

Randomness Deficiency Overlap

Samuel Epstein*

June 21, 2023

Abstract

In this paper we prove a lower bound on the computable measure of sets with high randomness deficiency with respect to two computable measures. We extend this result to computable metric spaces and universal uniform tests.

1 Introduction

In a previous paper, a lower bound was proved on the computable measure of sets with high randomness deficiency. The \mathbf{D} term is the infinite randomness deficiency function and \mathbf{K} is the prefix Kolmogorov complexity.

Theorem. [Eps22] *For computable measures μ and nonatomic λ over $\{0,1\}^\infty$ and $n \in \mathbb{N}$, $\lambda\{\alpha : \mathbf{D}(\alpha|\mu) > n\} > 2^{-n-\mathbf{K}(n,\mu,\lambda)-O(1)}$.*

This paper generalizes this result, in the context of overlap between the randomness deficiency function with respect to two different computable probability measures.

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

The $O(1)$ term is dependent solely on the choice of the universal Turing machine. With a bit of work, the theorem can be proved for potentially uncomputable λ that has finite mutual information with the halting sequence. It is possible to see this theorem being referenced in proofs of more sophisticated theorems. This theorem is of note because it factors out the mutual information with the halting sequence term that is so prevalent in the resultant theorems from similar proofs to the ones found in the paper. Lemma 3 is a reworking of Lemma 2 in [Eps23a], Lemma 5 is a reworking of Lemma 4 in [Eps23b], and Theorem 2 is a reworking of Theorem 3 in [Eps22]. The tight bounds of the main theorem derived from lemmas with looser bounds is achieved through relativization.

1.1 Universal Uniform Tests

The study of randomness of computable metric spaces can be seen in the works of [HR09, G21]. These spaces are important because physical random phenomena are modeled using infinite objects, and not the Cantor space. For definitions in this introduction, we use [HR09]. A computable metric space \mathcal{X} is a metric space with a dense set of ideal points on which the distance function is computable. A computable probability is defined by a computable sequence of converging points in the corresponding space of Borel probability measures, $\mathcal{M}(\mathcal{X})$, over \mathcal{X} . A uniform test T takes in a description of a probability measure μ and produces a lower computable μ test, with $\int_{\mathcal{X}} T^\mu d\mu \leq 1$.

*JP Theory Group. samepst@jpththeorygroup.org

There exists a universal test, \mathbf{t} , where for any uniform test T there is a $c_T \in \mathbb{N}$ where $c_T \mathbf{t} > T$. We extend the main theorem of this paper to computable metric spaces and universal uniform tests. Note that unlike the infinite sequence case, the constant depends on the computable metric space and computable probability measures. There doesn't appear to be a straightforward way to achieve the same bounds as in the sequences case.

Theorem. *Given computable probability measures μ , ρ , and λ , non-atomic λ , over a computable metric space \mathcal{X} and universal uniform test \mathbf{t} , there is $c \in \mathbb{N}$ where for all n , $\lambda(\{\alpha : \mathbf{t}_\mu(\alpha) > 2^n \text{ and } \mathbf{t}_\rho(\alpha) > 2^n\}) > 2^{-n-\mathbf{K}(n)-c}$.*

2 Conventions

The function $\mathbf{K}(x|y)$ is the conditional prefix Kolmogorov complexity. $\mathbf{m}(x|y)$ is the conditional algorithmic probability. The mutual information between two strings $x, y \in \{0, 1\}^*$, is $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$. For probability p over \mathbb{N} , randomness deficiency is $\mathbf{d}(a|p, b) = \lfloor -\log p(a) \rfloor - \mathbf{K}(a|p, b)$ and measures the extent of the refutation against the hypothesis p given the result a [G21]. $\mathbf{d}(a|p) = \mathbf{d}(a|p, \emptyset)$. The amount of information that the halting sequence $\mathcal{H} \in \{0, 1\}^\infty$ has about $a \in \{0, 1\}^*$, conditional to $y \in \{0, 1\}^*$ is $\mathbf{I}(a; \mathcal{H}|y) = \mathbf{K}(a|y) - \mathbf{K}(a|y, \mathcal{H})$. $\mathbf{I}(a; \mathcal{H}) = \mathbf{I}(a; \mathcal{H}|\emptyset)$. We use $<^+ f$ to denote $< f + O(1)$ and $<^{\log} f$ to denote $< f + O(\log(f+1))$. Stochasticity is $\mathbf{Ks}(a|b) = \min\{\mathbf{K}(Q|b) + 3 \log \max\{\mathbf{d}(a|Q, b), 1\} : Q \text{ has finite support and a range in } \mathbb{Q}\}$. $\mathbf{Ks}(a|b) < \mathbf{Ks}(a) + O(\log \mathbf{K}(b))$. For a mathematical statement A , let $[A] = 1$ if A is true and $[A] = 0$, otherwise. The chain rule gives $\mathbf{K}(x, y) =^+ \mathbf{K}(x|y, \mathbf{K}(y)) + \mathbf{K}(y)$. The randomness deficiency of $\alpha \in \{0, 1\}^\infty$ with respect to computable continuous probability measure P is $\mathbf{D}(\alpha|P) = \sup_n -\log P(\alpha[0..n]) - \mathbf{K}(\alpha[0..n]|\langle P \rangle)$. The term $\langle P \rangle$ is a program to compute P . The following definition is from [Lev74].

Definition 1 (Information) *For infinite sequences $\alpha, \beta \in \{0, 1\}^\infty$, their mutual information is defined to be $\mathbf{I}(\alpha : \beta) = \log \sum_{x, y \in \{0, 1\}^*} 2^{\mathbf{I}(x:y) - \mathbf{K}(x|\alpha) - \mathbf{K}(y|\beta)}$.*

There are many proofs in the literature that non-stochastic numbers have high mutual information with the halting sequence. One such detailed proof is in [Eps21].

Lemma 1 $\mathbf{Ks}(x) < \mathbf{I}(x; \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(x; \mathcal{H})))$.

Lemma 2 ([Eps22]) *For partial computable f , $\mathbf{I}(f(x) : \mathcal{H}) <^+ \mathbf{I}(x; \mathcal{H}) + \mathbf{K}(f)$.*

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

3 On Exotic Sets of Natural Numbers

Lemma 3 *For computable probabilities p, q over \mathbb{N} , $D \subset \mathbb{N}$, $|D| = 2^s$, $s < \max_{a \in D} \min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{I}(D; \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(D; \mathcal{H}), p, q, s))$.*

Proof. We relativize the universal Turing machine to $\langle s, p, q \rangle$. Let Q be a probability measure that realizes $\mathbf{Ks}(D)$, with $d = \max\{\mathbf{d}(D|Q), 1\}$. Let $F \subseteq \mathbb{N}$ be a random set where each element $a \in \mathbb{N}$ is selected independently with probability $cd2^{-s}$, where $c \in \mathbb{N}$ is chosen later. $\mathbf{E}[p(F)] = \mathbf{E}[q(F)] \leq cd2^{-s}$. Furthermore

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

Thus finite $W \subset \mathbb{N}$ can be chosen such that $p(W) \leq 4cd2^{-s}$, $q(W) \leq 4cd2^{-s}$, and $Q(\{G : |G| = 2^s, G \cap W = \emptyset\}) \leq e^{2^{-cd}}$. $D \cap W \neq \emptyset$, otherwise, using the Q -test, $t(G) = e^{cd-1}$ if $(|G| = 2^s, G \cap W = \emptyset)$ and $t(G) = 0$ otherwise, 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 . Thus there is an $a \in D \cap W$, where

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

Making the relativization of $\langle s, p, q \rangle$ explicit, and using Lemma 1 results in

$$\begin{aligned} s &<^+ \min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{K}s(D) + O(\mathbf{K}(s, p, q)) \\ s &< \max_{a \in D} \min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{K}s(D) + O(\mathbf{K}(s, p, q)) \\ s &< \max_{a \in D} \min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{I}(D; \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(D; \mathcal{H}), s, p, q)). \square \end{aligned}$$

4 On Exotic Sets of Reals

Let $\Omega = \sum\{2^{-\|p\|} : U(p) \text{ halts}\}$ be Chaitin's Omega, $\Omega_n \in \mathbb{Q}_{\geq 0}$ be the rational formed from the first n bits of Ω , and $\Omega^t = \sum\{2^{-\|p\|} : U(p) \text{ halts in time } t\}$. For $n \in \mathbb{N}$, let $\mathbf{bb}(n) = \min\{t : \Omega_n < \Omega^t\}$. $\mathbf{bb}^{-1}(m) = \arg \min_n \{\mathbf{bb}(n-1) < m \leq \mathbf{bb}(n)\}$. Let $\Omega[n] \in \{0, 1\}^*$ be the first n bits of Ω .

Lemma 4 For $n = \mathbf{bb}^{-1}(m)$, $\mathbf{K}(\Omega[n]|m, n) = O(1)$.

Proof. For a string x , let $BB(x) = \inf\{t : \Omega^t > 0.x\}$. Enumerate strings of length n , starting with 0^n , and return the first string x such that $BB(x) \geq m$. This string x is equal to $\Omega[n]$, otherwise let y be the largest common prefix of x and $\Omega[n]$. Thus $BB(y) = \mathbf{bb}(\|y\|) \geq BB(x) \geq m$, which means $\mathbf{bb}^{-1}(m) \leq \|y\| < n$, causing a contradiction. \square The following lemma, while lengthy, is a series of straightforward application of inequalities.

Lemma 5 For computable probabilities P, Q , over $\{0, 1\}^\infty$, $Z \subset \{0, 1\}^\infty$, $|Z| = 2^s$, $s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(\langle Z \rangle; \mathcal{H}) + O(\mathbf{K}(s, P, Q) + \log \mathbf{I}(\langle Z \rangle; \mathcal{H}))$.

Proof. We relativize the universal Turing machine to s , which can be done due to the precision of the theorem. Let $Z_n = \{\alpha[0..n] : \alpha \in Z\}$ and $m = \arg \min_m |Z_m| = |Z|$. Let $n = \mathbf{bb}^{-1}(m)$ and $k = \mathbf{bb}(n)$. Let p and q be probabilities over $\{0, 1\}^*$, where $p(x) = [\|x\| = k]P(x)$ and $\langle p \rangle = \langle k, P \rangle$ and let $q(x) = [\|x\| = k]Q(x)$ and $\langle q \rangle = \langle k, P \rangle$. Using $D = Z_k$, Lemma 3, relativized to k , produces $x \in Z_k$, where

$$\begin{aligned} s &< \min\{\mathbf{d}(x|p), \mathbf{d}(x|q)\} + \mathbf{I}(Z_k; \mathcal{H}|k) + O(\mathbf{K}(\mathbf{I}(Z_k; \mathcal{H}|k), q, p|k)) \\ &< \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{K}(Z_k|k) + \mathbf{K}(k) - \mathbf{K}(Z_k|k, \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(Z_k; \mathcal{H}|k), q, p|k)). \\ &< \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{K}(Z_k|k) + \mathbf{K}(k) - \mathbf{K}(Z_k|k, \mathcal{H}) + O(\mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H}|k)). \end{aligned}$$

Since $\mathbf{K}(k) <^+ n + \mathbf{K}(n)$, by the chain rule,

$$\begin{aligned} & \mathbf{K}(Z_k|k) + \mathbf{K}(k) \\ & <^+ \mathbf{K}(Z_k|k, \mathbf{K}(k)) + \mathbf{K}(\mathbf{K}(k)|k) + \mathbf{K}(k) \\ & < \mathbf{K}(Z_k, k) + O(\log n) \\ & < \mathbf{K}(Z_k) + O(\log n). \end{aligned}$$

So

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{K}(Z_k) - \mathbf{K}(Z_k|k, \mathcal{H}) + O(\log n + \mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H}|k)).$$

Since $\mathbf{K}(k|n, \mathcal{H}) = O(1)$, $\mathbf{K}(Z_k|\mathcal{H}) <^+ \mathbf{K}(Z_k|k, \mathcal{H}) + \mathbf{K}(n)$,

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k; \mathcal{H}) + O(\log n + \mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H}|k)).$$

Furthermore since $\mathbf{I}(Z_k; \mathcal{H}|k) + \mathbf{K}(k) < \mathbf{I}(Z_k; \mathcal{H}) + O(\log n)$,

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k; \mathcal{H}) + O(\log n + \mathbf{K}(P, Q)) + O(\log \mathbf{I}(Z_k; \mathcal{H})).$$

By Lemma 4, $\mathbf{K}(\Omega[n]|Z_k) <^+ \mathbf{K}(n)$ so by Lemma by 2,

$$n <^{\log} \mathbf{I}(\Omega[n]; \mathcal{H}) <^{\log} \mathbf{I}(Z_k; \mathcal{H}) + \mathbf{K}(n) <^{\log} \mathbf{I}(Z_k; \mathcal{H}).$$

The above equation used the common fact that the first n bits of Ω has $n - O(\log n)$ bits of mutual information with \mathcal{H} . So

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k; \mathcal{H}) + O(\mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H})).$$

By the definition of mutual information \mathbf{I} between infinite sequences

$$\mathbf{I}(Z_k; \mathcal{H}) <^+ \mathbf{I}(Z : \mathcal{H}) + \mathbf{K}(Z_k|Z) <^{\log} \mathbf{I}(Z : \mathcal{H}) + \mathbf{K}(k|Z).$$

Now m is simple relative to Z and by Lemma 4, $\Omega[n]$ is simple relative to m and n . Furthermore k is simple relative to $\Omega[n]$. Therefore $\mathbf{K}(Z_k|Z) <^+ \mathbf{K}(n)$. So

$$\begin{aligned} s & < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z : \mathcal{H}) + O(\log n) + O(\mathbf{K}(P, Q) + \log \mathbf{I}(Z; \mathcal{H})) \\ s & < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z : \mathcal{H}) + O(\mathbf{K}(s, P, Q) + \log \mathbf{I}(Z; \mathcal{H})). \end{aligned}$$

□

5 Asymptotic Properties of Randomness Deficiency

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

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

$$\begin{aligned} \prod_{\beta \in D} \lambda\{\mathbf{D}(\beta|\mu) \leq n \text{ or } \mathbf{D}(\beta|\rho) \leq n\} &\geq \\ (1 - |D|2^{-n-\mathbf{K}(n,\mu,\lambda,\rho)-c}) &\geq \\ (1 - 2^{n+\mathbf{K}(n,\mu,\lambda,\rho)+c-1}2^{-n-\mathbf{K}(n,\mu,\lambda,\rho)-c}) &\geq 1/2. \end{aligned}$$

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

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

Let v be a shortest program to compute $\langle n, \mu, \rho, \lambda \rangle$. By Theorem 1, with the universal Turing machine relativized to v ,

$$\lambda^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|v) > m\}) \stackrel{*}{<} 2^{-m+\mathbf{K}(\lambda^{n,c}|v)} \stackrel{*}{<} 2^{-m+\mathbf{K}(n,\mathbf{K}(n,\mu,\lambda,\rho),c,\lambda|v)} \stackrel{*}{<} 2^{-m+\mathbf{K}(c)}.$$

Therefore,

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

Thus, by probabilistic arguments, there exists $\alpha \in \{0,1\}^\infty$, such that $\alpha = \langle D \rangle$ is an encoding of $2^{n+\mathbf{K}(n,\mu,\rho,\lambda)+c-1}$ elements $\beta \in D \subset \{0,1\}^\infty$, where each β has $\mathbf{D}(\beta|\mu) \leq n$ or $\mathbf{D}(\beta|\rho) \leq n$ and $\mathbf{I}(\alpha : \mathcal{H}|v) <^+ \mathbf{K}(c)$. By Lemma 5, relativized to v , there are constants $d, f, g \in \mathbb{N}$ where

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

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

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

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

6 Computable Probability Spaces

The first main result of this paper uses computable metric spaces and computable probability measures from [HR09]. Some constructs need changes, which we present in the later section. But in this section we show the definitions, lemmas, and theorems that are directly taken from [HR09]. If a theorem or lemma is presented without a proof, then it can be found in [HR09]

Definition 2 *A computable metric space consists of a triple $(\mathcal{X}, \mathcal{S}, d)$, where*

- \mathcal{X} is a separable complete metric space.
- \mathcal{S} is an enumerable list of dense ideal points \mathcal{S} in \mathcal{X} .
- d is a distance metric that is uniformly computable over points in \mathcal{S} .

For $x \in \mathcal{X}$, $r \in \mathbb{Q}_{>0}$ a ball is $B(x, r) = \{y : d(x, y) < r\}$. The ideal points induce a sequence of enumerable ideal balls $B_i = \{B(s_i, r_j) : s_i \in \mathcal{S}, r_j \in \mathbb{Q}_{>0}\}$. A sequence of ideal points $\{x_n\} \subseteq X$ is said to be a fast Cauchy sequence if $d(x_n, x_{n+1}) < 2^{-n}$ for all $n \in \mathbb{N}$. A point x is computable there is a computable fast Cauchy sequence converging to x . Each computable function f between computable metric spaces \mathcal{X} and \mathcal{Y} has an algorithm \mathfrak{A} such that if $f(x) = y$ then for all fast Cauchy sequences \vec{x} for x , $\mathfrak{A}(\vec{x})$ outputs an encoding of a fast Cauchy sequence for y .

Definition 3 *Lower computable functions $f \in \mathcal{F}$ have algorithms that enumerate $\{(B_i, r_i)\}$, where B_i is an ideal ball and $r_i \in \mathbb{Q}_{>0}$, and $f(x) = \sup\{r_i : x \in B_i\}$.*

The computable metric space of all Borel probability measures over \mathcal{X} is $\mathcal{M}(\mathcal{X})$. If \mathcal{X} is separable and compact then so is $\mathcal{M}(\mathcal{X})$. The ideal points of $\mathcal{M}(\mathcal{X})$ are \mathcal{D} , the set of probability measures that are concentrated on finitely many points with rational values. The distance metric on $\mathcal{M}(\mathcal{X})$ is the *Prokhorov metric*, defined as follows.

Definition 4 (Prokhorov metric)

$$\pi(\mu, \nu) = \inf \{ \epsilon \in \mathbb{R}^+ : \mu(A) \leq \nu(A^\epsilon) \text{ for Borel set } A \},$$

where $A^\epsilon = \{x : d(x, A) < \epsilon\}$.

Theorem 3 *Given a probability measure $\mu \in \mathcal{M}(\mathcal{X})$, the following are equivalent.*

1. μ is computable.
2. $\mu(B_{i_1} \cup \dots \cup B_{i_k})$ is lower semi-computable uniformly in $\langle i_1, \dots, i_k \rangle$.
3. $\int d\mu : \mathcal{F} \rightarrow \mathbb{R}_{\geq 0}$ is lower semi-computable.

Definition 5

1. A constructive G_δ -set is a set of the form $\bigcap_n U_n$ where $\{U_n\}$ is a sequence of uniformly r.e. open sets.
2. A computable probability space is a pair (\mathcal{X}, μ) , where \mathcal{X} is a computable metric space and μ is a Borel probability measure on \mathcal{X} .
3. Let (\mathcal{X}, μ) be a computable probability space and \mathcal{Y} a computable metric space. A function $f : D_f \subset (\mathcal{X}, \mu) \rightarrow \mathcal{Y}$ is almost computable if it is computable on a constructive G_δ -set (D_f) of μ -measure one.

4. A morphism of computable probability spaces $Q : (\mathcal{X}, \mu) \rightarrow (\mathcal{Y}, \nu)$ is an almost computable measure-preserving function $Q : D_Q \subset \mathcal{X} \rightarrow \mathcal{Y}$, where $\mu(Q^{-1}(A)) = \nu(A)$ for all Borel sets A . An isomorphism (Q, R) is a pair of morphisms such that $Q \circ R = \text{id}$ on $R^{-1}(D_Q)$ and $R \circ Q = \text{id}$ on $Q^{-1}(D_R)$.
5. A binary representation of a computable probability space (\mathcal{X}, μ) is a pair (δ, μ_δ) where μ_δ is a computable probability measure on $\{0, 1\}^\infty$ and $\delta : (\{0, 1\}^\infty, \mu_\delta) \rightarrow (\mathcal{X}, \mu)$ is a surjective morphism such that, calling $\delta^{-1}(x)$ the set of expansions of $x \in X$:
 - There is a dense full-measure constructive G_δ -set D of points having a unique expansion.
 - $\delta^{-1} : D \rightarrow \delta^{-1}(D)$ is computable.
 - (δ, δ^{-1}) is an isomorphism.

Theorem 4 Every computable probability space (\mathcal{X}, μ) has a binary representation.

Definition 6 Given a probability measure $\mu \in \mathcal{M}(\mathcal{X})$, a μ -randomness test is a μ -constructive function $T \in \mathcal{F}$, such that $\int T d\mu \leq 1$. A uniform randomness test is a constructive function T from $\mathcal{M}(\mathcal{X})$ to \mathcal{F} such that $\int T^\mu d\mu \leq 1$.

Theorem 5

1. Let μ be a probability measure. For every μ -randomness test t , there is a uniform randomness test $T : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{F}$ with $T(\mu) = .5t$.
2. There is a universal uniform randomness test, that is a uniform test \mathbf{t} such that for every uniform test T , there is a constant $c > 0$ with $\mathbf{t} > cT$.

7 Multi Binary Representation

This paper introduces a new concept that is needed in the first theorem: multi binary representation. While a binary representation is a mapping from one computable probability space to the Cantor space, a multi binary representation maps three computable probability spaces to Cantor spaces, each sharing the same mapping.

Definition 7 A set A is almost decidable with respect to probability measures (μ, ν, ρ) if there are two r.e. open sets U and V such that $U \subset A$, $V \subseteq A^c$, $U \cup V$ is dense and has full μ , ν , and ρ measure. We say the elements of a sequence $\{A_i\}$ are uniformly almost decidable with respect to (μ, ν, ρ) if there are two sequences $\{U_i\}$ and $\{V_i\}$ of uniformly r.e. sets satisfying the above conditions.

Theorem 6 On a computable metric space, every dense constructive G_δ -set has a dense sequence of uniformly computable points.

Lemma 6 There is a sequence of $\{r_n\}$ of uniformly computable reals such that $\{B(s_i, r_n)\}_{i,n}$ is a basis of uniformly almost computable decidable balls, relative to (μ, ν, ρ) .

Proof. Define $U_{\langle i,k \rangle} = \{r \in \mathbb{R}_{>0} : \mu(\overline{B}(s_i, r)) < \mu(B(s_i, r)) + 1/k\}$. By computability of μ , this is a r.e. open subset of $\mathbb{R}_{>0}$ uniformly in $\langle i, k \rangle$. Let $W_{\langle i,k \rangle} = \{r \in \mathbb{R}_{>0} : \nu(\overline{B}(s_i, r)) < \nu(B(s_i, r)) + 1/k\}$, which is also an r.e. open subset of $\mathbb{R}_{>0}$. Let $X_{\langle i,j \rangle} = \{r \in \mathbb{R}_{>0} : \rho(\overline{B}(s_i, r)) < \rho(B(s_i, r)) + 1/k\}$, which is also an r.e. open subset of $\mathbb{R}_{>0}$. They both are dense in $\mathbb{R}_{>0}$. The spheres $S_r = \overline{B}(s_i, r) \setminus B(s_i, r)$ are disjoint for different radii and μ, ν , and ρ are finite, so the set of r for which $\mu(S_r) \geq 1/k$, $\nu(S_r) \geq 1/k$, or $\rho(S_r) > 1/k$ is finite. Let $V_{\langle i,j \rangle} = \mathbb{R}_{>0} \setminus \{d(s_i, s_j)\}$ be a dense r.e. open set, uniformly in $\langle i, j \rangle$. Then by Theorem 6, the dense constructive G_δ -set

$$\bigcap_{\langle i,k \rangle} U_{\langle i,k \rangle} \cap \bigcap_{\langle i,k \rangle} W_{\langle i,k \rangle} \cap \bigcap_{\langle i,j \rangle} X_{\langle i,j \rangle} \cap \bigcap_{\langle i,j \rangle} V_{\langle i,j \rangle}$$

contains a sequence $\{r_n\}$ of uniformly computable real numbers which is dense in $\mathbb{R}_{>0}$. For any s_i and r_n , $B(s_i, r_n)$ is almost decidable, relative to (μ, ν, ρ) . Thus $\{(B_i, r_n)\}_{i,n}$ is a basis of uniformly almost computable decidable balls, relative to (μ, ν, ρ) . \square

Definition 8 A multi probability space $(\mathcal{X}, \mu, \nu, \rho)$ is a computable metric space \mathcal{X} and three computable Borel probability measures, μ, ν , and ρ over \mathcal{X} .

Definition 9 A multi binary representation of a multi probability space $(\mathcal{X}, \mu, \nu, \rho)$ is a tuple $(\delta, \mu_\delta, \nu_\delta, \rho_\delta)$ where μ_δ, ν_δ , and ρ_δ are computable probability measures on $\{0, 1\}^\infty$ and $\delta : (\{0, 1\}^\infty, \mu_\delta) \rightarrow (\mathcal{X}, \mu)$, $\delta : (\{0, 1\}^\infty, \nu_\delta) \rightarrow (\mathcal{X}, \nu)$, and $\delta : (\{0, 1\}^\infty, \rho_\delta) \rightarrow (\mathcal{X}, \rho)$ are surjective morphisms. Denoting $\delta^{-1}(x)$ to be the set of expansion of $x \in X$:

- There is a dense full-measure constructive G_δ -set D of points have a unique expansion.
- $\delta^{-1} : D \rightarrow \delta^{-1}(D)$ is computable.
- (δ, δ^{-1}) is an isomorphism.

Theorem 7 Every multi probability space (\mathcal{X}, μ, ν) has a multi binary representation.

Proof. This proof follows identically to the proof of Theorem 5.1.1 in [HR09], using Lemma 6 instead of Lemma 5.1.1, and noting that $\mu_\delta = \mu \circ b^{-1}$, $\nu_\delta = \nu \circ b^{-1}$, and $\rho_\delta = \rho \circ b^{-1}$, where b is defined in the proof. \square

8 Universal Uniform Tests

Lemma 7 Let $Q : D \subset \mathcal{X} \rightarrow \mathcal{Y}$ be a morphism of computable probability spaces (\mathcal{X}, μ) and (\mathcal{Y}, ν) , with universal tests \mathbf{t}_μ and \mathbf{t}_ν . There is a $c \in \mathbb{N}$ with the following properties. If $x \in \mathcal{X}$ and $\mathbf{t}_\mu(x) < \infty$, then $Q(x)$ is defined and $\mathbf{t}_\nu(Q(x)) < c\mathbf{t}_\mu(x)$.

Proof. The proof is a slight modification to Proposition 6.2.1 in [HR09]. So, assuming $\mathbf{t}_\mu(x) < \infty$, then x is a random point then $x \in D$, because due to Lemma 6.2.1 in [HR09], every random point lies in every r.e. open set of full measure, and D is an intersection of full-measure r.e. open sets. Thus $Q(x)$ is defined.

Let \mathcal{A} be any algorithm lower semi-computing the function $\mathbf{t}_\nu \circ Q : D \rightarrow \mathbb{R}_{\geq 0}^\infty$. This algorithm can be converted into a lower computable function $f_\mathcal{A} : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\infty$ by feeding all finite prefixes of fast Cauchy sequences to Q and enumerating all resultant outputted ideal balls and seeing which outputted ideal balls are in the ideal balls of those enumerated by \mathbf{t}_ν . Since $\mu(D) = 1$, $\int \mathbf{t}_\nu \circ Q d\mu$ equals $\int f_\mathcal{A} d\mu$. As Q is measure-preserving, $\int \mathbf{t}_\nu \circ Q du = \int \mathbf{t}_\nu d\nu \leq 1$. Hence $f_\mathcal{A}$ is a μ -test, with $f_\mathcal{A} < c\mathbf{t}_\mu$ for some $c \in \mathbb{N}$. Thus $\mathbf{t}_\nu(Q(x)) = f_\mathcal{A}(x) < c\mathbf{t}_\mu(x)$. \square

Corollary 1 Let $(Q, R) : (\mathcal{X}, \mu) \rightleftharpoons (\mathcal{Y}, \nu)$ be an isomorphism of computable probability spaces, with universal tests \mathbf{t}_μ and \mathbf{t}_ν . Then there is a $c \in \mathbb{N}$ where $\mathbf{t}_\nu(Q(x)) = \mathbf{t}_\mu(x) \pm c$ and $\mathbf{t}_\mu(R(y)) = \mathbf{t}_\nu(y) \pm c$.

Claim 1 We recall that 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

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

We have $\mathbf{D}(\alpha|P) = \mathbf{D}(\alpha|P, \emptyset)$. By [G21], $2^{\mathbf{D}}$ is a lower-computable P -test, in that $\int_{\{0,1\}^\infty} 2^{\mathbf{D}(\alpha|P)} dP(\alpha) = O(1)$, lower computed by program of size $\mathbf{K}(\langle P \rangle)$. Thus since \mathbf{t}_P is a universal uniform test, $\mathbf{t}_P(\alpha) \geq^* \mathbf{m}(\langle P \rangle) 2^{\mathbf{D}(\alpha|P)}$.

Theorem 8 Given computable probability measures μ, ρ , and λ , non-atomic λ , over a computable metric space \mathcal{X} and universal uniform test \mathbf{t} , there is $c \in \mathbb{N}$ where for all n , $\lambda(\{\alpha : \mathbf{t}_\mu(\alpha) > 2^n \text{ and } \mathbf{t}_\rho(\alpha) > 2^n\}) > 2^{-n-\mathbf{K}(n)-c}$.

Proof. Assume not. For all $c \in \mathbb{N}$, there exist computable measures μ, ρ , and λ , non-atomic λ , over computable metric space \mathcal{X} and there exists n , where $\lambda\{\alpha : \mathbf{t}_\mu(\alpha) > 2^n \text{ and } \mathbf{t}_\rho(\alpha) > 2^n\} \leq 2^{-n-\mathbf{K}(n)-c}$. Sample $2^{n+\mathbf{K}(n)+c-1}$ elements $G \subset \mathcal{X}$ according to λ . The probability that all samples $\beta \in G$ have $\mathbf{t}_\mu(\beta) \leq 2^n$ or $\mathbf{t}_\rho(\beta) \leq 2^n$ is

$$\prod_{i=1}^{|G|} \lambda\{\beta : \mathbf{t}_\mu(\beta) \leq 2^n \text{ or } \mathbf{t}_\rho(\beta) \leq 2^n\} \geq (1 - |G| 2^{-n-\mathbf{K}(n)-c}) \geq (1 - 2^{n+\mathbf{K}(n)+c-1} 2^{-n-\mathbf{K}(n)-c}) \geq 1/2. \quad (2)$$

By Theorem 4, fix a multi binary representation $(\delta, \lambda_\delta, \rho_\delta, \mu_\delta)$ for multi probability space $(\mathcal{X}, \lambda, \mu, \rho)$. Note that δ is a measure-preserving transform, where $\lambda(A) = \lambda_\delta(\delta^{-1}(A))$ for all Borel sets A . Due to Lemma 7, there is some $c' \in \mathbb{N}$ where

$$\begin{aligned} & \lambda(\{\beta : \mathbf{t}_\mu(\beta) \leq 2^n \text{ or } \mathbf{t}_\rho(\beta) \leq 2^n\}) \\ &= \lambda_\delta(\delta^{-1}(\{\beta : \mathbf{t}_\mu(\beta) \leq 2^n \text{ or } \mathbf{t}_\rho(\beta) \leq 2^n\})) \\ &\leq \lambda_\delta(\delta^{-1}(\{\beta : \mathbf{t}_{\mu_\delta}(\delta^{-1}(\beta)) \leq 2^{n+c'} \text{ or } \mathbf{t}_{\rho_\delta}(\delta^{-1}(\beta)) \leq 2^{n+c'}\})) \\ &= \lambda_\delta(\delta^{-1}(\{\beta \in \delta(\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) \leq 2^{n+c'} \text{ or } \mathbf{t}_{\rho_\delta}(\alpha) \leq 2^{n+c'}\})\})) \\ &= \lambda_\delta(\delta^{-1}(\delta(\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) \leq 2^{n+c'} \text{ or } \mathbf{t}_{\rho_\delta}(\alpha) \leq 2^{n+c'}\}))) \\ &\leq \lambda_\delta(\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) \leq 2^{n+c'} \text{ or } \mathbf{t}_{\rho_\delta}(\alpha) \leq 2^{n+c'}\}). \end{aligned} \quad (3)$$

Combining Equations 2 and 3, and from Claim 1, we get, (by updating c')

$$\begin{aligned} 1/2 &\leq \prod_{i=1}^{|G|} \lambda_\delta\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) \leq 2^{n+c'} \text{ or } \mathbf{t}_{\rho_\delta}(\alpha) \leq 2^{n+c'}\}, \\ 1/2 &\leq \prod_{i=1}^{|G|} \lambda_\delta\{\alpha : \mathbf{D}(\alpha|\mu_\delta) \leq n + c' + \mathbf{K}(\mu_\delta) \text{ or } \mathbf{D}(\alpha|\rho_\delta) \leq n + c' + \mathbf{K}(\rho_\delta)\}. \end{aligned} \quad (4)$$

Let $\lambda_\delta^{n,c}$ be the probability of an encoding of $2^{n+\mathbf{K}(n)+c-1}$ infinite sequences each distributed independently according to λ_δ . Thus

$$\begin{aligned} & \lambda_\delta^{n,c}(\text{Encoding of } 2^{n+\mathbf{K}(n)+c-1} \text{ infinite sequences } \alpha, \text{ each having } \mathbf{D}(\alpha|\mu_\delta) \leq n + c' + \mathbf{K}(\mu_\delta) \\ & \text{ or } \mathbf{D}(\alpha|\rho_\delta) \leq n + c' + \mathbf{K}(\rho_\delta)) \geq 1/2. \end{aligned}$$

Let $n^* = \langle n, \mathbf{K}(n) \rangle$. By Theorem 1, with the universal Turing machine relativized to n^* ,

$$\lambda_\delta^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|n^*) > m\}) \stackrel{*}{<} 2^{-m+\mathbf{K}(\lambda_\delta^{n,c}|n^*)} \stackrel{*}{<} 2^{-m+\mathbf{K}(n^*,c,\lambda_\delta|n^*)} \stackrel{*}{<} 2^{-m+\mathbf{K}(c,\lambda_\delta)}.$$

Therefore,

$$\lambda_\delta^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|n^*) >^+ \mathbf{K}(c, \lambda_\delta)\}) < 1/4.$$

Thus, by probabilistic arguments, there exists $\alpha \in \{0,1\}^\infty$, such that $\alpha = \langle G \rangle$ is an encoding of $2^{n+\mathbf{K}(n)+c-1}$ elements $G \subset \{0,1\}^\infty$. Each $\gamma \in G$ has $\mathbf{D}(\gamma|\mu_\delta) < n + c' + \mathbf{K}(\mu_\delta)$ or $\mathbf{D}(\alpha|\rho_\delta) \leq n + c' + \mathbf{K}(\rho_\delta)$. and also $\mathbf{I}(\alpha : \mathcal{H}|n^*) <^+ \mathbf{K}(c, \lambda_\delta)$. By Lemma 5, relativized to n^* , there are constants $d, f, g \in \mathbb{N}$ where

$$\begin{aligned} m &= \log |G| < \max_{\beta \in G} \min\{\mathbf{D}(\beta|\mu_\delta, n^*), \mathbf{D}(\beta|\rho_\delta, n^*)\} + 2\mathbf{I}(G : \mathcal{H}|n^*) \\ &\quad + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*) + g\mathbf{K}(\rho_\delta|n^*) \\ m &<^+ \max_{\beta \in G} \min\{\mathbf{D}(\beta|\mu_\delta, n^*), \mathbf{D}(\beta|\rho_\delta, n^*)\} + \mathbf{K}(n) + 2\mathbf{I}(G : \mathcal{H}|n^*) \\ &\quad + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*) + g\mathbf{K}(\rho_\delta|n^*) \\ m &<^+ n + c' + \mathbf{K}(n) + 2\mathbf{I}(G : \mathcal{H}|n^*) + d\mathbf{K}(m|n^*) + (f+1)\mathbf{K}(\mu_\delta|n^*) + (g+1)\mathbf{K}(\rho_\delta|n^*) \\ m &<^+ n + c' + \mathbf{K}(n) + 2\mathbf{K}(c, \lambda_\delta) + d\mathbf{K}(m|n^*) + (f+1)\mathbf{K}(\mu_\delta|n^*) + (g+1)\mathbf{K}(\rho_\delta|n^*). \end{aligned} \quad (5)$$

Therefore:

$$\begin{aligned} m &= n + \mathbf{K}(n) + c - 1 \\ \mathbf{K}(m|n^*) &<^+ \mathbf{K}(c). \end{aligned} \quad (6)$$

Plugging Equation 6 back into Equation 5 results in

$$\begin{aligned} n + \mathbf{K}(n) + c &<^+ n + c' + \mathbf{K}(n) + (2+d)\mathbf{K}(c) + dO(1) + (f+1)\mathbf{K}(\mu_\delta) + (g+1)\mathbf{K}(\rho_\delta) + 2\mathbf{K}(\lambda_\delta) \\ c &<^+ c' + (2+d)\mathbf{K}(c) + dO(1) + (f+1)\mathbf{K}(\mu_\delta) + (g+1)\mathbf{K}(\rho_\delta) + 2\mathbf{K}(\lambda_\delta). \end{aligned}$$

This result is a contradiction for sufficiently large c solely dependent on \mathcal{X} , μ , ρ , λ , and the reference universal Turing machine. \square

References

- [Eps21] Samuel Epstein. All sampling methods produce outliers. *IEEE Transactions on Information Theory*, 67(11):7568–7578, 2021.
- [Eps22] S. Epstein. The outlier theorem revisited. *CoRR*, abs/2203.08733, 2022.
- [Eps23a] S. Epstein. On the existence of anomalies. *CoRR*, abs/2302.05972, 2023.
- [Eps23b] S. Epstein. On the existence of anomalies, the reals case. *CoRR*, abs/2303.05614, 2023.
- [G21] Peter Gács. Lecture notes on descriptonal complexity and randomness. *CoRR*, abs/2105.04704, 2021.
- [Gei12] Philipp Geiger. *Mutual information and Gödel incompleteness*. PhD thesis, Heidelberg University, 10 2012.

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