On Kolmogorov Structure Functions

Sam Epstein*

June 6, 2024

Abstract

All strings with low mutual information with the halting sequence will have flat Kolmogorov Structure Functions, in the context of Algorithmic Statistics. Assuming the Independence Postulate, strings with non-negligible information with the halting sequence are purely mathematical constructions, and cannot be found in nature. Thus Algorithmic Statistics does not study strings in the physical world. We also discuss issues with set-restricted Kolmogorov Structure Functions.

1 Introduction

In statistics, one tries to determine a model (such as a parameter for a distribution) from data which is assumed to have noise. In the Minimum Description Principle [Gru07], the model that describes information with the shortest code is assumed to be the best model. The data is described as a two part code, where the first part is the model and the second part is the noise. In one of his last works, Kolmogorov suggested a two part code for individual strings $x \in \{0,1\}^*$ based off Kolmogorov Complexity. The first part (the model) is a set D containing x, the second part (the noise) is the code of x given D, of size $\lceil \log |D| \rceil$. Other works examined probabilities and also total computable functions as models [Vit02]. Kolmogorov suggested the following *structure function* at the Tallinn conference in Estonia, 1973.

$$\mathbf{H}_k(x) = \min\{\log |S| : x \in S, \mathbf{K}(S) \leq k\}.$$

The function $\mathbf{K}$ is the prefix Kolmogorov complexity. Theorem 1 of [VS17] showed that any shape of the structure function is possible. This definition is used for the following function, which is a central definition of *Algorithmic Statistics* [VS15, VS17, VV04a],

$$k \mapsto k + \mathbf{H}_k(x) - \mathbf{K}(x).$$

This function's equivalence to several other definitions is the main theorem of Algorithmic Statistics [SSV24].

The structure function is flat for all strings with low mutual information with the halting sequence. Assuming the *Independence Postulate*, [Lev84, Lev13],

*samepst@jptheorygroup.org

strings with non-negligible mutual information with the halting sequence are exotic, in that they cannot be found in nature. Such strings are purely mathematical constructions.

2 Bounds

We review the results of [GTV01], in particular Theorem of III.24, which I don't think is widely known. $\mathbf{m}(x)$ is the algorithmic probability. The amount of information that the halting sequence $\mathcal{H} \in \{0,1\}^\infty$ has about $x \in \{0,1\}^*$ is $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$. We use $x <^+ y$, $x >^+ y$ and $x =^+ y$ to denote $x < y + O(1)$, $x + O(1) > y$ and $x = y \pm O(1)$, respectively. In addition, $x <^{\log} y$ and $x >^{\log} y$ denote $x < y + O(\log y)$ and $x + O(\log x) > y$, respectively. Furthermore, $\overset{*}{<} f$, $\overset{*}{>} f$ denotes $< O(1)f$ and $> f/O(1)$. For $x, y \in \{0,1\}^*$, $x \sqsubseteq y$ if $y = xz$ for some $z \in \{0,1\}^*$. $[A] = 1$ if mathematical statement A is true, and $[A] = 0$ otherwise.

Let $S_k = \{x : \mathbf{K}(x) \leq k\}$. Let $N_k = |S_k|$ where $\log N_k =^+ k - \mathbf{K}(k)$, due to [GTV01]. Let I_k^x be the index of x in an enumeration of S_k. For $\mathbf{K}(x) = k$, let m_x be the longest joint prefix of I_k^x and N_k. So $m_x 0 \sqsubseteq I_k^x$ and $m_x 1 \sqsubseteq N_k$. Let $S_x = \{y : m_x 0 \sqsubseteq I_k^y\}$. So

$$\begin{aligned}
\log |S_x| &=^+ k - \mathbf{K}(k) - \|m_x\| \\
\mathbf{K}(S_x) &<^+ \mathbf{K}(k) + \mathbf{K}(m_x) <^+ \mathbf{K}(k) + \|m_x\| + \mathbf{K}(\|m_x\|).
\end{aligned}$$

Theorem 1 ([GTV01]).

$$\|m_x\| < \mathbf{K}(\mathbf{K}(x)) + \mathbf{I}(x;\mathcal{H}) + O(\log \mathbf{I}(x;\mathcal{H})).$$

Proof. Let $\nu(y) = c[\mathbf{K}(y) \leq k]\mathbf{m}(y)2^{\|m_y\|}/(\|m_y\|^2)$. For proper choice of c, ν is a semimeasure and computable relative to $\mathcal{H}$ and k. So $\mathbf{K}(x|\mathcal{H},k) <^+ -\log \nu(x) =^+ \mathbf{K}(x) - \|m_x\| + 2\log \|m_x\|$. $\square$

Note that with some additional effort, the $\mathbf{K}(\mathbf{K}(x))$ term can be eliminated.

Corollary 1. *For $x \in \{0,1\}^*$, $n = \mathbf{K}(x)$, for all $m \leq n$, $m \in \mathbb{W}$, there is a set $S \ni x$ such that $|S| = 2^m$ and $\mathbf{K}(S) + m <^{\log} n + \mathbf{I}(x;\mathcal{H})$.*

Claim 1. *Thus there exists a set $S \ni x$ such that $\mathbf{K}(S) <^{\log} 2\mathbf{K}(\mathbf{K}(x)) + \mathbf{I}(x;\mathcal{H})$ and $\mathbf{K}(S) + \log |S| <^+ \mathbf{K}(x) + \mathbf{K}(\mathbf{K}(x)) + O(\log(\mathbf{I}(x;\mathcal{H}) + \mathbf{K}(\mathbf{K}(x))))$. This fact combined with the following proposition characterizes the Kolmogorov Structure Function.*

Proposition 1. *Let $S \ni x$. For all $s < \log |S|$ there exists a set $S' \ni x$ such that $|S'| \leq |S| 2^{-s}$ and $\mathbf{K}(S') <^+ \mathbf{K}(S) + s + \mathbf{K}(s)$.*

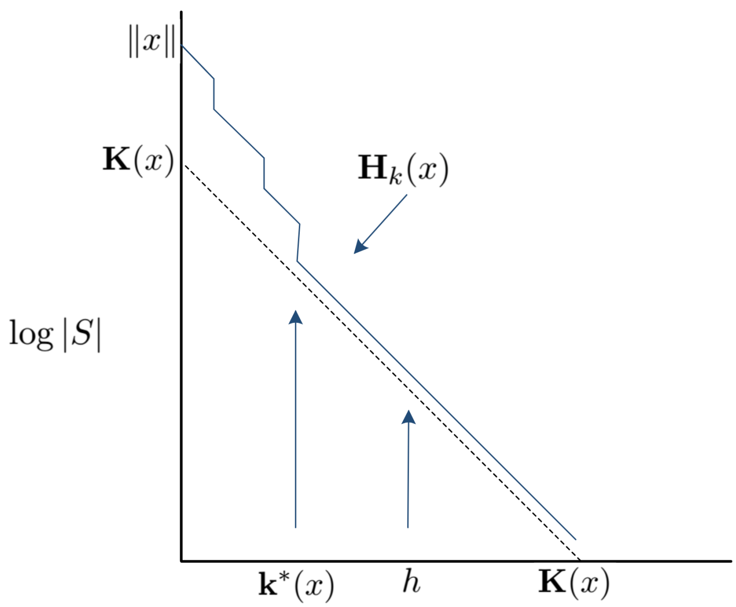

Figure 1: A visual representation of the Kolmogorov Structure Function $\mathbf{H}_k(x)$. The amount of information that the halting sequence has about x is $h = \mathbf{I}(x;\mathcal{H})$. Since h is negligible for almost all x, the Kolmogorov Structure Function is almost always flat.

The minimal sufficient statistic for $x \in \{0,1\}^*$ is

$$\mathbf{k}^*(x) = \min\{k : \mathbf{H}_k(x) + k = \mathbf{K}(x)\}.$$

This is the location in which the Kolmogorov Structure Function reaches the boundary point and becomes flat. Due to Theorem 1, $\mathbf{k}^*(x) <^{\log} \mathbf{K}(\mathbf{K}(x)) + \mathbf{I}(x;\mathcal{H})$ (but note that the $\mathbf{K}(\mathbf{K}(x))$ term can be eliminated). A visualization of the Kolmogorov Structure Function can be seen in Figure 1.

3 Set-Restricted Structure Functions

One potential method to create strings with non-simple Kolmogorov Structure Functions is to restrict the sets under consideration. Thus for a set of sets $\mathcal{S}$,

$$\mathbf{H}_k^{\mathcal{S}}(x) = \min\{\log|S| : x \in S \in \mathcal{S}, \mathbf{K}(S) \leq k\}.$$

This would banish the pesky set S_x defined in the last section. This was studied in Section 6 of [VS15]. However there is an inherent obstacle to proving such functions can have any shape. Proofs to statements (such as Theorem 10 in [VS15]) of such effect use a shape function R to (non-recursively) construct a string x whose structure function has that shape R (up to a degree of precision depending on $\mathcal{S}$). Thus the proof can be thought of as a program to produce x given R and $\mathcal{H}$, with $\mathbf{K}(x|\mathcal{H}) <^+ \mathbf{K}(R)$. Thus proofs saying that for every shape R there is a set x such that $\mathbf{H}^{\mathcal{S}}_k(x)$ has shape R (up to a certain precision) also implies that $\mathbf{I}(x;\mathcal{H}) >^+ \mathbf{K}(x) - \mathbf{K}(R)$. However, this obstacle does not preclude a proof of the existence of a large number of strings with profile R, which could potentially overcome the barrier described in this section.

In general, the Independence Postulate states if a string can be described by a small mathematical statement but has high Kolmogorov complexity then it cannot be found in the physical world. This presents an obstacle for constructive proofs in Algorithmic Information Theory.

4 $\mathbf{I}(x;\mathcal{H})$ as an Error Term

The Independence Postulate states one cannot find strings x with nonnegligible $\mathbf{I}(x;\mathcal{H})$. Thus the term $\mathbf{I}(x;\mathcal{H})$ serves as a very good error term. Furthermore, as shown in the following conservation law, no amount of deterministic or probabilistic processing will increase $\mathbf{I}(x;\mathcal{H})$.

Lemma.

- [Eps22a] For partial computable f, $\mathbf{I}(f(x);\mathcal{H}) <^+ \mathbf{I}(x;\mathcal{H}) + \mathbf{K}(f)$.
- [Eps22b] For probability P over $\mathbb{N}$ computed by program q, $\Pr_{a\sim P}[\mathbf{I}(a;\mathcal{H}) > \mathbf{I}(q;\mathcal{H}) + m] \overset{*}{<} 2^{-m}$.

In addition, there are many provable statements about a mathematical construct C with the following form

$$\mathbf{K}(x(C)) <^{\log} y(C) + \mathbf{I}(C;\mathcal{H}).$$

The term $x(C)$ is some string associated with C. The term $y(C)$ is some property about C. The term $\mathbf{I}(C;\mathcal{H})$ is the information $\mathcal{H}$ has about the entire encoding of C. For example, as seen in, [Eps24b], let $C = \{(a_i, b_i)\}$ be a finite set of pairs of numbers, $x(C)$ be the simpliest total computable function consistent with C and $y(C) = \sum_i \mathbf{K}(b_i|a_i)$. One gets the following characterization of regression:

$$\mathbf{K}(x(C)) <^{\log} \sum_i \mathbf{K}(b_i|a_i) + \mathbf{I}(C;\mathcal{H}).$$

5 Discussion

The Independence Postulate [Lev84, Lev13] states:

IP: *Let α be a sequence defined with an n-bit mathematical statement (e.g., in PA or set theory), and a sequence β can be located in the physical world with a k-bit instruction set (e.g., ip-address). Then $\mathbf{I}(\alpha : \beta) < k + n + c$ for some small absolute constant c.*

When I first learned of **IP**, I didn't realize how much of impact it could have on different fields of study. For example, **IP** and the Many Worlds Theory [Eve57] are in conflict because measuring the spin of a million electrons results in the creation of a world where a large prefix of Chaitin's Omega, Ω, is found at a small address. Furthermore, **IP** causes issues in Constructor Theory [Deu13], which characterizes tasks in physics as either possible or impossible. This raises the question: "Is it possible or impossible to find large prefixes of Ω?". The answer causes trouble for either Constructor Theory or **IP**.

This note reiterates that **IP** implies Algorithmic Statistics does not study strings in the physical world. Thus the unrestricted structure function really doesn't say anything about good or bad models for a string. The set-restricted structure function might, but there are obstacles to showing this, as seen in Section 3. This makes the connection between Algorithmic Statistics and the Minimum Description Length Principle [Gru07] tenuous.

The intention is not to denigrate the theory; a majority of my work (including [Eps24a, Eps23c, Eps23b, Eps24b, Eps23a]) is descendent from Algorithmic Statistics, particularly [VV04b]. My interpretation of the Kolmogorov Structure Function is that it (and its equivalent definitions) provide a means to know that a string x has high $\mathbf{I}(x; \mathcal{H})$.

References

[Deu13] D. Deutsch. Constructor theory. *Synthese*, 190(18):4331–4359, 2013.

[Eps22a] S. Epstein. 22 examples of solution compression via derandomization. *CoRR*, abs/2208.11562, 2022.

[Eps22b] S. Epstein. The kolmogorov birthday paradox. *CoRR*, abs/2208.11237, 2022.

[Eps23a] S. Epstein. The EL Theorem, 2023.

[Eps23b] Samuel Epstein. Kolmogorov Derandomization. 2023. HAL Archive, hal-04292439, `https://hal.science/hal-04292439`.

[Eps23c] Samuel Epstein. On Outliers. 2023. HAL Archive, hal-04285958, `https://hal.science/hal-04285958`.

[Eps24a] S. Epstein. Algorithmic Physics. `http://www.jptheorygroup.org/doc/APhysics.pdf`, 2024.

[Eps24b] S. Epstein. On Exotic Sequences. `http://www.jptheorygroup.org/doc/OnExoticSequences.pdf`, 2024.

[Eve57] Hugh Everett. ”relative state” formulation of quantum mechanics. *Rev. Mod. Phys.*, 29, 1957.

[Gru07] P. Grunwald. *The Minimum Description Length Principle*. The MIT Press, 2007.

[GTV01] P. Gács, J. Tromp, and P. Vitányi. Algorithmic statistics. *Information Theory, IEEE Transactions on*, 47:2443 − 2463, 2001.

[Lev84] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Lev13] L. A. Levin. Forbidden information. *J. ACM*, 60(2), 2013.

[SSV24] A. Semenov, A. Shen, and N. Vereshchagin. Kolmogorov's Last Discovery? (Kolmogorov and Algorithmic Statistics). *Theory of Probability & Its Applications*, 68(4):582–606, 2024.

[Vit02] P. Vitányi. Meaningful information. In *Algorithms and Computation*, pages 588–599, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[VS15] N. Vereshchagin and A. Shen. *Algorithmic Statistics Revisited*, pages 235–252. Springer International Publishing, 2015.

[VS17] N. Vereshchagin and A. Shen. *Algorithmic Statistics: Forty Years Later*, pages 669–737. Springer International Publishing, 2017.

[VV04a] N. Vereshchagin and P. Vitanyi. Kolmogorov's structure functions and model selection. *IEEE Transactions on Information Theory*, 50(12):3265–3290, 2004.

[VV04b] N. Vereshchagin and P. Vitányi. Rate Distortion and Denoising of Individual Data Using Kolmogorov Complexity. *IEEE Transactions on Information Theory*, 56:3438–3454, 2004.