Outliers in Dynamics

Samuel Epstein*

June 30, 2022

Abstract

We show that outliers occur almost surely in computable dynamics over infinite sequences. Ever greater outliers can be found as the number of visited states increase.

1 Introduction

An outlier is an observation whose value lies outside the set of values considered likely according to some hypothesis (usually one based on other observations); an isolated point. In the realm of algorithmic information theory, outliers are modeled using the deficiency of randomness. The randomness deficiency of an infinite sequence $\alpha \in \Sigma^\infty$ with respect to a computable probability measure P is $\mathbf{D}(\alpha|P) = \sup_n -\log P(\alpha[0..n]) - \mathbf{K}(\alpha[0..n])$. The term $\mathbf{K}$ is the prefix free Kolmogorov complexity.

The definition of a randomness deficiency is a very useful mathematical definition, and can be used to show the ubiquity of outliers in different constructs. A sampling method is a probabilistic program that when given n, outputs with probability 1, n unique elements, either numbers or infinite sequences. In [Eps21], it was proven that sampling methods produce outliers. In [Eps22], it was proven that for ergodic dynamical systems, ever greater outlying states occur with diminishing measure. In this paper, we show that arbitrary (potentially non-ergodic) computable dynamical systems over Σ^∞ produce outliers, where as more states are visited, more and greater outliers are guaranteed to occur.

A computable dynamical system (λ, δ), over Σ^∞ consists of a computable starting state probability λ over Σ^∞ and a computable transition function $\delta : \Sigma^\infty \to \Sigma^\infty$. We assume that the dynamical system is non-degenerate, in that for λ-a.e. starting states α, an infinite number of states is visited using δ.

Theorem. *There is a $d \in \mathbb{N}$, where for computable probability μ over Σ^∞ and computable dynamics (λ, δ) over Σ^∞, for λ-a.e. starting states $\alpha \in \Sigma^\infty$, there exists $c_\alpha \in \mathbb{N}$, where among the first 2^m states visited, for any $n < m$, there are at least 2^n states β with $\mathbf{D}(\beta|\mu) > m - n - d\log m - c_\alpha$. Furthermore, for the smallest such c_α, $\mathbf{E}_{\alpha\sim\lambda}[c_\alpha - O(\log c_\alpha)] <^+ \mathbf{K}(\lambda) + \mathbf{K}(\delta) + \mathbf{K}(\mu)$.*

The proof technique is two stages. First is to prove that certain finite sets of natural numbers or infinite sequences have high mutual information, $\mathbf{I}$, with the halting sequence. This is represented in Theorems 1 and 2. The second step is to use conservation properties of $\mathbf{I}$, shown in Theorem 3, to achieve the main theorem of the paper. This is compatible with the Independence Postulate [Lev13, Lev84]. It remains to be seen whether there is a more direct method to proving the results.

*JP Theory Group. samepst@jptheorygroup.org

In fact, as shown in Section 7, the above theorem can be generalized to aribtrary (i.e. uncomputable) dynamics. The above theorem holds if the (potentially infinite) encoding of the dynamical system has finite mutual information with the halting sequence. This generalization is made possible due the two step process described above.

2 Conventions

We use $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, Σ, Σ^*, and Σ^∞ to represent natural numbers, rational numbers, reals, bits, finite strings, and infinite strings. The removal of the last bit of a string is denoted by $(p0^-)$=$(p1^-)$=p, for $p \in \Sigma^*$. We use $\Sigma^{*\infty}$ to denote $\Sigma^* \cup \Sigma^\infty$, the set of finite and infinite strings. For $x \in \Sigma^{*\infty}$, $y \in \Sigma^{*\infty}$, we say $x \sqsubseteq y$ if $x = y$ or $x \in \Sigma^*$ and $y = xz$ for some $z \in \Sigma^{*\infty}$. Also $x \sqsubset y$ if $x \sqsubseteq y$ and $x \neq y$. The indicator function of a mathematical statement A is denoted by $[A]$, where if A is true then $[A] = 1$, otherwise $[A] = 0$. For sets Z of infinite strings, $Z_n = \{\alpha[0..n] : \alpha \in Z\}$ and $\langle Z\rangle = \langle Z_1\rangle\langle Z_2\rangle\langle Z_3\rangle \dots$.

As is typical of the field of algorithmic information theory, the theorems in this paper are relative to a fixed universal machine, and therefore their statements are only relative up to additive and logarithmic precision. For positive real functions f the terms $<^+ f$, $>^+ f$, $=^+ f$ represent $< f + O(1)$, $> f - O(1)$, and $= f \pm O(1)$, respectively. In addition $\overset{*}{<} f$, $\overset{*}{>} f$ denote $< f/O(1)$, $> f/O(1)$. The terms $\overset{*}{=} f$ denotes $\overset{*}{<} f$ and $\overset{*}{>} f$. For nonnegative real function f, the terms $<^{\log} f$, $>^{\log} f$, $=^{\log} f$ represent the terms $< f + O(\log(f+1))$, $> f - O(\log(f+1))$, and $= f \pm O(\log(f+1))$, respectively. A discrete measure is a nonnegative function $Q : \mathbb{N} \to \mathbb{R}_{\geq 0}$ over natural numbers. The support of a measure Q is the set of all elements whose Q value is positive, with $\mathrm{Supp}(Q) = \{a : Q(a) > 0\}$. A measure is elementary if its support is finite and its range is a subset of $\mathbb{Q}$. We say that Q is a probability measure if $\sum_a Q(a) = 1$.

$T_y(x)$ is the output of algorithm T (or $\perp$ if it does not halt) on input $x \in \Sigma^*$ and auxiliary input $y \in \Sigma^{*\infty}$. T is prefix-free if for all $x, s \in \Sigma^*$ with $s \neq \emptyset$, and $y \in \Sigma^{*\infty}$, either $T_y(x) = \perp$ or $T_y(xs) = \perp$. The complexity of $x \in \Sigma^*$ with respect to T_y is $\mathbf{K}_T(x|y) = \min\{\|p\| : T_y(p) = x\}$.

There exists optimal for $\mathbf{K}$ prefix-free algorithm U, meaning that for all prefix-free algorithms T, there exists $c_T \in \mathbb{N}$, where $\mathbf{K}_U(x|y) \leq \mathbf{K}_T(x|y) + c_T$ for all $x \in \Sigma^*$ and $y \in \Sigma^{*\infty}$. For example, one can take a universal prefix-free algorithm U, where for each prefix-free algorithm T, there exists $t \in \Sigma^*$, with $U_y(tx) = T_y(x)$ for all $x \in \Sigma^*$ and $y \in \Sigma^{*\infty}$. The function $\mathbf{K}(x|y)$, defined to be $\mathbf{K}_U(x|y)$, is the Kolmogorov complexity of $x \in \Sigma^*$ relative to $y \in \Sigma^{*\infty}$. When we say that a universal Turing machine is relativized to an object, this means that an encoding of the object is provided to the universal Turing machine on an auxiliary tape.

The chain rule for Kolmogorov complexity is $\mathbf{K}(x, y) =^+ \mathbf{K}(x) + \mathbf{K}(y|\langle x, \mathbf{K}(x)\rangle)$. The mutual information in finite strings x and y relative to $z \in \Sigma^*$ is $\mathbf{I}(x : y \,|\, z) = \mathbf{K}(x|z) + \mathbf{K}(y|z) - \mathbf{K}(\langle x, y\rangle|z) =^+ \mathbf{K}(x|z) - \mathbf{K}(x|\langle y, \mathbf{K}(y|z), z\rangle)$. The universal probability of a number $a \in \mathbb{N}$ is $\mathbf{m}(a|y) = \sum_z [U_y(z) = a] 2^{-\|z\|}$. The coding theorem states $-\log \mathbf{m}(a|y) =^+ \mathbf{K}(a|y)$.

For computable probability measure λ over Σ^∞, $\mathbf{K}(\lambda)$ is the size of the shortest program that can compute $\lambda(x\Sigma^\infty)$, for all $x \in \Sigma^*$. For computable function $\delta : \Sigma^\infty \to \Sigma^\infty$, $\mathbf{K}(\delta)$ is the size of the shortest program on the input tape that outputs $\delta(\alpha)$ when α is on the auxilliary tape. The halting sequence $\mathcal{H} \in \Sigma^\infty$ is the infinite string where $\mathcal{H}[i] = [U(i) \neq \perp]$ for all $i \in \mathbb{N}$. The amount of information that $a \in \mathbb{N}$ has with $\mathcal{H}$ is denoted by $\mathbf{I}(a : \mathcal{H}) = \mathbf{K}(a) - \mathbf{K}(a|\mathcal{H})$.

The deficiency of randomness of x with respect to elementary measure Q and $v \in \mathbb{N}$ is $\mathbf{d}(x|Q, v) = \lfloor -\log Q(x) \rfloor - \mathbf{K}(x|\langle Q\rangle, v)$. The stochasticity of $a \in \mathbb{N}$, conditional to $b \in \mathbb{N}$, is measured by $\mathbf{Ks}(a|b) = \min\{\mathbf{K}(Q|b) + 3\log\max\{\mathbf{d}(a|Q, b), 1\} : Q$ is an elementary probability measure$\}$. We have $\mathbf{Ks}(a) = \mathbf{Ks}(a|\emptyset)$, with $\mathbf{Ks}(a|b) < \mathbf{Ks}(a) + O(\log \mathbf{K}(b))$.

3 Sets of Numbers

Theorem 1 *For computable probability p over $\mathbb{N}$ and for $D \subset \Sigma^*$, $|D| = 2^s$, $m \in [0, s-1]$, there are 2^m elements $a \in D$, with $s - m < \mathbf{d}(a|p) + \mathbf{Ks}(D) + \mathbf{K}(p) + O(\log s + \log \mathbf{K}(p))$.*

Proof. We relativize the universal Turing machine U to p and s for the duration of the proof, which can be done as the theorem has precision $O(\log s)$. Let Q be an elementary probability distribution that realizes $\mathbf{Ks}(D)$. Let $d = \mathbf{d}(D|Q)$ be the deficiency of randomness of D with respect to Q. Let V be the combined elements of encoded sets of size 2^s in the support of Q. We create an algorithm, that given Q and s produces 2^{s-1} sets $F_i \subseteq V$. We start with the first round. Suppose each element of V is selected independently with probability $cd2^{-s}$, where c is a constant to be chosen later. The selected set is F_1, and $\mathbf{E}[p(F_1)] \le cd2^{-s}$. Furthermore

$$\mathbf{E}[Q(\{G : |G| = 2^s, G \cap F_1 = \emptyset\})] \le \sum_G Q(G)(1 - cd2^{-s})^{2^s} < e^{-cd}.$$

Thus a finite set F_1 can be chosen such that $p(F_1) \le 2cd2^{-s}$ and $Q(\{G : |G| = 2^s, G \cap F_1 = \emptyset\}) \le e^{1-cd}$.

Now it must be that $D \cap F_1 \neq \emptyset$. Otherwise, using the Q-test, $t(G) = [|G| = 2^s, G \cap F_1 = \emptyset]e^{cd-1}$, we have

$$\begin{aligned}\mathbf{K}(D|Q, d, c) &<^+ -\log Q(D) - (\log e)cd \\ (\log e)cd &<^+ -\log Q(D) - \mathbf{K}(D|Q) + \mathbf{K}(d, c) \\ (\log e)cd &<^+ d + \mathbf{K}(d, c),\end{aligned}$$

which is a contradiction for large enough c solely dependent on the universal Turing machine U. Thus there is an $a \in D \cap F_1$, where

$$\begin{aligned}\mathbf{K}(a) &<^+ -\log p(a) + \log d - s + \mathbf{K}(d) + \mathbf{K}(Q) \\ s &<^+ \mathbf{d}(a|p) + \mathbf{Ks}(D).\end{aligned}$$

Removing the relativization of p and s for just the following 2 equations,

$$\begin{aligned}s &<^+ -\log p(a) - \mathbf{K}(a|p) + \mathbf{Ks}(D|p) + O(\log s), \\ s &<^+ \mathbf{d}(a|p) + \mathbf{Ks}(D) + \mathbf{K}(p) + O(\log s + \log \mathbf{K}(p)).\end{aligned}$$

We define the construction of set F_i given that the first $i-1$ rounds have already occured. Let $F_{<i} = \bigcup_{j=1}^{i-1} F_j$. A set G is eligible if $|G| = 2^s$, and $|G \setminus F_{<i}| \ge 2^s - (i-1)$. Set F_i is selected at random from V, with each element selected at random with probability $cd_i 2^{-s}$, with $d_i = d \log i$. $\mathbf{E}[p(F_i)] \le cd_i 2^{-s}$.

$$\begin{aligned}&\mathbf{E}[Q(\{G : G \text{ is eligible }, (G \setminus F_{<i}) \cap F_i = \emptyset\}) \\ \le &\sum_{\text{eligible G}} Q(G)(1 - cd_i 2^{-s})^{2^s - (i-1)} \\ \le &e^{-cd_i 2^{-s}(2^s - (i-1))} \le e^{-cd_i/2}.\end{aligned}$$

Thus a finite set F_i can be chosen such that $p(F_i) \le 2cd_i 2^{-s}$ and $Q(\{G : G \text{ is eligible }, (G \setminus F_{<i}) \cap F_i = \emptyset] \le e^{-cd_i/2+1}$. It must be that on the rounds i that D is eligible, $(D \setminus F_{<i}) \cap F_i \neq \emptyset$. Otherwise

one can create a Q-test $t_i(G) = [G \text{ is eligible}, (G \setminus F_{<i}) \cap F_i = \emptyset] e^{cd_i/2-1}$. Thus $t_i(D) \neq 0$ and

$$\begin{aligned}
\mathbf{K}(D|Q, d_i, i, c) &<^+ -\log Q(D) - (\log e)cd_i/2 \\
.5(\log e)cd\log i &<^+ -\log Q(D) - \mathbf{K}(D|Q) + \mathbf{K}(d_i, i, c) \\
.5(\log e)cd\log i &<^+ d + \mathbf{K}(d, \mathbf{K}(i), i, c).
\end{aligned}$$

This is a contradiction for large enough c dependent solely on the universal Turing machine U. Thus on rounds i where D is eligible, there exist an $a \in (D \setminus F_{<i}) \cap F_i$, with

$$\begin{aligned}
\mathbf{K}(a) &<^+ -\log p(a) + \log d_i - s + \mathbf{K}(d_i) + \mathbf{K}(i) + \mathbf{K}(Q) \\
s &< \mathbf{d}(a|p) + \log i + O(\log\log i) + \log d + \mathbf{K}(d) + \mathbf{K}(Q) \\
s - \log i &< \mathbf{d}(a|p) + \mathbf{Ks}(D) + O(\log s).
\end{aligned}$$

Removing the relativization of p (and s) results in

$$\begin{aligned}
s - \log i &< -\log p(a) - \mathbf{K}(a|p) + \mathbf{Ks}(D|p) + O(\log s), \\
s - \log i &< \mathbf{d}(a|p) + \mathbf{Ks}(D) + \mathbf{K}(p) + O(\log s + \log \mathbf{K}(p)). \qquad (1)
\end{aligned}$$

On rounds i in which D is not eligible, then there exist rounds $j < i$ where $|(D \setminus F_{<j}) \cap F_j| > 1$. And for each such element in the intersection, a bound on their deficiency of randomness similar to Equation 1 can be proven. □

4 Left-Total Machines

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 $y \in \Sigma^{*\infty}$ if 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 $\alpha \in \Sigma^{*\infty}$ is to the "left" of $\beta \in \Sigma^{*\infty}$, and use the notation $\alpha \lhd \beta$, if there exists $x \in \Sigma^*$ 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. An example left-total machine can be seen in Figure 1.

For the remaining part of this paper, we can and will change the universal self delimiting machine U into an optimal left-total machine U'. For a detailed explanation on how to construct a left-total universal Turing machine, we refer readers to [Eps21].

Without loss of generality, the complexity terms of this paper are defined with respect to the optimal left total machine U. 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. The term "border" is used because for any string $x \in \Sigma^*$, $x \lhd \mathcal{B}$ implies that x total with respect to U and $\mathcal{B} \lhd x$ implies that U will never halt when given x as an initial input. Figure 2 shows the domain of U with respect to $\mathcal{B}$.

5 Sets of Infinite Strings

For total string b, let $\mathbf{bb}(b) = \max\{U(p) : p \lhd b \text{ or } p \sqsubseteq b\}$ be the largest number produced by a program that extends b or is to the left of b.

Definition 1 ([Lev74]) *The mutual information between two infinite sequences α and β is*

$$\mathbf{I}(\alpha : \beta | z) = \log \sum_{x, y \in \Sigma^*} \mathbf{m}(x|z, \alpha)\mathbf{m}(y|z, \beta) 2^{\mathbf{I}(x:y|z)}.$$

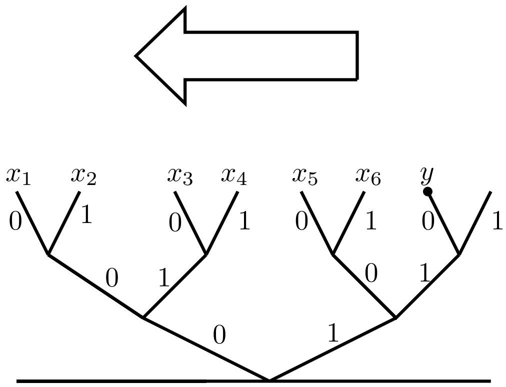

Figure 1: The above diagram represents the domain of a left total machine T with the 0 bits branching to the left and the 1 bits branching to the right. For $i \in \{1, \ldots, 5\}$, $x_i \lhd x_{i+1}$ and $x_i \lhd y$. Assuming $T(y)$ halts, each x_i is total. This also implies each x_i^- is total as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

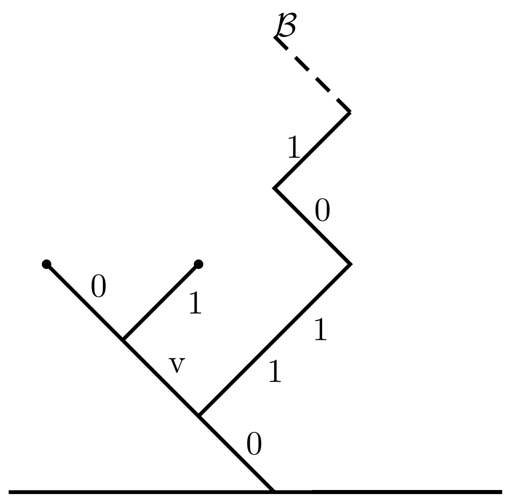

Figure 2: The above diagram represents the domain of the optimal left-total algorithm U, with the 0 bits branching to the left and the 1 bits branching to the right. The strings in the above diagram, $0v0$ and $0v1$, are halting inputs to U with $U(0v0) \neq \perp$ and $U(0v1) \neq \perp$. So $0v$ is 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.

Lemma 1 **([EL11])** *For $x \in \mathbb{N}$, $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x : \mathcal{H})$.*

Lemma 2 **([Eps21])** *If $b \in \Sigma^*$ is total and b^- is not, and $x \in \Sigma^*$, then $\mathbf{K}(b) + \mathbf{I}(x : \mathcal{H}|b) <^{\log} \mathbf{I}(x : \mathcal{H}) + \mathbf{K}(b|\langle x, \|b\| \rangle)$.*

Lemma 3 **([Eps21])** *If $b \in \Sigma^*$ is total and b^- is not, and for $x \in \Sigma^*$, $\mathbf{K}(b|\langle x, |b\| \rangle) = O(1)$, then $\mathbf{K}(\|b\|) <^{\log} 2 \log \mathbf{I}(x : \mathcal{H})$.*

Theorem 2 *For computable probability P over Σ^∞ and $Z \subset \Sigma^\infty$, $|Z| = 2^s$, $m \in [0, s-1]$, there are 2^m elements $\alpha \in Z$, with $s - m < \mathbf{D}(\alpha|P) + \mathbf{I}(Z : \mathcal{H}) + \mathbf{K}(P) + O(\log s + \log \mathbf{I}(Z : \mathcal{H}) + \log \mathbf{K}(P))$.*

Proof. The proof of this theorem follows closely in form to the proof of Theorem 5 in [Eps21], except Theorem 1 is referenced. Fix $m \in [0, s-1]$. Let b be the shortest total string such that $|Z_{\mathbf{bb}(b)}| = 2^s$. Set $D = Z_{\mathbf{bb}(b)}$. Let $p(x) = [\|x\| = \mathbf{bb}(b)]P(\{\alpha : x \sqsubset \alpha\})$. Using Theorem 1, relativized to b, produces 2^m elements F such that for $x \in F$,

$$\begin{aligned}
\mathbf{K}(x|b) &< -\log p(x) - s + m + \mathbf{Ks}(D|b) + \mathbf{K}(p|b) + O(\log s + \log \mathbf{K}(p|b)),\\
\mathbf{K}(x|b) &< -\log p(x) - s + m + \mathbf{Ks}(D|b) + \mathbf{K}(P) + O(\log s + \log \mathbf{K}(P)).
\end{aligned}$$

Using Lemma 1, relativized to b,

$$\begin{aligned}
&\mathbf{K}(x|b) < -\log p(x) - s + m + \mathbf{I}(D : \mathcal{H}|b) + \mathbf{K}(P) + O(\log s + \log \mathbf{I}(D : \mathcal{H}|b) + \log \mathbf{K}(P))\\
&\quad s - m < \log(\mathbf{m}(x)/p(x)) + \mathbf{K}(b) + \mathbf{I}(D : \mathcal{H}|b) + \mathbf{K}(P) + O(\log s + \log(\mathbf{I}(D : \mathcal{H}|b) + \mathbf{K}(b)) + \mathbf{K}(P)).
\end{aligned}$$

By Lemma 2,

$$\begin{aligned}
s - m < {} & \log(\mathbf{m}(x)/p(x)) + \mathbf{I}(D : \mathcal{H}) + \mathbf{K}(b|D, \|b\|) + \mathbf{K}(P)\\
& + O(\log s + \log(\mathbf{I}(D : \mathcal{H}) + \mathbf{K}(b|D, \|b\|)) + \log \mathbf{K}(P)).
\end{aligned}$$

Since $D \subseteq \Sigma^{\mathbf{bb}(b)}$, $\mathbf{K}(b|D, \|b\|) = O(1)$, as a program can output the leftmost total string y of length $\|b\|$ such that $\mathbf{bb}(y)$ is the length of the strings in D. So

$$s - m < \log(\mathbf{m}(x)/p(x)) + \mathbf{I}(D : \mathcal{H}) + \mathbf{K}(P) + O(\log s + \log \mathbf{I}(D : \mathcal{H}) + \log \mathbf{K}(P)).$$

We have that $\mathbf{K}(D|\langle Z\rangle) <^+ \mathbf{K}(\|b\|) + \mathbf{K}(s)$, as D is computable from $\langle Z\rangle$, $\|b\|$, and s. This is because b is computable from its length, s, and $\langle Z\rangle$, and thus so is $D = Z_{\mathbf{bb}(b)}$. By the Definition 1 of the mutual information between infinite sequences,

$$\begin{aligned}
\mathbf{I}(D : \mathcal{H}) &<^+ \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \mathbf{K}(D|\langle Z\rangle)\\
&<^+ \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \mathbf{K}(\|b\|) + \mathbf{K}(s)\\
&<^+ \mathbf{I}(\langle Z\rangle : \mathcal{H}) + 2\log \mathbf{I}(D : \mathcal{H}) + \mathbf{K}(s) \qquad (2)\\
&<^{\log} \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \mathbf{K}(s),
\end{aligned}$$

where Equation 2 is due to Lemma 3, noting $\mathbf{K}(b|D, \|b\|) = O(1)$. So there is an $\alpha \in Z$, $x \sqsubset \alpha$, with

$$\begin{aligned}
s - m &< \log(\mathbf{m}(x)/p(x)) + \mathbf{I}(D : \mathcal{H}) + \mathbf{K}(P) + O(\log s + \log \mathbf{I}(D : \mathcal{H}) + \log \mathbf{K}(P))\\
s - m &< \mathbf{D}(\alpha|P) + \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \mathbf{K}(P) + O(\log s + \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \log \mathbf{K}(P)).
\end{aligned}$$

□

6 Outliers in Dynamics

This section contains the main result of the paper, that computable dynamical systems will exhibit outliers. To prove this fact, Theorem 3 will be leveraged, which details the conservation properties of the halting sequence $\mathcal{H}$.

Theorem 3 ([Ver21, Lev74, Gei12])

- $\Pr_\lambda(\mathbf{I}(\alpha:\mathcal{H})>n)\overset{*}{<}2^{-n+\mathbf{K}(\lambda)}$.
- $\mathbf{I}(f(\alpha):\mathcal{H})<^+\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(f)$.
- $\mathbf{E}_{\alpha\sim\lambda}\left[2^{\mathbf{I}(\alpha:\mathcal{H})}\right]\overset{*}{<}2^{\mathbf{K}(\lambda)}$.

Theorem 4 *There is a $d\in\mathbb{N}$, where for computable probability μ over Σ^∞ and computable dynamics (λ,δ) over Σ^∞, for λ-a.e. starting states $\alpha\in\Sigma^\infty$, there exists $c_\alpha\in\mathbb{N}$, where among the first 2^m states visited, for any $n<m$, there are at least 2^n states β with $\mathbf{D}(\beta|\mu)>m-n-d\log m-c_\alpha$. Furthermore, for the smallest such c_α, $\mathbf{E}_{\alpha\sim\lambda}[c_\alpha-O(\log c_\alpha)]<^+\mathbf{K}(\lambda)+\mathbf{K}(\delta)+\mathbf{K}(\mu)$.*

Proof. Fix a starting state $\alpha\in\Sigma^\infty$ and fix $d\in\mathbb{N}$. Assume α has property A, in which for all $c_\alpha\in\mathbb{N}$, there exists $m,n\in\mathbb{N}$, $m<n$, where the first 2^n states $Z\subset\Sigma^\infty$ visited has less than 2^m states $\beta\in Z$, with

$$\mathbf{D}(\beta|\mu)>n-m-d\log n-c_\alpha.$$

Therefore, due to Theorem 2 there exists a state $\beta\in Z$, with

$$\mathbf{D}(\beta|\mu)\leq n-m-d\log n-c_\alpha$$

and

$$n-m<\mathbf{D}(\beta|\mu)+\mathbf{I}(Z:\mathcal{H})+\mathbf{K}(\mu)+O(\log\mathbf{K}(\mu)+\log n+\log\mathbf{I}(Z:\mathcal{H})).$$

Due to Theorem 3, we have

$$\mathbf{I}(Z:\mathcal{H})<^+\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(n)+\mathbf{K}(\delta),$$

so

$$n-m<\mathbf{D}(\beta|\mu)+\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(\delta)+\mathbf{K}(\mu)+O(\log\mathbf{K}(\delta)+\log\mathbf{K}(\mu)+\log n+\log(\mathbf{I}(\alpha:\mathcal{H}))).$$

So

$$\begin{aligned}n-m<&n-m-d\log n-c_\alpha+\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(\delta)+\mathbf{K}(\mu)\\&+O(\log\mathbf{K}(\delta)+\log\mathbf{K}(\mu)+\log n+\log(\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(\delta))),\end{aligned}$$

implying

$$d\log n+c_\alpha<\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(\delta)+\mathbf{K}(\mu)+O(\log\mathbf{K}(\delta)+\log\mathbf{K}(\mu)+\log n+\log\mathbf{I}(\alpha:\mathcal{H})).$$

Thus for large enough d, dependent solely on the universal Turing machine U, $\mathbf{I}(\alpha:\mathcal{H})=\infty$. Thus by Theorem 3, λ-a.e. states α do not have the property A.

By the reasoning above, the smallest such c_α has $c_\alpha<^{\log}\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(\delta)+\mathbf{K}(\mu)$. So $c_\alpha-O(\log c_\alpha)<\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(\delta)+\mathbf{K}(\mu)$. So $\mathbf{E}_{\alpha\sim\lambda}\left[2^{c_\alpha}c_\alpha^{-O(1)}\right]<\mathbf{E}_{\alpha\sim\lambda}\left[2^{\mathbf{I}(\alpha:\mathcal{H})+\mathbf{K}(\delta)+\mathbf{K}(\mu)}\right]$ and Theorem 3 implies $\mathbf{E}_{\alpha\sim\lambda}[c_\alpha-O(\log c_\alpha)]<^+\mathbf{K}(\lambda)+\mathbf{K}(\delta)+\mathbf{K}(\mu)$. $\square$

7 Arbitrary Dynamical Systems

For a continuous function $\delta : \Sigma^\infty \to \Sigma^\infty$, $\langle\delta\rangle$ is any infinite sequence $\delta' \in \Sigma^\infty$, such that if $\alpha \in \Sigma^\infty$ is on auxilliary tape of U and δ' is on the input tape, U outputs $\delta(\alpha)$ on the output tape, without halting. Similarly for arbitrary (i.e. uncomputable) probability measure λ over Σ^∞, $\langle\lambda\rangle$ is any infinite sequence λ' such that if $x \in \Sigma^*$ is on the auxillary tape and λ' is on the input tape of U, then U outputs $\lambda(x\Sigma^\infty)$. We say that for dynamical system (λ,δ), $\mathbf{I}((\lambda,\delta):\mathcal{H}) = \inf_{\langle\lambda\rangle,\langle\delta\rangle} \mathbf{I}(\langle\lambda,\delta\rangle:\mathcal{H})$, using Definition 1 of mutual information of infinite sequences. Theorem 6 is a proper generalization of Theorem 4.

Theorem 5 ([Ver21, Lev74, Gei12]) $\mathbf{E}_{\alpha\sim\lambda}\left[2^{\mathbf{I}(\alpha:\mathcal{H})}\right] \stackrel{*}{<} 2^{\mathbf{I}(\lambda:\mathcal{H})}$.

Theorem 6 *There is a $d \in \mathbb{N}$, where for probability μ over Σ^∞ and dynamics (λ,δ) over Σ^∞, with $\mathbf{I}((\lambda,\delta):\mathcal{H}) \neq \infty$, for λ-a.e. starting states $\alpha \in \Sigma^\infty$, there exists $c_\alpha \in \mathbb{N}$, where among the first 2^m states visited, for any $n < m$, there are at least 2^n states β with $\mathbf{D}(\beta|\mu) > m - n - d\log m - c_\alpha$. Furthermore, for the smallest such c_α, $\mathbf{E}_{\alpha\sim\lambda}[c_\alpha - O(\log c_\alpha)] <^+ \mathbf{I}((\lambda,\delta):\mathcal{H}) + \mathbf{K}(\mu)$.*

Proof. The proof follows similarly to that of the proof of Theorem 4, but we reproduce it in its entirety. Fix a starting state $\alpha \in \Sigma^\infty$ and fix $d \in \mathbb{N}$. Assume α has property A, in which for all $c_\alpha \in \mathbb{N}$, there exists $m, n \in \mathbb{N}$, $m < n$, where the first 2^n states $Z \subset \Sigma^\infty$ visited has less than 2^m states $\beta \in Z$, with

$$\mathbf{D}(\beta|\mu) > n - m - d\log n - c_\alpha.$$

Therefore, due to Theorem 2 there exists a state $\beta \in Z$, with

$$\mathbf{D}(\beta|\mu) \leq n - m - d\log n - c_\alpha$$

and

$$n - m < \mathbf{D}(\beta|\mu) + \mathbf{I}(Z:\mathcal{H}) + \mathbf{K}(\mu) + O(\log\mathbf{K}(\mu) + \log n + \log\mathbf{I}(Z:\mathcal{H})).$$

Due to Theorem 3, we have

$$\mathbf{I}(Z:\mathcal{H}) <^+ \mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H}) + \mathbf{K}(n),$$

so

$$n - m < \mathbf{D}(\beta|\mu) + \mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H}) + \mathbf{K}(\mu) + O(\log\mathbf{K}(\mu) + \log n + \log(\mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H})).$$

So

$$n - m < n - m - d\log n - c_\alpha + \mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H}) + \mathbf{K}(\mu) + O(\log\mathbf{K}(\mu) + \log n + \log\mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H})),$$

implying

$$d\log n + c_\alpha < \mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H}) + \mathbf{K}(\mu) + O(\log\mathbf{K}(\mu) + \log n + \log\mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H})).$$

Thus for large enough d, dependent solely on the universal Turing machine U, $\mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H}) = \infty$. Let $\gamma(\langle\alpha,\beta\rangle) = \lambda(\alpha)[\beta = \langle\delta\rangle]$. By Theorem 5, $\mathbf{E}_{\alpha\sim\lambda}[2^{\mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H})}] = \mathbf{E}_{\xi\sim\gamma}[2^{\mathbf{I}(\xi:\mathcal{H})}] \stackrel{*}{<} 2^{\mathbf{I}(\gamma:\mathcal{H})} \stackrel{*}{<} 2^{\mathbf{I}(\langle\lambda,\delta\rangle:\mathcal{H})} < \infty$. Thus by Theorem 3, λ-a.e. states α do not have the property A.

By the reasoning above, the smallest such c_α has $c_\alpha <^{\log} \mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H}) + \mathbf{K}(\mu)$. So $c_\alpha - O(\log c_\alpha) < \mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H}) + \mathbf{K}(\mu)$. So by Theorem 5, $\mathbf{E}_{\alpha\sim\lambda}\left[2^{c_\alpha} c_\alpha^{-O(1)}\right] \stackrel{*}{<} \mathbf{E}_{\alpha\sim\lambda}\left[2^{\mathbf{I}(\langle\alpha,\delta\rangle:\mathcal{H})+\mathbf{K}(\mu)}\right] \stackrel{*}{<} \mathbf{E}_{\xi\sim\gamma}\left[2^{\mathbf{I}(\xi:\mathcal{H})+\mathbf{K}(\mu)}\right]$ implies $\mathbf{E}_{\alpha\sim\lambda}[c_\alpha - O(\log c_\alpha)] <^+ \mathbf{I}(\gamma:\mathcal{H}) + \mathbf{K}(\mu) <^+ \mathbf{I}(\langle\lambda,\delta\rangle:\mathcal{H}) + \mathbf{K}(\mu)$. $\square$

8 Discussion

The paper proves results for discrete time dynamical systems. An open problem is how outliers manifest in continuous time dynamical systems. Another avenue of research involves the connections of outliers in dynamical systems to thermodynamics. A good starting point for research into this area would be the paper [Gac94], and the general research direction would be to prove properties of algorithmic fine grained entropy instead of randomness deficiency.

References

[EL11] Samuel Epstein and Leonid Levin. On sets of high complexity strings. *CoRR*, abs/1107.1458, 2011.

[Eps21] Samuel Epstein. All sampling methods produce outliers. *IEEE Transactions on Information Theory*, 67(11):7568–7578, 2021.

[Eps22] Samuel Epstein. The outliers theorem revisited. *CoRR*, abs/2203.08733, 2022.

[Gac94] P. Gacs. The boltzmann entropy and randomness tests. In *Proceedings Workshop on Physics and Computation. PhysComp '94*, pages 209–216, 1994.

[Gei12] Philipp Geiger. *Mutual information and Gödel incompleteness.* PhD thesis, Heidelberg University, 10 2012.

[Lev74] L. A. Levin. Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory. *Problemy Peredachi Informatsii*, 10(3):206–210, 1974.

[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.

[Ver21] N. Vereshchagin. Proofs of conservation inequalities for levin's notion of mutual information of 1974. *Theoretical Computer Science*, 856, 2021.