [3]:

Theorem 10 ([3]). *There are sequences of strings* $\{x_n\}, \{y_n\}$ *such that* $l(x_n) = l(y_n) = 2n + O(\log n)$, $K(x_n|y_n) = K(y_n|x_n) = n + O(\log n)$ *and for all but finitely many* n *there is no* z_n *satisfying the inequalities*

$$K(z_n) + K(x_n|z_n) + K(y_n|z_n) < 4n,$$
$$K(z_n) + K(x_n|z_n) < 3n, \quad K(z_n) + K(y_n|z_n) < 3n.$$

Let x_n, y_n be the sequences from Theorem 10. Assume that there is z_n satisfying (16). Then

$$\begin{aligned} K(z_n) &\leq (K(x_n) + K(y_n))/2 + o(n) \leq 2n + o(n), \\ K(x_n|z_n) &\leq K(x_n|y_n)/2 + o(n) \leq n/2 + o(n), \\ K(y_n|z_n) &\leq K(y_n|x_n)/2 + o(n) \leq n/2 + o(n). \end{aligned}$$

and

$$\begin{aligned} K(z_n) + K(x_n|z_n) + K(y_n|z_n) &\leq 3n + o(n) \ll 4n \\ K(z_n) + K(x_n|z_n) &\leq 5n/2 + o(n) \ll 3n \\ K(z_n) + K(y_n|z_n) &\leq 5n/2 + o(n) \ll 3n. \square \end{aligned}$$

6 Conclusion and open problems

We have shown that Equation (15) does not imply Equation (10) and that Equation (10) does not imply Equation (11). Are the inverse implications always true? The implication (11) $\Rightarrow$ (15) is straightforward. Can it be split into two implications: (11) $\Rightarrow$ (10) $\Rightarrow$ (15)?

References

1. M.Sipser, "Expanders, randomness, or time versus space", *J. Comput. and System Sci.*, 36 (1988) 379–383.
2. C.H. Bennett, P. Gács, M. Li, P. Vitányi and W. Zurek, "Information Distance", *IEEE Transactions on Information Theory* 44:4 (1998) 1407–1423.
3. A. Chernov, An. Muchnik, A. Romashchenko, A. Shen, and N. Vereshchagin, "Upper semi-lattice of binary strings with the relation 'x is simple conditional to y' ", *Theoretical Computer Science* 271 (2002) 69–95. Preliminary version in: *14th Annual IEEE Conference on Computational Complexity*, Atlanta, May 4-6, 1999, 114–122.
4. D. Hammer, A. Romashchenko, A. Shen, and N. Vereshchagin, "Inequalities for Shannon entropy and Kolmogorov complexity", *Journal of Computer and Systems Sciences* 60 (2000) 442–464.
5. M. Li and P.M.B. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*, Springer-Verlag, New York, 2nd Edition, 1997.
6. A.N. Kolmogorov, "Three approaches to the quantitative definition of information", *Problems Inform. Transmission* 1:1 (1965) 1–7.
7. D. Slepian and J.K. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Trans. Inform. Theory* IT-19 (1973) 471–480.
8. A.A. Muchnik, "Conditional complexity and codes", *Theoretical Computer Science* 271 (2002) 97–109.
9. C. E. Shannon, "A mathematical theory of communication", *Bell Sys. Tech. J.* 27 (1948) 379–423 and 623–656.
10. R.J. Solomonoff, "A formal theory of inductive inference", Part 1 and Part 2, *Information and Control* 7 (1964) 1–22 and 224–254.
11. V. A. Uspensky, A. Shen. "Relations Between Varieties of Kolmogorov Complexities", *Mathematical Systems Theory* 29(3) (1996) 271–292.