The Kolmogorov Birthday Paradox

Samuel Epstein*

July 28, 2022

Abstract

We prove 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

We prove a Kolmogorov complexity version of the birthday paradox. If you randomly select $2^{n/2}$ strings of length n, then, with overwhelming probability, you will have selected at least two strings x and y with low $\mathbf{K}(x/y)$. This is true for all probabilities with low mutual information with the halting sequence. The function $\mathbf{K}$ is the prefix free Kolmogorov complexity.

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 $2^k < |X|$. Bunches were introduced in [13], but we use a slightly different definition. Though bunches have only two parameters, they exhibit many interesting properties. Both [13] and [12] 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 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 string (or any object which it is represented by) has high mutual information with the halting sequence.

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

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 with a uniform probability. 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 probability over sets D consisting of $2^{n/2}$ (possibly repeated) strings of length n. Since $D \subset \{0,1\}^n$, for all D, $\max_{x,y\in D}\mathbf{K}(x/y) <^{+} n$. By Corollary 2 in Appendix A, $\Pr_{D\sim P}\left[\mathbf{I}(D;\mathcal{H}) > \mathbf{I}(P;\mathcal{H}) + m\right] \overset{*}{<} 2^{-m}$. Combining these facts with the above theorem, with $l = n + O(1)$ and $k = .5n - 1$, we get the following result.

*JP Theory Group. samepst@jptheorygroup.org

Figure 1: The domain of a Turing machine T can be interpreted as the $[0,1]$ interval, and the strings for which T halts can be seen as a collection of dyadic subintervals. A left-total machine L has the property that if L halts on a string x then it will halt on a string y whose binary interval is smaller (i.e. to the left of) x. The infinite sequence B is called the border sequence and is the binary expansion of Chaitin's Omega. This paper uses a left-total universal Turing machine.

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

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]

2 Related Work

The study of Kolmogorov complexity originated from the work of [7]. The canonical self-delimiting form of Kolmogorov complexity was introduced in [22] and treated later in [1]. The universal probability $\mathbf{m}$ was introduced in [17]. More information about the history of the concepts used in this paper can be found the textbook [11].

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 [18].

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 randomness of a number $a\in\mathbb{N}$ with respect to a probability P is $\mathbf{d}(a|P) = -\log P(a) - \mathbf{K}(a/\langle P\rangle)$. It is a measure of the extent of the refutation against the hypothesis P given the result a [6]. 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 [14], and also studied in [8, 20, 15]. Aspects involving stochastic objects were studied in [14, 15, 20, 21].

This work uses the notion of left total machine (see Figure 1) 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 [18, 5] introduced the notion of using the prefix of the border sequence to define strings into a two part code. This paper uses lemmas found in [2].

This paper can be seen as a conditional variant to the main result in [10]. In [10], it was proved for non exotic sets D, the a-priori probability, $\mathbf{m}$, of a set is concentrated on a single element.

Theorem. ([10]) $-\log\max_{x\in D}\mathbf{m}(x) <^{\log} -\log\sum_{x\in D}\mathbf{m}(x) + \mathbf{I}(D;\mathcal{H})$.

[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 $\log m < (n-c)/2 + O(\log n + \mathbf{I}(\langle P\rangle;\mathcal{H}))$.

There is a simple proof to this theorem in [16]. The proof of Theorem 1 is similar to that of the main result in [10], in that they both first prove stochasticity, $\mathbf{Ks}(O)$, of a an object O with certain properties and then show that this object has high $\mathbf{I}(O;\mathcal{H})$. In [10], O is equal to a set, and in this paper, O is equal to a (sub)graph. Theorem 2 is not direclty implied by the theorem in [10] because this paper deals with conditional complexities between elements of a set. In addition, Theorem 2 is not a generalization of the main theorem in [10] because it relies on the parameters of bunches and not the a-priori probability $\mathbf{m}$.

3 Conventions

We use $\{0,1\}$, $\{0,1\}^*$, $\{0,1\}^\infty$ $\mathbb{N}$, $\mathbb{Q}$, 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. $\{0,1\}^{*\infty} = \{0,1\}^* \cup \{0,1\}^\infty$. The positive part of a real is $\lceil a\rceil^+ = \max\{a,0\}$. For string $x \in \{0,1\}^*$, $x0^- = x1^- = x$. For $x \in \{0,1\}^*$ and $y \in \{0,1\}^{*\infty}$, we use $x \sqsubset y$ if there is some string $z \in \{0,1\}^{*\infty}$ where $xz = y$. The indicator 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 $x \in \{0,1\}^*$ is $\langle x\rangle = 1^{\|x\|}0x$. The encoding of (a possibly ordered) set $\{x_1,\dots,x_m\} \subset \{0,1\}^*$, is $\langle m\rangle\langle x_1\rangle\dots\langle x_m\rangle$.

Probability measures Q over numbers are elementary if $|\mathrm{Support}(Q)| < \infty$ and $\mathrm{Range}(Q) \subset Q_{\geq 0}$. Elementary probability measures Q with $\{x_1,\dots,x_m\} = \mathrm{Support}(Q)$ are encoded by finite strings, with $\langle Q\rangle = \langle\{x_1, Q(x_1),\dots,x_m,Q(x_m)\}\rangle$. 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, where if $x \in \{0,1\}^*$, $y \in \{0,1\}^{*\infty}$, then $\mathbf{K}(x/y) = \min\{\|p\| : U_y(p) = x\}$. 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 $\mathcal{H} \in \{0,1\}^\infty$ is the unique infinite sequence where $\mathcal{H}[i] = [U(i)\text{ halts}]$. The information that $x \in \{0,1\}^*$ has about $\mathcal{H}$, conditional to $y \in \{0,1\}^{*\infty}$, is $\mathbf{I}(x;\mathcal{H}/y) = \mathbf{K}(x/y) - \mathbf{K}(x/\langle y,\mathcal{H}\rangle)$. $\mathbf{I}(x;\mathcal{H}) = \mathbf{I}(x;\mathcal{H}/\emptyset)$.

This paper uses notions of stochasticity in the field of algorithmic statisics [19]. A string x is stochastic, i.e. has a low $\mathbf{Ks}(x)$ score, if it is typical of a simple probability distribution. The extended deficiency of randomness function of a string x with respect to an elementary probability measure P conditional to $y \in \{0,1\}^*$, is $\mathbf{d}(x|P,y) = \lfloor -\log P(x)\rfloor - \mathbf{K}(x/\langle P\rangle, y)$. $\mathbf{d}(x|P) = \lfloor -\log P(x)\rfloor - \mathbf{K}(x/\langle P\rangle)$

Definition 1 (Stochasticity) *For $x,y \in \{0,1\}^*$, $\mathbf{Ks}(x/y) = \min\{\mathbf{K}(P/y) + 3\log\max\{\mathbf{d}(x|P,y),1\} :$ P is an elementary probability measure$\}$. $\mathbf{Ks}(x) = \mathbf{Ks}(x/\emptyset)$.*

4 Labelled Graphs, Warmup

In Section 5, a property of a complete subgraph of a labelled 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 vertices, and subgraph $J = (J_E, V_V)$, Theorem 1 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 specify 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 complete subgraph $J \subseteq G$, then

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

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

The warm up arguments are as follows. Let $G = (G_E, G_V)$ 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 $l > 2k$. Each vertex $x \in G_V$ 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 $\Pr(\emptyset = L_x[i]) = 1 - \mathrm{OutDegree}(x)2^{-l}$. For $J \in \mathcal{J}$, indexed list $L_\bullet$,

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

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

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

Now assume that $|L_x| = b2^{l-2k}$ for all $x \in G_V$, i.e. b times more than before. It is not hard to see that $\Pr(\mathrm{Miss}(J, L_\bullet)) < e^{-b}$ for each $J \in \mathcal{J}$. We assume a uniform distribution $\mathcal{U}$ over $\mathcal{J}$. Under this assumption,

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

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 $\mathrm{Miss}(J, L'_\bullet)$. If $\mathrm{Miss}(J, L'_\bullet)$ is true for $J \in \mathcal{J}$, then it must be atypical of $\mathcal{U}$, because $\mathbf{E}_{J\sim\mathcal{U}}[\mathrm{Miss}(J, L'_\bullet)] < e^{-b}$. One can construct a $\mathcal{U}$-test using $L'_\bullet$. A $\mathcal{U}$ test is any function $t : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ such that $\sum_{J\in\mathcal{J}} t(J)\mathcal{U}(J) \leq 1$. Thus $t \cdot \mathcal{U}$ is a semi-measure and so

$$\mathbf{K}(J/t, \mathcal{U}) <^+ -\log t(J)\mathcal{U}(J). \tag{1}$$

Thus the function $t(J) = [\mathrm{Miss}(J, L'_\bullet)]e^b$ is a $\mathcal{U}$-test, with $\sum_{J\in\mathcal{J}} t(J)\mathcal{U}(J) < 1$. We set aside the parameters $(G, k, l, b, \mathcal{U})$ because they complicate the discussion. That is, we roll the parameters into the additive constants of the inequalities. By the definition of randomness deficiency,

$$
\begin{aligned}
\mathbf{d}(J|\mathcal{U}) &= -\log \mathcal{U}(J) - \mathbf{K}(J/\mathcal{U}) \\
&>^+ \log|\mathcal{J}| - \mathbf{K}(J/L'_\bullet) && (2)\\
&>^+ \log|\mathcal{J}| - \mathbf{K}(J/t) && (3)\\
&>^+ \log|\mathcal{J}| + \log t(J)\mathcal{U}(J) && (4)\\
&>^+ \log|\mathcal{J}| + \log t(J)|\mathcal{J}|^{-1} \\
&>^+ b\log e.
\end{aligned}
$$

Equation 2 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 G of size 2^k. The second term is due to the additive equalities

$$\mathbf{K}(J/L'_\bullet) = \mathbf{K}(J/L'_\bullet, G, k, l, b, \mathcal{U}) =^+ \mathbf{K}(J/G, k, l, b, \mathcal{U}) =^+ \mathbf{K}(J/\mathcal{U}),$$

because given all the hidden parameters $(G, k, l, b, \mathcal{U})$, one can compute $L'_\bullet$ using brute force search, as described above. Equation 3 is due to the fact that the test t is constructed from $L'_\bullet$ (and the hidden parameters). Equation 4 is due to the properties of tests, as shown in Equation 1.

Thus all complete subgraphs $J \in \mathcal{J}$ of G for which $\mathrm{Miss}(J, L'_\bullet)$ is true will be atypical of $\mathcal{U}$, with randomness deficiency $\mathbf{d}(J|\mathcal{U})$ 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 subgraphs $J \in \mathcal{J}$ will have

$$\min_{(x,y)\in J_E} \mathbf{K}(y/x) <^+ \log|L_x| <^+ l - 2k + \log b. \tag{5}$$

For Theorem 1, the uniform probability measure $\mathcal{U}$ is replaced by a special computable measure P that realizes the stochasticity, $\mathbf{Ks}$, 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 $\mathrm{Miss}(J)$ is false. This means Equation 5 holds for J. In addition, in the next section, the parameters (G, k, l, b) must be taken into account.

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 $(x, y) \in J_E$, such that the conditional complexity $\mathbf{K}(y/x)$ is high. The proof of the following theorem uses stochasticity $\mathbf{Ks}$. An example proof that uses $\mathbf{Ks}$ and mirrors the proof of Theorem 1 can be found in Appendix B. Note that the lemma in Appendix B is just an exercise to demonstrate reasoning with $\mathbf{Ks}$. The lemma is not used in the paper.

Theorem 1 *For graph G, complete subgraph J, if $2^l > \max \mathrm{Outdegree}(G)$, $2^k < |J|$, then we have $\min_{(x,y)\in J} \mathbf{K}(y/x) <^{\log} \lceil l - 2k\rceil^+ + \mathbf{I}(J; \mathcal{H}/G, k) + \mathbf{K}(G, k)$.*

Proof. We put (G, k) on an auxiliary tape to the universal Turing machine U. Thus all algorithms have access to (G, k) and all complexities implicitly have (G, k) as conditional terms.

Let $\ell = \max\{l, 2k\}$. Let P be the probability that realizes $\mathbf{Ks}(J)$ and the deficiency of randomness $d = \max\{\mathbf{d}(J|P), 1\}$. 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 $V(\emptyset|x) = 1 - \mathrm{OutDegree}(x)2^{-\ell}$. 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 $\kappa(L|x) = \prod_{y\in L} V(y|x)$. The constant $c \in \mathbb{N}$ will be determined later. Let $L_\bullet$ be an indexed

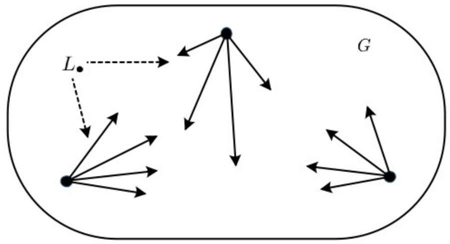

Figure 2: The above diagram is a graphical representation of κ and $L_\bullet$, assuming that $cd2^{\ell-2k} = 4$. Each vertex has four edges chosen at random, where each particular edge is chosen with probability $2^{-\ell}$.

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 $\kappa(L_\bullet) = \prod_{x\in G} \kappa(L_x|x)$. A graphical representation of κ and $L_\bullet$ can be found in Figure 2. For indexed list $L_\bullet$, graph $H = (H_E, H_V)$, we use the following indicator $\mathbf{i}(L_\bullet, H) = [\text{Complete } H \subseteq G, 2^k < |H_V|, \forall (x,y) \in H_E, y \notin L_x]$.

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

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

$$\mathbf{K}(L'_\bullet/P, c, d) = O(1). \tag{6}$$

Thus $t(H) = \mathbf{i}(L'_\bullet, H)e^{cd}$ is a P test, where $\mathbf{E}_{H\sim P}[t(H)] \leq 1$. This test t gives a high score to complete subgraphs of G of size $> 2^k$ which have no intersecting edges with $L'_\bullet$. A diagram of the components used in this proof can be found in Figure 3. Furthermore

$$\mathbf{K}(t|P, c, d) =^+ \mathbf{K}(t|L'_\bullet, P, c, d) = O(1).$$

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

$$\begin{aligned}
\mathbf{K}(J/P, c, d) &<^+ \mathbf{K}(J/t, P, c, d) \\
\mathbf{K}(J/P, c, d) &<^+ -\log t(J)P(J) \\
&<^+ -(\log e)cd - \log P(J) \\
(\log e)cd &<^+ -\log P(J) - \mathbf{K}(J/P, c, d) \\
(\log e)cd &<^+ -\log P(J) - \mathbf{K}(J/P) + \mathbf{K}(c, d) \\
(\log e)cd &<^+ d + \mathbf{K}(c, d),
\end{aligned} \tag{7}$$

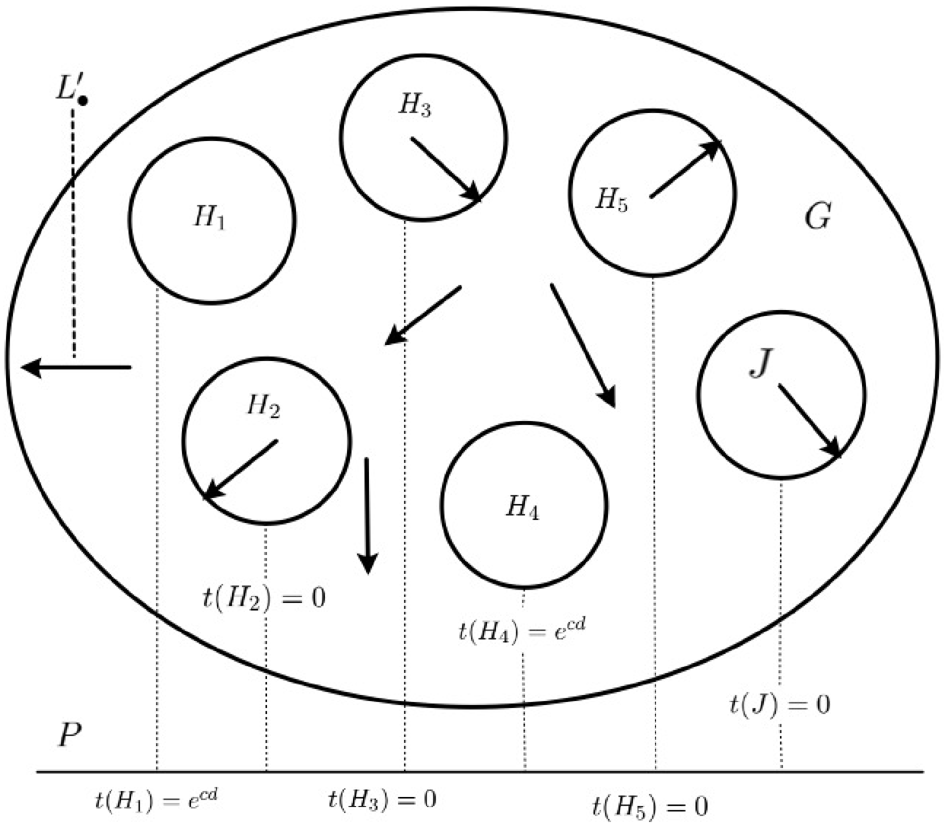

Figure 3: The above diagram is a graphical representation of the concepts used in the proof of Theorem 1. The main ellipse models the graph G and the circles in the graph represent complete subgraphs (labeled H_1 to H_5 and also J) with $> 2^k$ vertices. Each subgraph is in the support of probability 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 $L'_\bullet$.

which is a contradiction for large enough c solely dependent on the universal Turing machine U. Equation 7 is due to Equation 1. The constant c is folded into the additive constants of the inequalities of the rest of the proof. Thus since there exists $(x, y) \in J_E$ where $y \in L_x$,

$$\begin{aligned}
\mathbf{K}(y/x) &<^+ \log|L'_x| + \mathbf{K}(L'_\bullet) \\
&<^+ \lceil l - 2k\rceil^+ + \log d + \mathbf{K}(L'_\bullet/P, d) + \mathbf{K}(P, d) \\
&<^+ \lceil l - 2k\rceil^+ + \log d + \mathbf{K}(P, d) \\
&<^+ \lceil l - 2k\rceil^+ + 3\log d + \mathbf{K}(P) \\
&<^+ \lceil l - 2k\rceil^+ + \mathbf{Ks}(D)
\end{aligned} \tag{8}$$

Equation 8 is due to Equation 6. We now make the relativization of (G, k) explicit, with

$$\begin{aligned}
\mathbf{K}(y/x, G, k) &<^+ \lceil l - 2k\rceil^+ + \mathbf{Ks}(J/G, k) \\
&<^{\log} \lceil l - 2k\rceil^+ + \mathbf{I}(J; \mathcal{H}/G, k) \\
\mathbf{K}(y/x) &<^{\log} \lceil l - 2k\rceil^+ + \mathbf{I}(J; \mathcal{H}/G, k) + \mathbf{K}(G, k).
\end{aligned} \tag{9}$$

Equation 9 is due to Lemma 10 in [2], which states $\mathbf{Ks}(x) < \mathbf{I}(x:\mathcal{H}) + O(\mathbf{K}(\mathbf{I}(x:\mathcal{H})))$. $\square$

6 Warm Up for the 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 $X \subset \{0,1\}^*$ be a (k,l)-bunch, where $|X| > 2^k$, and $\max_{x,y\in X} \mathbf{K}(y/x) < 2^{-l}$.

Let $\mathbf{K}^r(x/y) = \min\{\|p\| : U_y(p) = x \text{ in time } r\}$ 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 $G^r = (G^r_E, G^r_V)$ be a graph defined by $(x,y) \in G^r_E$ iff $\mathbf{K}^r(x/y) < l$.

Let s be the smallest number where $\mathbf{K}^s(x/y) < l$, for all $x,y \in X$. Let $G = (G_E, G_V) = G^s$. Since X is a (k,l)-bunch, X can be viewed as a complete subgraph of G of size $>2^k$. Invoking Theorem 1, we get

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

We have $\mathbf{K}(s/G) <^+ \mathbf{K}(l)$ because $s = \min\{r : G = G^r\}$. So

$$\mathbf{K}(X/G) <^+ \mathbf{K}(X/s) + \mathbf{K}(s/G) <^+ \mathbf{K}(X/s) + \mathbf{K}(l). \tag{11}$$

Due to the definition of $G = G^s$,

$$\mathbf{K}(G/s) <^+ \mathbf{K}(l). \tag{12}$$

By the definition of $\mathbf{I}$,

$$\begin{aligned}
\mathbf{I}(X;\mathcal{H}/G,k) &= \mathbf{K}(X/G,k) - \mathbf{K}(X/G,k,\mathcal{H}) \\
&= \mathbf{K}(X/G) - \mathbf{K}(X/G,\mathcal{H}) + O(\mathbf{K}(k)) \\
&<^+ \mathbf{K}(X/s) - \mathbf{K}(X/G,\mathcal{H}) + O(\mathbf{K}(k,l)) \qquad (13) \\
&< \mathbf{K}(X/s) - \mathbf{K}(X/s,\mathcal{H}) + \mathbf{K}(G/s) + O(\mathbf{K}(k,l)) \\
&< \mathbf{I}(X;\mathcal{H}/s) + O(\mathbf{K}(k,l)). \qquad (14)
\end{aligned}$$

Equation 13 is due to Equation 11. Equation 14 is due to Equation 12. Using $\mathbf{K}(G) <^+ \mathbf{K}(s) + \mathbf{K}(l)$ and Equation 14, we get

$$\mathbf{I}(X;\mathcal{H}/G,k) + \mathbf{K}(G,k) < \mathbf{I}(X;\mathcal{H}/s) + \mathbf{K}(s) + O(\mathbf{K}(k,l)). \tag{15}$$

Combining Equations 10 and 15, we get

$$\min_{(x,y)\in J} \mathbf{K}(y/x) <^{\log} \lceil l - 2k\rceil^+ + \mathbf{I}(X;\mathcal{H}/s) + \mathbf{K}(s) + O(\mathbf{K}(k,l)). \tag{16}$$

This inequality is close to the form of Theorem 2. The main difference is that the number r appears in Equation 16. This can be rectified if we use a different notion of a computational resource. In the next section we introduce left-total universal machines, and the resource used is not a number s, 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 $x \in \{0,1\}^*$, $\Gamma_x = \{x\beta : \beta \in \{0,1\}^\infty\}$ is the interval of x. The notions of total strings and the "left-total" universal algorithm are needed in this paper. We say $x \in \{0,1\}^*$ 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 $y \in \{0,1\}^{*\infty}$ iff there exists a finite prefix free set of strings $Z \subset \{0,1\}^*$ 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 $\alpha \in \{0,1\}^{*\infty}$ is to the "left" of $\beta \in \{0,1\}^{*\infty}$, and use the notation $\alpha \lhd \beta$, if there exists a $x \in \{0,1\}^*$ such that $x0 \sqsubseteq \alpha$ and $x1 \sqsubseteq \beta$. A machine T is left-total if for all auxiliary strings $\alpha \in \{0,1\}^{*\infty}$ and for all $x, y \in \{0,1\}^*$ 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 [10]. An example can be seen in Figure 4.

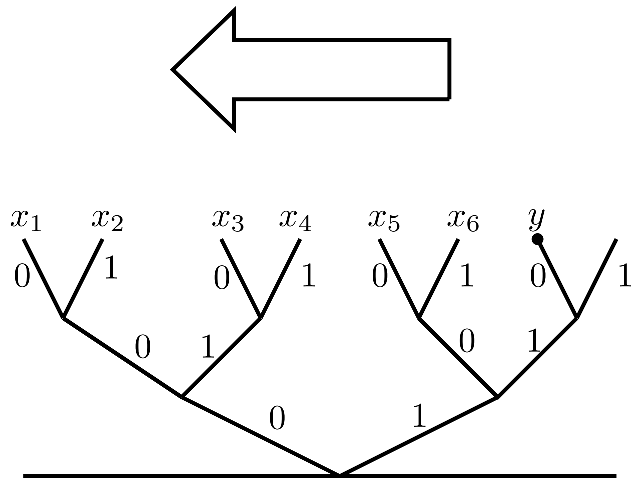

Figure 4: The above diagram represents the domain of a left total machine T with the 0 bits branching to the left and the 1 bits branching to the right. For $i \in \{1,\dots,5\}$, $x_i \lhd x_{i+1}$ and $x_i \lhd y$. Assuming $T(y)$ halts, each x_i is total. This also implies each x_i^- is total as well.

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 $p \in \{0,1\}^*$ 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 $p \in \{0,1\}^*$ and $q \in \{0,1\}^*$ 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 $p \in \{0,1\}^*$ 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 $p \in \{0,1\}^*$ 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 $\alpha \in \{0,1\}^{*\infty}$.

Recall that a function $f : \mathbb{N} \to \mathbb{N}$ is partial computable with respect to U if there is a string $t \in \{0,1\}^*$ 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 $t \in \{0,1\}^*$, 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 [10].

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 $t \in \{0,1\}^*$, 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 $\mathcal{B}$.*

3. *If a string b is total and b^- is not, then $b^- \sqsubset \mathcal{B}$.*

Proof.

1. Let $\Omega \in \mathbb{R}$ be the Chaitin's Omega, the probability that a random sequence of bits halts when given to U, with $\Omega = \sum_{p \in \{0,1\}^*} [U(p) \text{ halts}] 2^{-\|p\|}$. Thus Ω characterizes the domain of U, with $\bigcup_{p \in \{0,1\}^*} i_p = [0, \Omega)$. Let $\mathcal{B} \in \{0,1\}^\infty$ be the binary expansion of Ω, which is a ML random string. For each $n \in \mathbb{N}$, let $b_n \sqsubset \mathcal{B}$, $\|b_n\| = n$. Let $m \in \mathbb{W}$ be the smallest whole number such that $b_n 1^{(m)} 0 \sqsubset \mathcal{B}$. Then $b_n 1^{(m+1)}$ is a non-total string because $[0, \Omega] \cap \Gamma(b_n 1^{(m+1)}) = \emptyset$. Furthermore let $m \in \mathbb{W}$ be the smallest whole number such that $b_n 0^{(m)} 1 \sqsubset \mathcal{B}$. Then $b_n 0^{(m+1)}$ is a total string because $\Gamma(b_n 0^{(m+1)}) \subset [0, \Omega)$.

2. Assume there are two strings x and y of length n that have total and non-total extensions, with $x \lhd y$. Since y has total extensions, there exist z such that $U'(yz)$ halts. Since $x \lhd yz$, by the definition of left-total machines, x is total, causing a contradiction.

3. This is due to the fact that b^- has total and non-total extensions.

□

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 $\mathcal{H}$.

Lemma 1 ([2]) *If $b \in \{0,1\}^*$ is total and b^- is not, and $x \in \{0,1\}^*$, then* $\mathbf{K}(b) + \mathbf{I}(x; \mathcal{H}/b) <^{\log} \mathbf{I}(x\,; \mathcal{H}) + \mathbf{K}(b/\langle x, \|b\|\rangle)$.

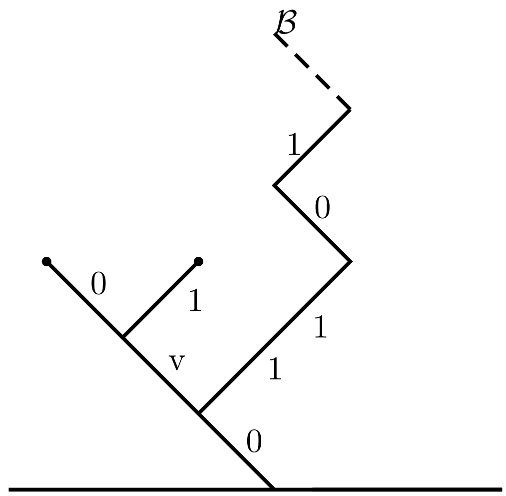

Figure 5: The above diagram represents the domain of the universal left-total algorithm U', with the 0 bits branching to the left and the 1 bits branching to the right. The strings in the above diagram, $0v0$ and $0v1$, are halting inputs to U' with $U(0v0) \neq\perp$ and $U(0v1) \neq\perp$. So $0v$ is a total string. The infinite border sequence $\mathcal{B} \in \{0,1\}^\infty$ 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.

We call this infinite sequence $\mathcal{B}$, "border" because for any string $x \in \{0,1\}^*$, $x \lhd \mathcal{B}$ 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 5 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 $y \in \{0,1\}^{*\infty}$, is $\mathbf{K}_b(x/y) = \min\{\|p\| : U_y(p) = x \text{ in } \mathbf{bb}(b) \text{ time and} \|y\| \leq \mathbf{bb}(b)\}$. If b and c are total, and $b \lhd c$, then $\mathbf{K}_b \geq \mathbf{K}_c$. In addition, if b and b^- are total, then $\mathbf{K}_b \geq \mathbf{K}_{b^-}$.

8 Minimum Conditional Complexity

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

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

$$\mathbf{K}(b/X) <^+ \mathbf{K}(\|b\|, l), \tag{17}$$

as there is a program that when enumerating total strings of length $\|b\|$ from left to right, returns the first one with the desired properties. The first total string found is b, as shown in Figure 6.

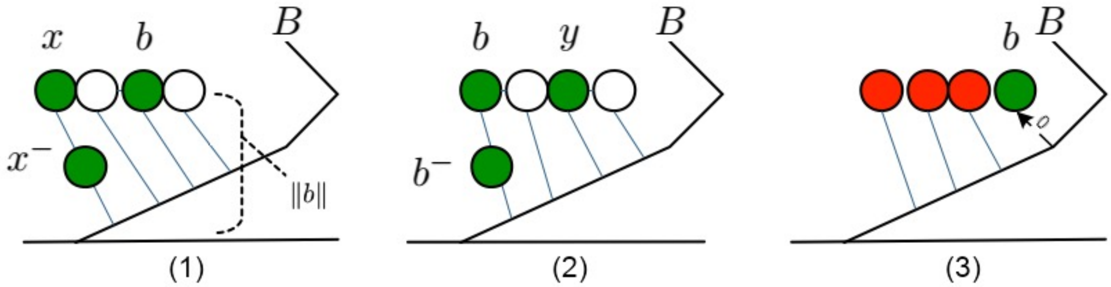

Figure 6: The above diagram represents the domain of the universal left-total Turing machine U and uses the same conventions as Figure 5, with 0s branching to the left and 1s branching to the right. It shows all the total strings of length $\|b\|$, including b. The large diagonal line is the border sequence, 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. If x is green and $x \lhd y$, then y is green, since $\mathbf{K}_x \geq \mathbf{K}_y$. Furthermore, if x is green and total and x^- is total, then xm is green, as $\mathbf{K}_x \geq \mathbf{K}_{x^-}$. It cannot be that there is a green $x \lhd b$ with $\|x\| = \|b\|$. Otherwise x^- is total, and thus it is green, causing a contradiction because it is shorter than b. This is shown in part (1). Furthermore, there can't be a green y, with $b \lhd y$ and $\|y\| = \|b\|$. Otherwise b^- is total and thus it is green, contradicting the definition of b. This is shown in part (2). Thus b is unique, and since b^- is not total, by Proposition 2, b^- is a prefix of border, as shown in part (3). Thus an algorithm returning a green string of length $\|b\|$ will return b.

Thus b^- is not total, and by Proposition 2, $b^- \sqsubset B$ 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 $G = (G_E, G_V) = G^b$. We have

$$\mathbf{K}(G/b) <^+ \mathbf{K}(l) \tag{18}$$
$$\mathbf{K}(b/G) <^+ \mathbf{K}(\|b\|, l). \tag{19}$$

Equation 18 is because $G = G^b$. Equation 19 is due to the existence of a program that enumerates total strings of length $\|b\|$ (from left to right) 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 the definition of b being the shortest total string with $G^b \supseteq X$. Thus using this impossibility and the reasoning detailed in Figure 6, where y is green if $G \subseteq G^y$, the program returns b. Theorem 1, gives $x, y \in X$, where

$$\mathbf{K}(y/x) <^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X; H/G, k) + \mathbf{K}(G, k) \tag{20}$$

The rest of the proof is a straightforward sequence of application of inequalities. We have

$$\begin{aligned}\mathbf{K}(X/G) &<^+ \mathbf{K}(X/b) + \mathbf{K}(b/G) \\ &<^+ \mathbf{K}(X/b) + \mathbf{K}(\|b\|, l),\end{aligned} \tag{21}$$

where Equation 21 is due to Equation 19. We also have

$$\begin{aligned}\mathbf{K}(X/b, \mathcal{H}) &< \mathbf{K}(X/G, \mathcal{H}) + \mathbf{K}(G/b, \mathcal{H}), \\ &< \mathbf{K}(X/G, \mathcal{H}) + \mathbf{K}(l),\end{aligned} \tag{22}$$

where Equation 22 is due to Equation 18. So

$$\begin{aligned}\mathbf{I}(X;\mathcal{H}/G) &= \mathbf{K}(X/G) - \mathbf{K}(X/G,\mathcal{H})\\ &<^{+} \mathbf{I}(X;\mathcal{H}/b) + \mathbf{K}(l) + \mathbf{K}(\|b\|,l). \qquad (23)\end{aligned}$$

Combining Equations 20 and 23,

$$\begin{aligned}\mathbf{K}(y/x) &<^{\log} \lceil l-2k\rceil^{+} + \mathbf{I}(X;\mathcal{H}/b) + \mathbf{K}(G) + \mathbf{K}(\|b\|) + O(\mathbf{K}(k,l))\\ &<^{\log} \lceil l-2k\rceil^{+} + \mathbf{I}(X;\mathcal{H}/b) + \mathbf{K}(b) + \mathbf{K}(\|b\|) + O(\mathbf{K}(k,l)) \qquad (24)\\ &<^{\log} \lceil l-2k\rceil^{+} + \mathbf{I}(X;\mathcal{H}/b) + \mathbf{K}(b) + O(\mathbf{K}(k,l)). \qquad (25)\end{aligned}$$

Equation 24 is due to Equation 18. In Equation 25 is due to the fact that the precision is ($<^{\log}$). Furthermore, since b is total and b^{-} is not, by Proposition 2, $b^{-} \sqsubset B$. The border B is the binary expansion of Chaitin's Omega (see Proposition 2), so b is random, with $\mathbf{K}(\|b\|) = O(\log \mathbf{K}(b))$. Using Lemma 1 on Equation 25, we get

$$\begin{aligned}\mathbf{K}(y/x) &<^{\log} \lceil l-2k\rceil^{+} + \mathbf{I}(X;\mathcal{H}) + \mathbf{K}(b/X,\|b\|) + O(\mathbf{K}(k,l))\\ &<^{\log} \lceil l-2k\rceil^{+} + \mathbf{I}(X;\mathcal{H}) + O(\mathbf{K}(k,l)) \qquad (26)\end{aligned}$$

where Equation 26 is due to Equation 17. Adding (k,l) to the conditional on all terms results in

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

□

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 [9], 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 program q that computes probability p over $\mathbb{N}$, $\mathbf{E}_{a\sim p}\left[2^{\mathbf{I}(\langle q,a\rangle;\mathcal{H})}\right] \overset{*}{<} 2^{\mathbf{I}(q;\mathcal{H})}$.*

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

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 ($\mathbf{K}(x,y) =^+ \mathbf{K}(x) + \mathbf{K}(y/\mathbf{K}(x),x)$) 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 $\mathbb{N}$, $\mathbf{E}_{a\sim p}[2^{\mathbf{I}(a;\mathcal{H})}] \overset{*}{<} 2^{\mathbf{I}(p;\mathcal{H})}$.*

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

B Warmup Exercise in Stochasticity

The following proof demonstrates how the stochasticity term $\mathbf{Ks}$ can be used in mathematical arguments. The general structure of the proof parallels the proof in Theorem 1. This lemma first appeared (in a slightly different form) as Lemma 5 in [4]. The lemma itself in just an exercise, and is not used in the paper.

Lemma 2 *For $D\subseteq\{0,1\}^n$, $|D|=2^s$, $\min_{x\in D}\mathbf{K}(x) <^{\log} n - s + \mathbf{Ks}(D) + O(\mathbf{K}(s,n))$.*

Proof. We put (n,s) on an auxiliary tape to the universal Turing machine U. Thus all algorithms have access to (n,s) and all complexities implicitly have (n,s) as conditional terms. This can be done because the precision of the lemma is $O(\mathbf{K}(s,n))$. Let Q realize $\mathbf{Ks}(D)$, with $d = \max\{\mathbf{d}(D|Q),1\}$. Thus Q is an elementary probability measure over $\{0,1\}^*$ and $D \in \mathrm{Support}(Q)$, with randomness deficiency d.

Let $F \subseteq \{0,1\}^n$ be a random set where each element $a \in \{0,1\}^n$ is selected independently with probability $cd2^{-s}$, where $c \in \mathbb{N}$ is chosen later. Let $\mathcal{U}_n$ be the uniform measure over $\{0,1\}^n$. $\mathbf{E}[\mathcal{U}_n(F)] \leq cd2^{-s}$. Furthermore

$$\mathbf{E}[Q(\{G : |G| = 2^s, G\cap F = \emptyset\})] \leq \sum_G Q(G)(1-cd2^{-s})^{2^s} < e^{-cd}.$$

Thus by the Markov inequality, $W \subseteq \{0,1\}^n$ can be chosen such that $\mathcal{U}_n(W) \leq 2cd2^{-s}$ and $Q(\{G : |G| = 2^s, G\cap W = \emptyset\}) \leq e^{1-cd}$.

$$\mathbf{K}(W/Q,d,c) = O(1). \tag{27}$$

It must $D\cap W \neq \emptyset$. Otherwise, we get a contradiction with the following reasoning. Let $t : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ be a Q-test, with $t(G) = [|G| = 2^s, G\cap W = \emptyset]e^{cd-1}$, and $\sum_G Q(G)t(G) \leq 1$. Thus

t gives a high score to sets G which do not intersect W. So $t(D) = e^{cd-1}$. We have

$$\begin{aligned}
\mathbf{K}(D/Q,d,c) &<^+ \mathbf{K}(D/W,Q,d,c) && (28)\\
&<^+ \mathbf{K}(D/t,W,Q,d,c) && (29)\\
&<^+ -\log Q(D)t(D) && (30)\\
&<^+ -\log Q(D) - (\log e)cd\\
(\log e)cd &<^+ -\log Q(D) - \mathbf{K}(D/Q) + \mathbf{K}(d,c)\\
&<^+ d + \mathbf{K}(d,c),
\end{aligned}$$

which is a contradiction for large enough c dependent solely on the universal Turing machine. Equation 28 is due to Equation 27. Equation 29 is because the test t can be computed from (W,c,d). Equation 30 is due to Equation 1. Thus there is an $x \in D \cap W$. Thus since $\mathcal{U}_n(W) \leq 2cd2^{-s}$, the function $q(a) = [a \in W](2^s/cd)\mathcal{U}_n(a)$ is a semi-measure. So we have

$$\mathbf{K}(x) <^+ -\log q(x) + \mathbf{K}(q) <^+ n + \log d - s + \mathbf{K}(d) + \mathbf{K}(Q) <^+ n - s + \mathbf{Ks}(D).$$

□

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] Samuel Epstein. A note on the outliers theorem. *CoRR*, 2203.08733, 2021. URL https://arxiv.org/abs/2203.08733. v2.

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

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

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

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

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

[10] 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.

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

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

[13] 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.

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

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

[16] 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.

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

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

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

[20] 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.

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

[22] 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.