The Outliers Theorem Revisited

Samuel Epstein*

August 11, 2022

Abstract

An outlier is a datapoint that is set apart from a sample population. The outlier theorem in algorithmic information theory states that given a computable sampling method, outliers must appear. We present a simple proof to the outliers theorem, with exponentially improved bounds. We extend the outlier theorem to ergodic dynamical systems which are guaranteed to hit ever larger outlier states with diminishing measures. We show how to construct deterministic functions from random ones, i.e. function derandomization. We also prove that all open sets of the Cantor space with large uniform measure will either have a simple computable member or high mutual information with the halting sequence.

1 Introduction

The deficiency of randomness of an infinite sequence $\alpha \in \{0,1\}^\infty$ with respect to a computable measure P over $\{0,1\}^\infty$ is defined to be $\mathbf{D}(\alpha|P) = \sup_n - \log P(\alpha[0..n]) - \mathbf{K}(\alpha[0..n])$. The term $\mathbf{K}$ is the prefix free Kolmogorov complexity.

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

This equation has special meaning when λ is the stationary measure of a dynamical system. The theorem was proven using a general template consistent with the Independence Postulate, [Lev13, Lev84]. This method involves first proving that an object has mutual information with the halting sequence. The second step involves removing the mutual information term from the inequality. The removal of the information term can be done in a number of ways, and the dynamical systems theorem represents one such example.

1.1 Outliers

In addition, we present a simple proof of the outliers theorem in [Eps21] with exponentially improved bounds. A sampling method A is a probabilistic function that maps an integer N with probability 1 to a set containing N different strings. Let $P = P_1, P_2, \dots$ be a sequence of measures over strings. For example, one may choose $P_1 = P_2 \dots$ or choose P_n to be the uniform measure over n-bit strings. A conditional probability bounded P-test is a function $t : \{0,1\}^* \times \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that for all $n \in \mathbb{N}$ and positive real number r, we have $P_n(\{x : t(x|n) \geq r\}) \leq 1/r$. If $P_1, P_2, \dots$ is uniformly computable, then there exists a lower-semicomputable such P-test t that is "maximal" (i.e., for which $t' \leq O(t)$ for every other such test t'). We fix such a t, and let $\overline{\mathbf{d}}_n(x|P) = \log t(x|n)$.

*JP Theory Group. samepst@jptheorygroup.org

Theorem B. *Let $P = P_1, P_2 \ldots$ be a uniformly computable sequence of measures on strings and let A be a sampling method. There exists $c \in \mathbb{N}$ such that for all n and k:*

$$\Pr\left(\max_{a \in A(2^n)} \overline{\mathbf{d}}_n(a|P) > n - k - c\right) \geq 1 - 2e^{-2^k}.$$

1.2 Function Derandomization

In this paper, we show how to construct deterministic functions from random ones. Random functions F over natural numbers are modeled by discrete stochastic processes indexed by $\mathbb{N}$, where each $F(t)$, $t \in \mathbb{N}$, is a random variable over $\mathbb{N}$. $\mathcal{F}$ is the set of all random functions. A random function $F \in \mathcal{F}$ is computable if there is a program that on input $(a_1, \ldots, a_n)$ lower computes $\Pr[F(1) = a_1 \cap F(2) = a_2 \cap \cdots \cap F(n) = a_n]$. Put another way, a random function $F \in \mathcal{F}$ is computable if $A = \Pr[F(a_1) = b_1 \cap \cdots \cap F(a_n) = b_n]$ is uniformly computble in $\{(a_i, b_i)\}_{i=1}^n$. The complexity $\mathbf{K}(F)$ of a random function $F \in \mathcal{F}$, is the smallest program that computes A. $\mathcal{G}$ is the set of all deterministic functions $G : \mathbb{N} \to \mathbb{N}$. A sample $S \in \mathcal{S}$ is a finite set of pairs $\{(a_i, b_i)\}_{i=1}^n$. $\mathcal{S}$ is the set of all samples. The encoding of a sample is $\langle S \rangle = \langle \{(a_i, b_i)\}_{i=1}^n \rangle$. We say $G(S)$ if G is consistent with S, with $G(a_i) = b_i$, $i = 1, \ldots, n$. For random functions, $F(S)$ is the event that F is consistent with S. The amount of information that a string has with the halting sequence is $\mathbf{I}(x; \mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$.

Theorem C. *For $F \in \mathcal{F}$, $S \in \mathcal{S}$, $\min_{G \in \mathcal{G}, G(S)} \mathbf{K}(G) <^{\log} \mathbf{K}(F) - \log \Pr[F(S)] + \mathbf{I}(\langle S \rangle; \mathcal{H})$.*

1.3 Open Sets

For $x \in \{0,1\}^*$ let $\Gamma_x = \{x\beta : \beta \in \{0,1\}^\infty\}$ be the interval of x. For open set $S \subseteq \{0,1\}^\infty$, let its encoding be $\langle S \rangle = \langle \{x : \Gamma_x \text{ is maximal in } S\} \rangle$. Arbitrary open sets $S \subset \{0,1\}^\infty$ can have infinite $\langle S \rangle$. The halting sequence is $\mathcal{H} \in \{0,1\}^\infty$. The Kolmogorov complexity of an infinite sequence $\alpha \in \{0,1\}^\infty$ is $\mathbf{K}(\alpha)$, the size of the smallest program to a universal Turing machine that will output, without halting, α on the output tape. Let μ be the uniform measure of the Cantor space. The information term between infinite sequences is $\mathbf{I}(\alpha : \beta) = \log \sum_{x,y \in \{0,1\}^*} \mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\mathbf{I}(x:y)}$, where $\mathbf{m}$ is the algorithmic probability [Lev74]. The mutual information between two finite strings is defined to be $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$.

Theorem D. *For open $S \subseteq \{0,1\}^\infty$, $\min_{\alpha \in S} \mathbf{K}(\alpha) <^{\log} -\log \mu(S) + \mathbf{I}(\langle S \rangle : \mathcal{H})$.*

1.4 Other Results

Theorem D is a variation of the main theorem in [Lev16, Eps19]. We discuss continuous sampling methods as well as sampling methods that can not halt with positive probability. We prove slightly stronger results to Theorem D for clopen sets. Derandomization can be generalized to sets of samples, and also to lower computable random functions. We apply function derandomization to games. A monotone complexity variant to the main theorem in [Lev16, Eps19] is proven. We also show that there is no equivalent to Theorem D for closed sets. Due to Anonymous, there exists closed sets $C \subset \{0,1\}^\infty$ with no computable members, $\mu(C) > 0$, and $\mathbf{I}(\langle C \rangle : \mathcal{H}) < \infty$.

2 Conventions

Let $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, $\{0,1\}$, $\{0,1\}^*$, and $\{0,1\}^\infty$ be the sets of natural numbers, rationals, real numbers, bits, finite strings, and infinite strings. We use $\langle x\rangle$ to represent a self-delimiting code for $x \in \{0,1\}^*$, such as $1^{\|x\|}0x$. The self-delimiting code for a finite set of strings $\{a_i\}_{i=1}^n$ is $\langle\{a_i\}_{i=1}^n\rangle = \langle n\rangle\langle a_1\rangle\langle a_2\rangle\dots\langle a_n\rangle$.

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 term $\overset{*}{=}f$ denotes $\overset{*}{<}f$ and $\overset{*}{>}f$. For the nonnegative real function f, the terms $<^{\log} f$, $>^{\log} f$, and $=^{\log} f$ represent the terms $<f+O(\log(f+1))$, $>f-O(\log(f+1))$, and $=f\pm O(\log(f+1))$, respectively.

A semi measure is a function $Q : \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that $\sum_{a\in\mathbb{N}} Q(a) \leq 1$. A probability measure is a semi measure such that $\sum_{a\in\mathbb{N}} Q(a) = 1$. A probability measure Q is elementary if $|\{a : Q(a) > 0\}| < \infty$ and $\text{Range}(Q) \subseteq \mathbb{Q}_{\geq 0}$. Elementary measures Q can be encoded into finite strings $\langle Q\rangle$.

The universal probability of a string $x \in \{0,1\}^*$, conditional on $y \in \{0,1\}^* \cup \{0,1\}^\infty$, is $\mathbf{m}(x|y) = \sum\{2^{-\|p\|} : U_y(p) = x\}$ The coding theorem states $-\log \mathbf{m}(x|y) =^+ \mathbf{K}(x|y)$. The mutual information of a string x with the halting sequence is $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$, where $\mathcal{H} \in \{0,1\}^\infty$ is the halting sequence.

This paper uses notions of stochasticity in the field of algorithmic statistics [VS17]. A string x is stochastic, i.e. has a low $\mathbf{Ks}(x)$ score, if it is typical of a simple probability distribution. The deficiency of randomness function of a string x with respect to an elementary probability measure P conditional to $y \in \{0,1\}^*$, is $\mathbf{d}(x|P,y) = \lfloor -\log P(x)\rfloor - \mathbf{K}(x|\langle P\rangle, y)$.

Definition 1 (Stochasticity) *For $x, y \in \{0,1\}^*$,*
$\mathbf{Ks}(x|y) = \min\{\mathbf{K}(P|y)+3\log\max\{\mathbf{d}(x|P,y),1\} : P$ *is an elementary probability measure*$\}$. $\mathbf{Ks}(x) = \mathbf{Ks}(x|\emptyset)$. $\mathbf{Ks}(a|b) < \mathbf{Ks}(a) + O(\log \mathbf{K}(b))$.

3 Dynamical Systems

In this section, we prove that dynamical systems will hit ever larger outliers with diminishing probability. To achieve this, we use the properties of the mutual information of an infinite sequence with the halting problem. The deficiency of randomness of an infinite sequence $\alpha \in \{0,1\}^\infty$ with respect to a computable probability measure P over $\{0,1\}^\infty$ is defined to be

$$\mathbf{D}(\alpha|P,x) = \log \sup_n \mathbf{m}(\alpha[0..n]|x)/P(\alpha[0..n]).$$

We have $\mathbf{D}(\alpha|P) = \mathbf{D}(\alpha|P,\emptyset)$. We require the following two theorems for the primary proof of this section.

Theorem 1 ([Ver21, Lev74, Gei12]) $\Pr_\mu(\mathbf{I}(\alpha : \mathcal{H}) > n) \overset{*}{<} 2^{-n+\mathbf{K}(\mu)}$.

Theorem 2 ([Eps21]) *For computable probability measure P over $\{0,1\}^\infty$, for $Z \subseteq \{0,1\}^\infty$, if $\mathbb{N} \ni s < \log \sum_{\alpha\in Z} 2^{\mathbf{D}(\alpha|P)}$, then $s < \sup_{\alpha\in Z} \mathbf{D}(\alpha|P) + \mathbf{I}(\langle Z\rangle : \mathcal{H}) + O(\mathbf{K}(s) + \log \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \mathbf{K}(P))$.*

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

Proof. We first assume not. For all $c \in \mathbb{N}$, there exist computable nonatomic measures μ, λ, and there exists n, where $\lambda\{\alpha : \mathbf{D}(\alpha|\mu) > n\} \leq 2^{-n-\mathbf{K}(n,\mu,\lambda)-c}$. Sample $2^{n+\mathbf{K}(n,\mu,\lambda)+c-1}$ elements $D \subset \{0,1\}^\infty$ according to λ. The probability that all samples $\beta \in D$ have $\mathbf{D}(\beta|\mu) \leq n$ is

$$\prod_{\beta \in D} \lambda\{\mathbf{D}(\beta|\mu) \leq n\} \geq (1 - |D|2^{-n-\mathbf{K}(n,\mu,\lambda)-c}) \geq (1 - 2^{n+\mathbf{K}(n,\mu,\lambda)+c-1}2^{-n-\mathbf{K}(n,\mu,\lambda)-c}) \geq 1/2.$$

Let $\lambda^{n,c}$ be the probability of an encoding of $2^{n+\mathbf{K}(n,\mu,\lambda)+c-1}$ elements each distributed according to λ. Thus

$$\lambda^{n,c}(\text{Encoding of } 2^{n+\mathbf{K}(n,\mu,\lambda)+c-1} \text{ elements } \beta, \text{ each having } \mathbf{D}(\beta|\mu) \leq n) \geq 1/2.$$

Let v be a shortest program to compute $\langle n, \mu, \lambda\rangle$. By Theorem 1, with the universal Turing machine relativized to v, $\lambda^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|v) > m\}) \stackrel{*}{<} 2^{-m+\mathbf{K}(n,\mathbf{K}(n,\mu,\lambda),c,\lambda|v)} \stackrel{*}{<} 2^{-m+\mathbf{K}(c)}$. Therefore, there is a constant $f \in \mathbb{N}$, with

$$\lambda^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|v) > \mathbf{K}(c) + f\}) \leq 1/4.$$

Thus, by probabilistic arguments, there exists $\alpha \in \{0,1\}^\infty$, such that α is an encoding of $2^{n+\mathbf{K}(n,\mu,\lambda)+c-1}$ elements $\beta \in D \subset \{0,1\}^\infty$, where each β has $\mathbf{D}(\beta|\mu) \leq n$ and $\mathbf{I}(\alpha : \mathcal{H}|v) <^+ \mathbf{K}(c)$. By Theorem 2, relativized to v, there are constants $d, f \in \mathbb{N}$ where

$$\begin{aligned} m = \log|D| &< \max_{\beta \in D} \mathbf{D}(\beta|\mu, v) + 2\mathbf{I}(D : \mathcal{H}|v) + d\mathbf{K}(m|v) + f\mathbf{K}(\mu|v) \\ &<^+ \max_{\beta \in D} \mathbf{D}(\beta|\mu) + \mathbf{K}(n,\mu,\lambda) + 2\mathbf{K}(c) + d\mathbf{K}(m|v) + f\mathbf{K}(\mu|v) \\ &<^+ n + \mathbf{K}(n,\mu,\lambda) + d\mathbf{K}(m|v) + 2\mathbf{K}(c). \end{aligned} \tag{1}$$

Therefore:

$$\begin{aligned} m &= n + \mathbf{K}(n,\mu,\lambda) + c - 1 \\ \mathbf{K}(m|v) &<^+ \mathbf{K}(c). \end{aligned}$$

Plugging the inequality for $\mathbf{K}(m|v)$ back into Equation 1 results in

$$\begin{aligned} n + \mathbf{K}(n,\mu,\lambda) + c &<^+ n + \mathbf{K}(n,\mu,\lambda) + 2\mathbf{K}(c) + d\mathbf{K}(c) \\ c &<^+ (2+d)\mathbf{K}(c). \end{aligned}$$

This result is a contradiction for sufficiently large c solely dependent on the universal Turing machine. □

Similar to the construction in the introduction, we can define a universal conditional lower computable integral test $T(\alpha|n)$ over a sequence of uniformly computable measures Q_1, Q_2, ... over $\{0,1\}^\infty$. We can also define the randomness deficiency to be $\mathbf{D}_n(\alpha|Q) = \log T(\alpha|n)$. The following corollary is derived from the fact that $\mathbf{D}_n(\alpha|\mu, n) = \mathbf{D}_n(\alpha|\mu)$.

Corollary 1 *For uniformly computable measures $\{\mu_i\}$ and nonatomic $\{\lambda_i\}$ over $\{0,1\}^\infty$, for all n, $\lambda_n\{\alpha : \mathbf{D}_n(\alpha|\mu) > n\} > 2^{-n-\mathbf{K}(\mu,\lambda)-O(1)}$.*

Theorem 3 can be extended to incomputable λ, which can be accomplished using a stronger version of Theorem 1. The term $\langle\lambda\rangle \in \{0,1\}^\infty$ in the following corollary represents any encoding of λ that can compute $\lambda(x\{0,1\}^\infty)$ for $x \in \{0,1\}^*$ up to arbitrary precision.

Corollary 2

- *For measures μ and λ over $\{0,1\}^\infty$, nonatomic λ, computable μ, for all n,* $\lambda\{\alpha : \mathbf{D}(\alpha|\mu) > n\} > 2^{-n-\mathbf{K}(n,\mu)-\mathbf{I}(\langle\lambda\rangle:\mathcal{H})-O(\log \mathbf{I}(\langle\lambda\rangle:\mathcal{H}))}$.
- *For measures μ and λ over $\{0,1\}^\infty$, nonatomic λ, computable μ, if for every $c \in \mathbb{N}$, there is an $n \in \mathbb{N}$, where $\lambda\{\alpha : \mathbf{D}(\alpha|\mu) > n\} < 2^{-n-\mathbf{K}(n)-c}$, then $\mathbf{I}(\langle\lambda\rangle : \mathcal{H}) = \infty$.*

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 $\{0,1\}^\infty$; $\mathcal{B}$ be the Borel σ-algebra on $\{0,1\}^\infty$; λ and μ be computable measures over $\{0,1\}^\infty$ and λ be nonatomic; and $(\{0,1\}^\infty, \mathcal{B}, \lambda)$ be a measure space and $T : \{0,1\}^\infty \to \{0,1\}^\infty$ be an ergodic measure preserving transformation. By the Birkoff theorem,

Corollary 3 *Starting λ-almost everywhere, $\stackrel{*}{>} \mathbf{m}(n,\mu,\lambda)2^{-n}$ states α visited by iterations of T have $\mathbf{D}(\alpha|\mu) > n$.*

4 Outliers Theorem

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

Lemma 1 *Let $P = P_1, P_2 \ldots$ be a uniformly computable sequence of measures on strings and let A be a sampling method. For all integers M and N, there exists a finite set $S \subset \{0,1\}^*$ such that $P(S) \leq 2M/N$, and with probability strictly more than $1 - 2e^{-M}$: $A(N)$ intersects S.*

Proof. We show that some possibly infinite set S satisfies the conditions, and thus, some finite subset also satisfies the conditions due to the strict inequality. We use the probabilistic method: we select each string to be in S with probability M/N and show that 2 conditions are satisfied with positive probability. The expected value of $P(S)$ is M/N. By the Markov inequality, the probability that $P(S) > 2M/N$ is at most $1/2$. For any set D containing N strings, the probability that S is disjoint from D is

$$(1 - M/N)^N < e^{-M}.$$

Let Q be the measure over N-element sets of strings generated by the sampling algorithm $A(N)$. The left-hand side above is equal to the expected value of

$$Q(\{D : D \text{ is disjoint from } S\}).$$

Again by the Markov inequality, with probability less than $1/2$, this measure is less than $2e^{-M}$. By the union bound, the probability that at least one of the conditions is violated is less than $1/2 + 1/2$. Thus, with positive probability a required set is generated, and thus such a set exists.□

Theorem 4 *Let $P = P_1, P_2 \ldots$ be a uniformly computable sequence of measures on strings and let A be a sampling method. There exists $c \in \mathbb{N}$ such that for all n and k:*

$$\Pr\left(\max_{a \in A(2^n)} \overline{\mathbf{d}}_n(a|P) > n - k - c\right) \geq 1 - 2e^{-2^k}.$$

Proof. We now fix a search procedure that on input N and M finds a set $S_{N,M}$ that satisfies the conditions of Lemma 1. Let $t'(a|n)$ be the maximal value of $2^n/2^{k+2}$ such that $a \in S_{2^n,2^k}$ for some integer k. By construction, t' is a computable probability bound test, because $P(\{x : t'(x|n) = 2^\ell\}) \leq 2^{-\ell-1}$, and thus $P(t'(x|n) \geq 2^\ell) \leq 2^{-\ell-1} + 2^{-\ell-2} + \ldots$ With the given probability, the set $A(2^n)$ intersects $S_{2^n,2^k}$. For any number a in the intersection, we have $t'(x|n) \geq 2^{n-k-2}$, thus by the optimality of t and definition of d, we have $\overline{\mathbf{d}}_n(a|P) > n - k - O(1)$. □

An incomplete sampling method A takes in a natural number $n \in \mathbb{N}$ and outputs, with probability $f(n)$, a set of n numbers. Otherwise A outputs $\perp$. f is computable.

Corollary 4 *Let $P = P_1, P_2 \ldots$ be a uniformly computable sequence of measures on strings and let A be an incomplete sampling method. There exists $c \in \mathbb{N}$ such that for all n and k:*

$$\Pr_{D=A(n)}\left(D \neq \perp \text{ and } \max_{a\in D} \overline{\mathbf{d}}_n(a|P) \leq n - k - c\right) < 2e^{-2^k}.$$

4.1 Continuous Sampling Method

For a mathematical statement W, $[W] = 1$ if W is true, and $[W] = 0$ otherwise. Let $\mu = \mu_1, \mu_2, \ldots$ be a uniformly computable sequence of measures over infinite sequences. Similar way as for strings, the randomness deficiency $\overline{\mathbf{D}}_n(\omega|\mu)$ for sequences ω is defined using lower-semicomputable functions $\{0,1\}^\infty \times \mathbb{N} \to \mathbb{R}_{\geq 0}$. A continuous sampling method C is a probabilistic function that maps, with probability 1, an integer N to an infinite encoding of N different sequences.

Theorem 5 *There exists $c \in \mathbb{N}$ where for all n:*

$$\Pr\left(\max_{\alpha\in C(2^n)} \overline{\mathbf{D}}_n(\alpha|\mu) > n - k - c\right) \geq 1 - 2.5e^{-2^k}.$$

Proof. For $D \subseteq \{0,1\}^\infty$, $D_m = \{x : \|x\| = m, x\omega \in D\}$. Let $g(n) = \arg\min_m \Pr_{D=C(n)}(|D_m| < n) < 0.5e^{-2^n}$ be the smallest number such that the initial m-segment of $C(n)$ are sets of n strings with very high probability. g is computable, because C outputs a set with probability 1. For probability ψ over $\{0,1\}^\infty$, let $\psi^m(x) = [|x| = m]\psi(\{\omega : x \sqsubset \omega\})$. Let $\mu^g = \mu_1^{g(1)}, \mu_2^{g(2)}, \ldots$ be a uniformly computable sequence of discrete probability measures and let A be a discrete incomplete sampling method, where for random seed $\omega \in \{0,1\}^\infty$, $A(n,\omega) = C(n,\omega)_{g(n)}$ if $|C(n,\omega)_{g(n)}| = n$; otherwise $A(n,\omega) = \perp$. Thus due to Corollary 4,

$$\begin{aligned}
&\Pr\left(\max_{\alpha\in C(2^n)} \overline{\mathbf{D}}_n(\alpha|\mu) \leq n - k - O(1)\right)\\
\leq &\Pr_{Z=C(2^n)}\left((|Z_{g(n)}| < 2^n) \text{ or } (|Z_{g(n)}| = 2^n \text{ and } \max_{\alpha\in Z} \overline{\mathbf{D}}_n(\alpha|\mu) \leq n - k - O(1)\right)\\
< &\Pr_{D\in A(2^n)}\left(D = \perp \text{ or } (D \neq \perp \text{ and } \max_{x\in D} \overline{\mathbf{d}}_n(x|\mu^g) \leq n - k - O(1))\right)\\
< &0.5e^{-2^n} + 2e^{-2^k}\\
\leq &2.5e^{-2^k}.
\end{aligned}$$

4.2 Alternative Proof to Theorem 3

Using the theorem of the previous section, one can produce a simple proof to a variant of Theorem 3. The longer proof was included due to of its tight error terms as well as its corollaries extending the results to incomputable measures. Let $\lambda = \lambda_1, \lambda_2, \ldots$ and $\mu = \mu_1, \mu_2, \ldots$ be uniformly computable sequences of measures over infinite sequences. Let λ_n be nonatomic.

Theorem 6 *There is a constant $c \in \mathbb{N}$, where for all $n \in \mathbb{N}$, $\lambda_n \left\{\alpha : \overline{\mathbf{D}}_n(\alpha|\mu) > n\right\} > 2^{-n-c}$.*

Proof. Let $d_n = \lambda_n\{\alpha : \overline{\mathbf{D}}_n(\alpha|\mu) > n - O(1)\}$. We define the continuous sampling method C, where on input n, randomly sample n elements from λ_n. By Theorem 5, where $k = 0$,

$$\begin{aligned} 1 - (1 - d_n)^{2^n} &> 1 - 2.5e^{-1} \\ 1 - 2^n d_n &< 2.5/e \\ d_n &> (1 - 2.5/e)2^{-n}. \end{aligned}$$

4.3 Necessity of Double Exponential

Theorem 4 showed that the probability that $A(2^n)$ contains no strings of randomness deficiency less than $n - k$ decreases double exponentially in k. We show that at least a double exponential probability is required for $k = n - O(1)$. Let P_n be the uniform measure on $(n+2)$-bit strings. The algorithm A that on input 2^n generates a random set of 2^n strings of length $n + 2$ satisfies

$$\Pr\left(\forall x \in A(2^n) : \overline{\mathbf{d}}_n(x|P) \leq 2\right) \geq 2^{-2^n}.$$

For at most a quarter of the $(n+2)$-bit strings, we have $\overline{\mathbf{d}}_n(x|P) \geq 3$, by definition of a probability bounded test t. A random selection of $N = 2^n$ different $(n+2)$-bit strings, contains no such string with a probability of at least 2^{-N}. We consider the following situation. In a bag with $4N$ balls, N balls are marked. One selects N balls one by one. We consider the probability that no marked ball is drawn if previously no marked ball was drawn. The smallest probability appears at the last draw when there are $T = 4N - (N - 1)$ balls in the bag. This probability is $(T - N)/T \geq 1/2$.

4.4 Partial Sampling Methods

A partial sampling method is a sampling method that can output with probability less than 1. Theorem 4 does not hold for partial sampling methods B. Let P_n be the uniform measure on $(n+1)$-bit strings. Let $\#B(N)$ represent the event that B halts and outputs a set of size N. We present a partial sampling method B for which

$$\Pr\left(\#B(2^n) \text{ and } \forall x \in B(2^n) : \overline{\mathbf{d}}_n(x|P) \leq 1\right) \geq 2^{-n}.$$

For at most half of the $(n+1)$-bit strings, we have $\overline{\mathbf{d}}_n(x|P) \geq 2$. On input 2^n,the partial sampling method B generates a random natural number s bounded by 2^n, searches for s strings x of length $n+1$ with $\overline{\mathbf{d}}_n(x|P) \geq 2$, and outputs 2^n other $(n+1)$-bit strings. For some s, this search may never terminate. If A chooses to be precisely equal to the number of strings satisfying the condition, then it outputs only strings with deficiency at most 1, and the claim is proven. However partial sampling methods do exhibit the following properties

Proposition 1 *Let $P = P_1, P_2, \ldots$ be a uniformly computable sequence of measures and B be a partial sampling method, where $\#B(N)$ represents the event that $B(N)$ terminates and outputs a set of N strings.*

$$\Pr\left(\#B(N) \text{ and } \forall x \in B(2^n) : \overline{\mathbf{d}}_n(x|P) \leq n-k\right) \leq O(k2^{-k}).$$

Proof. Let Q be the lower-semicomputable semimeasure over sets of size 2^n such that $Q(D)$ equals the probability that $B(N) = D$. We show that

$$\Pr\left(\#B(N) \text{ and } \forall x \in B(2^n) : \overline{\mathbf{d}}_n(x|P) \leq n-k+\log k + O(1)\right) \leq O(2^{-k}).$$

This result is followed by a redefinition of k. We write Q as a uniform mixture over at most 2^k measures Q_i with finite support, and one lower semi-computable semimeasure Q_*:

$$Q = 2^{-k}\left(Q_1 + Q_2 + \ldots Q_f + Q_*\right).$$

With $f \leq 2^k$, we assume that the finite descriptions of $Q_1, \ldots, Q_f$ are enumerated one by one by a program (that may never terminate). For each enumerated measure Q, we search for a set S_i that satisfies the conditions of Lemma 1 for $M = k$. Let $S = \bigcup_{i \leq f} S_i$.Also, $P(S) \leq k2^{k+1-n}$; thus every element in S satisfies $\overline{\mathbf{d}}_n(x|P) \geq n-k+\log k + O(1)$.

The probability that $A(2^n)$ produces a set that does not contain such an element is at most $2^{-k} + 2e^{-k}$ because we can equivalently generate a set D by randomly selecting j from the list $[1, \ldots, f, *, \infty]$ with probabilities $[2^{-k}, \ldots, 2^{-k}, 2^{-k}r, 1-(f+r)2^{-k}]$ and generating a random set D from Q_j if $j \neq \infty$ and letting D be undefined otherwise. The probability that D is defined and does not contain an element from S is at most the probability $j = *$, which is $\leq 2^{-k}$, plus the probability that $j \in \{1, \ldots, f\}$ times $2e^{-k}$. □

5 Function Derandomization

We recall the definitions from the introduction. Random functions F over natural numbers are modeled by discrete stochastic processes indexed by $\mathbb{N}$, where each $F(t)$, $t \in \mathbb{N}$, is a random variable over $\mathbb{N}$. $\mathcal{F}$ is the set of all random functions. A random function $F \in \mathcal{F}$ is computable if there is a program that on input $(a_1, \ldots, a_n)$ lower computes $\Pr[F(1) = a_1 \cap F(2) = a_2 \cap \cdots \cap F(n) = (a_n)]$. Put another way, a random function $F \in \mathcal{F}$ is computable if $A = \Pr[F(a_1) = b_1 \cap \cdots \cap F(a_n) = b_n]$ is uniformlly computble in $\{(a_i, b_i)\}_{i=1}^n$. The complexity $\mathbf{K}(F)$ of a random function $F \in \mathcal{F}$, is the smallest program that computes A. $\mathcal{G}$ is the set of all deterministic functions $G : \mathbb{N} \to \mathbb{N}$. A sample $S \in \mathcal{S}$ is a finite set of pairs $\{(a_i, b_i)\}_{i=1}^n$. The encoding of a sample is $\langle S\rangle = \langle\{(a_i, b_i\}_{i=1}^n\rangle$. $\mathcal{S}$ is the set of all samples. We say $G(S)$ if G is consistent with S, with $G(a_i) = b_i$, $i = 1, \ldots, n$. For random functions, $F(S)$ is the event that F is consistent with S.

To prove function derandomization, we leverage the Baire space $\mathbb{N}^{\mathbb{N}}$. Individual cylinders are $C_n[v] = \{(a_1, a_2, \ldots) \in \mathbb{N}^{\mathbb{N}} : a_n = v\}$. Cylinders are generators for cylinder sets. The cylinder sets $C \in \mathcal{C}$ consists of all intersections of a finite number of cylinders. If $C = \bigcap_{i \in I} C_i[v_i]$, then for all $i \in I$, we say $i \in \overline{C}$. The set of all such cylinder sets provides a basis for the product topology of $\mathbb{N}^{\mathbb{N}}$. The encoding of a cylinder set $C = \bigcap_{i \in I} C_i[v_i]$, is $\langle C\rangle = \langle\{i, v_i\}_{i \in I}\rangle$. The set of all Borel probability measures over $\mathbb{N}^{\mathbb{N}}$ is $\mathcal{P}$. A probability $P \in \mathcal{P}$ is computable if given an encoding of a cylinder set $C \in \mathcal{C}$, $P(C)$ is computable.

Proposition 2
For every $c, n \in \mathbb{N}$, if $x < y + c$ for some $x, y \in \mathbb{N}m$ then $x + n\mathbf{K}(x) < y + n\mathbf{K}(y) + O(n \log n) + 2c$.

Proof. $\mathbf{K}(x) <^+ \mathbf{K}(y) + \mathbf{K}(y-x)$ as x can be computed from y and $(y-x)$. Therefore $n\mathbf{K}(x) - n\mathbf{K}(y) < n\mathbf{K}(y-x) + dn$, for some $d \in \mathbb{N}$ dependent on U. We assume that this equation is not true; then, there exists $x, y, c \in \mathbb{N}$ where $x < y + c$, and $g \leq O(n \log n) + 2c$ where $y - x + g < n\mathbf{K}(x) - n\mathbf{K}(y) < n\mathbf{K}(y-x) + dn$, which is a contradiction for $g =^+ dn + 2c + \max_a\{2n \log a - a\} =^+ dn + 2c + 2n \log n$.

Theorem 7 *For $F \in \mathcal{F}$, $S \in \mathcal{S}$, if $s = \lceil -\log \Pr[F(S)] \rceil$ and $h = \mathbf{I}(\langle S \rangle; \mathcal{H})$, then $\min_{G \in \mathcal{G}, G(S)} \mathbf{K}(G) < \mathbf{K}(F) + s + h + O(\mathbf{K}(s, h))$.*

Proof. Each sample $S \in \mathcal{S}$ where $S = \{(i, v_i)\}_{i \in I}$ can be identified by a cylinder set $C_S \in \mathcal{C}$ where $C_S = \cap_{i \in I} c_i(v_i)$. For every $\alpha \in \mathbb{N}^{\mathbb{N}}$ there is a deterministic function $G_\alpha : \mathbb{N} \to \mathbb{N}$, where $G_\alpha(i) = \alpha[i]$. Furthermore if $\alpha \in C_S$, then for all $(i, v_i) \in S$, $G_\alpha(i) = v_i$. For each random function $F \in \mathcal{F}$, we can identify a probability $P_F \in \mathcal{P}$ such that for each sample $S = \{(i, v_i)\}_{i \in I} \in \mathcal{S}$, $\Pr[F(S)] = P_F(C_S)$. This is because random functions and Borel probability measures over $\mathbb{N}^{\mathbb{N}}$ have the same form. Furthermore, if F is computable, then P_F is computable, with

$$\mathbf{K}(P_F | F) = O(1) \tag{2}$$

This is because given an encoding $\langle F \rangle$ and an encoded cylinder set $\langle C \rangle$, one can compute $\Pr[F(C)]$, which is equal to $P_F(C)$. Thus given a random function $F \in \mathcal{F}$ and sample $S \in \mathcal{S}$, by Lemma 2 applied to $P_F \in \mathcal{P}$ and $C_S \in \mathcal{C}$, we get the following result, with $h_C = \mathbf{I}(\langle C_S \rangle; \mathcal{H})$, $h_s = \mathbf{I}(\langle S \rangle; \mathcal{H})$, and $s = -\log P_F(C_S)$,

$$\min_{\alpha \in C_S} \mathbf{K}(a) < \mathbf{K}(P_F) + s + h_C + O(\mathbf{K}(s)) + O(\mathbf{K}(h_C))$$

$$\min_{G \in \mathcal{G}: G(S)} \mathbf{K}(G) < \mathbf{K}(P_F) + s + h_C + O(\mathbf{K}(s)) + O(\mathbf{K}(h_C)) \tag{3}$$

$$\min_{G \in \mathcal{G}: G(S)} \mathbf{K}(G) < \mathbf{K}(F) + s + h_C + O(\mathbf{K}(s)) + O(\mathbf{K}(h_C)) \tag{4}$$

$$\min_{G \in \mathcal{G}: G(S)} \mathbf{K}(G) < \mathbf{K}(F) + s + h_S + O(\mathbf{K}(s)) + O(\mathbf{K}(h_S)) \tag{5}$$

$$\min_{G \in \mathcal{G}, G(S)} \mathbf{K}(G) < \mathbf{K}(F) - \log \Pr[F(S)] + \mathbf{I}(\langle S \rangle; \mathcal{H}) + O(\mathbf{K}(\lceil -\log \Pr[F(S)] \rceil, \mathbf{I}(\langle S \rangle; \mathcal{H}))). \tag{6}$$

Equation 3 is because for the $\alpha \in \mathbb{N}^{\mathbb{N}}$ that minimizes the leftmost term, $G_\alpha \in \mathcal{G}$, with $G_\alpha(S)$ and $\mathbf{K}(G_\alpha) <^+ \mathbf{K}(\alpha)$. Equation 4 is because P_F can be constructed from F, i.e. Equation 2. Equation 5 is due to Proposition 2, Lemma 3 and the fact that $\mathbf{K}(\langle C_S \rangle | \langle S \rangle) = O(1)$. Equation 6 is due to the definition of P_F, where $s = \lceil -\log P_F(C_S) \rceil = \lceil -\log \Pr[F(S)] \rceil$. □

Lemma 2 *For cylinder set $C \in \mathcal{C}$, computable probability $P \in \mathcal{P}$, if $s = \lceil -\log P(C) \rceil$ and $h = \mathbf{I}(\langle C \rangle; \mathcal{H})$, then $\min_{\alpha \in C} \mathbf{K}(\alpha) < \mathbf{K}(P) + s + h + O(\mathbf{K}(s, h))$.*

Proof. We put (s, P) on an auxiliary tape to the universal Turing machine U. Thus, all algorithms have access to (s, P), and all complexities implicitly have (s, P) as conditional terms.

Let Q be an elementary probability measure that realizes $\mathbf{Ks}(\langle C \rangle)$. Let $d = \max\{\mathbf{d}(\langle C \rangle | Q), 1\}$ and $c \in \mathbb{N}$ be a constant to be chosen later. Let $n = \max\{m : m \in \overline{W}, W \in \mathcal{C}, \langle W \rangle \in \mathrm{Supp}(Q)\}$. For a list L of a list of numbers and cylinder set $W \in \mathcal{C}$, we say $L \rtimes W$ is the set of all $x \in L$ with $x\mathbb{N}^{\mathbb{N}} \subseteq W$. We define a measure κ over $cd2^s$ lists of lists of n numbers L, where $\kappa(L) = \prod_{i=1}^{cd2^s} P(L[i]\mathbb{N}^{\mathbb{N}})$. Given a list of lists of n numbers L, $\kappa(L)$ is computable (as a program for P is

on an auxiliary tape). We use the indicator function $\mathbf{i}(L, W) = [W \in \mathcal{C}, P(W) \geq 2^{-s}, L \rtimes W = \emptyset]$. The function $\mathbf{i}$ is computable, because $P(W)$ and $L \rtimes W$ are computable for all $W \in \mathcal{C}$.

$$\begin{aligned}
\mathbf{E}_{L\sim\kappa}\mathbf{E}_{W\sim Q} &\leq \sum_W Q(W) \Pr_{L\sim\kappa}(W \in \mathcal{C}, P(W) \geq 2^{-s}, L \rtimes W = \emptyset) \\
&\leq \sum_W Q(W) \prod_{i=1}^{cd2^s} (1 - 2^{-s}) \\
&\leq \sum_W Q(W)(1 - 2^{-s})^{cd2^s} \\
&< e^{-cd}
\end{aligned}$$

Thus there exists a list L' of $cd2^s$ sequences of numbers of length n such that $\mathbf{E}_{W\sim Q}[\mathbf{i}(W, L')] = e^{-cd}$. Thus $t(W) = \mathbf{i}(W, L')e^{cd}$ is a Q-test, with $\sum_W Q(W)t(W) \leq 1$. It must be that $L \rtimes C \neq \emptyset$. Otherwise $t(C) = e^{cd}$, and

$$\begin{aligned}
\mathbf{K}(C|c, d, Q) &<^+ -\log t(C)Q(C) \\
&<^+ -\log Q(C) - (\lg e)cd \\
(\lg e)cd &<^+ -\log Q(C) - \mathbf{K}(C|P) + \mathbf{K}(d, c) \\
(\lg e)cd &< d + \mathbf{K}(d, c) + O(1).
\end{aligned}$$

This is a contradiction for c large enough solely dependent on the universal Turing machine. We roll c into the additive constants of the rest of the proof. Thus there exists $x \in L \rtimes C$ where

$$\begin{aligned}
\mathbf{K}(x) &<^+ \log|L| + \mathbf{K}(L) \\
&<^+ \log|L| + \mathbf{K}(d, Q) \\
&<^+ \log d + s + \mathbf{K}(d) + \mathbf{K}(Q) \\
&<^+ s + 3\log d + \mathbf{K}(Q) \\
&<^+ s + \mathbf{Ks}(C).
\end{aligned}$$

Thus making the relativization of (s, p) explicit,

$$\begin{aligned}
\min_{\alpha\in C} \mathbf{K}(\alpha|\langle P, s\rangle) <^+ \mathbf{K}(x|\langle P, s\rangle) <^+ s + \mathbf{Ks}(\langle C\rangle|\langle P, s\rangle) \\
\min_{\alpha\in C} \mathbf{K}(\alpha) < \mathbf{K}(P) + s + \mathbf{Ks}(\langle C\rangle) + O(\mathbf{K}(s) + \log\mathbf{K}(P)) \\
\min_{\alpha\in C} \mathbf{K}(\alpha) < \mathbf{K}(P) + s + \mathbf{I}(\langle C\rangle; \mathcal{H}) + O(\mathbf{K}(s, \mathbf{I}(\langle C\rangle; \mathcal{H})). \qquad (7)
\end{aligned}$$

Equation 7 follows from Lemma 10 in [Eps21], which states $\mathbf{Ks}(x) < \mathbf{I}(x; \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(x; \mathcal{H})))$. □

Theorem 7 can be readily extended to sets of samples $\mathfrak{S} = \{S_1, \ldots, S_n\}$, where for deterministic function $G : \mathbb{N} \to \mathbb{N}$, $G(\mathfrak{S})$ if $\bigcup_{i=1}^n G(S_i)$. For random function $F \in \mathcal{F}$, $F(\mathfrak{S})$ is the union of events $F(S_i)$, $i = 1, \ldots, n$. The proof of the following corollary follows almost identically to the proofs of Theorem 7 and Lemma 2, noting that $P(\mathfrak{S})$ is computable given a computable probability $P \in \mathcal{P}$ and a finite description of a set of samples $\mathfrak{S}$.

Corollary 5 *For $F \in \mathcal{F}$, if $s = \lceil -\log\Pr[F(\mathfrak{S})]\rceil$ and $h = \mathbf{I}(\langle\mathfrak{S}\rangle; \mathcal{H})$, then $\min_{G\in\mathcal{G}, G(\mathfrak{S})} \mathbf{K}(G) <^{\log} \mathbf{K}(F) + s + h + O(\mathbf{K}(s, h))$.*

Another generalization of Theorem 7 is in the usage of lower computable random functions V. They are discrete stochastic processes $V(t)$, indexed by $t \in \mathbb{N}$, where each $V(t)$ is a random variable over $\mathbb{N} \cup \infty$. Furthermore $\Pr(V(1) = a_1 \cap V(2) = a_2 \cap \cdots \cap V(n) = a_n)$ is lower computable, where $a_i \in \mathbb{N}$, $i = 1 \ldots n$. The proof is extensive, relying on left total machines, introduced in [Lev16, Eps19].

6 Open Sets

We recall that the Kolmogorov complexity of an infinite sequence $\alpha \in \{0,1\}^\infty$ is $\mathbf{K}(\alpha)$, the size of the smallest program to a universal Turing machine that will output, without halting, α on the output tape.

Theorem 8 *For clopen set $C \subseteq \{0,1\}^\infty$, if $s = \lceil -\log \mu(C) \rceil$ and $h = \mathbf{I}(\langle C \rangle; \mathcal{H})$, $\min_{\alpha \in C} \mathbf{K}(\alpha) < s + h + O(\mathbf{K}(s,h))$.*

Proof. We define a collection of samples $\mathfrak{S}$, where for each maximal interval $\Gamma_x \subseteq C$, $x \in \{0,1\}^*$, we add the sample $S = \{(i, x[i])\}_{i=1}^{\|x\|}$. Thus $\mathbf{K}(\langle \mathfrak{S} \rangle | \langle C \rangle) = O(1)$. Furthermore we define a stochastic process $F(t)$ over $\mathbb{N}$, indexed by $t \in \mathbb{N}$ using the uniform distribution μ over $\{0,1\}^\infty$, where $\Pr[F(1) = a_1, F(2) = a_2, \ldots, F(n) = a_n] = 2^{-n}[\text{Each } a_i \in \{0,1\}]$. Thus $s = \lceil -\log \Pr[F(\mathfrak{S}]] = \lceil -\log \mu(C) \rceil$. Using Corollary 5,

$$\min_{G \in \mathcal{G}, G(\mathfrak{S})} \mathbf{K}(G) < s + \mathbf{I}(\langle \mathfrak{S} \rangle; \mathcal{H}) + O(\mathbf{K}(s)) + O(\mathbf{I}(\langle \mathfrak{S} \rangle; \mathcal{H}))$$

$$\min_{G \in \mathcal{G}, G(\mathfrak{S})} \mathbf{K}(G) < s + \mathbf{I}(\langle C \rangle; \mathcal{H}) + O(\mathbf{K}(s)) + O(\mathbf{I}(\langle C \rangle; \mathcal{H})). \tag{8}$$

$$\min_{\alpha \in C} \mathbf{K}(\alpha) < s + \mathbf{I}(\langle C \rangle; \mathcal{H}) + O(\mathbf{K}(s, \mathbf{I}(\langle C \rangle; \mathcal{H}))). \tag{9}$$

Equation 8 is due to Lemma 3 and Proposition 2. Equation 9 comes from modifying G to having it output 0 whenever it would normally output a number $b \notin \{0,1\}$. This new function α can be thought of an infinite sequence in $\{0,1\}^\infty$, and since $G(\mathfrak{S})$, it must be that $\alpha \in C$. Furthermore $\mathbf{K}(\alpha | G) = O(1)$. □

Example 1 *Let clopen set $C \subset \{0,1\}^\infty$ be defined by $\bigcup\{\Gamma_x : \|x\| = n, \mathbf{K}(x) > n - c\}$, for some small $c \in \mathbb{N}$. Thus $\lceil -\log \mu(C) \rceil = O(1)$ and $\min_{\alpha \in C} \mathbf{K}(\alpha) >^+ n$ because if α is in C, then $\alpha[0..n]$ is a random string. Furthermore $\mathbf{I}(\langle C \rangle; \mathcal{H}) >^+ n - \mathbf{K}(n)$ because for the first interval Γ_x encoded in $\langle C \rangle$, $\mathbf{K}(\langle C \rangle) >^+ \mathbf{K}(\langle \Gamma_x \rangle) >^+ n$, and $\mathbf{K}(\langle C \rangle | \mathcal{H}) <^+ \mathbf{K}(n)$.*

Theorem 8 can be generalized to arbitrary open sets of the Cantor space. Such sets S can have encodings $\langle S \rangle$ that are infinite sequences. The Big Oh term O and the $<^+$ are dependent solely on the choice of the universal Turing machine.

Theorem 9 *For open set $S \subseteq \{0,1\}^\infty$, if $s = \lceil -\log \mu(S) \rceil$ and $h = \mathbf{I}(\langle S \rangle : \mathcal{H})$, then $\min_{\alpha \in S} \mathbf{K}(\alpha) < s + h + O(\mathbf{K}(s,h))$.*

Proof. Let $\{x_i\}_{i=1}^{n} = \{x : \Gamma_x \text{ is maximal in } S\}$, with $n \in \mathbb{N} \cup \infty$. Let $N \in \mathbb{N}$ be the smallest number such that $\sum_{i=1}^{N} 2^{-\|x_i\|} > 2^{-s-1}$. Let $C = \bigcup_{i=1}^{N} \Gamma_{x_i}$ be a clopen set with $C \subseteq S$. By Theorem 8,

$$\min_{\alpha \in C} \mathbf{K}(\alpha) < s + \mathbf{I}(\langle C \rangle; \mathcal{H}) + O(\mathbf{K}(s)) + O(\mathbf{K}(\mathbf{I}(\langle C \rangle; \mathcal{H}))). \tag{10}$$

Based on the definition of $\mathbf{I}$:

$$\begin{aligned} \mathbf{I}(\langle C \rangle; \mathcal{H}) &<^{+} \mathbf{I}(\langle S \rangle : \mathcal{H}) + \mathbf{K}(\langle C \rangle | \langle S \rangle) \\ &<^{+} \mathbf{I}(\langle S \rangle : \mathcal{H}) + \mathbf{K}(s). \end{aligned}$$

By Proposition 2, where $x = \mathbf{I}(\langle C \rangle; \mathcal{H})$, $y = \mathbf{I}(\langle S \rangle : \mathcal{H})$, and $c = \mathbf{K}(s) + O(1)$,

$$\mathbf{I}(\langle C \rangle; \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(\langle C \rangle; \mathcal{H}))) < \mathbf{I}(\langle S \rangle : \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(\langle S \rangle : \mathcal{H}))) + O(\mathbf{K}(s)). \tag{11}$$

Putting Equations 10 and 11 together results in

$$\min_{\alpha \in S} \mathbf{K}(\alpha) < s + \mathbf{I}(\langle S \rangle : \mathcal{H}) + O(\mathbf{K}(s, \mathbf{I}(\langle S \rangle : \mathcal{H}))).$$

7 Algorithmic Monotone Probability of Sets

In [Lev16, Eps19], the combined algorithmic probability $\sum_{x \in D} \mathbf{m}(x)$ of a non-exotic set D was shown to be close to $\max_{x \in D} \mathbf{m}(x)$. In this section, we prove an analogous theorem with the universal lower-computable continuous semi-measure $\mathbf{M}$.

A continuous semi-measure Q is a function $Q : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, such that $Q(\emptyset) = 1$ and for all $x \in \{0,1\}^*$, $Q(x) \geq Q(x0) + Q(x1)$. For prefix free set D, $Q(D) = \sum_{x \in D} Q(x)$. Let $\mathbf{M}$ be a largest, up to a multiplicative factor, lower semi-computable continuous semi-measure. That is, for all lower computable continuous semi-measures Q there is a constant $c \in \mathbb{N}$ where for all $x \in \{0,1\}^*$, $c\mathbf{M}(x) > Q(x)$. The monotone complexity of a finite prefix-free set G of finite strings is $\mathbf{Km}(G) \stackrel{\text{def}}{=} \min\{\|p\| : U(p) \in x \sqsupseteq y \in G\}$. Note that this differs from the usual definition of $\mathbf{Km}$, in that our definition requires U to halt.

A string-monotonic program is a total recursive Turing machine with an input tape, a work tape, and an output tape, where the tape heads of input tape and the output tape can only move in one direction. A total computable function $\nu : \{0,1\}^* \to \{0,1\}^*$ is string-monotonic iff for all strings x and y, $\nu(x) \sqsubseteq \nu(xy)$. Let $\overline{\nu} : \{0,1\}^{*\infty} \to \{0,1\}^{*\infty}$ be used to represent the unique extension of ν to infinite sequences. Its definition for all $\alpha \in \{0,1\}^{*\infty}$ is $\overline{\nu}(\alpha) = \sup\{\nu(\alpha_{\leq n}) : n \leq \|\alpha\|\}$, where the supremum is respect to the partial order derived with the $\sqsubseteq$ relation. The following theorem relates prefix monotone machines and continuous semi-measures. It is equivalent to Theorem 4.5.2 in [LV08], with the simple modification that the machine be total computable.

Theorem 10 *For each lower-computable continuous semi-measure σ over $\{0,1\}^{\infty}$, there is a string-monotonic function ν_σ, where for prefix free $G \subset \{0,1\}^*$, $\lceil -\log \sigma(G) \rceil =^{+} \lceil -\log \mu\{\alpha : \overline{\nu_\sigma}(\alpha) \sqsupseteq x \in G\} \rceil$.*

Since there is a universal lower-semicomputable continuous semi-measure $\mathbf{M}$, there exists a string-monotonic function $\nu_{\mathbf{M}}$, with the following property.

Corollary 6 *For finite prefix free set G, $-\log \mathbf{M}(G) =^{+} -\log \mu\{\alpha : x \sqsubseteq \overline{\nu_{\mathbf{M}}}(\alpha), \alpha \in \{0,1\}^{\infty}, x \in G\}$.*

The following corollary is equivalent to Theorem 8 in terms of finite strings instead of clopen sets. For finite prefix free set $G \subset \{0,1\}^*$, $\mu(G) = \sum_{x \in G} 2^{-\|x\|}$.

Corollary 7 *For finite prefix free $G \subset \{0,1\}^*$, $\min_{x \sqsupseteq y \in G} \mathbf{K}(x) <^{\log} -\log \mu(G) + \mathbf{I}(G;\mathcal{H})$.*

Lemma 3 *For partial computable $f : \mathbb{N} \to \mathbb{N}$, for all $a \in \mathbb{N}$, $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a:\mathcal{H}) + \mathbf{K}(f)$.*

Proof.

$$\mathbf{I}(a;\mathcal{H}) = \mathbf{K}(a) - \mathbf{K}(a|\mathcal{H}) >^+ \mathbf{K}(a, f(a)) - \mathbf{K}(a, f(a)|\mathcal{H}) - \mathbf{K}(f).$$

The chain rule applied twice results in

$$\begin{aligned}
\mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f) &>^+ \mathbf{K}(f(a)) + \mathbf{K}(a|f(a), \mathbf{K}(f(a))) - (\mathbf{K}(f(a)|\mathcal{H}) + \mathbf{K}(a|f(a), \mathbf{K}(f(a)|\mathcal{H}), \mathcal{H}) \\
&=^+ \mathbf{I}(f(a);\mathcal{H}) + \mathbf{K}(a|f(a), \mathbf{K}(f(a))) - \mathbf{K}(a|f(a), \mathbf{K}(f(a)|\mathcal{H}), \mathcal{H}) \\
&=^+ \mathbf{I}(f(a);\mathcal{H}) + \mathbf{K}(a|f(a), \mathbf{K}(f(a))) - \mathbf{K}(a|f(a), \mathbf{K}(f(a)), \mathbf{K}(f(a)|\mathcal{H}), \mathcal{H}) \\
&>^+ \mathbf{I}(f(a);\mathcal{H}).
\end{aligned}$$

□

Theorem 11 *For finite prefix-free set G, $\mathbf{Km}(G) <^{\log} -\log \mathbf{M}(G) + \mathbf{I}(G;\mathcal{H})$.*

Proof. Let $i = \lceil -\log \mathbf{M}(G) \rceil$. Let $F \subset \{0,1\}^*$ be finite prefix-free set, such that

1. $-\log \mu(F) \leq i + 1$
2. for all $x \in F$, $\nu_{\mathbf{M}}(x) \sqsupseteq z \in G$,
3. $\mathbf{K}(F|G) <^+ \mathbf{K}(i)$.

By Corollary 7, there exists $y \sqsupseteq x \in F$, with $\mathbf{K}(y) <^{\log} i + \mathbf{I}(F;\mathcal{H})$. Using Lemma 3, $\mathbf{K}(y) <^{\log} i + \mathbf{I}(G;\mathcal{H})$. Thus there is a program p of length $<^+ \mathbf{K}(y)$ that computes y and then outputs $\nu_{\mathbf{M}}(y) \sqsupseteq \nu_{\mathbf{M}}(x) \sqsupseteq z \in G$. So $\mathbf{Km}(G) \leq \|p\| <^+ \mathbf{K}(y) <^{\log} i + \mathbf{I}(G;\mathcal{H})$. □

Corollary 8 *For (potentially infinite) prefix-free set G, $\mathbf{Km}(G) <^{\log} -\log \mathbf{M}(G) + \mathbf{I}(\langle G \rangle : \mathcal{H})$.*

The proof of this corollary follows analogously to the proof of Theorem 9, except $\mathbf{M}$ is used instead of μ.

8 Closed Sets

There is no equivalent to Theorem 9 for closed sets. For closed sets $S \subseteq \{0,1\}^\infty$ of infinite strings $S_{\leq n} = \{\alpha[0..n] : \alpha \in S\}$ and $\langle S \rangle = \langle S_{\leq 1} \rangle \langle S_{\leq 2} \rangle \langle S_{\leq 3} \rangle \ldots$ The closed set theorem uses the following proposition of conservation of information with respect to a partial computable function. The complexity of a partial computable function f, is $\mathbf{K}(f)$, the minimal length of a U-program to compute f. A short proof can be found in [Gei12].

Proposition 3 *For $\alpha, \beta \in \{0,1\}^\infty$, partial computable $f : \{0,1\}^\infty \to \{0,1\}^\infty$, $\mathbf{I}(f(\alpha) : \beta) < \mathbf{I}(\alpha : \beta) + \mathbf{K}(f)$.*

Theorem 12 (Anonymous) *There exists a closed set $C \subset \{0,1\}^\infty$ consisting of solely uncomputable sequences, $\mu(C) > 0$, and $\mathbf{I}(\langle C \rangle : \mathcal{H}) < \infty$.*

Proof. Let d be any positive constant and $\alpha \in \{0,1\}^\infty$ be any uncomputable sequence such that $\mathbf{I}(\alpha : \mathcal{H}) < \infty$. We inductively define a total computable function f such that $f(\alpha) = \langle C \rangle$ for some closed set C. At round 0, $f(a)$ outputs $\langle C_{\leq 0} \rangle$. Assume $f(\alpha)$ has outputted $\langle C_{\leq 1} \rangle \langle C_{\leq 2} \rangle \dots \langle C_{\leq n-1} \rangle$.

Let $\mathbf{K}_t(x) = \min\{\|p\| : U(p) = x \text{ in } \leq t \text{ steps}\}$. Let S consist of the set $x \sqsubseteq y \in C_{\leq n-1}$ such that $\|x\| - \mathbf{K}_n(x) > d$. $C_{\leq n}$ is constructed in the following way. For each $x \in C_{\leq n-1}$, if there is a $y \in S$, with $y \sqsubseteq x$, then $x(\alpha[\|x\| - \|y\| + 1])$ is added to $C_{\leq n}$. Otherwise $x0$ and $x1$ is added to $C_{\leq n}$. The function f then appends $\langle C_{\leq n} \rangle$ to the output and proceeds to step $n+1$. The amount of mutual information that C has with $\mathcal{H}$ is $\mathbf{I}(\langle C \rangle : \mathcal{H}) <^+ \mathbf{I}(\alpha : \mathcal{H}) + \mathbf{K}(f) < \infty$. Furthermore $\mu(C) \geq \mu(\{\alpha : \mathbf{D}(\alpha) < d\}) > 0$, where $\mathbf{D}(\alpha) = \sup_{x \sqsubset \alpha} \|x\| - \mathbf{K}(x)$. Every $\alpha \in C$ either has $\mathbf{D}(\alpha) < d$ or is equal to $x\alpha_{>\|x\|}$ for some $x \in \{0,1\}^*$, and is thus uncomputable.

9 Games

Function derandomization has applications to derandomization in the cybernetic agent model, whose connection to Algorithmic Information Theory is studied extensively in [Hut05]. In this section we describe a simplified cybernetic agent model. The agent $\mathbf{p}$ and environment $\mathbf{q}$ are defined as follows. The agent is a function $\mathbf{p} : (\mathbb{N} \times \mathbb{N})^* \to \mathbb{N}$, where if $\mathbf{p}(w) = a$, $w \in (\mathbb{N} \times \mathbb{N})^*$ is a list of the previous actions of the agent and the environment, and $a \in \mathbb{N}$ is the action to be performed. The environment is of the form $\mathbf{q} : (\mathbb{N} \times \mathbb{N})^* \times \mathbb{N} \to \mathbb{N} \cup \{\mathbf{W}\}$, where if $\mathbf{q}(w, a) = b \in \mathbb{N}$, then b is $\mathbf{q}$'s response to the agent's action a, given history w, and the game continues. If $\mathbf{q}$ responds $\mathbf{W}$ then the agents wins and the game halts. The agent can be randomized. The game can continue forever, given certain agents and environments.

Theorem 13 ([Lev16, Eps19]) *For finite $D \subset \{0,1\}^*$, $-\log \max_{x \in D} \mathbf{m}(x) <^{\log} -\log \sum_{x \in D} \mathbf{m}(x) + \mathbf{I}(D; \mathcal{H})$.*

The following theorem is a game-theoretic interpretation of Lemma 6 in [VV10].

Theorem 14 *If 2^r deterministic agents of complexity $< k$ win against environment $\mathbf{q}$, then there is a deterministic agent $\mathbf{p}$ of complexity $<^{\log} k - r + \mathbf{I}(\langle r, k, \mathbf{q} \rangle; \mathcal{H})$ that wins against $\mathbf{q}$.*

Proof. Given $\langle r, k, \mathbf{q} \rangle$, one can construct a finite set D of encoded agents that win against $\mathbf{q}$ and D contains at least 2^r agents of complexity $< k$. Furthermore $\sum_{x \in D} \mathbf{m}(x) \stackrel{*}{>} 2^r 2^{-k}$, so using Theorem 13, there is an agent $\mathbf{p} \in D$, where, using Lemma 3, $\mathbf{K}(\mathbf{p}) <^{\log} -\log k - r + \mathbf{I}(D; \mathcal{H}) <^{\log} -\log k - r + \mathbf{I}(\langle r, k, \mathbf{q} \rangle; \mathcal{H})$.

Theorem 15 *If probabilistic agent $\mathbf{p}'$ wins against environment $\mathbf{q}$ with probability p, then there is a deterministic agent $\mathbf{p}$ of complexity $<^{\log} \mathbf{K}(\mathbf{p}') - \log p + \mathbf{I}(\langle p, \mathbf{p}', \mathbf{q} \rangle; \mathcal{H})$ that wins against $\mathbf{q}$.*

Proof. Let $\mathcal{I}$ be a set of interactions between an arbitrary agent and the environment $\mathbf{q}$ such that each interaction ends in $\mathbf{W}$ and with probability $> p/2$, $\mathbf{p}'$ will act according to an interaction in $\mathcal{I}$. Thus $\mathbf{K}(\mathcal{I} | p, \mathbf{p}', \mathbf{q}) = O(1)$. $\mathbf{p}'$ can be encoded into a random function F, where the domain $(\mathbb{N} \times \mathbb{N})^*$ of $\mathbf{p}'$ can be encoded into a single number $\mathbb{N}$. $\mathbf{K}(F | \mathbf{p}') = O(1)$. Similarly, $\mathcal{I}$ can be encoded into a set of samples $\mathfrak{C}$, where $\Pr[F(\mathfrak{C})] > p/2$ and $\mathbf{K}(\langle \mathfrak{C} \rangle | \langle \mathcal{I} \rangle) = O(1)$. Using Corollary 5,

there is a deterministic function $G : \mathbb{N} \to \mathbb{N}$, such that

$$\begin{aligned}
\mathbf{K}(G) &<^{\log} \mathbf{K}(F) - \log[F(\mathfrak{C}] + \mathbf{I}(\langle \mathfrak{C} \rangle; \mathcal{H}) \\
&<^{\log} \mathbf{K}(\mathbf{p}') - \log[F(\mathfrak{C}] + \mathbf{I}(\langle \mathfrak{C} \rangle; \mathcal{H}) \\
&<^{\log} \mathbf{K}(\mathbf{p}') - \log p + \mathbf{I}(\langle \mathfrak{C} \rangle; \mathcal{H}) \\
&<^{\log} \mathbf{K}(\mathbf{p}') - \log p + \mathbf{I}(\langle \mathcal{I} \rangle; \mathcal{H}) \qquad (12) \\
&<^{\log} \mathbf{K}(\mathbf{p}') - \log p + \mathbf{I}(\langle p, \mathbf{p}', \mathbf{q} \rangle; \mathcal{H}), \qquad (13)
\end{aligned}$$

Where Equations 12 and 13 are due to Lemma 3. The deterministic function G is an encoding of an agent, $\mathbf{p}$, proving the theorem. □

Imagine a modification to the game such that the environment gives a nonnegative rational penalty term to the agent at each round. Furthermore the environment specifies an end to the game without specifying a winner or loser.

Corollary 9 *If given probabilistic agent* $\mathbf{p}$*, environment* $\mathbf{q}$ *halts with probability 1, and* $\mathbf{p}$ *has expected penalty less than* $n \in \mathbb{N}$*, then there is a deterministic agent of complexity* $<^{\log} \mathbf{K}(\mathbf{p}) + \mathbf{I}(\langle \mathbf{p}, n, \mathbf{q} \rangle; \mathcal{H})$ *that receives penalty* $< 2n$ *against* $\mathbf{q}$*.*

Proof. We create a win/no-halt game from $\mathbf{q}$ where an agent wins if it gets a penalty less than $2n$. Thus $\mathbf{p}$ is a probabilistic agent that wins this new game with probability $> .5$. Theorem 15 then can be used to prove the corollary.

Example 2 *An example penalty game is as follows. The environment* $\mathbf{q}$ *plays a game for* N *rounds, for some very large* $N \in \mathbb{N}$*, with each round starting with an action by* $\mathbf{q}$*. At round* i*, the environment gives a program to compute a probability* P_i *over* $\mathbb{N}$*. The agent responds with a number* $a_i \in \mathbb{N}$*. The environment gives the agent a penalty of size* $T_i(a_i)$*, where* $T_i : \mathbb{N} \to \mathbb{Q}_{\geq 0}$ *is a computable test, with* $\sum_{a \in \mathbb{N}} P_i(a) T_i(a) < 1$*. After* N *rounds,* $\mathbf{q}$ *halts.*

A very successful probabilistic agent $\mathbf{p}$ *can be defined. Its algorithm is simple. On receipt of a program to compute* P_i*, the agent randomly samples a number* $\mathbb{N}$ *according to* P_i*. At each round the expected penalty is* $\sum_{a_i} P_i(a_i) T_i(a_i) < 1$*, so the expected penalty of* $\mathbf{p}$ *for the entire game is* $< N$*. Thus by Corollary 9, there is a deterministic agent* $\mathbf{p}'$ *such that*

1. $\mathbf{p}'$ *receives a penalty of* $< 2N$*,*
2. $\mathbf{K}(\mathbf{p}') <^{\log} \mathbf{I}(\mathbf{q}; \mathcal{H})$*.*

Let $\mathbf{q}$ *be defined so that* $P_i(a) = [a \leq 2^i] 2^{-i}$ *and* $T_i = [a \leq 2^i] 2^{i - \mathbf{K}(a|i)}$*. Thus each* T_i *is a randomness deficiency function. The probabilistic algorithm* $\mathbf{p}$ *will receive an expected penalty* $< N$*. However any deterministic agent* $\mathbf{p}'$ *that receives a penalty* $< 2N$ *must be very complex, as it must select many numbers with low randomness deficiency. Thus, by the bounds above,* $\mathbf{I}(\mathbf{q}; \mathcal{H})$ *must be very high. This makes sense because* $\mathbf{q}$ *encodes* N *randomness deficiency functions.*

10 Discussion

In the proof of Theorem 3, a relativization technique can be used to convert an $O(\mathbf{K}(x))$ error term to a $\mathbf{K}(x)$ error term, which allows the removal of quantifiers from the theorem statement. This technique can be performed by first relativizing inequalities to a shortest program that computes

all the relevant parameters μ, λ, and n. Then the next part is to reconfigure all terms that have the parameters as conditional information, in this case the deficiency of randomness $\mathbf{D}(\alpha|\mu)$. This technique was also used in [Eps22a, Eps22b].

References

[Eps19] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

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

[Eps22a] Samuel Epstein. A note on the outliers theorem. *CoRR*, abs/2203.08733, 2022.

[Eps22b] Samuel Epstein. On the kolmogorov complexity of binary classifiers. *CoRR*, abs/2201.12374, 2022.

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

[Hut05] Ml Hutter. *Universal Artificial Intelligence*. Texts in Theoretical Computer Science. An EATCS Series. Springer, Berlin and Heidelberg, 2005.

[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.

[Lev16] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016. And also: S. Epstein and L.A. Levin, Sets have simple members, arXiv preprint arXiv:1107.1458, 2011.

[LV08] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Publishing Company, Incorporated, 3 edition, 2008.

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

[VS17] Nikolay K. Vereshchagin and Alexander Shen. Algorithmic statistics: Forty years later. In *Computability and Complexity*, pages 669–737, 2017.

[VV10] N. Vereshchagin and P. Vitányi. Rate Distortion and Denoising of Individual Data using Kolmogorov Complexity. *IEEE Transactions on Information Theory*, 56, 2010.