

AIT Blog

A New Proof to the Sets Have Simple Members Theorem

Samuel Epstein*

October 9, 2022

This blog entry presents a new proof to the Sets Have Simple Members Theorem [Lev16, Eps19]. The first game theoretic proof, and currently the most accessible proof in the literature, can be found in [She12]. The proof in this entry follows analogously to the proof in [Lev16] except left-total machines are not used. We also provide a proof that non-stochastic elements have high mutual information with the halting sequence, a well known result in the literature. This proof also does not rely on left-total machines. However in a later blog, I'll show the utility of left-total machines. The quest for more proofs is in part motivated by the question of whether there is a resource bounded version of the Sets Have Simple Members Theorem.

Theorem. For finite $D \subset \{0,1\}^*$, $\min_{x \in D} \mathbf{K}(x) <^{\log} -\log \mathbf{m}(D) + \mathbf{I}(D; \mathcal{H})$.

The term $\mathbf{m}(x) = 2^{-\mathbf{K}(x)}$ is the algorithmic probability, where \mathbf{K} is the prefix complexity. $\mathbf{m}(D) = \sum_{x \in D} \mathbf{m}(x)$. $\mathbf{I}(x|\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$ is the mutual information of x with the halting sequence \mathcal{H} .

Motivation

We revisit the original motivation of the theorem. One common goal in computer science is to find the hidden part of the environment, this task has been called Inductive Inference, Extrapolation, Passive Learning, etc. The complete environment can be represented as a huge string $x \in \{0,1\}^*$. The known observations restrict it to a set $D \subset \{0,1\}^*$. For example in thermodynamics, the environment x can be seen as a record of every particle's position and velocity in a closed box. An observation of some macro parameters, such as pressure and temperature, restrict the possible environments to a set D of hypotheses consistent with the observation.

One method used to select a hypothesis (i.e. environment) is to leverage an *a priori* distribution over the environment space. This distribution p encodes any knowledge about the environment known before the observation is made. Then selection of the hypothesis is

$$\arg \max_{x \in D} p(x).$$

Note in AIT, for enumerable distributions (i.e. generatable as outputs of randomized algorithms), there is a universal apriori distribution $\mathbf{m}(x)$. This is because $O(1)\mathbf{m} > p$, for all enumerable p . Furthermore, for all $x \in \{0,1\}^*$, $\mathbf{d}(x|\mathbf{m}) = O(1)$, where \mathbf{d} is deficiency of randomness; so

*JP Theory Group. samepst@jpththeorygroup.org

there is no lower computable refutation to the statement: “ x is generated from \mathbf{m} ”. Thus when the universal prior is used, inductive inference becomes an exercise of Occam’s razor:

$$\arg \min_{x \in D} \mathbf{K}(x).$$

However there exists a potential complication. It could be there is a collection $G \subset D$ of hypotheses representing a concept (such as a more detailed description of particles) where its combined apriori measure is greater than that of the simplest element x , with $\mathbf{m}(x) \ll \mathbf{m}(G)$. Or, making the endeavor more murkier, it could be that G is just the set of all complicated hypothesis and G has greater combined apriori measure than the simplest element. In this case, which explanation does one choose?

The Sets Have Simple Members Theorem shows that this dilemma is purely a mathematical construction. All the universal apriori measure of an observation D is concentrated on its simplest member. This is true for all non-exotic set D with low mutual information with the halting sequence. There are no (randomized) algorithmic means of creating D with arbitrarily high $\mathbf{I}(D; \mathcal{H})$.

New Proof

A probability is *elementary*, if it has finite support and rational values. The deficiency of randomness of x relative to a elementary probability measure Q is $\mathbf{d}(x|Q) = -\log Q(x) - \mathbf{K}(x|Q)$.

Definition 1 (Stochasticity) *A string x is (α, β) -stochastic if there exists an elementary probability measure Q such that*

$$\mathbf{K}(Q) \leq \alpha \text{ and } \mathbf{d}(x|Q) \leq \beta.$$

Theorem 1 ([Lev16, Eps19]) *Let P be a lower-semicomputable semimeasure and c be a large constant. Every (α, β) -stochastic set D with $s = \lceil -\log P(D) \rceil$ contains an element x with*

$$\mathbf{K}(x) < s + \alpha + 2 \log \beta + \mathbf{K}(s) + 2 \log \mathbf{K}(s) + c.$$

The theorem is directly implied by the following proposition.

Proposition 1 *Let P be a lower-semicomputable semimeasure and c be a large constant. If a set D is (α, β) -stochastic relative to an integer $s = \lceil -\log P(D) \rceil$, then D contains an element x with*

$$\mathbf{K}(x) < s + \alpha + \log \beta + \mathbf{K}(\log \beta) + \mathbf{K}(s) + c.$$

Note that if y is (α, β) -stochastic relative to s , then it is $(\alpha, \beta + \mathbf{K}(s))$ -stochastic. Hence the proposition implies the theorem.

Lemma 1 *Let P be a discrete measure and Q be a measure on sets. There exists a set S of size $\lceil \beta/\gamma \rceil$ such that*

$$Q(\{D : P(D) \geq \gamma \text{ and } D \text{ is disjoint from } S\}) \leq \exp(-\beta).$$

Proof. We use the probabilistic method, and show that if we draw $\lceil \beta/\gamma \rceil$ elements according to the distribution P , then the obtained set S satisfies the inequality with positive probability. The probability that a fixed set D with $P(D) \geq \gamma$ is disjoint from S is

$$\leq (1 - \gamma)^{\beta/\gamma} \leq \exp(-\beta).$$

Hence the expected Q -measure of such a D is at most $\exp(-\beta)$ and the required set S exists. \square

Proof of Proposition 1 for computable P . Let Q be an elementary probability measure with $\mathbf{K}(Q) \leq \alpha$ and $\mathbf{d}(D|Q, s) \leq \beta$. Without loss of generality, we assume that β is large positive power of 2. Fix a search procedure that on input Q , β , and $\gamma = 2^{-s}$ finds a set satisfying the conditions of Lemma 1.

For large β , the set D must intersect the obtained set S . Indeed, consider the Q -test $g(X|Q, s)$ that is equal to $\exp(\beta)$ if X is disjoint from S , and is zero otherwise. This is indeed a test, because the above lemma implies that its expected value for $X \sim Q$ is bounded by 1. Since the test is also computable, it is a lower bound to the optimal test $\mathbf{t}(X|Q, s)$, up to a constant factor. By stochasticity of the set D , $g(D|Q, s) < O(1)\mathbf{t}(D|Q, s) < O(2^\beta)$, because $2^{\mathbf{d}(X|Q, s)}$ is an optimal Q test relative to s . Thus for large enough β , D intersects Q .

It remains to construct a description of each element in S of the size given in the proposition. We construct a special decompressor that assigns short description to each element in S . On input of a string, the decompressor interprets the string as a concatenation of 4 parts:

1. A prefix-free description of Q of size at most α .
2. A prefix-free description of $\log \beta$ of size $\mathbf{K}(\log \beta)$.
3. A prefix-free description of s of size $\mathbf{K}(s)$.
4. An integer of bitsize $\log(\beta/\gamma) = s + \log \beta$.

It interprets the last integer as the index of an element in the set S of size $\lceil \beta/\gamma \rceil$ that is computed by the search procedure on input Q , β , and γ . The element is the output of the decompressor. The proposition is proven for computable P . \square

Remark 1 *If P is computable, a set S satisfying the conditions of the lemma can be easily searched. But if P is not computable, then the collection of sets D with $P(D) \geq \gamma$ grows over time. Thus after constructing a good S , it can happen that a large Q -measure of sets D appears that does not contain an element from S , and that new elements to S need to be added. This type of interactive construction leads to an equivalent characterization of the problem in terms of a game which is shown in [She12]. Below, another proof is presented. It might turn out to be useful to derive resource bounded versions of the theorem.*

Proof of Proposition 1 for lower-semicomputable P . We still assume that β is a large power of 2. Let $\gamma = 2^{-s}/2$. We can rewrite $P = \frac{\gamma}{\beta}(P_1 + \dots + P_f + P_*)$, with $f \leq \beta/\gamma$, such that P_1, \dots, P_f are probability measures with finite support obtained by a lower semi-computable approximation of P , and P_* is a lower-semicomputable semimeasure.

Construction of a lower-semicomputable test g over sets. We first construct tests g_1, \dots, g_f together with a list of strings z_1, \dots, z_f . Let $g_0(X) = 1$. Assume we already constructed z_1, \dots, z_{i-1} and g_{i-1} for some $i = 1, \dots, f$. Choose z_i such that the test

$$g_i(X) = \begin{cases} g_{i-1}(X) & \text{if } g_{i-1}(X) \geq \exp(\beta) \\ \exp(P_i(X))g_{i-1}(X) & \text{if } g_{i-1}(X) < \exp(\beta) \text{ and } X \text{ is disjoint from } \{z_1, \dots, z_i\} \\ 0 & \text{otherwise.} \end{cases}$$

satisfies $\mathbf{E}g_i(X) \leq \mathbf{E}g_{i-1}(X)$ where the expectations are taken for $X \sim Q$. Let $g(X)$ be equal to $\exp(\beta)$ if there exists an i such that $g_i(X) \geq \exp \beta$, otherwise let $g(X) = 0$. *End of construction*

We first show that each required string z_i in the construction exists. Suppose z_1, \dots, z_{i-1} and g_{i-1} have already been constructed. We show the existence of z_i using the probabilistic method. If we draw z_i according to P_i , then for each set X for which the second condition of g_i is satisfied, we have

$$\mathbf{E}_{z_i \sim P_i} g_i(X) \leq (1 - P_i(X))g_{i-1}(X) \exp P_i(X) \leq g_{i-1}(X),$$

because of the inequality $1 + r \leq \exp(r)$ for all reals r . If X satisfies the first or third condition, then $\mathbf{E}g_i(X) \leq \mathbf{E}g_{i-1}(X)$ is trivially true. So

$$\begin{aligned} \mathbf{E}_{X \sim Q} \mathbf{E}_{z_i \sim P_i} g_i(X) &\leq \mathbf{E}_{X \sim Q} g_{i-1}(X), \\ \mathbf{E}_{z_i \sim P_i} \mathbf{E}_{X \sim Q} g_i(X) &\leq \mathbf{E}_{X \sim Q} g_{i-1}(X), \end{aligned}$$

and the required z_i exists.

We have $G(x) \leq O(\mathbf{t}(X|Q, (\gamma, \beta)))$, where \mathbf{t} is the optimal test because the construction implies $\mathbf{E}g \leq 1$ and is effective, thus g is lower semicomputable. Every set X with $P(X) \geq 2^{-s} = 2\gamma$ satisfies $P_1(X) + \dots + P_f(X) \geq \frac{\beta}{\gamma}P(D) - 1 \geq 2\beta - 1 \geq \beta$ by choice of P_i . Any such X that is disjoint from the set $\{z_1, \dots, z_f\}$ satisfies

$$g_f(X) = \exp(P_1(X)) \exp(P_2(X)) \dots \exp(P_f(X)) \geq \exp(\beta).$$

This implies $\mathbf{d}(X|Q, s) > \beta$ for large β , because up to $O(1)$ constants, we have

$$1.44\beta \leq \log g(X) \leq \mathbf{d}(X|Q, (\beta, \gamma)) \leq \mathbf{d}(X|Q, s) + 2 \log \beta.$$

By the assumption on (α, β) -stochasticity of D , we have $\mathbf{d}(D|Q, s) \leq \beta$ and hence D must contain some z_j . The theorem follows by constructing a description for each string z_i of bitsize $s + \alpha + \log \beta + \mathbf{K}(\log \beta) + \mathbf{K}(s)$ in a similar way as above. \square

Non-Stochastic Objects

The stochasticity of an object can be measured by

$$\mathbf{Ks}(x) = \min\{\mathbf{K}(P) + O(\log \max\{\mathbf{d}(x|P), 1\}) : P \text{ is an elementary probability measure}\}.$$

This term combines the complexity of the model P with how well it fits x , i.e. the randomness deficiency \mathbf{d} . It is well known in the literature that non-stochastic objects have high mutual information with the halting sequence [VS17]. In the following lemma, we reprove this fact, without using left-total machines, which was used in the original proof.

Lemma 2 $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x; \mathcal{H})$.

Proof. We dovetail all programs to the universal Turing machine U . For $p \in \text{Domain}(U)$, $n(p) \in \mathbb{N}$ is the position in which the program $p \in \{0, 1\}^*$ terminates. Let $\Omega^n = \sum_{p: n(p) < n} 2^{-\|p\|}$ and $\Omega = \Omega^\infty$ be Chaitin's Omega. Let Ω_t^n be Ω^n restricted to the first t digits. Let $x^* \in \{0, 1\}^{\mathbf{K}(x)}$, with $U(x^*) = x$ with minimum $n(x^*)$. Let $k(p) = \max\{\ell : \Omega_\ell^{n(p)} = \Omega_\ell\}$ and $k = k(x^*)$. We define the elementary probability measure $Q(x) = \max\{2^{-\|p\|+k} : k(p) = k, U(p) = x\}$, $Q(\emptyset) = 1 - Q(\{0, 1\}^* \setminus \{x^*\})$.

$$\begin{aligned} \mathbf{d}(x|Q) &= -\log Q(x) - \mathbf{K}(x|Q) <^+ (\mathbf{K}(x) - k) - \mathbf{K}(x|\Omega_k) \\ &<^+ (\mathbf{K}(x|\Omega_k) + \mathbf{K}(\Omega_k) - k) - \mathbf{K}(x|\Omega_k) <^+ (k + \mathbf{K}(k)) - k \\ &<^+ \mathbf{K}(k). \end{aligned}$$

$$\begin{aligned} \mathbf{K}(x|\mathcal{H}) &<^+ \mathbf{K}(x|Q) + \mathbf{K}(Q|\mathcal{H}) <^+ \mathbf{K}(x|Q) + \mathbf{K}(\Omega_k|\mathcal{H}) \\ &<^+ -\log Q(x) + \mathbf{K}(k) <^+ (\mathbf{K}(x) - k) + \mathbf{K}(k) \\ k &<^{\log} \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H}) \end{aligned}$$

$$\mathbf{Ks}(x) <^+ \mathbf{K}(Q) + O(\max\{\mathbf{d}(x|P), 1\}) <^+ k + O(\mathbf{K}(k)) <^{\log} \mathbf{I}(x; \mathcal{H}).$$

□

The following corollary comes from Theorem 1 and Lemma 2

Corollary 1 (Sets Have Simple Members)

For finite $D \subset \{0, 1\}^*$, $\min_{x \in D} \mathbf{K}(x) <^{\log} -\log \mathbf{m}(D) + \mathbf{I}(D; \mathcal{H})$.

References

- [Eps19] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.
- [Lev16] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016.
- [She12] 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.
- [VS17] Nikolay K. Vereshchagin and Alexander Shen. Algorithmic statistics: Forty years later. In *Computability and Complexity*, pages 669–737, 2017.