The Kolmogorov Birthday Paradox

Samuel Epstein*

July 25, 2022

Abstract

We prove bounds for a Kolmogorov complexity variant of the birthday paradox. Sufficiently sized random subsets of strings are guaranteed to have two members x and y with low $\mathbf{K}(x/y)$. To prove this, we first show that the minimum conditional Kolmogorov complexity between members of finite sets is very low if they are not exotic. Exotic sets have high mutual information with the halting sequence.

1 Introduction

In this paper, we prove a Kolmogorov complexity version of the birthday paradox. That is, if you randomly select $2^{n/2}$ strings of length n, then, with overwhelming probability, you have two selected strings x and y with low $\mathbf{K}(x/y)$. Otherwise the probability is exotic, in that its encoding has high mutual information with the halting sequence.

To prove this fact, we first prove an interesting property about bunches of finite strings. A (k,l)-bunch is a finite set of strings X where $2^l > \max_{x,y\in X} \mathbf{K}(y/x)$ and $2^k < |X|$. The function $\mathbf{K}$ is the prefix free Kolmogorov complexity. Bunches were introduced in [12], but we use a slightly different definition. Though bunches have only two parameters, they exhibit many interesting properties. In the literature, there have been properties proven about the common information of bunches. Both [12] and [11] proved the existence of strings simple to each member of the bunches. That is, there exists a string z such that $\mathbf{K}(z/x) < O(l-k) + \mathbf{K}(l)$ and $\mathbf{K}(x/z) < l + O(l-k) + \mathbf{K}(l)$, for all $x \in X$. In [3], it was proven that each bunch has a member that is simple relative to all members of the bunch, similar to the above definition. If not, then the bunch has high mutual information with the halting sequence. The mutual information between a string (and the bunch it encodes) and the halting sequence is $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x/\mathcal{H})$. We prove that if a non exotic bunch X has many members and low $\max_{x,y\in X, x\neq y} \mathbf{K}(y/x)$ then it will have two elements x, y with very low $\mathbf{K}(y/x)$. An exotic bunch has high mutual information with the halting sequence.

Theorem. *For (k,l)-bunch X, $\min_{x,y\in X} \mathbf{K}(y/x) <^{\log} \lceil l-2k \rceil^+ + \mathbf{I}(X;\mathcal{H}) + 2\mathbf{K}(l,k)$.*

The Kolmogorov Birthday Paradox. Lets say we select a random subset D of size $2^{n/2}$ consisting of (possibly repeated) strings of length n, where each string is selected independently. For the simple Kolmogorov birthday paradox, with overwhelming probability, there are two (possibly the same) strings $x, y \in D$, such that $\mathbf{K}(x/y) = O(1)$, for a large enough constant. This is due to reasoning from the classical birthday paradox. We now prove the general Kolmogorov birthday paradox. Let P be any measure over sets D consisting of $2^{n/2}$ (possibly repeated) strings of length n. Since $D \subset \Sigma^n$, for all D, $\max_{x,y\in D} \mathbf{K}(x/y) <^+ n$. By Corollary 2 in the Appendix, $\Pr_{D\sim P}[\mathbf{I}(D;\mathcal{H}) > \mathbf{I}(P;\mathcal{H}) + m] \stackrel{*}{<} 2^{-m}$. Using the above theorem, we get the following result, with

*JP Theory Group. samepst@jptheorygroup.org

$l = n + O(1)$, $k = .5n - 1$. Note that if D has repeat members, then x could equal y, and trivially $\mathbf{K}(x|y) = O(1)$. Obviously, the bound loosens if P samples sets of smaller size, mirroring the classical birthday paradox.[1]

Corollary. $\Pr_{D\sim P}\left[\min_{x,y\in D} \mathbf{K}(x/y) <^{\log} \mathbf{I}(P;\mathcal{H}) + 2\mathbf{K}(n) + c\right] > 1 - 2^{-c}$.

2 Related Work

The study of Kolmogorov complexity originated from the work of [6]. The canonical self-delimiting form of Kolmogorov complexity was introduced in [21] and treated later in [1]. The universal probability $\mathbf{m}$ was introduced in [16]. More information about the history of the concepts used in this paper can be found the textbook [10].

The main result of this paper is an inequality including the mutual information of the encoding of a finite set with the halting sequence. A history of the origin of the mutual information of a string with the halting sequence can be found in [17].

A string is stochastic if it is typical of a simple elementary probability distribution. A string is typical of a probability measure if it has a low deficiency of randomness. The deficiency of the randomness of a number $a \in \mathbb{N}$ with respect to a probability P is $\mathbf{d}(a|p) = -\log P(a) - \mathbf{K}(a)$. It is a measure of the extent of the refutation against the hypothesis p given the result a [5]. Thus the stochasticity of a string a is, roughly, $\min_{\text{probability } p} \mathbf{K}(p) + O(\log \mathbf{d}(a|P))$.

In the proof of Theorem 1, the stochasticity measure of encodings of finite sets is used. The notion of the deficiency of randomness with respect to a measure follows from the work of [13], and also studied in [7, 19, 14]. Aspects involving stochastic objects were studied in [13, 14, 19, 20].

This work uses the notion of left total machine and the notion of the infinite "border" sequence, which is equal to the binary expansion of Chaitin's Omega, (see Section 7). The works of [17, 4] introduced the notion of using the prefix of the border sequence to define strings into a two part code. This paper uses theorems and lemmas found in [2].

This paper can be seen as a conditional variant to the main result in [9]. In [9], it was proven for non exotic sets D, the a-priori probability of a set is concentrated on a single element.

Theorem. ([9]) $-\log \max_{x\in D} \mathbf{m}(x) <^{\log} -\log \sum_{x\in D} \mathbf{m}(x) + \mathbf{I}(D;\mathcal{H})$.

There is a simple proof to this theorem in [15]. The proof in this paper is similar to that in [9], in that they both first prove stochasticity of a an object O and then show that this object has high $\mathbf{I}(O;\mathcal{H})$. In [9] O is equal to a set, and in this paper O is equal to a (sub)graph. Theorem 2 is not proved in [9] because it deals with conditional complexities. In addition, Theorem 2 is not a generalization of the main theorem in [9] because it relies on the parameters of bunches and not the a-priori probabilities.

3 Conventions

We use Σ, Σ^*, Σ^∞ $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{R}$ to denote bits, finite strings, infinite sequence, natural numbers, rationals, and reals. Let $X_{\geq 0}$ and $X_{>0}$ be the sets of non-negative and of positive elements of X. $\Sigma^{*\infty} = \Sigma^* \cup \Sigma^\infty$. The positive part of a real is $\lceil a \rceil^+ = \max\{a, 0\}$. For string $x \in \Sigma^*$, $x0^- = x1^- = x$. For $x \in \Sigma^*$ and $y \in \Sigma^{*\infty}$, we use $x \sqsubset y$ if there is some string $z \in \Sigma^{*\infty}$ where $xz = y$. The indicator

[1] Formulated a different way, if probability P samples $m < 2^{n/2}$ strings of length n and $\mathbf{E}_{D\sim P}[\min_{x,y\in D} \mathbf{K}(y|x)] > c$, then $\log m < (n-c)/2 + O(\log n + \mathbf{I}(P;\mathcal{H}))$.

function of a mathematical statement A is denoted by $[A]$, where if A is true then $[A] = 1$, otherwise $[A] = 0$. The self delimiting code of a string $x \in \Sigma^*$ is $\langle x \rangle = 1^{\|x\|}0x$. The encoding of (a possibly ordered) set $\{x_1, \ldots, x_m\} \subset \Sigma^*$, is $\langle m \rangle \langle x_1 \rangle \ldots \langle x_m \rangle$.

Probability measures Q over numbers are elementary if $|\text{Support}(Q)| < \infty$ and $\text{Range}(Q) \subset Q_{\geq 0}$. Elementary probability measures Q with $\{x_1, \ldots, x_m\} = \text{Support}(Q)$ are encoded by finite strings, with $\langle Q \rangle = \langle \{x_1, Q(x_1), \ldots, x_m, Q(x_m)\} \rangle$.

For nonnegative real function f, we use $<^+ f$, $>^+ f$, $=^+ f$ to denote $< f + O(1)$, $> f - O(1)$, and $= f \pm O(1)$. We also use $<^{\log} f$ and $>^{\log} f$ to denote $< f + O(\log(f+1))$ and $> f - O(\log(f+1))$.

We use a universal prefix free algorithm U, where we say $U_\alpha(x) = y$ if U, on main input x and auxiliary input α, outputs y. We define Kolmogorov complexity with respect to U, with for $x \in \Sigma^*$, $y \in \Sigma^{*\infty}$, $\mathbf{K}(x/y) = \min\{\|p\| : U_y(p) = x\}$. The universal probability $\mathbf{m}$ is defined as $\mathbf{m}(x/y) = \sum_p [U_y(p) = x] 2^{-\|p\|}$. By the coding theorem $\mathbf{K}(x/y) =^+ -\log \mathbf{m}(x/y)$. By the chain rule, $\mathbf{K}(x, y) =^+ \mathbf{K}(x) + \mathbf{K}(y/x, \mathbf{K}(x))$. The halting sequence $\mathcal{H} \in \Sigma^\infty$ is the unique infinite sequence where $\mathcal{H}[i] = [U(i) \text{ halts}]$. The information that $x \in \Sigma^*$ has about $\mathcal{H}$, conditional to $y \in \Sigma^{*\infty}$, is $\mathbf{I}(x; \mathcal{H}/y) = \mathbf{K}(x/y) - \mathbf{K}(x/\langle y, \mathcal{H} \rangle)$.

This paper uses notions of stochasticity in the field of algorithmic statistics [18]. A string x is stochastic, i.e. has a low $\mathbf{Ks}(x)$ score, if it is typical of a simple probability distribution. The deficiency of randomness function of a string x with respect to an elementary probability measure P conditional to $y \in \Sigma^*$, is $\mathbf{d}(x|P, y) = \lfloor -\log P(x) \rfloor - \mathbf{K}(x/\langle P \rangle, y)$.

Definition 1 (Stochasticity) *For $x, y \in \Sigma^*$, $\mathbf{Ks}(x/y) = \min\{\mathbf{K}(P/y) + 3 \log \max\{\mathbf{d}(x|P, y), 1\} : P$ is an elementary probability measure$\}$. $\mathbf{K}(s) = \mathbf{K}(s/\emptyset)$.*

4 Labelled Graphs, Warmup

In Section 5, a property of a complete subgraph of a label graph is proven. A labelled graph is a directed graph such that each vertex has a unique string attached to it. Given certain properties of the graph $G = (G_E, G_V)$, where G_E are the directed edges and G_V are the directed vertices, and subgraph $J = (J_E, V_V)$, the theorem in Section 5 proves J is guaranteed to have an edge $(x, y) \in J_E$ with low $\mathbf{K}(x|y)$. In this section, we describe the overall arguments in the proof of this theorem.

We use a vertex interchangeably with the string assigned to it. The general argument for the proof of Theorem 1 is as follows. Given a labelled graph G, if there is a random subgraph $F = (F_E, F_V)$ that is large enough, then it will probably share an edge with most large complete subgraphs J of G. Thus large complete subgraphs of G with empty intersection with F will be considered atypical. If F shares an edge with subgraph $J \subseteq G$, then

$$\min_{(x,y) \in J_E} \mathbf{K}(y/x) \lessapprox \log \max_{x \in F_V} \text{OutDegree}(x) + \mathbf{K}(F).$$

This inequality follows from that fact that given a description of F in terms of $\{(x, y) : (x, y) \in F_E\}$, and an $x \in F$, each $y \in \{y : (x, y) \in F_E\}$ can be described relative to x, in $\log\big(\text{OutDegree}(x)\big)$ bits. In this section, instead of using random subgraphs, we use random lists of vertices $L_\bullet$, indexed by $x \in G$. Thus for each $x \in G$, L_x is a list of vertices, possibly with repetition. This allows for easier manipulation.

The warm up arguments are as follows. Let G be a graph of max degree 2^l and $\mathcal{J}$ be the set of complete subgraphs of G of size 2^k. We assume $l > 2k$. For each vertex $x \in G$ has a random list L_x of 2^{l-2k} vertices where for $i \in [1, 2^{l-2k}]$, $\Pr(y = L_x[i]) = [(x, y) \in G_E] 2^{-l}$ and $\Pr(\emptyset = L_x[i]) = 1 - \text{OutDegree}(x) 2^{-l}$. For $J \in \mathcal{J}$, indexed list $L_\bullet$,

$$\text{Miss}(J) \text{ is true iff } \forall x, \forall y \in J_V, y \notin L_x.$$

For each $J \in \mathcal{J}$,

$$\begin{aligned}
\Pr(\text{Miss}(J)) &= \prod_{x \in J_V} \Pr(\forall y \in J_V, y \notin L_x) \\
&\leq \prod_{x \in J_V} \left(1 - 2^{k-l}\right)^{|L_x|} \\
&\leq \prod_{x \in J_V} \left(1 - 2^{k-l}\right)^{2^{l-2k}} \\
&\leq \left((1 - 2^{k-l})^{2^{l-2k}}\right)^{|J|} \\
&\leq \left(e^{-2^{-k}}\right)^{|J|} < e^{-1} < 1.
\end{aligned}$$

It is not hard to see that $\Pr(\text{Miss}(J)) < e^{-b}$ for each $J \in \mathcal{J}$. We assume a uniform distribution $\mathcal{U}$ over $\mathcal{J}$. Under this assumption,

$$\mathbf{E}\Big[[\text{Miss}(J)]\Big] < \sum_{J \in \mathcal{J}} |\mathcal{J}|^{-1} e^{-b} = e^{-b}.$$

Thus given all the parameters, G, k, l, and b, using brute force search, one can find a set of lists $L_\bullet$ of size $b2^{l-2k}$ indexed by $x \in G_V$, such that less than e^{-b} of members J of $\mathcal{J}$ have $\text{Miss}(J)$. Thus $t(J) = [\text{Miss}(J)]e^b$ is a $\mathcal{U}$-test, with $\sum_{J \in \mathcal{J}} t(J)\mathcal{U}(J) < 1$.

We set aside the parameters (G, k, l, b) because they complicate the discussion. That is, we roll the parameters into the additive constants of the inequalities. If $\text{Miss}(J)$ is true for $J \in \mathcal{J}$, then

$$\begin{aligned}
\mathbf{d}(J|\mathcal{U}) &= -\log \mathcal{U}(J) - \mathbf{K}(J/\mathcal{U}) \\
&>^+ \log |\mathcal{J}| - \mathbf{K}(J/L_\bullet, \mathcal{U}) \\
&>^+ \log |\mathcal{J}| - \log t(J)\mathcal{U}(J) \\
&>^+ b \log e.
\end{aligned} \tag{1}$$

Equation 1 has two components. The first term $\log |\mathcal{J}|$ is equal to $-\log \mathcal{U}(J)$ because $\mathcal{U}$ is the uniform distribution over all $\mathcal{J} \ni J$, the set of all complete subgraphs of of G of size 2^k. The second term $-\mathbf{K}(J/L_\bullet, \mathcal{U})$ is equal to $-\mathbf{K}(J/\mathcal{U})$, because given all the hidden parameters (G, k, l, b), one can compute $L_\bullet$ using brute force search, as described above. Thus all subgraphs of G for which Miss is true will be atypical of $\mathcal{U}$, with randomness deficiency greater than b. Thus if a subgraph $J \in \mathcal{J}$ is b-typical, then there exists $(x, y) \in J_E$, with $y \in L_x$. So b-typical subbgraphs $J \in \mathcal{J}$ will have

$$\min_{(x,y) \in J_E} \mathbf{K}(y/x) <^+ \log \left|L_x\right| <^+ l - 2k + \log b. \tag{2}$$

For Theorem 1, the uniform probability measure $\mathcal{U}$ is replaced by a special computable measure P that realizes the stochasticity of the subgraph J. In addition, b is chosen to equal $b \approx \mathbf{d}(J|P)$ so that the subgraph J is guaranteed to be typical of P, so $\text{Miss}(J)$ is false. This means Equation 2 holds for J. In addition, in the next section, the parameters (G, k, l, b) must be taken into account, and to do so, a special constant c is introduced.

5 Labelled Graphs

In this section, we study exotic subgraphs of simple labelled graphs. A subgraph J is exotic if it has a lot of labelled edges $(x, y) \in J$, such that the conditional complexity $\mathbf{K}(y/x)$ is high. This theorem may be of independent interest.

Theorem 1 *For graph $G = (G_E, G_V)$, complete subgraph $J = (G_E, G_V)$, if $2^l > \max \text{Outdegree}(G)$, $2^k < |J|$, $h = \mathbf{I}(J; \mathcal{H}/G, k)$, then $\min_{(x,y)\in J_E} \mathbf{K}(y/x) < \lceil l - 2k \rceil^+ + h + \mathbf{K}(G, k) + O(\mathbf{K}(h))$.*

Proof. Let $\ell = \max\{l, 2k\}$. Let P realize $\mathbf{Ks}(J/G, k)$ and $d = \max\{\mathbf{d}(J|P, G, k), 1\}$. Let $V : G \times G \to \mathbb{R}_{\geq 0}$ be a conditional probability measure where $V(y|x) = [(x, y) \in G_E]2^{-\ell}$ and $V(\emptyset|x) = 1 - \text{OutDegree}(x)2^{-\ell}$. We define a conditional probability measure over lists L of $cd2^{\ell-2k}$ vertices of G, with $\kappa : G \times G^{cd2^{\ell-2k}} \to \mathbb{R}_{\geq 0}$, where $\kappa(L|x) = \prod_{y\in L} V(y|x)$. Let $L_\bullet$ be an indexed list of $cd2^{\ell-2k}$ elements, indexed by $x \in G$, where each list is denoted by L_x for $x \in G_V$. Let $\kappa(L_\bullet) = \prod_{x\in G} \kappa(L_x|x)$. For indexed list $L_\bullet$, graph $H = (H_E, H_V)$, we use the following indicator $\mathbf{i}(L_\bullet, H) = [\text{Complete } H \subseteq G, 2^k < |H|, \forall (x, y) \in H_E, y \notin L_x]$.

$$\begin{aligned}
\mathbf{E}_{L_\bullet\sim\kappa}\mathbf{E}_{H\sim P}[\mathbf{i}(L_\bullet, H)] &\leq \sum_H P(H) \prod_{x\in H_V} (1 - 2^{k-\ell})^{|L_x|} \\
&\leq \sum_H P(H) \prod_{x\in H_V} (1 - 2^{k-\ell})^{cd2^{\ell-2k}} \\
&\leq \sum_H P(H) \prod_{x\in H_V} e^{-cd2^{-k}} \\
&< \sum_H P(H) e^{-cd} \\
&= e^{-cd}.
\end{aligned}$$

Thus there exists an $L_\bullet$ such that $\mathbf{E}_{H\sim P}[\mathbf{i}(L_\bullet, H)] < e^{-cd}$. This $L_\bullet$ can be found with brute force search with all the parameters, with $\mathbf{K}(L_\bullet/G, P, k, c, d) = O(1)$. Thus $t(H) = \mathbf{i}(L_\bullet, H)e^{cd}$ is a P test, where $\mathbf{E}_{H\sim P}[t(H)] \leq 1$. A diagram of the components can be found in Figure 1

It must be that there is an $(x, y) \in J_E$ where $y \in L_x$. Otherwise $t_{L_\bullet}(J) = e^{cd}$ and

$$\begin{aligned}
\mathbf{K}(J/P, G, k, c, d) &<^+ -\log t(J)P(J) \\
&<^+ -(\log e)cd - \log P(J) \\
(\log e)cd &<^+ -\log P(J) - \mathbf{K}(J/P, G, k, c, d) \\
(\log e)cd &<^+ -\log P(J) - \mathbf{K}(J/P, G, k) + \mathbf{K}(c, d) \\
(\log e)cd &<^+ d + \mathbf{K}(c, d),
\end{aligned}$$

which is a contradiction for large enough c solely dependent on the universal Turing machine U. The constant c is folded into the additive constants of the inequalities of the rest of the proof. Thus

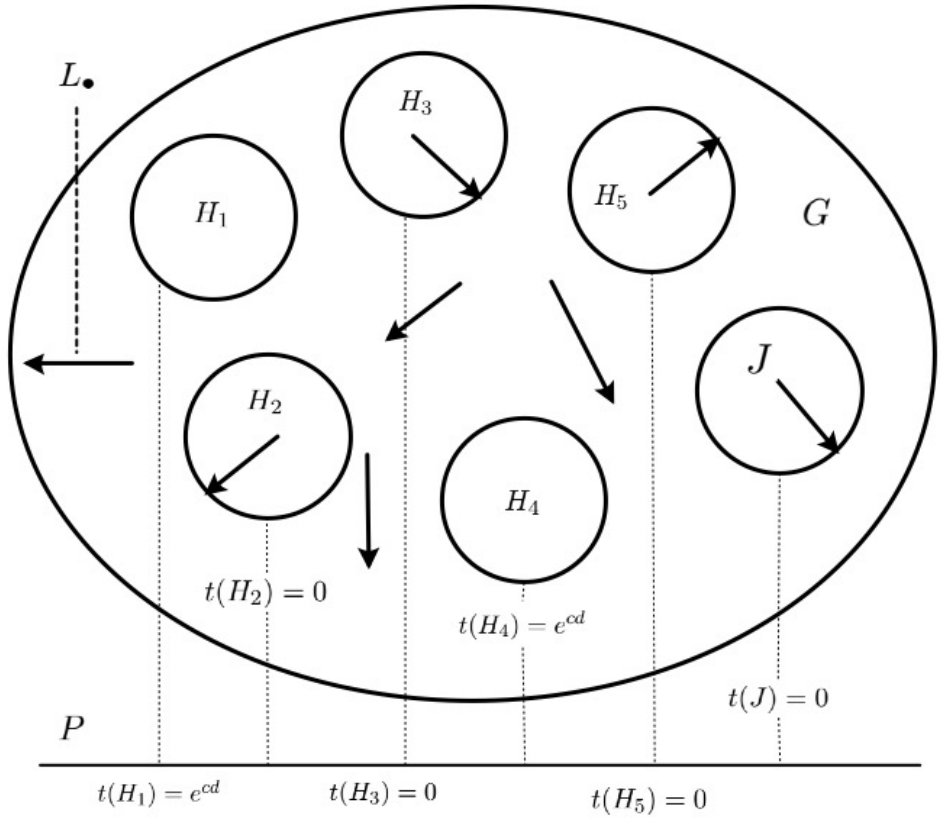

Figure 1: The above diagram is a graphical representation of the concepts used in the proof of Theorem 1. The main ellipse models the graph G and the circles in the graph represent complete subgraphs (labeled H_1 to H_6). Each subgraph is in the support of P, represented by the dotted lines. The set $L_\bullet$ represents a collection of selected edges. If a subgraph H_i does not contain an edge in $L_\bullet$, then H_i is *atypical* and has a high score $t(H_i)$. By design, J is typical, thus shares an edge with $L_\bullet$.

since exists $(x, y) \in J_E$ where $y \in L_x$,

$$\begin{aligned}
\mathbf{K}(y/x, G, k) &<^+ \log |L_x| + \mathbf{K}(L_\bullet/G, k) \\
&<^+ \lceil l - 2k \rceil^+ + \log d + \mathbf{K}(L_\bullet/G, P, k, d) + \mathbf{K}(P, d/G, k) \\
&<^+ \lceil l - 2k \rceil^+ + \log d + \mathbf{K}(P, d/G, k) \\
&<^+ \lceil l - 2k \rceil^+ + 3 \log d + \mathbf{K}(P/G, k) \\
&<^+ \lceil l - 2k \rceil^+ + \mathbf{Ks}(J/G, k) \\
&<^+ \lceil l - 2k \rceil^+ + \mathbf{I}(J : \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(J : \mathcal{H}))) \qquad (3) \\
\mathbf{K}(y/x) &<^+ \lceil l - 2k \rceil^+ + \mathbf{I}(J : \mathcal{H}) + \mathbf{K}(G, k) + O(\mathbf{K}(\mathbf{I}(J : \mathcal{H}))).
\end{aligned}$$

□

Equation 3 is due to Lemma 10 in [2], which states $\mathbf{Ks}(x) < \mathbf{I}(x : \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(x : \mathcal{H})))$.

6 Warm Up for Main Theorem of Paper

Theorem 1 can be used to prove results about the minimum conditional complexity between two elements of a bunch. This sections gives a broad overview of the arguments used in the proof of Theorem 2. Let $X \subset \Sigma^*$ be a (k, l)-bunch, where $|X| > 2^k$, and $\max_{x,y \in X} \mathbf{K}(y/x) < 2^{-l}$.

Let $\mathbf{K}^r(x/y)$ be the conditional complexity of x given y in time r. So given a number r, $\mathbf{K}^r$ is computable. We also assume $\mathbf{K}^r(x/y) = \infty$ if $\|y\| > r$ to make sure that $\mathbf{K}^r$ has finite $\{(x, y) : \mathbf{K}^r(x/y) < \infty, x, y \in \mathbb{N}\}$ for each r. Let r be the smallest number where $\mathbf{K}^r(x/y) < l$, for all $x, y \in X$. Let G be the labelled graph where $(x, y) \in G$ iff $\mathbf{K}^r(y/x) < l$. X can be viewed as a complete subgraph of G of size $>2^k$. Invoking Theorem 1, we get

$$\min_{(x,y) \in X} \mathbf{K}(y/x) \lessapprox \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}/G) + \mathbf{K}(G, k). \qquad (4)$$

We have $\mathbf{K}(r/G) <^+ \mathbf{K}(l)$ because r is the smallest number such that G can be constructed from l and $\mathbf{K}^r$. So

$$\begin{aligned}
\mathbf{K}(X/G) &<^+ \mathbf{K}(X/r) + \mathbf{K}(r/G) \\
&<^+ \mathbf{K}(X/r) + \mathbf{K}(l).
\end{aligned}$$

Now $\mathbf{K}(G/r) <^+ \mathbf{K}(l)$ by the definition of G. By the definition of $\mathbf{I}$,

$$\begin{aligned}
\mathbf{I}(X : \mathcal{H}/G) &= \mathbf{K}(X/G) - \mathbf{K}(X/G, \mathcal{H}) \\
&<^+ \mathbf{K}(X/r) - \mathbf{K}(X/G, \mathcal{H}) + \mathbf{K}(l) \\
&<^+ \mathbf{K}(X/r) - \mathbf{K}(X/r, \mathcal{H}) + \mathbf{K}(G/r) + \mathbf{K}(l) \\
&<^+ \mathbf{I}(X : \mathcal{H}/r) + O(\mathbf{K}(l)). \qquad (5)
\end{aligned}$$

Using $\mathbf{K}(G) <^+ \mathbf{K}(r) + \mathbf{K}(l)$ and combining Equations 4 and 5, we get

$$\min_{(x,y) \in J} \mathbf{K}(y/x) \lessapprox \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}/r) + \mathbf{K}(r) + O(\mathbf{K}(l, k)). \qquad (6)$$

This inequality is close to the form of Theorem 2. The main difference is that the number r appears in Equation 6. This can be rectified if we use a different notion of computational resource. In the next section we introduce left-total universal machines, and the resource used is not a number r but a so-called total string b. Then Lemma 1, defined in Section 7, can be used to remove the b factor from the final inequality.

7 Left-Total Machines

We recall that for $x \in \Sigma^*$, $\Gamma_x = \{x\beta : \beta \in \Sigma^\infty\}$ is the interval of x. The notions of total strings and the "left-total" universal algorithm are needed in this paper. We say $x \in \Sigma^*$ is total with respect to a machine if the machine halts on all sufficiently long extensions of x. More formally, x is total with respect to T_y for some $y \in \Sigma^{*\infty}$ iff there exists a finite prefix free set of strings $Z \subset \Sigma^*$ where $\sum_{z \in Z} 2^{-\|z\|} = 1$ and $T_y(xz) \neq \perp$ for all $z \in Z$. We say (finite or infinite) string $\alpha \in \Sigma^{*\infty}$ is to the "left" of $\beta \in \Sigma^{*\infty}$, and use the notation $\alpha \lhd \beta$, if there exists a $x \in \Sigma^*$ such that $x0 \sqsubseteq \alpha$ and $x1 \sqsubseteq \beta$. A machine T is left-total if for all auxiliary strings $\alpha \in \Sigma^{*\infty}$ and for all $x, y \in \Sigma^*$ with $x \lhd y$, one has that $T_\alpha(y) \neq \perp$ implies that x is total with respect to T_α. Left-total machines were introduced in [9]. An example can be seen in Figure 2.

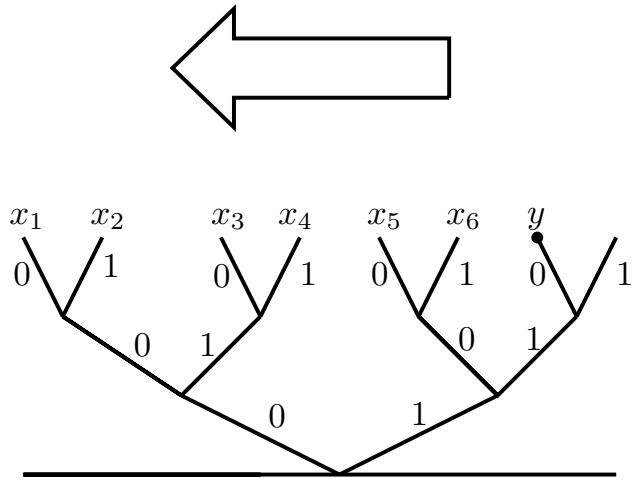

Figure 2: The above diagram represents the domain of a left total machine T with the 0 bits branching to the left and the 1 bits branching to the right. For $i \in \{1, \ldots, 5\}$, $x_i \lhd x_{i+1}$ and $x_i \lhd y$. Assuming $T(y)$ halts, each x_i is total. This also implies each x_i^- is total as well.

For the remaining of this paper, we can and will change the universal self delimiting machine U into a universal left-total machine U' by the following definition. The algorithm U' orders all strings $p \in \Sigma^*$ by the running time of U when given p as an input. Then U' assigns each p an interval $i_p \subseteq [0,1]$ of width $2^{-\|p\|}$. The intervals are assigned "left to right", where if $p \in \Sigma^*$ and $q \in \Sigma^*$ are the first and second strings in the ordering, then they will be assigned the intervals $[0, 2^{-\|p\|}]$ and $[2^{-\|p\|}, 2^{-\|p\|} + 2^{-\|q\|}]$.

Let the target value of $p \in \Sigma^*$ be $(p) \in \mathbb{W}$, which is the value of the string in binary. For example, the target value of both strings 011 and 0011 is 3. The target value of 0100 is 4. The target interval of $p \in \Sigma^*$ is $\Gamma(p) = ((p)2^{-\|p\|}, ((p)+1)2^{-\|p\|})$.

The universal machine U' outputs $U(p)$ on input p' if the intervals $\Gamma(p')$ are strictly contained in i_pm with $\Gamma(p') \subset i_p$ and $\Gamma(p'^-)$ are not strictly contained in i_p, with $\Gamma(p'^-) \not\subset i_p$. The same definition applies for the machines U'_α and U_α, over all $\alpha \in \Sigma^{*\infty}$.

Recall that a function $f : \mathbb{N} \to \mathbb{N}$ is partial computable with respect to U if there is a string $t \in \Sigma^*$ such that $f(x) = U(t\langle x \rangle)$ when $f(x)$ is defined and $U(t\langle x \rangle)$ does not halt otherwise. Similarly a function $f : \mathbb{N} \to \mathbb{N}$ is partial computable with respect to U' if there is $t \in \Sigma^*$, that whenever $f(x)$ is defined, there is an interval $i_{t\langle x \rangle}$ and for any string p where $\Gamma(p)$ and not that of $\Gamma(p^-)$, is

contained in $i_{t\langle x\rangle}$, then $U'(p) = f(x)$. Otherwise, if $f(x)$ is not defined, there does not exist the interval $i_{t\langle x\rangle}$. The following proposition was used without being proven in [9].

Proposition 1 $\mathbf{K}_U(x/y) =^+ \mathbf{K}_{U'}(x/y)$.

Proof. It must be that $\mathbf{K}_U(x/y) <^+ \mathbf{K}_{U'}(x/y)$, because there is a Turing machine that computes U'. Therefore, due to the universality of U, there is a $t \in \Sigma^*$, such that $U_y(tx) = U'_y(x)$, thus proving the minimality of $\mathbf{K}_U$. It must be that $\mathbf{K}_{U'}(x/y) <^+ \mathbf{K}_U(x/y)$. This is because if $U(x) = z$, then there is interval i_x such that for all strings p where $\Gamma(p)$ and not that of $\Gamma(p^-)$ that are strictly contained in i_x has $U'_y(p) = U_y(x)$. Thus we have that $\|p\| \leq \|x\| + 2$. This implies that $\mathbf{K}_{U'}(x/y) \leq \mathbf{K}_U(x/y) + 2$. □

For the rest of the paper, we now set U to be equal U', so the universal Turing machine can be considered to be left-total. Without loss of generality, as shown in Proposition 1 the complexity terms of this paper are defined with respect to the universal left total machine U.

Proposition 2 *There exists a unique infinite sequence $\mathcal{B}$ with the following properties.*

1. *All the finite prefixes of $\mathcal{B}$ have total and non-total extensions.*
2. *If a finite string has total and non-total extensions then it is a prefix of $\mathcal{B}$.*

Proof. The border sequence $\mathcal{B}$ is the binary expansion of Chaitin's Omega for machine U, because the probability that a random infinite sequence contains a prefix that is a halting program is precisely the probability that the random sequence is at the left of the border sequence. If $b \in \Sigma^*$ is total and b^- is not, then b^- has a total extension b^-0 and a non total extension b^-1, thus by the definition of the border sequence, $b^- \sqsubset \mathcal{B}$. □

The following lemma shows that if a prefix of the border sequence is simple relative to a string x, then it will be the common information between x and the halting sequence $\mathcal{H}$. Note that if a string b is total and b^- is not, then $b^- \sqsubset \mathcal{B}$, due to the fact that b^- has total and non-total extensions.

Lemma 1 ([2]) *If $b \in \Sigma^*$ is total and b^- is not, and $x \in \Sigma^*$,*
then $\mathbf{K}(b) + \mathbf{I}(x : \mathcal{H}/b) <^{\log} \mathbf{I}(x : \mathcal{H}) + \mathbf{K}(b/\langle x, \|b\|\rangle)$.

We call this infinite sequence $\mathcal{B}$, "border" because for any string $x \in \Sigma^*$, $x \lhd \mathcal{B}$ implies that x is total with respect to U and $\mathcal{B} \lhd x$ implies that U will never halt when given x as an initial input. Figure 3 shows the domain of U' with respect to $\mathcal{B}$. We now set U to be equal U'. Without loss of generality, as shown in Proposition 1 the complexity terms of this paper are defined with respect to the universal left total machine U.

For total string b, we define the busy beaver function, $\mathbf{bb}(b) = \max\{\|x\| : U(p) = x, p \lhd b \text{ or } p \sqsupseteq b\}$. For total string b, the b-computable complexity of a string x with respect to a string $y \in \Sigma^{*\infty}$, is $\mathbf{K}_b(x/y) = \min\{\|p\| : U_y(p) = x \text{ in } \mathbf{bb}(b) \text{ time and} \|y\| \leq \mathbf{bb}(b)\}$.

8 Minimum Conditional Complexity

We define a (k, l)-bunch X to be a finite set of strings, where $2^k < |X|$ and for all $x, x' \in X$, $\mathbf{K}(x/x') < l$. If $l \gg k$, such as the (k, l)-bunch consisting of two large independent random strings, then it is difficult to prove properties about it. If $l \approx 2k$, then interesting properties emerge.

Theorem 2 *For (k, l)-bunch X,* $\min_{x,y \in X} \mathbf{K}(y/x) <^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X; \mathcal{H}) + 2\mathbf{K}(l, k)$.

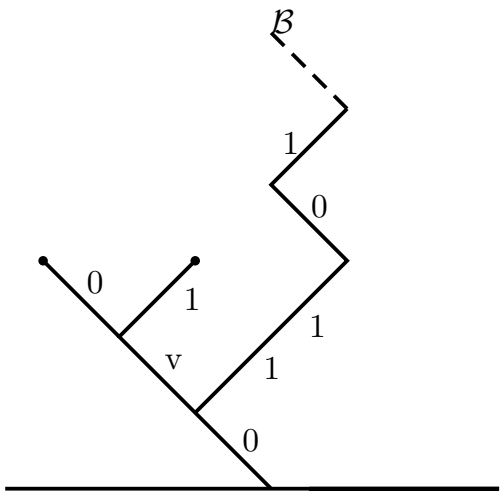

Figure 3: The above diagram represents the domain of the universal left-total algorithm U', with the 0 bits branching to the left and the 1 bits branching to the right. The strings in the above diagram, $0v0$ and $0v1$, are halting inputs to U' with $U(0v0) \neq \perp$ and $U(0v1) \neq \perp$. So $0v$ is a total string. The infinite border sequence $\mathcal{B} \in \Sigma^\infty$ represents the unique infinite sequence such that all its finite prefixes have total and non total extensions. All finite strings branching to the right of $\mathcal{B}$ will cause U' to diverge.

Proof. We assume that the universal Turing machine U is left-total. Let b be the shortest total string such that $\mathbf{K}_b(y/x) < l$ for all $x, y \in X$. We have

$$\mathbf{K}(b/X) <^+ \mathbf{K}(\|b\|, l), \tag{7}$$

as there is a program that when enumerating total strings of length $\|b\|$, returns the first one with the desired properties. The first total string found is b, as shown in Figure 4.

Thus b^- is not total, and b^- is a prefix of border. For total string c, let G_c be the graph defined by $(x, y) \in G$ iff $\mathbf{K}_c(y/x) < l$. Let $G = G_b$. We have

$$\mathbf{K}(G/b) <^+ \mathbf{K}(l) \tag{8}$$
$$\mathbf{K}(b/G) <^+ \mathbf{K}(\|b\|, l). \tag{9}$$

TEquation 9 is due to the existence of a program that enumerates total strings of length $\|b\|$ and returns the first total string c such that $G \subseteq G_c$. It cannot be that there is a total string c shorter than b with $G \subseteq G_c$. Otherwise $G_c \supseteq G \supseteq X$, contradicting b being the shortest total string with $G_b \supseteq X$. Thus using this impossibility and the reasoning detailed in Figure 4, the program returns b. Theorem 1, gives $x, y \in X$, where

$$\mathbf{K}(y/x) <^+ \lceil l - 2k \rceil^+ + \mathbf{I}(X; H/G, k) + \mathbf{K}(G, k). \tag{10}$$

We have

$$\begin{aligned} \mathbf{K}(X/G) &<^+ \mathbf{K}(X/b) + \mathbf{K}(b/G) \\ &<^+ \mathbf{K}(X/b) + \mathbf{K}(\|b\|, l), \end{aligned} \tag{11}$$

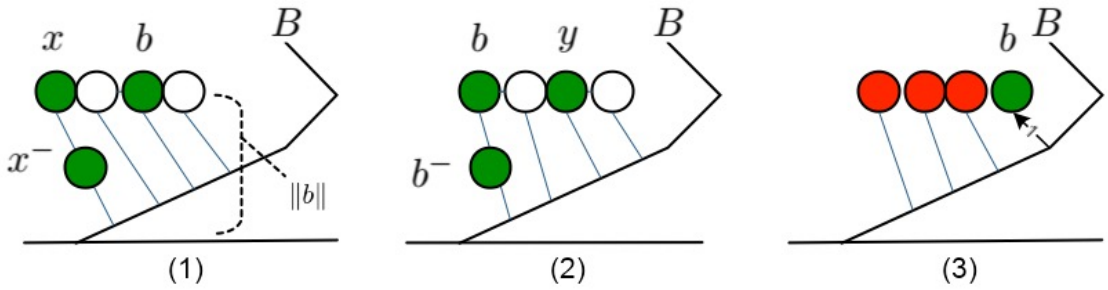

Figure 4: The above diagram represents the domain of the universal Turing machine U and uses the same conventions as Figure 3, with 0s branching to the left and 1s branching to the right. It shows all the total strings of length $\|b\|$, including b. The large diagonal line is the border sequence, B. A string c is marked green if $\mathbf{K}_c(y/x) < l$ for all $x, y \in X$. By definition, b is a shortest green string. By the definition of $\mathbf{K}_b$, if x is green and $x \lhd y$, then y is green. Furthermore, if x is green and total and x^- is total, then x^- is green. It cannot be that there is a green $x \lhd b$. Otherwise x^- is green, which is shorter than b. This is shown in part (1). Furthermore, there can't be a green y, with $b \lhd y$. Otherwise b^- is green, contradicting the definition of b. This is shown in part (2). Thus b is unique, and since b^- is not total, b^- is a prefix of border, as shown in part (3). Thus an algorithm returning a green string of length $\|b\|$ will return b.

where Equation 11 is due to Equation 9. We also have

$$\begin{aligned}\mathbf{K}(X/b, \mathcal{H}) &< \mathbf{K}(X/G, \mathcal{H}) + \mathbf{K}(G/b, \mathcal{H}),\\ &< \mathbf{K}(X/G, \mathcal{H}) + \mathbf{K}(l),\end{aligned} \tag{12}$$

where Equation 12 is due to Equation 8. So

$$\begin{aligned}\mathbf{I}(X : \mathcal{H}/G) &= \mathbf{K}(X/G) - \mathbf{K}(X/G, \mathcal{H})\\ &<^+ \mathbf{I}(X : \mathcal{H}/b) + \mathbf{K}(l) + \mathbf{K}(\|b\|, l).\end{aligned} \tag{13}$$

Combining Equations 10 and 13,

$$\begin{aligned}\mathbf{K}(y/x) &<^+ \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}/b) + \mathbf{K}(G) + \mathbf{K}(\|b\|) + O(\mathbf{K}(k, l))\\ &<^+ \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}/b) + \mathbf{K}(b) + \mathbf{K}(\|b\|) + O(\mathbf{K}(k, l)) \qquad (14)\\ &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}/b) + \mathbf{K}(b) + O(\mathbf{K}(k, l)). \qquad (15)\end{aligned}$$

Equation 14 is due to Equation 8. In Equation 15, the term $\mathbf{K}(\|b\|)$ is removed because the equation is switched to logarithmic precision. Since $b^- \sqsubset B$ and the border B is the binary expansion of Chaitin's Omega (see Proposition 2), $\mathbf{K}(\|b\|) = O(\log \mathbf{K}(B))$. Using Lemma 1, we get

$$\begin{aligned}\mathbf{K}(y/x) &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}) + \mathbf{K}(b/X, \|b\|) + O(\mathbf{K}(k, l))\\ &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}) + O(\mathbf{K}(k, l)) \qquad (16)\\ \mathbf{K}(y/x, k, l) &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}/k, l)\\ \mathbf{K}(y/x) &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X : \mathcal{H}) + 2\mathbf{K}(k, l).\end{aligned}$$

where Equation 16 is due to Equation 7. □

A Conservation Inequalities

The following section presents some conservation inequalities for support of the main result of this paper, which is the corollary in the introduction. The results and proofs are similar to that of [8], except we use $\mathbf{I}(a;\mathcal{H})$ instead of $\mathbf{I}(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$.

Theorem 3 *For computable probability p over $\mathbb{N}$, $\mathbf{E}_{a\sim p}\left[2^{\mathbf{I}(\langle p,a\rangle;\mathcal{H})}\right] \stackrel{*}{<} 2^{\mathbf{I}(p;\mathcal{H})}$.*

Proof. $\sum_a p(a)\mathbf{m}(a,p/\mathcal{H})/\mathbf{m}(a,p) \stackrel{*}{<} \mathbf{m}(p/\mathcal{H})/\mathbf{m}(p)$. Some reworking implies the following inequality, with $\sum_a \big(\mathbf{m}(p)p(a)/\mathbf{m}(a,p)\big)\big(\mathbf{m}(a,p/\mathcal{H})/\mathbf{m}(p/\mathcal{H})\big) \stackrel{*}{<} 1$. The term $\mathbf{m}(p)p(a)/\mathbf{m}(a,p) \stackrel{*}{<} 1$ because $\mathbf{K}(p) - \log p(a) >^+ \mathbf{K}(a,p)$. Furthermore, it follows directly that $\sum_a \mathbf{m}(a,p/\mathcal{H})/\mathbf{m}(p/\mathcal{H}) \stackrel{*}{<} 1$. □

Theorem 4 *For partial computable $f : \mathbb{N} \to \mathbb{N}$, for all $a \in \mathbb{N}$, $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f)$.*

Proof.

$$\begin{aligned}\mathbf{I}(a;\mathcal{H}) &= \mathbf{K}(a) - \mathbf{K}(a/\mathcal{H}) \\ &>^+ \mathbf{K}(a,f(a)) - \mathbf{K}(a,f(a)/\mathcal{H}) - \mathbf{K}(f)\end{aligned}$$

The chain rule applied twice results in

$$\begin{aligned}\mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f) &>^+ \mathbf{K}(f(a)) + \mathbf{K}(a/f(a),\mathbf{K}(f(a))) - (\mathbf{K}(f(a)/\mathcal{H}) + \mathbf{K}(a/f(a),\mathbf{K}(f(a)/\mathcal{H}),\mathcal{H}) \\ &=^+ \mathbf{I}(f(a);\mathcal{H}) + \mathbf{K}(a/f(a),\mathbf{K}(f(a))) - \mathbf{K}(a/f(a),\mathbf{K}(f(a)/\mathcal{H}),\mathcal{H}) \\ &=^+ \mathbf{I}(f(a);\mathcal{H}) + \mathbf{K}(a/f(a),\mathbf{K}(f(a))) - \mathbf{K}(a/f(a),\mathbf{K}(f(a)),\mathbf{K}(f(a)/\mathcal{H}),\mathcal{H}) \\ &>^+ \mathbf{I}(f(a);\mathcal{H}).\end{aligned}$$

□

Corollary 1 *For computable probability p over $\mathbb{N}$, $\mathbf{E}_{a\sim p}[2^{\mathbf{I}(a;\mathcal{H})})] \stackrel{*}{<} 2^{\mathbf{I}(p;\mathcal{H})}$.*

Corollary 2 *For computable probability p over $\mathbb{N}$, $\Pr_{a\sim p}\left[\mathbf{I}(a;\mathcal{H}) > \mathbf{I}(p;\mathcal{H}) + m\right] \stackrel{*}{<} 2^{-m}$.*

Acknowledgements. The author thanks the anonymous referees of the Theoretical Computer Science journal for their careful review of the paper and insightful comments.

References

[1] G. J. Chaitin. A Theory of Program Size Formally Identical to Information Theory. *Journal of the ACM*, 22(3):329–340, 1975.

[2] Samuel Epstein. All sampling methods produce outliers. *IEEE Transactions on Information Theory*, 67(11):7568–7578, 2021. doi: 10.1109/TIT.2021.3109779.

[3] Samuel Epstein. On the conditional complexity of sets of strings. *CoRR*, 1907.01018, 2021. URL https://arxiv.org/abs/1907.01018.

[4] P. Gács, J. Tromp, and P. Vitányi. Algorithmic Statistics. *IEEE Transactions on Information Theory*, 47(6):2443–2463, 2001.

[5] Peter Gács. Lecture notes on descriptional complexity and randomness. *CoRR*, abs/2105.04704, 2021. URL https://arxiv.org/abs/2105.04704.

[6] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems in Information Transmission*, 1:1–7, 1965.

[7] A. N. Kolmogorov and V. A. Uspensky. Algorithms and Randomness. *SIAM Theory of Probability and Its Applications*, 32(3):389–412, 1987.

[8] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[9] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016. And also: S. Epstein and L.A. Levin, Sets have simple members, arXiv preprint arXiv:1107.1458, 2011.

[10] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Publishing Company, Incorporated, 3 edition, 2008.

[11] A. Romashchenko. Clustering with respect to the information distance. *Theoretical Computer Science*, 2022. URL https://www.sciencedirect.com/science/article/pii/S0304397522004133.

[12] Andrei E. Romashchenko. Extracting the mutual information for a triple of binary strings. In *IEEE Conference on Computational Complexity*, pages 221–229. IEEE Computer Society, 2003.

[13] A. Shen. The concept of (alpha,beta)-stochasticity in the Kolmogorov sense, and its properties. *Soviet Mathematics Doklady*, 28(1):295–299, 1983.

[14] A. Shen. Discussion on Kolmogorov Complexity and Statistical Analysis. *The Computer Journal*, 42(4):340–342, 1999.

[15] A. Shen. Game Arguments in Computability Theory and Algorithmic Information Theory. In *Proceedings of 8th Conference on Computability in Europe* , volume 7318 of *LNCS*, pages 655–666, 2012.

[16] R. J. Solomonoff. A Formal Theory of Inductive Inference, Part l. *Information and Control*, 7:1–22, 1964.

[17] N. Vereshchagin and P. Vitányi. Kolmogorov's Structure Functions and Model Selection. *IEEE Transactions on Information Theory*, 50(12):3265 – 3290, 2004.

[18] Nikolay K. Vereshchagin and Alexander Shen. Algorithmic statistics: Forty years later. In *Computability and Complexity*, pages 669–737, 2017.

[19] V.V. V'Yugin. On Randomness Defect of a Finite Object Relative to Measures with Given Complexity Bounds. *SIAM Theory of Probability and Its Applications*, 32:558–563, 1987.

[20] V.V. V'Yugin. Algorithmic complexity and stochastic properties of finite binary sequences. *The Computer Journal*, 42:294–317, 1999.

[21] A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, page 11, 1970.