Research Overview

Sam Epstein

January 12, 2025

On Anomalies

Anyone can attest to the fact that statistical outliers are common in the outputs of systematic methods that produce datapoints. However it may come as a surprise that using algorithmic randomness, one can prove such outliers have to occur in datasets. The anomaly score used is a universal lower computable test, where the probability is over natural numbers, infinite sequences, or computable metric spaces. The larger the sample, the larger the outlier in the dataset.

Entropy Synchronization

Two separate and isolated physical systems evolving over time cannot have thermodynamic entropies that are in synch. This is true for two systems[1] in the same room or even spread across the Milky Way galaxy[2]. Thermodynamic entropy is measured algorithmically and the systems need not be computable. Systems that are in synch have negative infinite entropy or infinite mutual information with the halting sequence, where both properties are considered to be unphysical.

Semi-Classical Subspaces

Partial signals and partial information cloning can be obtained on quantum states in semi-classical subspaces. Complete and no signals/cloning can be obtained from quantum states in classical and purely quantum subspaces, respectively. The wavefunction collapse from measurements can cause transitions from purely quantum to semi-classical and then classical subspaces. Quantum operations cannot cause such transitions.

Kolmogorov Derandomization

If the existence of an object can be proved using the probabilistic method, then upper bounds on its Kolmogorov complexity can be proved as well. This result has been applied to many problems such as GRAPH-COLORING, k-SAT, MAX-CUT, etc. Game Derandomization states if a simple probabilistic player wins a certain naturally defined game with high probability, then there is a simple deterministic player that can win the game.

---

[1] The lack of simultaneity of events in systems with spacelike separations in special relativity is resolved by the discreteness of the dynamics. Thus the countably infinite sets of events for two systems are aligned in the natural way.

[2] There is nothing nonlocal going on here. In general, upper computing the algorithmic thermodynamic entropy of one system reveals no information about any other system.

CPU Trace Recompilation

We introduce a new type of algorithm with $O(n \log n)$ time complexity that recompiles long sequences C of assembly instruction traces of length n into a novel code called SECODE, short for side effect code. When C is encountered again, SECODE is run, producing the same effects as C. SECODE also matches and reproduces the side effects of instruction traces that have registers pointing to different memory locations than C. The only two data structures of SECODE is a graph for matching and a table for the changes. During operation, a match produces a single array which is applied to the table. If C uses s non-transient heap space and t non-transient stack operations, then the size of SECODE is $O(s+t)$, and it runs in time $O(s \log s + t)$, with small constants. No code annotations are needed; it works on all compiled code.

Independence Postulate

If one were to measure the spin of a very large number of electrons, there is a chance that one gets a large prefix of the halting sequence. So for the *Many Worlds Theory*, one can create a branching world that violates the *Independence Postulate*. *Constructor Theory* consists of counterfactuals, decreeing that certain processes can or cannot occur. However this theory needs reconciliation with the *Independence Postulate*, where information is found/created with an address system.

On Classification

Suppose we have the task of training a classifier over natural numbers, where there is n training samples T. There exists a classifier that is completely consistent with T using $n + \epsilon$ bits of information. The term ϵ is the information the halting sequence has about T, which is negligible except in exotic cases.

On Regression

Suppose we have the task of training a function between natural numbers, where the training set T is $\{(a_i, b_i)\}_{i=1}^{n}$. There exists a function completely consistent with T using $\sum_{i=1}^{n} \mathbf{K}(b_i|a_i) + \epsilon$ bits. The term $\mathbf{K}$ is the conditional prefix Kolmogorov complexity and ϵ is the information the halting sequence has about T.

Müller's Theorem

Müller's Theorem states the Kolmogorov complexity of a string is equal to its quantum Kolmogorov complexity. Thus there is no benefit to using quantum mechanics to compress classical information. The quantitative amount of information in classical sources is invariant to the physical model used to measure it. The original proof is quite extensive, totaling 30 pages. We present two simple proofs, each consisting of two pages.