Uniform Tests and Algorithmic Thermodynamic Entropy

Samuel Epstein*

January 21, 2023

Abstract

We prove that given a computable metric space and two computable measures, the set of points that have high universal uniform test scores with respect to the first measure will have a lower bound with respect to the second measure. This result is transferred to thermodynamics, showing that algorithmic thermodynamic entropy must oscillate in the presence of dynamics.

1 Introduction

The study of randomness of computable metric spaces can be seen in the works of [HR09, Gá21]. These spaces are important because physical random phenomena are modeled using infinite objects, and not the Cantor space. In this paper, we prove a result in computable metric spaces and then transfer it to physics, in particular, thermodynamics. 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 $\int_{\mathcal{X}} T^\mu d\mu \leq 1$. 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 Theorem 3 of [Eps22] to computable metric spaces.

Theorem. *Given computable probability measures μ 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, $\lambda(\{\alpha : \mathbf{t}_\mu(\alpha) > 2^n\}) > 2^{-n-\mathbf{K}(n)-c}$.*

1.1 Thermodynamics

The above theorem can be applied to thermodynamics. Classical thermodynamics is the study of substances and changes to their properties such as volume, temperature, and pressure. Substances, such as a gas or a liquid, is modeled as a point in a phase space. The phase space, $\mathcal{X}$, is modeled by a computable metric space, [HR09], and a volume measure μ, is modeled by a computable (not necessarily probabilistic) positive measure over $\mathcal{X}$. The dynamics are modeled by a one dimensional transformation group G^t, indexed by $t \in \mathbb{R}$. Due to Louville's theorem, the dynamics are measure-preserving, where $\mu(G^t A) = \mu(A)$, for all Borel sets $A \subseteq \mathcal{X}$.

Whether quantum or classical, the known laws of physics are reversible. Thus the dynamics G of our system are also reversible, in that if $\beta = G^t \alpha$, then there is some t' such that the original state can found with $\alpha = G^{t'} \beta$. Thus if given a set of particles with position and velocity, by reversing the velocities, a previous state can be found. This is contradiction to the second law of thermodynamics, which states,

---

*JP Theory Group. samepst@jptheorygroup.org

The total entropy of a system either increases or remains constant in any spontaneous process; it never decreases.

This conforms to our experiences of broken vases never reforming. To reconcile this difference, Boltzmann introduced *macro-states*, Π_i, indexed by $i \in \mathbb{N}$, which groups states together by macroscopic parameters, with corresponding Boltzmann entropy $S(\Pi_i) = k_B \ln \mu(\Pi_i)$. By definition, a vast majority of typical states will experience an increase in Boltzmann entropy.

In [Gac94], coarse grained entropy was introduced as an algorithmic update to Boltzmann entropy. This formulation was made to be independent of the choice of parameters of the macro state. In this paper, we introduce a modified version of coarse grained entropy. We also model the thermodynamic entropy of a micro-state with algorithmic methods. The micro-state of a system contains the information of the entire physical state. For example, the microstate of a system of N molecules is a point

$$(q_1, \ldots, q_{3N}, p_1, \ldots, p_{3N}) \in \mathbb{R}^{6N}$$

where q_i are the position coordinates and p_i are the momentum coordinates. The set of states, $\mathbb{R}^{3N}$ is a computable metric space. To model the entropy of the state, we use slight variant to algorithmic fine-grained entropy H_μ in [Gac94], using symbol $\mathbf{H}_\mu$. This entropy measure captures the level of disorder of the state. Continuing the example above, if all the particles are at rest, then the thermodynamic entropy of the state of

$$(q_1, \ldots, q_{3N}, 0, \ldots, 0)$$

is expected to be very low.

The evolution of the system will be thermodynamic like if it spends most of the time close to its maximum value, from which it exhibits frequent small fluctuations and rarer large fluctuations.

In this paper, using the algorithmic definition of thermodynamic entropy, $\mathbf{H}_\mu$, we prove that such fluctuations *have to* occur, and the greater the fluctuation, the lesser its measure. The symbol μ is a Borel measure representing volume of the phase space. In this paper, we show that thermodynamic entropy $\mathbf{H}_\mu$ must oscillate in the presence of dynamics.

Theorem. *Let L be the Lebesgue measure over $\mathbb{R}$ and $\alpha \in \mathcal{X}$, with finite mutual information with the halting sequence. For transformation group G^t acting on computable metric space $\mathcal{X}$, there are constants c_1 and c_2 with $2^{-n-\mathbf{K}(n)-c_1} < L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - n\} < 2^{-n+c_2}$.*

2   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 later sections. 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 1. 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 $\overrightarrow{x}$ for x, $\mathfrak{A}(\overrightarrow{x})$ outputs an encoding of a fast Cauchy sequence for y.

Definition 2. 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 3 (Prokhorov metric).

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

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

Theorem 1. *Given a probability measure $\mu \in \mathcal{M}(\mathcal{X})$, the following are equivalent.*

1. *μ is computable.*

2. *$\mu(B_{i_1} \cup \cdots \cup B_{i_k})$ is lower semi-computable uniformly in $\langle i_1, \ldots, i_k \rangle$.*

3. *$\int d\mu : \mathcal{F} \to \mathbb{R}_{\geq 0}$ is lower semi-computable.*

Definition 4.

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 $(\mathcal{X}, \mu)$, where $\mathcal{X}$ is a computable metric space and μ is a Borel probability measure on $\mathcal{X}$.

3. Let $(\mathcal{X}, \mu)$ be a computable probability space and $\mathcal{Y}$ a computable metric space. A function $f : D_f \subset (\mathcal{X}, \mu) \to \mathcal{Y}$ 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 $Q : (\mathcal{X}, \mu) \to (\mathcal{Y}, \nu)$ is an almost computable measure-preserving function $Q : D_Q \subset \mathcal{X} \to \mathcal{Y}$, 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 $(\mathcal{X}, \mu)$ is a pair (δ, μ_δ) where μ_δ is a computable probability measure on $\{0,1\}^\infty$ and $\delta : (\{0,1\}^\infty, \mu_\delta) \to (\mathcal{X}, \mu)$ 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.
   - $\delta^{-1} : D \to \delta^{-1}(D)$ is computable.
   - (δ, δ^{-1}) is an isomorphism.

Theorem 2. *Every computable probability space $(\mathcal{X}, \mu)$ has a binary representation.*

Definition 5. 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 3.

1. *Let μ be a probability measure. For every μ-randomness test t, there is a uniform randomness test $T : \mathcal{M}(\mathcal{X}) \to \mathcal{F}$ 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 $\mathbf{t} > cT$.*

3   Dual Binary Representation

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

Definition 6. A set A is *almost decidable* with respect to probability measures (μ, ν) if there are two. r.e. open sets U and V such that $U \subset A$, $V \subseteq A^\mathscr{C}$, $U \cup V$ is dense and has full μ and ν measure. We say the elements of a sequence $\{A_i\}$ are uniformly almost decidable with respect to (μ, ν) if there are two sequences $\{U_i\}$ and $\{V_i\}$ of uniformly r.e. sets satisfying the above conditions.

Theorem 4. *On a computable metric space, every dense constructive G_δ-set has a dense sequence of uniformly computable points.*

Lemma 1. *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).*

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 $\mathbb{R}_{>0}$. 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 and ν are finite, so the set of r for which $\mu(S_r) \geq 1/k$ or $\nu(S_r) \geq 1/k$ 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 4, the dense constructive G_δ-set

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

contains a sequence $\{r_n\}$ of uniformly computable reals 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). Thus $\{(B_i, r_n)\}_{i,n}$ is a basis of uniformly almost computable decidable balls, relative to (μ, ν). $\square$

Definition 7. A dual probability space $(\mathcal{X}, \mu, \nu)$ is a computable metric space $\mathcal{X}$ and two computable Borel probability measures, μ and ν, over $\mathcal{X}$.

Definition 8. A dual binary representation of a dual probability space $(\mathcal{X}, \mu, \nu)$ is a tuple $(\delta, \mu_\delta, \nu_\delta)$ where μ_δ and ν_δ are computable probability measures on $\{0,1\}^\infty$ and $\delta : (\{0,1\}^\infty, \mu_\delta) \to (\mathcal{X}, \mu)$ and $\delta : (\{0,1\}^\infty, \nu_\delta) \to (\mathcal{X}, \nu)$ 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.

- $\delta^{-1} : D \to \delta^{-1}(D)$ is computable.

- (δ, δ^{-1}) is an isomorphism.

Theorem 5. *Every dual probability space $(\mathcal{X}, \mu, \nu)$ has a dual binary representation.*

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

4 Universal Uniform Tests

This section contains the first main result of the paper, Theorem 8. This section also includes further results needed for the proof of this theorem. Lemma 2, is derived from Proposition 6.2.1 from [HR09]. We also include definitions and theorems about the Cantor space, including Theorem 7.

Lemma 2. *Let $Q : D \subset \mathcal{X} \to \mathcal{Y}$ be a morphism of computable probability spaces $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$, with universal tests $\mathbf{t}_\mu$ and $\mathbf{t}_\nu$. 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 $\mathbf{t}_\nu(Q(x)) < \mathbf{t}_\mu(x) + c$.*

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 $\mathscr{A}$ be any algorithm lower semi-computing the function $\mathbf{t}_\nu \circ Q : D \to \mathbb{R}_{\geq 0}^\infty$. This algorithm can be converted into a lower computable function $f_\mathscr{A} : \mathcal{X} \to \mathbb{R}_{\geq 0}^\infty$ 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 $\mathbf{t}_\nu$. Since $\mu(D) = 1$, $\int \mathbf{t}_\nu \circ Q d\mu$ equals $\int f_\mathscr{A} d\mu$. As Q is measure-preserving, $\int \mathbf{t}_\nu \circ Q du = \int \mathbf{t}_\nu d\nu \leq 1$. Hence $f_\mathscr{A}$ is a μ-test, with $f_\mathscr{A} < c\mathbf{t}_\mu$ for some $c \in \mathbb{N}$. Thus $\mathbf{t}_\nu(Q(x)) = f_\mathscr{A}(x) < c\mathbf{t}_\mu(x)$. $\qquad\square$

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

The algorithmic probability is $\mathbf{m}(x|y)$. The conditional prefix free complexity is $\mathbf{K}(x|y)$. Symmetric mutual information is $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$. The halting sequence is $\mathcal{H} \in \{0,1\}^\infty$. For positive real functions f, by $<^+ f$, $>^+ f$, $=^+ f$, denote $\leq f + O(1)$, $\geq f - O(1)$, $= f \pm O(1)$, respectively. In addition $\overset{*}{<} f$, $\overset{*}{>} f$, and $\overset{*}{=} f$ denote $< f/O(1)$, $> f/O(1)$ and $= f * /O(1)$, respectively.

Definition 9. For computable probability P, $\langle P \rangle \in \{0,1\}^*$ is the smallest program that can compute $P(x\{0,1\}^\infty)$, uniformly computable $x \in \{0,1\}^*$. 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 [Gá21], $2^{\mathbf{D}}$ 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, $\mathbf{t}_P(\alpha) \overset{*}{>} \mathbf{m}(\langle P \rangle)2^{\mathbf{D}(\alpha|P)}$.

Definition 10. The information term between infinite sequences is
$\mathbf{I}(\alpha : \beta) = \log \sum_{x,y \in \{0,1\}^*} \mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\mathbf{I}(x:y)}$ [Lev74]. $\mathbf{I}(f(\alpha) : \beta) <^+ \mathbf{I}(\alpha : \beta) + \mathbf{K}(f)$.

Theorem 6 ([Ver21, Lev74, Gei12]). $\Pr_\mu(\mathbf{I}(\alpha : \mathcal{H}) > n) \overset{*}{<} 2^{-n+\mathbf{K}(\mu)}$, where $\mathbf{K}(\mu)$ is the size of the smallest program that can compute $\mu(x\{0,1\}^\infty)$, uniformly in $x \in \{0,1\}^*$.

Theorem 7 ([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 8. *Given computable probability measures μ 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, $\lambda(\{\alpha : \mathbf{t}_\mu(\alpha) > 2^n\}) > 2^{-n-\mathbf{K}(n)-c}$.*

Proof. Assume not. For all $c \in \mathbb{N}$, there exist computable non-atomic measures μ, λ over computable metric space $\mathcal{X}$ and there exists n, where $\lambda\{\alpha : \mathbf{t}_\mu(\alpha) > 2^n\} \leq 2^{-n-\mathbf{K}(n)-c}$. 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 $\mathbf{t}_\mu(\beta) \leq 2^n$ is

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

By Theorem 2, fix a dual binary representation $(\delta, \lambda_\delta, \mu_\delta)$ for dual probability space $(\mathcal{X}, \lambda, \mu)$. Note that δ is a measure-preserving transform, where $\lambda(A) = \lambda_\delta(\delta^{-1}(A))$ for all Borel sets A. Due to Lemma 2, there is some $c' \in \mathbb{N}$ where

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

Combining Equations 1 and 2, and from Definition 9, we get, (by updating c')

$$1/2 \leq \prod_{i=1}^{|G|} \lambda_\delta\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) \leq 2^{n+c'}\},$$

$$1/2 \leq \prod_{i=1}^{|G|} \lambda_\delta\{\alpha : \mathbf{D}(\alpha|\mu_\delta) \leq n + c' + \mathbf{K}(\mu_\delta)\}. \quad (3)$$

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

$$\lambda_\delta^{n,c}(\text{Encoding of } 2^{n+\mathbf{K}(n)+c-1} \text{ infinite sequences } \alpha, \text{ each having } \mathbf{D}(\alpha|\mu_\delta) \le n + c' + \mathbf{K}(\mu_\delta)) \ge 1/2.$$

Let $n^* = \langle n, \mathbf{K}(n) \rangle$. By Theorem 6, with the universal Turing machine relativized to n^*,

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

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 $\mathbf{D}(\gamma|\mu_\delta) < n + c' + \mathbf{K}(\mu_\delta)$ and also $\mathbf{I}(\alpha : \mathcal{H}|n^*) <^+ \mathbf{K}(c,\lambda_\delta)$. By Theorem 7, relativized to n^*, there are constants $d, f \in \mathbb{N}$ where

$$m = \log|G| < \max_{\beta \in G} \mathbf{D}(\beta|\mu_\delta, n^*) + 2\mathbf{I}(G : \mathcal{H}|n^*) + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*)$$
$$m <^+ \max_{\beta \in G} \mathbf{D}(\beta|\mu_\delta) + \mathbf{K}(n) + 2\mathbf{I}(G : \mathcal{H}|n^*) + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*)$$
$$m <^+ \max_{\beta \in G} \mathbf{D}(\beta|\mu_\delta) + \mathbf{K}(n) + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*) + 2\mathbf{K}(c,\lambda_\delta)$$
$$m <^+ n + c' + \mathbf{K}(n) + d\mathbf{K}(m|n^*) + (f+1)\mathbf{K}(\mu_\delta) + 2\mathbf{K}(c,\lambda_\delta). \tag{4}$$

Therefore:

$$m = n + \mathbf{K}(n) + c - 1$$
$$\mathbf{K}(m|n^*) <^+ \mathbf{K}(c). \tag{5}$$

Plugging Equation 5 back into Equation 4 results in

$$n + \mathbf{K}(n) + c <^+ n + c' + \mathbf{K}(n) + (2+d)\mathbf{K}(c) + dO(1) + (f+1)\mathbf{K}(\mu_\delta) + 2\mathbf{K}(\lambda_\delta)$$
$$c <^+ c' + (2+d)\mathbf{K}(c) + dO(1) + (f+1)\mathbf{K}(\mu_\delta) + 2\mathbf{K}(\lambda_\delta).$$

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

5   Computable Measure Theory

In thermodynamics, the measure function representing the volume is not necessarily a probability measure. Thus the results of Section 2 needs to be extended to nonnegative measures of arbitrary size to prove a result about thermodynamics. Let $(\mathbb{R}_{\ge 0}, \mathbb{Q}_{\ge 0}, d_{\mathbb{R}})$ be the computable metric space where $\mathbb{R}_{\ge 0}$ is the complete separable metric space and nonnegative rationals $\mathbb{Q}_{\ge 0}$ consists of the ideal points. The distance function is $d_{\mathbb{R}}(x,y) = |x-y|$, which is obviously computable over $\mathbb{Q}_{\ge 0}$. The space of nonnegative Borel measures over a computable metric space is the space $\mathfrak{M}(\mathcal{X}) = \mathcal{M}(X) \times \mathbb{R}_{\ge 0}$, the product space of the space of probability measures of $\mathcal{X}$, $\mathcal{M}(\mathcal{X})$, with the space of nonnegative reals. The distance function of $\mathfrak{M}$ is $d_{\mathfrak{M}}((\mu, m), (\nu, n)) = \max\{\pi(\mu, \nu), d_{\mathbb{R}}(n, m)\}$, where π is the Prokhorov metric (see Definition 3). The ideal points of $\mathfrak{M}(\mathcal{X})$ is the set of all finite points with nonnegative rational values. This definition is different from the ideal points in $\mathcal{M}(\mathcal{X})$ in that they don't have to sum to 1. The computable measures of $\mathfrak{M}(\mathcal{X})$ are its constructive points, with respect to a fast Cauchy description. From this definition, the results of Theorem 1 apply directly to arbitrary measures $\mu \in \mathfrak{M}(\mathcal{X})$.

Corollary 2. *Given an arbitrary measure $\mu \in \mathfrak{M}(\mathcal{X})$, the following are equivalent.*

1. *μ is computable.*

2. *$\mu(B_{i_1} \cup \cdots \cup B_{i_k})$ is lower semi-computable uniformly in $\langle i_1, \ldots, i_k \rangle$.*

3. *$\int d\mu : \mathcal{F} \to \mathbb{R}_{\geq 0}$ is lower semi-computable.*

Definition 11.

1. A *computable measure space* is a pair $(\mathcal{X}, \mu)$, where $\mathcal{X}$ is a computable metric space, and μ is a Borel nonnegative measure on $\mathcal{X}$.

2. Let $(\mathcal{X}, \mu)$ be a computable measure space and $\mathcal{Y}$ a computable metric space. A function $f : D_f \subset (\mathcal{X}, \mu) \to \mathcal{Y}$ is *almost computable* if it is computable on a constructive G_δ-set (D_f) of measure $\mu(\mathcal{X})$.

3. A *morphism* of computable measure spaces $Q : (\mathcal{X}, \mu) \to (\mathcal{Y}, \nu)$ is an almost computable measure-preserving function $Q : D_Q \subset \mathcal{X} \to \mathcal{Y}$. An *isomorphism* (Q, R) is a pair of morphisms such that $Q \circ R = id$ on $R^{-1}(D_Q)$ and $R \circ Q = id$ on $Q^{-1}(D_R)$.

4. A *binary representation* of a computable measure space $(\mathcal{X}, \mu)$ is a pair (δ, μ_δ) where μ_δ is a computable measure on $\{0, 1\}^\infty$ and $\delta : (\{0, 1\}^\infty, \mu_\delta) \to (\mathcal{X}, \mu)$ is a surjective morphism such that, calling $\delta^{-1}(x)$ the set of expansions of $x \in X$:

   - There is a dense constructive G_δ-set D of points having a unique expansion and $\mu(D) = \mu(\mathcal{X})$.
   - $\delta^{-1} : D \to \delta^{-1}(D)$ is computable.
   - (δ, δ^{-1}) is an isomorphism.

Corollary 3. *Every computable measure space $(\mathcal{X}, \mu)$ has a binary representation.*

This corollary has an identical proof to Theorem 2, which is from Theorem 5.1.1 in [HR09].

Definition 12. Given an arbitrary 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 9.

1. *Let μ be a measure. For every μ-randomnness test t, there is a uniform randomness test $T : \mathcal{M}(\mathcal{X}) \to \mathcal{F}$ 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 $\mathbf{t} > cT$.*

6   Algorithmic Thermodynamic Entropy

In this section, we prove the second main result of the paper, that thermodynamic entropy must oscillate in the presence of dynamics. Algorithmic thermodynamic entropy, $\mathbf{H}$, is formally defined as well as the transform group G^t representing dynamics. Theorem 10 is a property about conservation of information with halting sequence. We also provide a new proof to a result in [Gac94] that states thermodynamics will decrease in only small measure.

Definition 13 (Algorithmic Thermodynamic Entropy). Given a computable metric space $\mathcal{X}$ and a nonnegative measure $\mu \in \mathfrak{M}(\mathcal{X})$ the algorithmic thermodynamic entropy is $\mathbf{H}_\mu(\alpha) = -\log \mathbf{t}_\mu(\alpha)$.

Definition 14 (Mutual Information with the Halting Sequence). An encoding of a fast Cauchy sequence $\overrightarrow{x}$ is $\langle \overrightarrow{x} \rangle \in \{0,1\}^\infty$, with $\langle \overrightarrow{x} \rangle = \langle x_1 \rangle \langle x_2 \rangle \dots$. Each point $x \in \mathcal{X}$ has a certain mutual information with the halting sequence $\mathbf{I}(x : \mathcal{H}) = \inf\{\mathbf{I}(\langle \overrightarrow{x} \rangle : \mathcal{H}) : \langle \overrightarrow{x} \rangle$ is a fast Cauchy sequence for $x\}$.

Definition 15 (Computable Transformation Group). A one dimensional transformation group G^t, parameterized by $t \in \mathbb{R}$ over a measure space $(\mathcal{X}, \mu)$ where each G^t is a homeomorphism of $\mathcal{X}$ onto itself, where $G^t(G^s(x) = G^{t+s}(x)$. And $G^t x$ is continuously simultaneously in x and t. G is measure preserving, where $\mu(G^t(A)) = \mu(A)$, for all Borel sets A. Furthermore there is a program that when given an encoding of a fast Cauchy sequence of $t \in \mathbb{R}$ and $x \in \mathcal{X}$, outputs an encoding of a fast Cauchy sequence of $G^t x$.

Theorem 10 ([Ver21, Lev74]). *Let P_ρ, be a family of probability distributions over $\{0,1\}^\infty$, indexed by $\rho \in \{0,1\}^\infty$. Assume that there is a Turing machine T such that for all $\rho \in \{0,1\}^\infty$ computes P_ρ having oracle access to ρ. By "compute" we mean all the measures of the cylinder sets $P_\rho(x\{0,1\}^\infty)$, can be computed, uniformly in $x \in \{0,1\}^*$. Then there is a constant $c_T > 0$ solely dependent on T such that*

$$P_\rho\{\gamma : \mathbf{I}(\langle \gamma, \rho \rangle : \mathcal{H}) > m\} < 2^{\mathbf{I}(\rho:\mathcal{H})-m+c_T}.$$

Lemma 3. *Let $Q : D \subset \mathcal{X} \to \mathcal{Y}$ be a morphism of computable measure spaces $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$, with universal tests $\mathbf{t}_\mu$ and $\mathbf{t}_\nu$. There is a $c \in \mathbb{N}$ with the following properties. If $x \in X$ and $\mathbf{t}_\mu(x) < \infty$, then $Q(x)$ is defined and $\mathbf{t}_\nu(Q(x)) < \mathbf{t}_\mu(x) + c$.*

The proof for this lemma is identical to that of the proof of Lemma 2.

Definition 16. We update Definition 9 to arbitrary measures. For computable measure μ, $\langle \mu \rangle \in \{0,1\}^*$ is the smallest program that can compute $\mu(x\{0,1\}^\infty)$, uniformly computable $x \in \{0,1\}^*$. The deficiency of randomness of an infinite sequence $\alpha \in \{0,1\}^\infty$ with respect to a computable measure μ over $\{0,1\}^\infty$ is defined to be

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

By [G21], $2^\mathbf{D}$ is a lower-computable μ-test, in that $\int_{\{0,1\}^\infty} 2^{\mathbf{D}(\alpha|\mu)} d\mu(\alpha) = O(1)$, lower computed by program of size $\mathbf{K}(\langle \mu \rangle)$. Thus since $\mathbf{t}_\mu$ is a universal uniform test, $\mathbf{t}_\mu(\alpha) \stackrel{*}{>} \mathbf{m}(\langle \mu \rangle) 2^{\mathbf{D}(\alpha|\mu)}$.

Theorem 11 (Oscillation of Thermodynamic Entropy). *Let L be the Lebesgue measure over $\mathbb{R}$, $(\mathcal{X}, \mu)$ be a computable measure space, $\alpha \in \mathcal{X}$, with finite $\mathbf{I}(\alpha : \mathcal{H})$. For transformation group G^t acting on $\mathcal{X}$, there is a constant c with $L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log \mu(X) - n\} > 2^{-n-\mathbf{K}(n)-c}$.*

Proof. We first assume not. There exists $(G^t, \mathcal{X})$ and computable measure space $(\mathcal{X}, \mu)$ and there exists $\alpha \in X$ such that for all $c \in \mathbb{N}$, there exist exists n, where

$$L(\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log \mu(\mathcal{X}) - n\}) < 2^{-n-\mathbf{K}(n)-c}$$

$$L(\{t \in [0,1] : n - \log \mu(\mathcal{X}) < \log \mathbf{t}_\mu(G^t\alpha)\}) < 2^{-n-\mathbf{K}(n)-c}.$$

We sample $2^{n+\mathbf{K}(n)+c-1}$ elements F by choosing a time t uniformly between $[0,1]$. The probability that all samples $\beta \in F$ have $\mathbf{t}_\mu(G^\beta\alpha) \leq n - \log\mu(\mathcal{X})$ is

$$\prod_{i=1}^{|F|} L\{t \in [0,1] : \log\mathbf{t}_\mu(G^t\alpha) \leq n - \log\mu(\mathcal{X})\}$$
$$\geq (1 - |F|2^{-n-\mathbf{K}(n)-c})$$
$$\geq (1 - 2^{n+\mathbf{K}(n)+c-1}2^{-n-\mathbf{K}(n)-c})$$
$$\geq 1/2.$$

Let $(\{0,1\}^\infty, \Gamma)$ be the Cantor space with the uniform measure. The binary representation (see Definition 4) creates an isomorphism (ϕ, ϕ^{-1}) of computable probability spaces between the spaces $(\{0,1\}^\infty, \Gamma)$ and $([0,1], L)$. Thus for all Borel sets $A \subseteq [0,1]$, $\Gamma(\phi^{-1}(A)) = L(A)$. Since $\{t \in [0,1] : \log\mathbf{t}_\mu(G^t\alpha) \leq n - \log\mu(\mathcal{X})\}$ is closed,

$$L\{t \in [0,1] : \log\mathbf{t}_\mu(G^t\alpha) \leq n - \log\mu(\mathcal{X})\} = \Gamma\{\gamma \in \{0,1\}^\infty : \log\mathbf{t}_\mu(G^{\phi(\gamma)}\alpha) \leq n - \log\mu(\mathcal{X})\}.$$

So

$$1/2 \leq \prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \log\mathbf{t}_\mu(G^{\phi(\gamma)}\alpha) \leq n - \log\mu(\mathcal{X})\}.$$

Let (δ, μ_δ) be a binary representation (see Definition 11), for the computable measure space $(\mathcal{X}, \mu)$. Thus μ_δ is a computable (not necessarily probability) measure over $\{0,1\}^\infty$. By Lemma 3, there is a $c' > 0$, where

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \log\mathbf{t}_{\mu_\delta}(\delta^{-1}(G^{\phi(\gamma)}\alpha)) \leq n - \log\mu(\mathcal{X}) + c'\} \geq 1/2.$$

Let $f : \{0,1\}^\infty \times \{0,1\}^\infty \to \{0,1\}^\infty$, where $f(\gamma, \langle\overrightarrow{\zeta}\rangle) = \delta^{-1}(G^{\phi(\gamma)}\zeta)$. Note, $f(\gamma, \langle\overrightarrow{\zeta}\rangle)$ can be undefined when $\mathbf{t}_\mu(G^{\phi(\gamma)}\zeta) = \infty$, because the morphism δ^{-1} is only proven to be defined on a constructive G_δ set of full measure which includes random points. Let $\xi = \langle\overrightarrow{\alpha}\rangle$ be an encoding of a fast Cauchy sequence $\overrightarrow{\alpha}$ such that $\mathbf{I}(\xi : \mathcal{H}) < \infty$. The sequence ξ is guaranteed to exist because the assumption of the theorem statement. So

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \log\mathbf{t}_{\mu_\delta}(f(\gamma, \xi)) \leq n - \log\mu(\mathcal{X}) + c'\} \geq 1/2.$$

By Definition 16, (and also updating c')

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \mathbf{D}(f(\gamma, \xi)|\mu_\delta) \leq n - \log\mu(\mathcal{X}) + c' + \mathbf{K}(\mu_\delta)\} \geq 1/2.$$

Let $\overline{\mu}_\delta(\alpha) = \mu_\delta(\alpha)/\mu_\delta(\{0,1\}^\infty)$, which is a computable probability measure over $\{0,1\}^\infty$.

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \mathbf{D}(f(\gamma, \xi)|\overline{\mu}_\delta) \leq n + c' + \mathbf{K}(\mu_\delta)\} \geq 1/2.$$

Let Γ^{n+c} be a computable distribution over the product of $1 + 2^{n+\mathbf{K}(n)+c-1}$ independent probability measures over $\{0,1\}^\infty$, encoding into a $\{0,1\}^\infty$ in the standard way. The first probability distribution gives measure 1 to ξ and the last $2^{n+\mathbf{K}(n)-c}$ probability measures are the uniform distribution Γ over $\{0,1\}^\infty$. So

$$\Gamma^{n+c}(\text{Encoding of } 1 + 2^{n+\mathbf{K}(n)+c-1} \text{ elements with the first encoded sequence being } \xi$$
$$\text{and the rest of encoded sequences } \beta \text{ has } \mathbf{D}(f(\beta,\xi)|\overline{\mu}_\delta) \le n + c' + \mathbf{K}(\mu_\delta)) \ge 1/2.$$

Let $n^* = \langle n, \mathbf{K}(n) \rangle$. There is an infinite sequence $\eta = \langle n, \mathbf{K}(n), c \rangle \xi$ and a Turing machine T, such that T computes Γ^{n+c} when given oracle access to η. By Theorem 10, with the universal Turing machine relativized to n^*, and folding the constants together,

$$\Gamma^{n+c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|n^*) > m\})$$
$$< \Gamma^{n+c}(\{\gamma : \mathbf{I}(\langle \gamma, \eta \rangle : \mathcal{H}|n^*) >^+ m\})$$
$$\overset{*}{<} 2^{-m+\mathbf{I}(\eta:\mathcal{H}|n^*)+c_T}$$
$$\overset{*}{<} 2^{-m+\mathbf{K}(n,\mathbf{K}(n),c|n^*)+\mathbf{I}(\xi:\mathcal{H}|n^*)+c_T}$$
$$\overset{*}{<} 2^{-m+\mathbf{K}(c)}.$$

Therefore,

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

Thus, by probabilistic arguments, there exists $\kappa \in \{0,1\}^\infty$, such that $\kappa = \langle D, \xi \rangle$, where $D \subset \{0,1\}^\infty$ and $|D| = 2^{n+\mathbf{K}(n)+c-1}$ and each $\beta \in D$ has $\mathbf{D}(f(\beta,\xi)|\overline{\mu}_\delta) \le n+c'+\mathbf{K}(\mu_\delta)$ and $\mathbf{I}(\kappa : \mathcal{H}|n^*) <^+ \mathbf{K}(c)$. Thus since $\mathbf{K}(f(D,\xi)|\kappa,n^*) = O(1)$ we have $\mathbf{I}(f(D,\xi) : \mathcal{H}|n^*) <^+ \mathbf{I}(\kappa : \mathcal{H}|n^*) <^+ \mathbf{K}(c)$. By Theorem 7, relativized to n^*, on the set $D' = f(D,\xi)$ and probability $\overline{\mu}_\delta$, there exists constants $d, f \in \mathbb{N}$ where

$$m = \log|D| < \max_{\beta \in D'} \mathbf{D}(\beta|\overline{\mu}_\delta, n^*) + 2\mathbf{I}(D' : \mathcal{H}|n^*) + d\mathbf{K}(m|v) + f\mathbf{K}(\overline{\mu}_\delta|n^*)$$
$$m < \max_{\beta \in D'} \mathbf{D}(\beta|\overline{\mu}_\delta) + \mathbf{K}(n) + 2\mathbf{I}(D' : \mathcal{H}|n^*) + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*)$$
$$<^+ \max_{\beta \in D'} \mathbf{D}(\beta|\overline{\mu}_\delta) + \mathbf{K}(n) + 2\mathbf{K}(c) + d\mathbf{K}(m|v) + f\mathbf{K}(\mu_\delta|n^*)$$
$$<^+ n + \mathbf{K}(n) + d\mathbf{K}(m|v) + 2\mathbf{K}(c) + (f+1)\mathbf{K}(\mu_\delta). \qquad (6)$$

Therefore:

$$m = n + \mathbf{K}(n) + c - 1$$
$$\mathbf{K}(m|n^*) <^+ \mathbf{K}(c). \qquad (7)$$

Plugging Equation 7 back into Equation 6 results in

$$n + \mathbf{K}(n) + c <^+ n + \mathbf{K}(n) + 2\mathbf{K}(c) + d(\mathbf{K}(c) + O(1)) + (f+1)\mathbf{K}(\mu_\delta)$$
$$c <^+ (2+d)\mathbf{K}(c) + dO(1) + (f+1)\mathbf{K}(\mu_\delta).$$

This result is a contradiction for sufficiently large c solely dependent $\mathcal{X}$, G, μ, and the universal Turing machine. $\qquad \square$

Let $(\mathcal{X}, \mu)$ be a computable probability space. Let $([0,1], L)$ be the computable probability space, where L is the Lebesgue measure. $(\mathcal{X} \times [0,1], \mu \times L)$ is a product computable probability space, which we will be using in the following lemma. It provides a different proof to the same lemma in [Gac94].

Lemma 4. *Let L be the Lebesgue measure over $\mathbb{R}$, $(\mathcal{X}, \mu)$ be a computable measure space, and $\alpha \in \mathcal{X}$. For transformation group G^t acting on $\mathcal{X}$, there is a constant c where $L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - m\} < 2^{-m+c}$.*

Proof. Since

$$\int_\mathcal{X} \int_{[0,1]} 2^{-\mathbf{H}_{\mu \times L}(\alpha, t)} dL(t) d\mu(\alpha) = \int_\mathcal{X} \int_{[0,1]} \mathbf{t}_{\mu \times L}(\alpha, t) dL(t) d\mu(\alpha) \le 1,$$

the function $f(\alpha) = \int_{[0,1]} 2^{-\mathbf{H}_{\mu \times L}(\alpha, t)} dL(t)$ is a μ-test. So

$$\int_{[0,1]} 2^{-\mathbf{H}_{\mu \times L}(\alpha, t)} dt = f(\alpha) \overset{*}{<} \mathbf{t}_\mu(\alpha) \overset{*}{=} 2^{-\mathbf{H}_\mu(\alpha)}.$$

So

$$\{t \in [0,1] : 2^{-\mathbf{H}_{\mu \times L}(\alpha, t)} > 2^{m - \mathbf{H}_\mu(\alpha)}\} \overset{*}{<} 2^{-m}$$

$$\{t \in [0,1] : \mathbf{H}_{\mu \times L}(\alpha, t) < \mathbf{H}_\mu(\alpha) - m\} \overset{*}{<} 2^{-m}$$

$\mathbf{H}_{\mu \times L}(\alpha, t) <^+ \mathbf{H}_\mu(G^t\alpha)$ because

$$\int_{[0,1]} \int_\mathcal{X} \mathbf{t}_\mu(G^t\alpha) d\mu(\alpha) dL(t)$$

$$= \int_{[0,1]} \int_\mathcal{X} \mathbf{t}_\mu(\alpha) d\mu(G^{-t}\alpha) dL(t)$$

$$= \int_{[0,1]} \int_\mathcal{X} \mathbf{t}_\mu(\alpha) d\mu(\alpha) dL(t),$$

$$= \int_{[0,1]} 1 dL(t)$$

$$\le 1,$$

which means $\mathbf{t}_\mu(G^t\alpha) \overset{*}{<} \mathbf{t}_{\mu \times L}(\alpha, t)$ and thus $2^{-\mathbf{H}_\mu(G^t\alpha)} \overset{*}{<} 2^{-\mathbf{H}_{\mu \times L}(\alpha, t)}$. Thus

$$\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - m\} \overset{*}{<} 2^{-m}.$$

Corollary 4. *Let L be the Lebesgue measure over $\mathbb{R}$, $(\mathcal{X}, \mu)$ be a computable measure space, and $\alpha \in \mathcal{X}$, with finite $\mathbf{I}(\alpha : \mathcal{H})$. For transformation group G^t acting on $\mathcal{X}$, there are constants $c_1, c_2 > 0$ with $2^{-n - \mathbf{K}(n) - c_1} < L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - n\} < 2^{-n + c_2}$.*

7 Discrete Dynamics

Theorem 8 has a discrete version. Discrete dynamics is modeled by a transform group G^t from Definition 15, but with $t \in \mathbb{Z}$, being an integer. We assume there no $\alpha \in \mathcal{X}$ with a finite orbit. Discrete dynamics will visit states with ever increasing $\mathbf{t}_\mu$ score. Given a finite set $D \subset \mathcal{X}$, with $D = \{\alpha_i\}_{i=1}^n$, its mutual information with the halting sequence is defined by $\mathbf{I}(D : \mathcal{H}) = \inf_{\overrightarrow{\alpha_1},\dots,\overrightarrow{\alpha_n}} \mathbf{I}(\langle \overrightarrow{\alpha_1}, \dots, \overrightarrow{\alpha_n} \rangle : \mathcal{H})$, which is the infimum over all encoded fast Cauchy sequences to members of D. Thus $\mathbf{I}(\Omega(\alpha, n) : \mathcal{H}) <^+ \mathbf{I}(\alpha : \mathcal{H}) + \mathbf{K}(n, f)$, using Definition 14.

Lemma 5. *Given computable measure space $(\mathcal{X}, \mu)$, there is a constant $c_{\mathcal{X},\mu}$, with universal uniform test $\mathbf{t}_\mu$, for a finite set $Z \subset \mathcal{X}$ with $n = \lceil \log |Z| \rceil$, $n < \log \max_{\alpha \in Z} \mathbf{t}_\mu(\alpha) + \mathbf{I}(\langle Z \rangle : \mathcal{H}) + O(\log \mathbf{I}(\langle Z \rangle : \mathcal{H}) + \mathbf{K}(n) + c_{X,\mu})$.*

Proof. Let $(\{0,1\}^\infty, \mu_\delta)$ be a binary representation that is isomorphic to computable measure space $(\mathcal{X}, \mu)$, with $\delta : (\{0,1\}^\infty, \mu_\delta) \to (\mathcal{X}, \mu)$. If $\max_{\alpha \in Z} \mathbf{t}_\mu(\alpha) = \infty$, then the lemma is proven. Thus $\delta^{-1}(\alpha)$ is defined for all $\alpha \in Z$. Let $W = \delta^{-1}(Z) \subset \{0,1\}^\infty$. By Theorem 7 applied to W and μ_δ, with $s = n - O(1) < \log \sum_{\alpha \in W} 2^{\mathbf{D}(\alpha|\mu_\delta)}$, gives

$$s < \max_{\alpha \in W} \mathbf{D}(\alpha|\mu_\delta) + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(s) + c_{\mathcal{X},\mu}).$$

Due to Definition 16,

$$n < \max_{\alpha \in W} \log \mathbf{t}_{\mu_\delta}(\alpha) + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

Since $(\{0,1\}^\infty, \mu_\delta)$ is isomorphic to $(\mathcal{X}, \mu)$,

$$n < \max_{\alpha \in Z} \log \mathbf{t}_\mu(\alpha) + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

Given any encoding of the fast Cauchy sequences of the members of Z, one can compute W with δ^{-1}, thus $\mathbf{K}(W|Z) = O(1)$, so

$$n < \max_{\alpha \in Z} \log \mathbf{t}_\mu(\alpha) + \mathbf{I}(Z : \mathcal{H}) + O(\log \mathbf{I}(Z : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

$\square$

Theorem 12. *Let $(\mathcal{X}, \mu)$ be a computable measure space and $\alpha \in \mathcal{X}$, with finite $\mathbf{I}(\alpha : \mathcal{H})$. For discrete time dynamics G^t, there is a c such that $\max_{\gamma \in G^{\{1,\dots,2^n\}}\alpha} \mathbf{t}_\mu(\gamma) > 2^{n - O(\mathbf{K}(n)) - c}$.*

Proof. Let $Z_n = G^{\{1,\dots,2^n\}}\alpha$. Lemma 5, applied to $(\mathcal{X}, \mu)$ and Z_n, results in $\gamma \in Z_n$ such that

$$n < \log \mathbf{t}_\mu(\gamma) + \mathbf{I}(Z_n : \mathcal{H}) + O(\log \mathbf{I}(Z_n : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu,\alpha}).$$

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

$$n < \log \mathbf{t}_\mu(\gamma) + \mathbf{I}(\alpha : \mathcal{H}) + O(\log \mathbf{I}(\alpha : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu,\alpha,G}).$$

The theorem is proven by noting $\mathbf{I}(\alpha : \mathcal{H}) < \infty$. $\square$

8   Coarse Grained Entropy

Coarse grained entropy was introduced in [Gac94] as an update to Boltzmann entropy. The goal was a parameter independent formulation of entropy. It was defined using cells. In this section we define coarse grained entropy with respect to open sets, leveraging the results of [HR09]. Let $\Pi(\cdot)$ be a set of disjoint uniformly enumerable open sets in the computable metric space $\mathcal{X}$.

Definition 17 (Algorithmic Coarse Grained Entropy). $\mathbf{H}_\mu(\Pi_i) = \mathbf{K}(i|\mu) + \log \mu(\Pi_i)$.

Coarse grained entropy is an excellent approximation of fine grained entropy, as shown by the following two results.

Proposition 1. *Let $(\mathcal{X}, \mu)$ be a computable measure space. If $\mu(\Pi_i)$ is uniformly computable and $\alpha \in \Pi_i$ then $\mathbf{H}_\mu(\alpha) <^+ \mathbf{H}_\mu(\Pi_i) + \mathbf{K}(\Pi)$.*

Proof.   Let $t(\alpha) = [\alpha \in \Pi_i] \mathbf{m}(\Pi_i)/\mu(\Pi_i)$. $t \in \mathcal{F}$ is lower semi-computable and $\int_{\mathcal{X}} t(\alpha) d\mu(\alpha) = \sum_i \int_{\Pi_i} (\mathbf{m}(\Pi_i)/\mu(\Pi_i)) d\mu(\alpha) = \sum_i \mathbf{m}(\Pi_i) \leq 1$. Thus $\mathbf{t}_\mu(\alpha) \overset{*}{>} t(\alpha)$.

Lemma 6. *For computable measure space $(\mathcal{X}, \mu)$, for lower computable function $f \in \mathcal{F}$, and enumerable open set U, $\int_U f d\mu$ is lower computable.*

Proof.   For a finite union of balls $V = \bigcup_{j=1}^n B_{i_j}$ and an enumerable open set $W = \bigcup_{j=1}^\infty B_{k_j}$ and a computable measure μ, the term $\mu(V \cap W)$ is lower computable. Due to Corollary 2, the term $\mu(\bigcup \{B : \exists_{s,t} \text{ such that } B \subseteq B_{i_s} \text{ and } B \subseteq B_{k_t}\}) = \mu(V \cap W)$ is lower computable.

The integral of a finite supremum of step functions over U is lower computable by induction. For the base case $\int_U f_{i,j} d\mu = q_j \mu(B_i \cap U)$ is lower computable by the above reasoning. For the inductive step

$$\int_U \sup\{f_{i_1,j_1}, \dots f_{i_k,j_k}\} d\mu = q_{j_m} \mu \left((B_{i_1} \cup \cdots \cup B_{i_k}) \cap U\right) + \int_U \sup\{f_{i_1,j_1'}, \dots f_{i_k,j_k'}\} d\mu,$$

where q_{j_m} is minimal among $\{q_{j_1}, \dots, q_{j_k}\}$ and $q_{j_1'} = q_{j_1} - q_{i_m}, \dots, q_{j_k'} = q_{j_k} - q_{i_k}$. The first term on the right is lower-computable and by the induction assumption, the last term on the right is lower-computable. $\square$

The following lemma is an update to the Stability Theorem 5 in [Gac94], using open sets instead of cells.

Lemma 7. *For computable measure space $(\mathcal{X}, \mu)$, $\mu\{\alpha \in \Pi_i : \mathbf{H}_\mu(\alpha) < \mathbf{H}_\mu(\Pi_i) - \mathbf{K}(\Pi) - m\} \overset{*}{<} 2^{-m} \mu(\Pi_i)$.*

Proof.   Let $f(i) = \int_{\Pi_i} \mathbf{t}_\mu(\alpha) d\mu(\alpha)$. By Lemma 6, the function $f(i)$ is lower computable, and $\sum_i f(i) \leq 1$. Thus $f(i) \overset{*}{<} \mathbf{m}(i)/\mathbf{m}(\Pi)$. So

$$\mu(\Pi_i)^{-1} \int_{\Pi_i} 2^{-\mathbf{H}_\mu(\alpha)} d\mu(\alpha) \overset{*}{<} 2^{-\mathbf{H}(\Pi_i)+\mathbf{K}(\Pi)}.$$

By Markov inequality,

$$\mu\{\alpha \in \Pi_i : \mathbf{H}_\mu(\alpha) < \mathbf{H}_\mu(\Pi_i) - \mathbf{K}(\Pi) - m\} \overset{*}{<} 2^{-m} \mu(\Pi_i).$$

Corollary 5. *For computable measure space $(\mathcal{X}, \mu)$, $\mu\{\alpha : \mathbf{H}_\mu(\alpha) < \log \mu(\mathcal{X}) - m\} \overset{*}{<} 2^{-m} \mu(\mathcal{X})$.*

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.

[Ǵ21] Peter Gács. Lecture notes on descriptional complexity and randomness. *CoRR*, abs/2105.04704, 2021.

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

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

[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.

[MZ55] D. Montgomery and L. Zippin. *Topological Transformation Groups*. Interscience tracts in pure and applied mathematics. Interscience Publishers, 1955.

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