AIT Blog

Outliers from Algorithms and Dynamics

Samuel Epstein*

September 30, 2022

1 Randomness Deficiencies

Outliers or anomalies are an ubiquitous phenomenon. In this entry, we trace the occurrence of anomalies in algorithmic sampling methods, measurements procured from the real world, and dynamics over the Cantor space and computable metric spaces. In algorithmic information theory, outliers are measured by the randomness deficiency function. The deficiency of randomness $\mathbf{d}$ of a string $x \in \{0,1\}^*$ with respect a computable probability measure p over strings is

$$\mathbf{d}(x|p) = \lfloor -\log p(x) \rfloor - \mathbf{K}(x|p).$$

The condition prefix complexity is $\mathbf{K}$. Randomness deficiency can also be defined over infinite sequences $\alpha \in \{0,1\}^\infty$ and computable probability measures P over $\{0,1\}^\infty$.

$$\mathbf{D}(\alpha|P) = \sup_n -\log P(\alpha[0..n]) - \mathbf{K}(\alpha[0..n]|P).$$

For more information about randomness deficiency, we refer readers to [G13́].

2 Sampling Methods

A discrete sampling method A is a probabilistic function that maps an integer N with probability 1 to a set containing N different strings.

Theorem 1 *Let p be a computable probability measure over $\mathbb{N}$. Let A be a sampling method. There exists $c \in \mathbb{N}$ such that for all n and k:*

$$\Pr(\max_{a \in A(2^n)} \mathbf{d}(a|p) > n - k - c\log n) \geq 1 - 2e^{-2^k}.$$

There is also emergent outliers in continuous sampling methods. A continuous sampling method C is a probabilistic function that maps, with probability 1, an integer N to an infinite encoding of N different inftinite sequences.

Theorem 2 *Let P be a computable probability measure over $\{0,1\}^\infty$. Let C be a continuous sampling method. There exists $c \in \mathbb{N}$ such that for all n and k,*

$$\Pr(\max_{\alpha \in C(2^n)} \mathbf{D}(\alpha|P) > n - k - c\log n) \geq 1 - 2.5e^{-2^k}.$$

*JP Theory Group. samepst@jptheorygroup.org

3 Probabilistic Algorithms

Outliers can be found in more general constructs than sampling methods. Given a computable measure μ over $\{0,1\}^\infty$ Any probabilistic algorithm that outputs an infinite sequence (with no individual sequence with positive probability) is guaranteed to produce ever larger μ-outlier sequences with diminishing measure. More formally, the following theorem encodes this fact.

Theorem 3 *For computable measures μ and non-atomic λ over $\{0,1\}^\infty$ and $n \in \mathbb{N}$, $\lambda\{\alpha : \mathbf{D}(\alpha|\mu) > n\} > 2^{-n-\mathbf{K}(n,\mu,\lambda)-O(1)}$.*

4 Outliers in the Physical World

In the prvious sections, it is proven that algorithmic sampling methods have to produce anomalies. However some sampling methods are too complex to be considered algorithmic. One example is your local weather forecast. Using the Independence Postulate, which is a finitary Church-Turing thesis, this open issue is addressed. Outliers must occur in the physical world.

The Independence Postulate (**IP**), [Lev84, Lev13], is an unprovable inequality on the information content shared between two sequences, postulating that certain infinite and finite sequences cannot be found in nature, a.k.a. have high "physical addresses". In this paper we show that **IP** explains why outliers are found in the physical world. The approach in **IP** is different from that of the previous sections. While the latter shows that computable consructs produce outliers with high probability, the former states that individual sequences without outliers have high addresses, i.e. are hard to find nature.

The information between two strings x, y is $\mathbf{I}(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$. The algorithmic probability is $\mathbf{m}$. The information between two infinite sequences $\alpha, \beta \in \{0,1\}^\infty$ is defined to be [Lev74]

$$\mathbf{I}(\alpha:\beta) = \log \sum_{x,y\in\{0,1\}^*} \mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\mathbf{I}(x:y)}.$$

The Independence Postulate states [Lev13]:

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

Let $\tau \in \mathbb{N}^\mathbb{N}$ represent a series of observations. In reality, observed information is finite. But observations can be considered to be potentially infinite, and represented by never-ending sequences. τ is assumed to have an infinite number of unique observations. $\tau(n)$ is the first 2^n unique numbers of τ. For a probability p over $\mathbb{N}$, let $s_{\tau,p} = \sup_n \left(n - 3\mathbf{K}(n) - \max_{a\in\tau(n)} \mathbf{d}(a|p)\right)$. It is a score of the level of outliers in τ. If $s_{\tau,p}$ is large then τ can be considered to have low level of outliers. if $s_{\tau,p}$ is infinite, then τ has bounded level of outliers.

Theorem 4 *$s_{\tau,p} <^{\log} \mathbf{I}(\langle\tau\rangle : \mathcal{H}) + O(\log \mathbf{K}(p))$.*

Let k be a physical address of τ. The halting sequence $\mathcal{H}$ can be described by a small mathematical statement. By Theorem 4 and IP,

$$s_{\tau,p} <^{\log} \mathbf{I}(\langle\tau\rangle : \mathcal{H}) + O(\log \mathbf{K}(p)) <^{\log} k + c + O(\log \mathbf{K}(p)).$$

It's hard to find observations with small anomalies and impossible to find observations with no anomalies.

5 Dynamics in the Cantor Space

Sampling algorithms and probabilistic algorithms in Section 3 have outliers that are guaranteed to appear. In the next two sections, we show that outliers are emergent in dynamics. This appears in the natural setting of dynamics over the Cantor space as well as the more general setting of dynamics in computable metric spaces.

We define a metric g on $\{0,1\}^\infty$ with $g(\alpha,\beta) = 1/2^k$, where k is the first place where α and β disagree. Let $\mathfrak{F}$ be the topology induced by g on $\{0,1\}^\infty$; $\mathcal{B}(\mathfrak{F})$ be the Borel σ-algebra on $\{0,1\}^\infty$; λ and μ be computable measures over $\{0,1\}^\infty$ and λ be nonatomic; and $(\{0,1\}^\infty, \mathcal{B}(\mathfrak{F}), \lambda)$ be a measure space and $T : \{0,1\}^\infty \to \{0,1\}^\infty$ be an ergodic measure preserving transformation. By the Birkoff theorem,

Theorem 5 *Starting λ-almost everywhere, $\overset{*}{>} \mathbf{m}(n,\mu,\lambda)2^{-n}$ states α visited by iterations of T have $\mathbf{D}(\alpha|\mu) > n$.*

A computable dynamical system (λ, δ) consists of a computable starting state probability λ over $\{0,1\}^\infty$ and a computable transition function $\delta : \{0,1\}^\infty \to \{0,1\}^\infty$. 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 6 *There exists $d \in \mathbb{N}$, where for computable probability μ over $\{0,1\}^\infty$ and computable dynamics (λ, δ), for λ-a.e. starting states $\alpha \in \{0,1\}^\infty$, there exists $s_\alpha \in \mathbb{N}$, where among the first 2^m states visited, for any $n < m$, there are at least 2^n states β with $\mathbf{D}(\beta|\mu) > m - n - d\log m - s_\alpha$. Furthermore, for the smallest such s_α, $\mathbf{E}_{\alpha\sim\lambda}[s_\alpha - O(\log s_\alpha)] < \mathbf{K}(\lambda,\delta) + \mathbf{K}(\mu)$.*

Dynamics in Computable Metric Spaces

Results on outliers can be applied to more general spaces than the Cantor space. In this section, we show how outliers are emergent in dynamics over computable metric spaces.

Definition 1 (Coputable Metric Space) *A computable metric space is a triple $\mathfrak{X} = (X, d, \mathfrak{I})$, where*

- *(X, d) is a separable complete metric space.*
- *$\mathfrak{I} = \{s_i : i \in \mathbb{N}\}$ is a countable dense subset of X.*
- *The real numbers $d(s_i, s_j)$ are all computable, uniformly in $\langle i, j\rangle$.*

Definition 2 (Ideal Balls) *Let $B(x, r)$ be the metric ball $\{y \in X, d(x, y) < r\}$. The numbered sets $\mathfrak{I}$, and $\mathbb{Q}_{>0}$ induced the numbered set of ideal balls $\mathfrak{B} = \{B(s_i, q_j) : s_i \in \mathfrak{I}, q_j \in \mathbb{Q}_{>0}\}$. We write $B_{\langle i,j\rangle}$ for $B(s_i, q_j)$.*

Definition 3 (Computable Measures) *A measure μ over $\mathfrak{X}$ is computable if $\mu(B_{i_1} \cup \cdots \cup B_{i_k})$ is lower semi-computanle uniformly in $\langle i_1, \ldots, i_k\rangle$.*

Definition 4 (Computable Probability Space) *A computable probability space is a pair $(\mathfrak{X}, \mu)$ where $\mathfrak{X}$ is a computable metric space and μ a computable Borel probability measure space on X.*

Definition 5 (Uniform Tests of Randomness) *A uniform test t over computable probability space $(\mathfrak{X}, \mu)$ is a lower semi-computable function over $\mathfrak{X}$ and a description of the probability measure μ. It is beyond the scope of this blog to describe how probability measures can be sent as input to a function. We refer readers to [HR09, G1́3]. In general, it is a fast Cauchy sequence converging to a point in a space where every point is a probability measure. There exists a universal uniform test of randomness $\mathbf{t}$, where for every uniform test of randomness t, there is a constant $c \in \mathbb{N}$ where $c\mathbf{t} > t$. We denote the universal uniform test by $\mathbf{t}_\mu(x)$. The deficiency of randomness is $\mathbf{d}_\mu(x) = \log t_\mu(x)$.*

Definition 6 (Transformation Group) *Dynamics are represented by one-dimensional transformation groups. For computable metric space $\mathfrak{X}$, a topological group G is defined such that each element is a homeomorphism of $\mathfrak{X}$ onto itself:*

$$f(g; x) = g(x) = x' \in X; g \in G, x \in X.$$

The symbol G will be called a topological transformation group if for every pair of elements $g1, g2$ of G, and every $x \in X$, $g1(g2(x)) = (g1g2)(x)$ and if

$$x' = g(x) = f(g; x)$$

is continuous simultaneously in $x \in X$ and $g \in G$.

Theorem 7 (Dynamics over Computable Measure Spaces) *Let L be the Lebesgue measure over $\mathbb{R}$. For one dimensional computable topological transformation group G^t acting on computable probability space $(\mathfrak{X}, \mu)$, for all $\alpha \in X$, $L\{t \in [0,1] : \mathbf{d}_\mu(G^t\alpha) > n\} > 2^{-n-O(\log n)-c(\alpha,\mathfrak{X},\mu)}$. The term $c(\alpha, \mathfrak{X}, \mu)$ is a constant dependent solely on α and $(\mathfrak{X}, \mu)$.*

References

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

[G1́3] P. Gács. Lecture notes on descriptional complexity and randomness, 2013.

[HR09] M. Hoyrup and C. Rojas. Computability of probability measures and martin-löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009.

[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.