On the Algorithmic Information between Probabilities

Samuel Epstein
samepst@jptheorygroup.org

April 19, 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 natural numbers, 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 then 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, GÍ3]. 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 space 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 algorithm information.

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 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.

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 $\mathbf{I}(x:y)=\mathbf{K}(x)+\mathbf{K}(y)-\mathbf{K}(x,y)$. $[A]=1$ 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 $\{0,1\}^*$, $\mathbf{I}(p:q)=\log\sum_{x,y}2^{\mathbf{I}(x:y)}p(x)q(y)$.

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

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

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

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 $\sum_c \mathbf{m}(a,b,c) <^* \mathbf{m}(a,b)$. □

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 computable probabiliy p and an input x, outputs $p(x)$. $\mathbf{I}(p:q)=\log\sum_{x,y}2^{\mathbf{I}(p:q)}T_p(x)T_q(y)$. Using Lemma 2 on p and q, $\mathbf{I}(p:q) <^+ \log\sum_y 2^{\mathbf{I}(\langle p\rangle:y)}q(y)/\mathbf{m}(T) <^+ \log 2^{\mathbf{I}(\langle p\rangle:\langle q\rangle)}/\mathbf{m}(T)^2 <^+ \mathbf{I}(\langle p\rangle:\langle q\rangle)$. □

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 $\mathbf{I}(p:p)=\mathbf{K}(x)$.*
- *For 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, $\mathbf{I}(fp:q) = \log\sum_{x,y} 2^{\mathbf{I}(x:y)}\sum_z f(x|z)p(z)q(y)$
$<^+ \log\sum_{y,z} q(y)p(z)\sum_x 2^{\mathbf{I}((x,z):y)}f(x|z)$. Using Lemma 2, $\mathbf{I}(fp:q) <^+ \log\sum_{z,y} q(y)p(z)2^{\mathbf{I}(z:y)}$
$=^+ \mathbf{I}(p:q)$. □

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

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

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

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

□

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 $\mathbf{I}(\langle P\rangle : y) > a$, 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])**.** $\int 2^{\mathbf{I}(\langle\alpha,\gamma\rangle:\beta)} dP_\gamma(\alpha) < 2^{\mathbf{I}(\gamma:\beta)} + c_P$.

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,

$$
\begin{aligned}
&2^{\mathbf{I}(P:Q)} \\
&= \int 2^{\mathbf{I}(\alpha:\beta)} dT_{\langle P\rangle}(\alpha) dT_{\langle Q\rangle}(\beta) \\
&<^* \int 2^{\mathbf{I}(\langle \alpha, P\rangle:\beta)} dT_{\langle P\rangle}(\alpha) dT_{\langle Q\rangle}(\beta) \\
&<^* \int 2^{\mathbf{I}(\langle P\rangle:\beta)} dT_{\langle Q\rangle}(\beta) \\
&<^* \int 2^{\mathbf{I}(\langle P\rangle:\langle \beta, Q\rangle)} dT_{\langle Q\rangle}(\beta) \\
&<^* 2^{\mathbf{I}(\langle P\rangle:\langle Q\rangle)}.
\end{aligned}
$$

The theoreom follows from Proposition 2. □

Definition 5. *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 over the Borel algebra of $\{0,1\}^\infty$, $\Lambda(B|\cdot)$ is a measurable function. For random transition Λ, probability P, $\Lambda P(\alpha) = \int \Lambda(\alpha|\beta) dP(\beta)$. A random transition Λ transistion 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.

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

Using Theorems 2 and 3,

$$
\begin{aligned}
&2^{\mathbf{I}(\Lambda P:Q)} \\
&<^* \int_\beta \left(\int_\alpha \left(\int_\gamma 2^{\mathbf{I}(\langle \alpha,\gamma\rangle:\beta)} \Lambda(\alpha|\gamma) dP(\gamma) \right) d\alpha \right) dQ(\beta) \\
&=^* \int_\beta \left(\int_\gamma \left(\int_\alpha 2^{\mathbf{I}(\langle \alpha,\gamma\rangle:\beta)} d\Lambda(\alpha|\gamma) \right) dP(\gamma) \right) dQ(\beta) \\
&<^* \int_\beta \int_\gamma 2^{\mathbf{I}(\gamma:\beta)} dP(\gamma) dQ(\beta) \\
&<^* 2^{\mathbf{I}(P:Q)}.
\end{aligned}
$$

□

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 set 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

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

□

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 $\log \int 2^{\mathbf{I}(\pi(\alpha):\pi(\beta))} d\mathcal{P}(\alpha) d\mathcal{Q}(\beta)$.*

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

Proof. By Claim 1,

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

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 $\overline{\Lambda}(M|\beta) = 0$. 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 $f(\alpha) = \sup_{x \sqsubset \alpha} g(x)$.

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

$\square$

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

Proof.

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

Using Proposition 5,

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

Using Theorem 2,

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

Using Theorem 3,

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

$\square$

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, $\mathbf{I}(p : q) = \log \sum_{x,y} 2^{\mathbf{I}(x:y)} p(x)q(y) <^+ \log \sum_y 2^{\mathbf{I}(\langle p\rangle:y)} q(y) <^+ \log 2^{\mathbf{K}(p)}$. $\square$

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 $\mathbf{I}(\mathcal{N}(\mu,\sigma^2):\mathcal{N}(\mu,\sigma^2)) <^+ \mathbf{K}(\mu,\sigma^2)$. 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 $\mathcal{F}:\mathbb{R} \to \mathbb{R}_{\geq 0}$, $\mathbf{I}(\mathcal{P} \star \mathcal{F} : \mathcal{P} \star \mathcal{F}) <^+ \mathbf{I}(\mathcal{P}:\mathcal{P})$.*

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 Computable Covers

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

Theorem 6. *If computable measures $\mathcal{M}$ and $\mathcal{R}$ cover probability measures $\mathcal{P}$, $\mathcal{Q}$, then $\mathbf{I}(\mathcal{P}:\mathcal{Q}) <^+ \mathbf{I}(\langle\mathcal{M}\rangle : \langle\mathcal{R}\rangle) + \log \mathcal{M}(X)\mathcal{R}(X)$.*

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

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

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

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

Let F_M and F_R be programs that compute $\mathcal{M}$ and $\mathcal{R}$ and minimize $\mathbf{I}(F_M : F_R)$. 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,

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

□

Corollary 2. *If computable measure μ covers probability measure $\mathcal{P}$, then for probability measure $\mathcal{Q}$, then $\mathbf{I}(\mathcal{P}:\mathcal{Q}) <^+ \mathbf{K}(\mu) + \log\mu(X)$.*

Corollary 3. *For semi measures p and q over $\{0,1\}^*$, if computable measures w and r over $\{0,1\}^*$ have $p \leq w$ and $q \leq r$, then $\mathbf{I}(p:q) <^+ \log w(\{0,1\}^*)r(\{0,1\}^*) + \mathbf{I}(\langle w\rangle : \langle r\rangle) + \mathbf{K}(\lceil w(\{0,1\}^*)\rceil) + \mathbf{K}(\lceil r(\{0,1\}^*)\rceil)$.*

Example 5.

- *If probability $\mathcal{P}$ over $\mathbb{R}$ has support limited to (a,b) with $\hat{c} \geq \sup_{a\in\mathbb{R}}\mathcal{P}(a)$, then for $\hat{a} \leq a < b \leq \hat{b}$, $\mathbf{I}(\mathcal{P}:\mathcal{P}) <^+ \log\hat{c}(\hat{b}-\hat{a}) + \mathbf{K}(\hat{a},\hat{b},\hat{c})$.*
- *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 Q_i (such as Gaussian or exponential distribution), then $\mathbf{I}(\mathcal{P}:\mathcal{P}) <^+ \log\sum_i^n c_i + \mathbf{K}(\{c_i,\mathcal{Q}_i\}_{i=1}^n)$.*
- *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.*

5 Averaged Information

The averged 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 7. *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 $\mathbf{E}_{i,j\sim p}[2^{\mathbf{I}(\mu_i:\mu_j)}] = O(1)$.*

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, $\mathbf{I}(\mu:\mu) <^+ \mathbf{K}(\mu) <^+ \mathbf{K}(p,\mathcal{E}) = O(1)$. This implies $\mathbf{E}_{i,j\sim P}[2^{\mathbf{I}(\mu_i:\mu_j)}] = O(1)$, because

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

□

Example 6. *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 7, $\mathbf{E}_{i,j\sim p}[2^{\mathbf{I}(\mathcal{N}(i,1):\mathcal{N}(j,1))}] = O(1)$.*

For the continuous case, we use random transition between two different measure spaces. This differs from other approaches such as [? GÍ3] 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 8. *For computable $\mathcal{P}$, $\mathbf{E}_{\alpha,\beta\sim\mathfrak{M}}[2^{\mathbf{I}(\Gamma_\alpha:\Gamma_\beta)}] = O(1)$.*

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

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

□

Example 7. *We let X be the real line and the random transition be $\Gamma(\cdot|u)$ 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 $\mathbf{E}_{a,b\sim\mathcal{U}[0,1]}\left[2^{\mathbf{I}(\mathcal{N}(a,1):\mathcal{N}(b,1))}\right] = O(1)$.*

6 Quantum Measurements

Quantum information theory studies the limits of communicating through quantum channels. This section shows the limitations of the algorithmic content of measurements of pure quantum states. 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 $E\sigma(k) = \mathrm{Tr}E_k\sigma$. 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 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 9. *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 [Eps]. 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.

References

[Eps] Samuel Epstein. On the algorithmic content of quantum measurements.

[GÍ3] P. Gács. Lecture notes on descriptional complexity and randomness, 2013.

[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.