On the Algorithmic Information Between Probabilities

Samuel Epstein*

April 21, 2022

Abstract

We extend algorithmic conservation inequalities to probability measures. The amount of self information of a probability measure cannot increase when submitted to randomized processing. This includes (potentially non-computable) measures over finite sequences, infinite sequences, and T_0, second countable topologies. One example is the convolution of signals over real numbers with probability kernels. Thus the smoothing of any signal due We show that given a quantum measurement, for an overwhelming majority of pure states, no meaningful information is produced.

Contents

1 Introduction

We prove conservation of probabilities over successively general spaces. This includeds finite sequences, infinite sequences, and T_0, second countable topologies. Conservation of probabilities over the case of finite and infinite sequences follow directly from conservation inequalities over random processing in individual sequences [Lev84, Lev74, Ver21, Gʹ21]. However there is benefit in revisiting these results in the context of manipulations of probabilities. This is particular true when the results are generalized to arbitrary topologies. Information between probability measures is achieved through a mapping from the general topology to infinite sequences and then applying the information function between individual sequences. We use the set of reals as an example and then

---

show conservation of information over computable convolutions. One example is the smoothing of a signal due to a Gaussian function, which results in degradation of self algorithmic information.

We also show how to lower bound information between probabilities over general spaces with information between probabilities over finite sequences using uniformly enumerable disjoint open sets. We provide an means to upper bound the probabilities between general spaces using computable non-probabilistic measure covers. We look at the average information between measures by using probability measures over spaces of measures.

The advantage to the topological approach used in this paper is that a very general topology can be used. The only assumption needed is that the topology needs to have the T_0 property and a computable countable basis. Typical requirements in computability theory such as compactness or metrizability are not needed. In addition this work deals with all measures, not just computable ones. This is analogous to how the mutual information term between infinite sequences is well defined over uncomputable inputs.

$\imath(x:y)$	Information between finite sequences
$\mathbf{i}(p:q)$	Information between probabilities over finite sequences
$\mathbf{I}(\alpha:\beta)$	Information between infinite sequences
$\mathbf{I}(P:Q)$	Information between probabilities over infinite sequences
$\mathcal{I}(\mathcal{P}:\mathcal{Q})$	Information between probabilities over general spaces

Figure 1: Information terms.

2 Probabilities over Sequences

The function $\mathbf{K}(x|y)$ is the conditional prefix free Kolmogorov complexity. The algorithmic probability is $\mathbf{m}(x|y)$. The mutual information of two finite sequences is $\imath(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$. $[A] = 1$ if the mathematical statement A is true. Otherwise $[A] = 0$. Let $\langle x \rangle = 1^{\|x\|}0x$ be a self delimiting encoding of x.

Definition 1 (Information, Discrete Semi-Measures)**.**
For semi-measures p and q over $\{0,1\}^*$, $\mathbf{i}(p:q) = \log \sum_{x,y \in \{0,1\}^*} 2^{\imath(x:y)} p(x) q(y)$.

The previous definition also applies to semi-measures over $\mathbb{N}$.

Lemma 1 ([Lev84])**.** *For partial recursive function $f : \{0,1\}^* \to \{0,1\}^*$, $\imath(f(x):y) <^+ \imath(x:y) + \mathbf{K}(f)$.*

Lemma 2. *Let ψ_a be an enumerable semi-measure, semi-computable relative to a.*
$\sum_c 2^{\imath(\langle a,c \rangle:b)} \psi_a(c) <^* 2^{\imath(a:b)}/\mathbf{m}(\psi)$.

Proof. This requires a slight modification of the proof of Proposition 2 in [Lev84], by requiring ψ to have a as auxilliary information. For completeness, we reproduce the proof. We need to show $\mathbf{m}(a,b)/(\mathbf{m}(a)\mathbf{m}(b)) >^* \sum_c (\mathbf{m}(a,b,c)/(\mathbf{m}(b)\mathbf{m}(a,c)))\mathbf{m}(\psi)\psi_a(c)$, or $\sum_c (\mathbf{m}(a,b,c)/\mathbf{m}(a,c))\mathbf{m}(c|a) <^* \mathbf{m}(a,b)/\mathbf{m}(a)$, since $\mathbf{m}(c|a) >^* \mathbf{m}(\psi)\psi_a(c)$. Rewrite it $\sum_c \mathbf{m}(c|a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) <^* \mathbf{m}(a,b)/\mathbf{m}(a)$ or $\sum_c \mathbf{m}(c|a)\mathbf{m}(a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) <^* \mathbf{m}(a,b)$. The latter is obvious since $\mathbf{m}(c|a)\mathbf{m}(a) <^* \mathbf{m}(a,c)$ and $\sum_c \mathbf{m}(a,b,c) <^* \mathbf{m}(a,b)$. □

2

Proposition 1. *For enumerable semi-measures p, q, $\mathbf{i}(p:q) <^+ \mathbf{I}(\langle p \rangle : \langle q \rangle)$.*

Proof. Let T be a Turing machine, that when given an encoding of a lower semi-computable probabiliy p and an input x, lower enumerates $p(x)$. $\mathbf{i}(p:q) = \log \sum_{x,y} 2^{\mathbf{I}(x:y)} T_p(x) T_q(y)$. Using Lemmas 2 and 1,

$$\mathbf{i}(p:q)$$
$$<^+ \log \sum_{x,y} 2^{\mathbf{I}(\langle x,p \rangle : y)} T_p(x) T_q(y)$$
$$<^+ \log \sum_{y} 2^{\mathbf{I}(\langle p \rangle : y)} q(y) / \mathbf{m}(T)$$
$$<^+ \log \sum_{y} 2^{\mathbf{I}(\langle p \rangle : \langle y,q \rangle)} q(y) / \mathbf{m}(T)$$
$$<^+ \log 2^{\mathbf{I}(\langle p \rangle : \langle q \rangle)} / \mathbf{m}(T)^2$$
$$<^+ \mathbf{I}(\langle p \rangle : \langle q \rangle).$$

$\square$

Example 1.

- *In general, a probability p, will have low $\mathbf{i}(p:p)$ if it has large measure on simple strings, or low measure on a large number of complex strings, or some combination of the two.*

- *If probability p is concentrated on a single string x, then $\mathbf{i}(p:p) = \mathbf{K}(x)$.*

- *The uniform distribution U_n over strings of length n has self information equal to (up to an additive constant) $\mathbf{K}(n)$. This is because due to Proposition 1, $\mathbf{i}(U_n : U_n) <^+ \mathbf{K}(n)$ and using Lemma 1, $\mathbf{i}(U_n : U_n) = \log \sum_{x,y \in \{0,1\}^n} 2^{\mathbf{I}(x:y)} 2^{-2n} >^+ \log \sum_{x,y \in \{0,1\}^n} 2^{\mathbf{I}(n:n)} 2^{-2n} >^+ \mathbf{K}(n)$.*

- *There are semi-measures that have infinite self information. Let α_n be the n bit prefix of a Martin Löf random sequence α and $n \in [2, \infty)$. Semi-measure $p(x) = [x = \alpha_n] n^{-2}$ has $\mathbf{i}(p:p) = \infty$.*

- *The universal semi-measure $\mathbf{m}$ has no self information.*

- *Another example is a probability p where for some $x \in \{0,1\}^n$, $p(xy) = 2^{-n}$ if $\|y\| = n$, and 0 otherwise. Using Proposition 1, $\mathbf{i}(p:p) <^+ \mathbf{K}(\langle p \rangle) <^+ \mathbf{K}(x)$. In addition, using Lemma 1, $\mathbf{i}(p:p) = \log \sum_{xy,xz} 2^{\mathbf{I}(xy:xz)} 2^{-2n} >^+ \log \sum_{xy,xz} 2^{\mathbf{I}(x:x)} 2^{-2n} =^+ \mathbf{K}(x)$. So the self information of p is equal to $\mathbf{K}(x)$.*

- *In general the information between probabilities can be arbitrarily smaller than the information between their encodings. For example, take an arbitrarily large random string x, and the probability $p(0) = 0.x$, and $p(1) = 1 - p(0)$. Thus $\mathbf{i}(p:p) \ll \mathbf{K}(p)$.*

- *There exists probabilities p and q such that $\mathbf{i}(p:p) \ll \mathbf{i}(p:q)$. Take a large random string y and let $p(0) = 0.5$ and $p(y) = 0.5$ and $q(y) = 1$.*

Definition 2 (Channel)**.** *A channel* $f : \{0,1\}^* \times \{0,1\}^* \to \mathbb{R}_{\geq 0}$ *has* $f(\cdot|x)$ *being a probability measure over* $\{0,1\}^*$ *for each* $x \in \{0,1\}^*$. *For probability* p, *channel* f, $fp(x) = \sum_z f(x|z)p(z)$.

Example 2 (Uniform Spread)**.** *An example channel* f *has* $f(\cdot|x)$ *be the uniform distribution over strings of length* $\|x\|$. *This is a cannonical spread function. Thus if* p *is a probability measure concentrated on a single string, then* $\mathbf{i}(p : p) = \mathbf{K}(x)$, *and* $\mathbf{i}(fp : fp) =^+ \mathbf{K}(\|x\|)$. *Thus* f *results in a decrease of self-information of* p. *This decrease of information occurs over all probabilities and computable channels.*

Theorem 1. *For probabilities* p *and* q *over* $\{0,1\}^*$, *computable channel* f, $\mathbf{i}(fp : q) <^+ \mathbf{i}(p : q)$.

Proof. Using Lemma 1, $\mathbf{i}(fp : q) = \log \sum_{x,y} 2^{\imath(x:y)} \sum_z f(x|z)p(z)q(y)$
$<^+ \log \sum_{y,z} q(y)p(z) \sum_x 2^{\imath((x,z):y)} f(x|z)$. Using Lemma 2, $\mathbf{i}(fp : q) <^+ \log \sum_{z,y} q(y)p(z)2^{\imath(z:y)}$
$=^+ \mathbf{i}(p : q)$. $\qquad\square$

Definition 3 (Information, Infinite Sequences, [Lev74])**.**
For $\alpha, \beta \in \{0,1\}^\infty$, $\mathbf{I}(\alpha : \beta) = \log \sum_{x,y \in \{0,1\}^*} \mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\imath(x:y)}$.

Proposition 2 (Folklore)**.** *For* $x, y \in \{0,1\}^*$, $\mathbf{I}(\langle x\rangle 0^\infty : \langle y\rangle 0^\infty) =^+ \imath(x : y)$.

Proof. $\mathbf{I}(\langle x\rangle 0^\infty : \langle y\rangle 0^\infty) > \log \mathbf{m}(x|\langle x\rangle 0^\infty)\mathbf{m}(y|\langle y\rangle 0^\infty)2^{\imath(x:y)} =^+ \imath(x : y)$. For the other direction, using Lemmas 1 and 2,

$$
\begin{aligned}
&\mathbf{I}(\langle x\rangle 0^\infty) : \langle y\rangle 0^\infty)) \\
&= \log \sum_{c,d} \mathbf{m}(c|\langle x\rangle 0^\infty)\mathbf{m}(d|\langle y\rangle 0^\infty)2^{\imath(c:d)} \\
&=^+ \log \sum_{c,d} \mathbf{m}(c|x)\mathbf{m}(d|y)2^{\imath(c:d)} \\
&<^+ \log \sum_{c,d} \mathbf{m}(c|x)\mathbf{m}(d|y)2^{\imath(\langle c,x\rangle:d)} \\
&<^+ \log \sum_d \mathbf{m}(d|y)2^{\imath(x:d)} \\
&<^+ \log \sum_d \mathbf{m}(d|y)2^{\imath(x,\langle d,y\rangle)} \\
&<^+ \log 2^{\imath(x:y)}.
\end{aligned}
$$

$\qquad\square$

Definition 4 (Information, Probabilities over Infinite Sequences)**.** *For probabilities* P, Q *over infinite sequences.* $\mathbf{I}(P : Q) = \log \int 2^{\mathbf{I}(\alpha:\beta)} dP(\alpha)dQ(\beta)$.

By Carathéodory's theorem, a measure over $\{0,1\}^\infty$ can be associated with a function $F : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, where $F(\emptyset) = 1$ and $F(x) = F(x0) + F(x1)$. A probability P is computable if its corresponding function F is computable. The encoding of a computable probability P, is equal to $\langle P\rangle = \langle F\rangle$. This term means every possible encoding of $\langle F\rangle$, over all F that computes P. Thus if we say $\imath(\langle P\rangle : y) > a$, then this means all encoding of P have at least a mutual information with y.

Example 3 (Information over Cylinders). *Let P be the measure defined by $F(y) = [x \sqsubseteq y]2^{-\|y\|+\|x\|}$. Thus P is the uniform measure over all sequences that start with x. Let $\mathcal{U}$ be the uniform measure over $\{0,1\}^\infty$. $\mathbf{I}(P : P) = \log \int 2^{\mathbf{I}(\alpha:\beta)}dP(\alpha)dP(\beta) = \int 2^{\mathbf{I}(x\alpha:x\beta)}d\mathcal{U}(\alpha)d\mathcal{U}(\beta)$. For all $x \in \{0,1\}^*$, $\alpha, \beta \in \{0,1\}^\infty$, $\mathbf{I}(x\alpha : x\beta) > \log \mathbf{m}(x|x\alpha)\mathbf{m}(x|x\beta)2^{\mathbf{I}(x:x)} >^+ \mathbf{K}(x) - 2\mathbf{K}(\|x\|)$. Thus $\mathbf{I}(P : P) >^+ \mathbf{K}(x) - 2\mathbf{K}(\|x\|)$. This inequality holds for any probability P whose support is restriced to the cylinder set $x\{0,1\}^\infty$.*

Theorem 2 ([Ver21, Lev74]). *$\mathbf{I}(A(\alpha) : \beta) <^+ \mathbf{I}(\alpha : \beta)$, where A is an algorithm and $A(\alpha)$ produces an infinite sequence.*

Theorem 3 ([Ver21, Lev74]). *$\int 2^{\mathbf{I}(\langle(\alpha,\gamma):\beta\rangle)}dP_\gamma(\alpha) < 2^{\mathbf{I}(\gamma:\beta)} + c_P$.*

Proposition 3. *For computable probabilities P and Q over $\{0,1\}^\infty$, $\mathbf{I}(P : Q) <^+ \mathbf{\i}(\langle P \rangle : \langle Q \rangle)$.*

Proof. Let T be a program that on input $\langle R \rangle$ for some computable probability R, and some string x, outputs $R(x)$ to arbitrary precision. Using Theorems 2 and 3,

$$2^{\mathbf{I}(P:Q)}$$

$$= \int 2^{\mathbf{I}(\alpha:\beta)}dT_{\langle P \rangle}(\alpha)dT_{\langle Q \rangle}(\beta)$$

$$<^* \int 2^{\mathbf{I}(\langle\alpha,P\rangle:\beta)}dT_{\langle P \rangle}(\alpha)dT_{\langle Q \rangle}(\beta)$$

$$<^* \int 2^{\mathbf{I}(\langle P\rangle:\beta)}dT_{\langle Q \rangle}(\beta)$$

$$<^* \int 2^{\mathbf{I}(\langle P\rangle:\langle\beta,Q\rangle)}dT_{\langle Q \rangle}(\beta)$$

$$<^* 2^{\mathbf{I}(\langle P\rangle:\langle Q\rangle)}.$$

The theoreom follows from Proposition 2. $\qquad\square$

Definition 5. *A random transition is of the form $\Lambda : \{0,1\}^\infty \times \{0,1\}^\infty \to \mathbb{R}_{\geq 0}$ where each $\Lambda(\cdot|\alpha)$ is a semi-measure over $\{0,1\}^\infty$ for each $\alpha \in \{0,1\}^\infty$ and for each measurable set B in the Borel algebra of $\{0,1\}^\infty$, $\Lambda(B|\cdot)$ is a measurable function. For random transition Λ, probability P, $\Lambda P(\alpha) = \int_{\{0,1\}^\infty} \Lambda(\alpha|\beta)dP(\beta)$. A random transition Λ is computable if the semi-measure $\Lambda(\cdot|\alpha)$ is uniformly computable given oracle access to α.*

Theorem 4. *For probabilities P, Q, computable random transition Λ, $\mathbf{I}(\Lambda P : Q) <^+ \mathbf{I}(P : Q)$.*

Proof.

$$2^{\mathbf{I}(\Lambda P:Q)}$$

$$= \int_\beta \int_\alpha 2^{\mathbf{I}(\alpha:\beta)}d\Lambda P(\alpha)dQ(\beta)$$

$$= \int_\beta \left(\int_\alpha \left(\int_\gamma 2^{\mathbf{I}(\alpha:\beta)}\Lambda(\alpha|\gamma)dP(\gamma) \right) d\alpha \right) dQ(\beta).$$

Using Theorems 2 and 3,

$$2^{\mathbf{I}(\Lambda P:Q)}$$

$$<^* \int_\beta \left(\int_\alpha \left(\int_\gamma 2^{\mathbf{I}(\langle\alpha,\gamma\rangle:\beta)} \Lambda(\alpha|\gamma) dP(\gamma) \right) d\alpha \right) dQ(\beta)$$

$$=^* \int_\beta \left(\int_\gamma \left(\int_\alpha 2^{\mathbf{I}(\langle\alpha,\gamma\rangle:\beta)} d\Lambda(\alpha|\gamma) \right) dP(\gamma) \right) dQ(\beta)$$

$$<^* \int_\beta \int_\gamma 2^{\mathbf{I}(\gamma:\beta)} dP(\gamma) dQ(\beta)$$

$$<^* 2^{\mathbf{I}(P:Q)}.$$

$\square$

3 Probabilities over General Spaces

We extend conservation to Borel measures over T_0, second countable topologies. We restrict our attention to such topologies which can be represented by a tuple $(X, \sigma, \mathcal{B}, \nu)$ where X is a set, σ is a countable basis for X where $\sigma = (\nu(1), \nu(2), \dots)$, and $\mathcal{B}$ is the Borel algebra formed from the topology. Because of the T_0 property, each point $x \in X$ is uniquely defined by the basis sets which contain it.

Definition 6. *We define the following measurable injection π from X to $\{0,1\}^\infty$. For $\alpha \in X$, let $\pi(\alpha)_i = [\alpha \in \sigma(i)]$. For $x \in \{0,1\}^*$, let $\sigma(x) = \pi^{-1}(x\{0,1\}^\infty)$ be the corresponding measurable set in X associated with $x \in \{0,1\}^*$.*

Let $\mathcal{R}$ be the smallest ring formed from σ. By Carathéodory's theorem, we can associate each pre-measure μ over $\mathcal{R}$ with a unique Borel measure $\mathcal{P}$ over $\mathcal{B}$, such that its restriction to $\mathcal{R}$ is equal to μ. Thus for each measure $\mathcal{P}$, we can associate a function $F : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, such that $F(x) = \mathcal{P}(\sigma(x))$. The probability measure $\overline{\mathcal{P}}$ over $\{0,1\}^\infty$ associated with F, is called the dual of $\mathcal{P}$. The probability measure $\mathcal{P}$ is computable if $\overline{\mathcal{P}}$ is computable. If $\mathcal{P}$ is computable, then $\langle \mathcal{P} \rangle = \langle \overline{\mathcal{P}} \rangle$.

Claim 1. *For lower semi-continuous $f : \{0,1\}^\infty \to \mathbb{R}_{\geq 0} \cup \infty$, probability measure $\mathcal{P}$,*
$\int_X f(\pi(\alpha)) d\mathcal{P}(\alpha) = \int_{\{0,1\}^\infty} f(\alpha) d\overline{\mathcal{P}}(\alpha).$

Proof. Let $g : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, such that $g(x) = \min_{x \sqsubset \alpha} f(\alpha)$ and $f(\alpha) = \sup_{x \sqsubset \alpha} g(x)$. By the definition of integration

$$\int_{\{0,1\}^\infty} f(\alpha) d\overline{\mathcal{P}}(\alpha) = \lim_{n\to\infty} \sum_{x\in\{0,1\}^n} g(x)\overline{\mathcal{P}}(x) = \lim_{n\to\infty} \sum_{x\in\{0,1\}^n} g(x)\mathcal{P}(\sigma(x)) = \int_X f(\pi(\alpha)) d\mathcal{P}(\alpha).$$

$\square$

Definition 7 (Information of Probabilities, General Topology)**.** *Given two measures $\mathcal{P}$ and $\mathcal{Q}$ over topology $(X, \sigma, \mathcal{B}, \nu)$, their mutual information is $\mathcal{I}(\mathcal{P} : \mathcal{Q}) = \log \int 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} d\mathcal{P}(\alpha) d\mathcal{Q}(\beta).$*

Proposition 4. *If $\mathcal{P}$ and $\mathcal{Q}$ are computable, then $\mathcal{I}(\mathcal{P}:\mathcal{Q}) <^+ {}_1(\langle\mathcal{P}\rangle : \langle\mathcal{Q}\rangle)$.*

Proof. By Claim 1,

$$\mathcal{I}(\mathcal{P}:\mathcal{Q}) = \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))}d\mathcal{P}(\alpha)d\mathcal{Q}(\beta) = \int_{\{0,1\}^\infty} \int_{\{0,1\}^\infty} 2^{\mathbf{I}(\alpha:\beta)}d\overline{\mathcal{P}}(\alpha)d\overline{\mathcal{Q}}(\beta).$$

The proposition then follows from Proposition 3.   □

A random transition $\Lambda : X \times X \to \mathbb{R}_{\geq 0}$, is a functiom such that $\Lambda(\cdot|\beta)$ is a semi-measure over X for each $\beta \in X$, and $\Lambda(B|\cdot)$ is a measurable function for each measurable set $B \in \mathcal{B}$. A random transition Λ has a dual $\overline{\Lambda}$ random transition in the Cantor space where measurable set $M \subseteq \{0,1\}^\infty$, $\beta \in \{0,1\}^\infty$, $\overline{\Lambda}(M|\beta) = \Lambda(\pi^{-1}(M)|\pi^{-1}(\beta))$. If $\pi^{-1}(\beta)$ doesn't exist, then $\overline{\Lambda}(M|\beta) = 0$. The set of all such β with no $\pi^{-1}(\beta)$ is a measurable set of $\{0,1\}^\infty$. A random transition is computable if its dual is computable.

Proposition 5. *For lower semi-continuous $f : \{0,1\}^\infty \to \mathbb{R}_{\geq 0} \cup \infty$, $\gamma \in X$,*
$\int_X f(\pi(\alpha))d\Lambda(\alpha|\gamma) = \int_{\{0,1\}^\infty} f(\alpha)d\overline{\Lambda}(\alpha|\pi(\lambda))$.

Proof. Let $g : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, such that $g(x) = \min_{x \sqsubset \alpha} f(\alpha)$ and $f(\alpha) = \sup_{x \sqsubset \alpha} g(x)$.

$$\int_X f(\pi(\alpha))d\Lambda(\alpha|\gamma)$$
$$= \lim_{n\to\infty} \sum_{x\in\{0,1\}^n} g(x)\Lambda(\sigma(x)|\gamma)$$
$$= \lim_{n\to\infty} \sum_{x\in\{0,1\}^n} g(x)\overline{\Lambda}(x|\pi(\gamma))$$
$$= \int_{\{0,1\}^\infty} f(\alpha)d\overline{\Lambda}(\alpha|\pi(\gamma)).$$

    □

Theorem 5. *For probabilities $\mathcal{P}$, $\mathcal{Q}$, computable random transition Λ, $\mathcal{I}(\Lambda\mathcal{P}:\mathcal{Q}) <^+ \mathcal{I}(\mathcal{P}:\mathcal{Q})$.*

Proof.

$$2^{\mathcal{I}(\Lambda\mathcal{P}:\mathcal{Q})} = \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))}d\Lambda\mathcal{P}(\alpha)d\mathcal{Q}(\beta)$$
$$= \int_X \left(\int_X \left(\int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))}\Lambda(\alpha|\gamma)d\mathcal{P}(\gamma) \right) d\alpha \right) d\mathcal{Q}(\beta)$$
$$= \int_X \left(\int_X \left(\int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))}d\Lambda(\alpha|\gamma) \right) d\mathcal{P}(\gamma) \right) d\mathcal{Q}(\beta)$$

Using Proposition 5,

$$2^{\mathcal{I}(\Lambda P:Q)} = \int_X \left(\int_X \left(\int_{\{0,1\}^\infty} 2^{\mathbf{I}(\alpha:\pi(\beta))}d\overline{\Lambda}(\alpha|\pi(\gamma)) \right) d\mathcal{P}(\gamma) \right) d\mathcal{Q}(\beta).$$

7

Using Theorem 2,

$$2^{\mathcal{I}(\Lambda\mathcal{P}:\mathcal{Q})} <^* \int_X \left(\int_X \left(\int_{\{0,1\}^\infty} 2^{(\langle\alpha,\pi(\gamma)\rangle:\pi(\beta))} d\overline{\Lambda}(\alpha|\pi(\gamma)) \right) d\mathcal{P}(\gamma) \right) d\mathcal{Q}(\beta)\beta.$$

Using Theorem 3,

$$2^{\mathcal{I}(\Lambda\mathcal{P}:\mathcal{Q})} <^* \int_X \int_X 2^{\mathbf{I}(\pi(\gamma):\pi(\beta))} d\mathcal{P}(\gamma) d\mathcal{Q}(\beta) =^* 2^{\mathcal{I}(\mathcal{P}:\mathcal{Q})}.$$

$\square$

Claim 2. *For probability measures p, q over finite sequences, infinite sequences, or general spaces, if p is computable,* $\mathbf{I}(p:q) <^+ \mathbf{K}(p)$.

Proof. We prove it for the finite sequence case, and the other cases follow similarly. By Lemma 2, $\mathbf{i}(p:q) = \log \sum_{x,y} 2^{\mathrm{i}(x:y)} p(x)q(y) <^+ \log \sum_y 2^{\mathrm{i}(\langle p\rangle:y)} q(y) <^+ \log 2^{\mathbf{K}(p)}$. $\square$

Example 4 (Convolutions on the real line). *One example topology to be used throughout this paper is the real line,* $(\mathbb{R}, \sigma, \mathcal{B}_{\mathbb{R}}, \nu)$*, where* σ *is the set of all intervals* $\{(a,b)\}$*,* $a, b \in \mathbb{Q}$*, and* ν *is some computable ordering of* σ*. For such topology, the Gaussian distribution* $\mathcal{N}(\mu, \sigma^2)$ *is computable, and thus by Theorem 4, has self information bounded by* $\mathcal{I}(\mathcal{N}(\mu, \sigma^2) : \mathcal{N}(\mu, \sigma^2)) <^+ \mathbf{K}(\mu, \sigma^2)$*. Similarly the self-information of parameterized distributions over* $\mathbb{R}$ *will be less than the complexity of their encoded parameters.*

A convolution of a probability $\mathcal{P}$ *over* $\mathbb{R}$ *with a probability kernel* $\mathcal{F}$ *produces a new probability* $(\mathcal{P} \star \mathcal{F})(x) = \int_{-\infty}^{\infty} \mathcal{P}(y)\mathcal{F}(x-y)dy$*. Convolution is a random transition, and it is computable if* $\mathcal{F} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ *is computable. The pdf of the sum of two random variables is the convolution of their respective pdfs.*

Corollary 1. *For probability* $\mathcal{P}$ *over* $\mathbb{R}$*, computable probability kernel* $\mathcal{F} : \mathbb{R} \to \mathbb{R}_{\geq 0}$*,* $\mathcal{I}(\mathcal{P} \star \mathcal{F} : \mathcal{P} \star \mathcal{F}) <^+ \mathcal{I}(\mathcal{P} : \mathcal{P})$.

Let $\mathcal{G} \sim \mathcal{N}(0, \sigma^2)$ be a Gaussian distribution over $\mathbb{R}$. Thus convolution of a signal (probability measure) $\mathcal{P}$ with $\mathcal{G}$ results in smoothing of $\mathcal{P}$, proportional to σ^2. By the above corollary, a smoothing of any signal (computable or not) will result in a decrease of self information.

4   Information Transfer

In this section, we show how information between probabilities on general spaces can be related to the information between probabilities of finite sequences through the use of disjoint, enumerable open sets.

Definition 8 (Information, Infinite Sequences, Extended). *For* $x, y \in \{0,1\}^*$*,* $\overline{\mathbf{m}}(x|y) = \sum\{2^{-\|p\|} : U_y(p) = x$ *and at most* $\|y\|$ *bits are read from the auxilliary tape*$\}$*. For* $x \in \{0,1\}^*$*,* $\alpha \in \{0,1\}^\infty$*,* $\overline{\mathbf{m}}(x|\alpha) = \mathbf{m}(x|\alpha)$*. For* $\alpha, \beta \in \{0,1\}^* \cup \{0,1\}^\infty$*,* $\overline{\mathbf{I}}(\alpha : \beta) = \log \sum_{x,y \in \{0,1\}^*} \overline{\mathbf{m}}(x|\alpha)\overline{\mathbf{m}}(y|\beta)2^{\mathrm{i}(x:y)}$*.*

If $x \sqsubseteq y$, $\overline{\mathbf{I}}(x : \beta) \leq \overline{\mathbf{I}}(y : \beta)$ and $\lim_{n\to\infty} \overline{\mathbf{I}}(\alpha[0..n] : \beta) = \overline{\mathbf{I}}(\alpha : \beta) = \mathbf{I}(\alpha : \beta)$.

Definition 9 (Prefix monotone function). $\sqsubseteq$-sup *is the supremum under the partial order of* $\sqsubseteq$ *on* $\{0,1\}^{*\infty}$. *Say function* $\nu : \{0,1\}^* \to \{0,1\}^*$ *has the property that for all* $p, q \in \{0,1\}^*$, $\nu(p) \sqsubseteq \nu(pq)$. *Then* $\overline{\nu} : \{0,1\}^{*\infty} \to \{0,1\}^{*\infty}$ *denotes the unique extension of* ν, *where* $\overline{\nu}(p) = \sqsubseteq$-$\sup \{\nu(p_{\le n}) : n \le \|p\|, n \in \mathbb{N}\}$ *for all* $p \in \{0,1\}^{*\infty}$. *These extensision are called prefix monotone functions.*

Claim 3. *For computable prefix-monotone function* A, $\alpha, \beta \in \{0,1\}^* \cup \{0,1\}^\infty$, $\overline{\mathbf{I}}(A(\alpha) : \beta) <^+ \overline{\mathbf{I}}(\alpha : \beta)$.

Definition 10. *A series of open sets* $\{\phi_i\}$ *is enumerable if* $\{(i,j) : \sigma(j) \subseteq \phi_i\}$ *is enumerable.*

Theorem 6. *For enumeration* $\{\phi_i\}$ *of disjoint open elements, if two probability measures* $\mathcal{P}$ *and* $\mathcal{Q}$ *have corresponding semi-measures* p *and* q *over* $\mathbb{N}$, *where* $p(i) \le \mathcal{P}(\phi_i)$ *and* $q(i) \le \mathcal{Q}(\phi_j)$, *then* $\mathcal{I}(\mathcal{P} : \mathcal{Q}) >^+ \mathbf{i}(p : q)$.

Proof. For $j \in \mathbb{N}$, let $\Omega_j = \{0,1\}^{j-1}1\{0,1\}^\infty$. For each i, let $O_i = \bigcup\{\Omega_j : \sigma(j) \subseteq \phi_i\}$. Note that if $i \ne j$, $O_i \cap O_j = \emptyset$. Let $A : \{0,1\}^* \cup \{0,1\}^\infty \to \{0,1\}^* \cup \{0,1\}^\infty$, be a computable prefix monotone function mapping each $\alpha \in O_i$ to $\langle i \rangle$ and mapping all other sequences to $\emptyset$. The function A is computable because each open set $\phi_i \subseteq X$ is uniformly enumerable. By Claims 1 and 3,

$$
2^{\mathcal{I}(\mathcal{P}:\mathcal{Q})}
$$

$$
= \int_{\{0,1\}^\infty} \int_{\{0,1\}^\infty} 2^{\overline{\mathbf{I}}(\alpha:\beta)} d\overline{\mathcal{P}}(\alpha)\overline{\mathcal{Q}}(\beta)
$$

$$
>^* \int_{\{0,1\}^\infty} \int_{\{0,1\}^\infty} 2^{\overline{\mathbf{I}}(A(\alpha):A(\beta))} d\overline{\mathcal{P}}(\alpha)\overline{\mathcal{Q}}(\beta)
$$

$$
=^* \sum_{i,j\in\mathbb{N}} \int_{O_j} \int_{O_i} 2^{\overline{\mathbf{I}}(\langle i\rangle:\langle j\rangle)} d\overline{\mathcal{P}}(\alpha)\overline{\mathcal{Q}}(\beta)
$$

$$
=^* \sum_{i,j\in\mathbb{N}} 2^{\overline{\mathbf{I}}(\langle i\rangle:\langle j\rangle)} d\overline{\mathcal{P}}(O_i) d\overline{\mathcal{Q}}(O_j)
$$

$$
=^* \sum_{i,j\in\mathbb{N}} 2^{\overline{\mathbf{I}}(\langle i\rangle:\langle j\rangle)} \mathcal{P}(\phi_i)\mathcal{Q}(\phi_j)
$$

$$
>^* \sum_{i,j} 2^{\mathbf{i}(i:j)} p(i)q(j)
$$

$$
=^* 2^{\mathbf{i}(p:q)}.
$$

$\square$

The above theorem is true for the case of infinite sequences. Let $\{\phi_i\}$, $\{\theta_i\}$ be enumerations of disjoint open sets on $\{0,1\}^\infty$, and P, Q, be probability measures on $\{0,1\}^\infty$. Then for semi measures $p(i) \le P(\phi_i)$ and $q(i) \le Q(\theta_i)$, we have $\mathbf{I}(P : Q) >^+ \mathbf{i}(p : q)$.

Example 5. *For any probability measure* $\mathcal{P}$ *over* $\mathbb{R}$, $\mathcal{I}(\mathcal{P} : \mathcal{P}) >^+ \mathbf{K}(n) + 2\log \mathcal{P}(\mathcal{U}(n, n+1))$, *where* $\mathcal{U}(n, n+1)$ *is the uniform distribution. This is due to the enumeration of* $\{(n, n+1)\}$.

Example 6 (Random Pulse). *Let $\alpha \in \{0,1\}^\infty$ be a Martin Löf random sequence, where $\mathbf{K}(\alpha[0..n]) >^+ n$. Using the real line topology of Example 4, let $\mathcal{N}_n = \mathcal{N}(0.\alpha, n^{-2})$ be a Gaussian distribution centered at a random point between 0 and 1 and with variance inversely proportional to n. We show that $\lim_{n\to\infty} \mathcal{I}(\mathcal{N}_n : \mathcal{N}_n) = \infty$. This limit is known as the (offseted) Dirac delta function.*

Let $\nu \in (0, 0.5)$ and $\{\phi_i^m\}$ be 2^m open intervals where $\phi_i^m = ((i-1)2^{-m}, i2^{-m})$, for $i \in [1, 2^m]$. Each i can be associated with a string x of length m, where $s(x) = i$. For each m, there is an n, such that $\mathcal{N}_n(\phi_{s(\alpha[m])}^m) > \nu$. Thus, using Theorem 6 (noting that the additive constant is on the order of the complexity of the enumeration) for each m, there is an n, where $\mathcal{I}(\mathcal{N}_n, \mathcal{N}_n) >^+ \mathbf{K}(s(\alpha[m])) + 2\log\nu - O(\mathbf{K}(\{\phi_i^m\})) >^+ m + 2\log\nu - O(\mathbf{K}(m))$. Thus $\lim_{n\to\infty} \mathcal{I}(\mathcal{N}_n : \mathcal{N}_n) = \infty$.

Example 7. *For rational σ^2, we show that for $n \in \mathbb{N}$, $\mathcal{I}(\mathcal{N}(n, \sigma^2) : \mathcal{N}(n, \sigma^2)) =^+ \mathbf{K}(n)$. Let $\mathcal{N}_n = \mathcal{N}(n, \sigma^2)$. Indeed, due to Proposition 4, $\mathcal{I}(\mathcal{N}_n : \mathcal{N}_n) <^+ \mathbf{K}(n, \sigma^2) =^+ \mathbf{K}(n)$. For the other direction let $\{\phi_i\}$ be an enumeration of open intervals, where $\phi_i = (i, i+1)$. Then $\mathcal{N}_n(\phi_n) > c_\sigma$, for some $c_\sigma \in \mathbb{R}_{\geq 0}$, dependent solely on σ^2. Due to Theorem 6 on $\mathcal{N}_n$ there is a semi-measure p over $\mathbb{N}$, where $p(n) = c_\sigma$, and $\mathcal{I}(\mathcal{N}_n : \mathcal{N}_n) >^+ \mathbf{i}(p : p) >^+ \mathbf{K}(n) + 2\log p(n) =^+ \mathbf{K}(n)$.*

5   Computable Covers

Given a topology $(X, \sigma, \mathcal{B}, \nu)$, a (not necessarily probability) measure ρ covers measure μ if $\rho(B) \geq \mu(B)$ for all measurable sets $B \in \mathcal{B}$.

Theorem 7. *If computable measures $\mathcal{M}$ and $\mathcal{R}$ cover probability measures $\mathcal{P}$, $\mathcal{Q}$, then $\mathcal{I}(\mathcal{P} : \mathcal{Q}) <^+ \imath(\langle\mathcal{M}\rangle : \langle\mathcal{R}\rangle) + \log\mathcal{M}(X)\mathcal{R}(X)$.*

Proof. We define computable probability measures, $m = \mathcal{M}/\mathcal{M}(X)$ and $r = \mathcal{R}/\mathcal{R}(X)$. Then

$$\mathcal{I}(\mathcal{P} : \mathcal{Q})$$
$$\leq \log \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} d\mathcal{M}(\alpha) d\mathcal{R}(\beta)$$
$$<^+ \log \mathcal{M}(X)\mathcal{R}(X) + \log \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} dm(\alpha) dr(\beta)$$
$$=^* \log \mathcal{M}(X)\mathcal{R}(X) + \mathcal{I}(m : r).$$

Let F_m and F_r be the sets of programs that compute m and r, respectively. Using Proposition 4, for all $f_m \in F_m$, $f_r \in F_r$.

$$\mathcal{I}(\mathcal{P} : \mathcal{Q}) <^* \log \mathcal{M}(X)\mathcal{R}(X) + \imath(\langle f_m \rangle : \langle f_r \rangle).$$

Let F_M and F_R be programs that compute $\mathcal{M}$ and $\mathcal{R}$ and minimize $\imath(\langle F_M \rangle : \langle F_R \rangle)$. Let $f'_m \in F_m$ and $f'_r \in F_R$ be programs that compute m and r by first computing $F_M(x)$ then dividing by $F_M(\emptyset)$, and similarly for F_R. Thus it must be that $\mathbf{K}(\langle f'_m \rangle | \langle F_M \rangle) = O(1)$ and similarly, $\mathbf{K}(\langle f'_r \rangle | \langle F_R \rangle) = O(1)$. Using Lemma 1,

$$\mathcal{I}(\mathcal{P} : \mathcal{Q}) <^+ \log \mathcal{M}(X)\mathcal{R}(X) + \imath(\langle m \rangle : \langle r \rangle)$$
$$<^+ \log \mathcal{M}(X)\mathcal{R}(X) + \imath(\langle f'_m \rangle : \langle f'_r \rangle)$$
$$<^+ \log \mathcal{M}(X)\mathcal{R}(X) + \imath(\langle F_M \rangle : \langle F_R \rangle)$$
$$<^+ \log \mathcal{M}(X)\mathcal{R}(X) + \imath(\langle\mathcal{M}\rangle : \langle\mathcal{R}\rangle).$$

$\square$

Corollary 2. *For semi measures p and q over $\{0,1\}^*$, if computable measures w and r over $\{0,1\}^*$ have $p \le w$ and $q \le r$, then $\mathbf{i}(p:q) <^+ \log w(\{0,1\}^*)r(\{0,1\}^*) + \iota(\langle w \rangle : \langle r \rangle) + \mathbf{K}(\lceil w(\{0,1\}^*) \rceil) + \mathbf{K}(\lceil r(\{0,1\}^*) \rceil)$.*

Example 8.

- *If probability $\mathcal{P}$ over $\mathbb{R}$ has support limited to (a,b) with $\hat{c} \ge \sup_{a \in \mathbb{R}} \mathcal{P}(a)$, then for $\hat{a} \le a < b \le \hat{b}$, $\mathcal{I}(\mathcal{P}:\mathcal{P}) <^+ 2 \log \hat{c}(\hat{b} - \hat{a}) + \mathbf{K}(\hat{a}, \hat{b}, \hat{c})$.*

- *If probability $\mathcal{P}$ over $\mathbb{R}$ is less than the weighted combination $\sum_{i=1}^{n} c_i \mathcal{Q}_i$ of parametric distributions $\mathcal{Q}_i$ (such as a Gaussian or a exponential distribution), then $\mathcal{I}(\mathcal{P}:\mathcal{P}) <^+ 2 \log \sum_{i}^{n} c_i + \mathbf{K}(\{c_i, \mathcal{Q}_i\}_{i=1}^{n})$.*

- *Let $\mathcal{U}(a,b)$ be the uniform measure over (a,b). If an infinite sequence $\alpha \in \{0,1\}^\infty$ is encoded into a semi-measure of the form $\mathcal{P}(x) = \sum_{n=2}^{\infty} \alpha_n n^{-2} \mathcal{U}(n, n+1)(x)$ then it is covered by $\mathcal{M}(x) = \sum_{n=2}^{\infty} n^{-2} \mathcal{U}(n, n+1)(x)$ which has negligible self-information. However if $\mathcal{P}$ is normalized to a probability measure, it can have arbitrarily high self-information. An example is $\alpha \in \{0,1\}^\infty$, $\alpha[i] = [i = n]$, where $n \in \mathbb{N}$ is a large random number.*

6 Averaged Information

The average information between probability measures is small, less than the complexity of the averaging. This is true in the discrete and continuous case. For the discrete case, an enumerable sequence of uniformly computable probability measures over a general space is a sequence of measures $\{\mu_i\}$ such that $\overline{\mu}_i(x\{0,1\}^\infty)$ is uniformly computable with respect to i, for all $x \in \{0,1\}^*$.

Theorem 8. *Let $\mathcal{E} = \{\mu_i\}$ be an enumerable sequence of uniformly computable probability measures over a general space. Let p be a computable measure over $\mathbb{N}$. Then $\mathbf{E}_{i,j \sim p}[2^{\mathcal{I}(\mu_i : \mu_j)}] = O(1)$.*

Proof. The measure $\mu = \sum_i p_i \mu_i$ is computable because for each ϵ, $\overline{\mu}(x)$ can be computed to within ϵ. Thus by Proposition 4, $\mathcal{I}(\mu : \mu) <^+ \mathbf{K}(\mu) <^+ \mathbf{K}(p, \mathcal{E}) = O(1)$. This implies $\mathbf{E}_{i,j \sim P}[2^{\mathcal{I}(\mu_i : \mu_j)}] = O(1)$, because

$$2^{\mathcal{I}(\mu:\mu)} = \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} d\mu(\alpha)d\mu(\beta) = \sum_{i,j} p_i p_j \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} d\mu_i(\alpha)d\mu_j(\beta) = \sum_{i,j} p_i p_j 2^{\mathcal{I}(\mu_i:\mu_j)}.$$

$\square$

Example 9. *Let $\mathcal{E}$ consist of 2^n Gaussians $\mathcal{N}(u,1)$ for $u \in \{1, .., 2^n\}$. Let p be the uniform distribution over the first 2^n natural numbers. Then by Theorem 8, $\mathbf{E}_{i,j \sim p}[2^{\mathcal{I}(\mathcal{N}(i,1):\mathcal{N}(j,1))}] = O(1)$.*

For the continuous case, we use a random transition between two different measure spaces. This differs from other approaches such as [HR09, Gź1] which defines a metric space of measures. We recall that an random transition from one topology $(X_M, \sigma_M, \mathcal{B}_M, \nu_M)$ to another $(X, \sigma, \mathcal{B}, \nu)$ is a measurable function $\Gamma : X_M \times X \to [0,1]$, such that $\Gamma(\cdot | x_M)$ is a probability measure for each $x_M \in X_M$ and $\Gamma(B | \cdot)$ is a measurable function for all measurable sets $B \in \mathcal{B}$. Thus topology X_M can be seen as a space of probability measures, with each point representing a probability measure. A probability measure $\mathfrak{M}$ over X_M produces an averaged (over X_M) probability over X, with $\mathcal{P}(\sigma(x)) = \int_{X_M} \Gamma(\sigma(x) | \alpha) d\mathfrak{M}(\alpha)$.

Theorem 9. *For computable $\mathcal{P}$, $\mathbf{E}_{\alpha,\beta\sim\mathfrak{M}}[2^{\mathcal{I}(\Gamma_\alpha:\Gamma_\beta)}] = O(1)$.*

Proof. For computable $\mathcal{P}$, by Proposition 4, $\mathcal{I}(\mathcal{P}:\mathcal{P}) <^+ \mathbf{K}(\mathcal{P}) = O(1)$. This implies $\mathbf{E}_{\alpha,\beta\sim\mathfrak{M}}[2^{\mathcal{I}(\Gamma_\alpha:\Gamma_\beta)}] = O(1)$ because

$$2^{\mathcal{I}(\mathcal{P}:\mathcal{P})}$$
$$= \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} d\mathcal{P}(\alpha) d\mathcal{P}(\beta)$$
$$= \int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} \left(\int_{X_M} \Gamma(\alpha|\xi) d\mathfrak{M}(\xi)\right) d\alpha \left(\int_{X_M} \Gamma(\beta|\zeta) d\mathfrak{M}(\zeta)\right) d\beta$$
$$= \int_{X_M} \int_{X_M} \left(\int_X \int_X 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} \Gamma(\alpha|\xi)\Gamma(\beta|\zeta) d\alpha d\beta\right) d\mathfrak{M}(\xi) d\mathfrak{M}(\zeta)$$
$$= \int_{X_M} \int_{X_M} 2^{\mathcal{I}(\Gamma_\xi:\Gamma_\zeta)} d\mathfrak{M}(\xi) d\mathfrak{M}(\zeta)$$
$$= \mathbf{E}_{\alpha,\beta\sim\mathfrak{M}} \left[2^{\mathcal{I}(\Gamma_\alpha:\Gamma_\beta)}\right].$$

$\square$

Example 10. *We let X be the real line and the random transition $\Gamma(\cdot|u)$ be a Gaussian $\mathcal{N}(u,1)$ with mean u. The space of measures is $X_M = [0,1]$, representing all Gaussians with 1 variance with means between 0 and 1. Then for $\mathfrak{M} \sim \mathcal{U}[0,1]$ being the uniform measure between 0 and 1, we have $\mathbf{E}_{a,b\sim\mathcal{U}[0,1]} \left[2^{\mathcal{I}(\mathcal{N}(a,1):\mathcal{N}(b,1))}\right] = O(1)$.*

7 Quantum Mechanics

Quantum information theory studies the limits of communicating through quantum channels. This section shows the limitations of the algorithmic content of [ure states and their measurements. Given a measurement apparatus E, there is only a tiny fraction of quantum pure states on which E's application produces coherent information. This is independent of the number of measurement outcomes of E.

In quantum mechanics, given a quantum state $|\psi\rangle$, a measurement, or POVM, E produces a probability measure $E|\psi\rangle$ over strings. This probability represents the classical information produced from the measurement. More formally, a POVM E is a finite set of positive definite matrices $\{E_k\}$ such that $\sum_k E_k = I$. For a given density matrix σ, a POVM E induces a probability measure over strings, where $E\sigma(k) = \text{Tr}E_k\sigma$. This can be seen as the probability of seeing measurement k given quantum state σ and measurement E.

Given a measurement E, for an overwhelming majority of pure quantum states $|\psi\rangle$, the probability produced will have no meaningful information, i.e. $\mathbf{i}(E|\psi\rangle : E|\psi\rangle)$ is negligible.

Theorem 10. *Let Λ be the uniform distribution on the unit sphere of an n qubit space. For the universal Turing machine relativized to an encoding of POVM E, $\int 2^{\mathbf{i}(E|\psi\rangle:E|\psi\rangle)}d\Lambda = O(1)$.*

The proof for this theorem can be found in [Eps21]. Its form is rather bizarre, in that it uses *upper semi-computable* tests, which is most likely the only place in the algorithmic information theory literature where this occurs.

Another interesting property of quantum mechanics is that the vast majority of quantum pure states themselves will have negligible algorithmic self information $\mathbf{I}_{\mathcal{Q}}(|\psi\rangle : |\psi\rangle)$. For this definition we use the information term introduced in [Eps19]. From this reference, we get the following result.

Theorem 11. *Let Λ be the uniform distribution on the unit sphere of an n qubit space.*
$\int 2^{\mathbf{I}_{\mathcal{Q}}(|\psi\rangle : |\psi\rangle)} d\Lambda = O(1)$.

It is an open question as to what spaces of measures will have a majority of its members have negligible self information.

References

[Eps19] Samuel Epstein. Algorithmic no-cloning theorem. *IEEE Transactions on Information Theory*, 65(9):5925–5930, 2019.

[Eps21] Samuel Epstein. On the algorithmic content of quantum measurements. *CoRR*, abs/2102.03905, 2021.

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

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

[Lev74] L. A. Levin. Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory. *Problemy Peredachi Informatsii*, 10(3):206–210, 1974.

[Lev84] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

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