AIT Blog

Conservation of Information

Samuel Epstein[*]

November 8, 2022

In [Lee06], four pillars of Kolmogorov Complexity were stated:

1. Coding Theorem.
2. Incompressibility.
3. Language Compression.
4. Symmetry of Information.

In my opinion, the fifth pillar should be conservation of information, which states that deterministic or randomized processing cannot increase target information. Thus for some partial computable function, A, and information term $\mathbf{I}$ over finite or infinite sequences, the deterministic non growth law is

$$\mathbf{I}(A(a):b) <^+ \mathbf{I}(a:b).$$

Given a probability γ over finite or infinite sequences computed by program p, the randomized non-growth law is

$$\mathbf{E}_{a\sim\gamma}[2^{\mathbf{I}((a,p):b)}] \overset{*}{<} 2^{\mathbf{I}(p:b)}.$$

Note that there several other formulations of randomized non-growth laws, some which use tests. These inequalities have been proven for several different information measures and it proves the impossibility of exotic situations such as accurately guessing bits of the halting sequence. Conservation of information is an essential component of derandomization, in the sense of [Eps22a].

In this blog I'll introduce a new information measure and show that this information term can be used to produce better bounds of the main result in [Lev13]. Before hand, I'll go over known information non growth laws. This is not an exhaustice list; there are several notions of information which I do not go over.

Symmetric Information over Strings

Symmetric information is defined for $x, y \in \{0,1\}^*$, is $\mathbf{I}(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$, where $\mathbf{K}$ is the prefix-free Kolmogorov complexity. From [Lev84], one has the following non-growth laws.

[*] JP Theory Group. samepst@jptheorygroup.org

Theorem 1

- *For partial computable* $f : \{0,1\}^* \to \{0,1\}^*$, $\mathbf{I}(f(a) : b) <^+ \mathbf{I}(a : b) + \mathbf{K}(f)$.
- *For probability* q *over* $\{0,1\}^*$ *computed by program* p, $\mathbf{E}_{a\sim q}[2^{\mathbf{I}((a,p):b)}] \overset{*}{<} 2^{\mathbf{I}(p:b)}$.

Mutual Information with the Halting Sequence

The amount of information that $x \in \{0,1\}^*$ has with the halting sequence $\mathcal{H} \in \{0,1\}^\infty$ is defined to be $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$. From [Eps22c, Eps22b], the non-growth laws are as follows.

Theorem 2

- *For partial computable* $f : \{0,1\}^* \to \{0,1\}^*$, $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f)$.
- *For probability* q *over* $\{0,1\}^*$ *computed by program* p, $\mathbf{E}_{a\sim q}[2^{\mathbf{I}((a,p);\mathcal{H})}] <^+ 2^{\mathbf{I}(a;\mathcal{H})}$.

Information over Infinite Sequences

The information between sequences $\alpha, \beta \in \{0,1\}^\infty$ is $\mathbf{I}(\alpha : \beta) = \log \sum_{x,y\in\{0,1\}^*} \mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\mathbf{I}(x:y)}$, where $\mathbf{m}$ is the algorithmic probability [Lev74]. The non-growth laws (and later explicit proof) are due to [Lev74, Ver21].

Theorem 3

- *For partial computable* $f : \{0,1\}^\infty \to \{0,1\}^\infty$, $\mathbf{I}(f(\alpha) : \beta) <^+ \mathbf{I}(\alpha : \beta) + \mathbf{K}(f)$.
- *Let* $\gamma \in \{0,1\}^\infty$ *compute* $\mu(x\{0,1\}^\infty)$ *for some probability* μ *over* $\{0,1\}^\infty$, *over all* $x \in \{0,1\}^*$. *Then* $\mathbf{E}_{\alpha\sim\mu}\left[2^{\mathbf{I}((\alpha,\gamma):\beta)}\right] \overset{*}{<} 2^{\mathbf{I}(\gamma:\beta)}$.

New Information Term

This new term $\mathbf{I}(\alpha;\mathcal{H})$, measures the amount of information between a finite or infinite sequences $\alpha \in \{0,1\}^* \cup \{0,1\}^\infty$ and the halting sequence $\mathcal{H}$. This term is larger than the information term between infinite sequences defined earlier, I will show how it can be used achieve better bounds in theorems. We will prove information non-growth for deterministic processing.

A function $Q : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ is a semi measure if $Q(\emptyset) \leq 1$ and $Q(x) \geq Q(x0) + Q(x1)$. We use $\mathbf{M}$ to denote a majorant (up to a multiplicative constant) lower semi-computable semi measure. For a prefix free set $D \subset \{0,1\}^*$, we have that $Q(D) = \sum_{x\in D} Q(x)$. Similarly, for an open set $S \subseteq \{0,1\}^\infty$, $Q(S) = Q(\{x : \Gamma_x \text{ is a maximal interval in } S\})$.

$\sqsubseteq$-sup is the supremum under the partial order of $\sqsubseteq$ on $\{0,1\}^* \cup \{0,1\}^\infty$. A function $\nu : \{0,1\}^* \to \{0,1\}^*$ is prefix-monotone iff 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 $\overline{\nu}(p) = \sqsubseteq\text{-sup}\,\{\nu(p_{\leq n}) : n \leq \|p\|, n \in \mathbb{N}\}$ for all $p \in \{0,1\}^{*\infty}$. The set of all extensions $\overline{\nu}$, of prefix-monotone functions ν that are computable relative to $\alpha \in \{0,1\}^\infty$, is $\mathcal{D}^\alpha$. $\mathcal{D} = \mathcal{D}^\emptyset$. For $x \in \{0,1\}^*$, $\xi \in \mathcal{D}^\alpha$, $\xi^{-1}(x) = \{y : y \in \{0,1\}^*, x \sqsubseteq \xi(y), x \not\sqsubseteq \xi(y^-)\}$. For semi measure Q, $\xi \in \mathcal{D}^\alpha$, let $\xi Q(x) = Q(\xi^{-1}(x))$.

Let $(x0)^- = (x1)^- = x$. For semi measure Q, we say that $t : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ is a Q test if for $x, y \in \{0,1\}^*$, $t(x) \leq t(xy)$, and for each $n \in \mathbb{N}$ where $D_{t,n} = \{x : t(x) > 2^n, t(x^-) \leq 2^n\}$, $\sum_{x \in D_{t,n}} Q(x) < 2^{-n}$. The domain of tests are extended to infinite sequences $\alpha \in \{0,1\}^\infty$ by $t(\alpha) = \sup\{t(x) : x \sqsubset \alpha, x \in \{0,1\}^*\}$. For a group of tests T, we say $t \in T$ is majorant if for all $g \in T$, there exists $c \in \mathbb{R}_{>0}$ such that for all $x \in \{0,1\}^*$, $t(x) > cg(x)$.

Theorem 4 *For computable semi measure Q, there exists a majorant lower semicomputable Q test.*

This can be proved by enumerating all lower semicomputable Q tests and then summing them up in the standard way. Let $\mathbf{h}$ be a majorant, lower semicomputable relative to $\mathcal{H}$, $\mathbf{M}$ test.

Definition 1 *The term $\mathbf{I}(\alpha;\mathcal{H}) = \log \mathbf{h}(\alpha)$ represents the information $\mathcal{H}$ has about α.*

Note that for any r.e., relative to $\mathcal{H}$, a $\mathbf{M}$ test, t, has $\log t(\alpha) <^+ \mathbf{I}(\alpha;\mathcal{H}) + \mathbf{K}(t/\mathcal{H})$. For $A \in \mathcal{D}^{\mathcal{H}}$, let $\mathbf{h}_A$ be a majorant, lower semicomputable relative to $\mathcal{H}$, $A\mathbf{M}$ test.

Theorem 5 *For $A \in \mathcal{D}^{\mathcal{H}}$, $\mathbf{h}_A(A\cdot)$ is a lower semicomputable, relative to $\mathcal{H}$, $\mathbf{M}$ test.*

Proof. Otherwise there exists an $n \in \mathbb{N}$ such that $\sum_{x \in D_{\mathbf{h}_A(A\cdot),n}} \mathbf{M}(x) \geq 2^{-n}$. So for each $x \in D_{\mathbf{h}_A(A\cdot),n}$, $\mathbf{h}_A(Ax) > 2^n$ and $\mathbf{h}_A(Ax^-) \leq 2^n$. So

$$\begin{aligned}
&\sum_{x \in D_{\mathbf{h}_A,n}} A\mathbf{M}(x) \\
=&\sum_{x \in D_{\mathbf{h}_A,n}} \mathbf{M}(A^{-1}x) \\
=&\sum_{x \in D_{\mathbf{h}_A,n}} \sum_{x \subseteq Ay, x \not\subseteq Ay^-} \mathbf{M}(y) \\
=&\sum_{y : \mathbf{h}_A(Ay) > 2^n, \mathbf{h}_A(Ay^-) \leq 2^n} \mathbf{M}(y) \\
=&\sum_{x \in D_{\mathbf{h}_A(A\cdot),n}} \mathbf{M}(x) \\
\geq& 2^{-n},
\end{aligned}$$

causing a contradiction, because $\mathbf{h}_A$ is a $A\mathbf{M}$ test.

Corollary 1 *For $A \in \mathcal{D}^{\mathcal{H}}$, $\mathbf{h}_A(A\alpha) <^+ \mathbf{h}(\alpha) + \mathbf{K}(A|\mathcal{H})$.*

Theorem 6 (Information Conservation) *For $A \in \mathcal{D}$ and $\alpha \in \{0,1\}^\infty$, $\mathbf{I}(A\alpha;\mathcal{H}) <^+ \mathbf{I}(\alpha;\mathcal{H}) + 3\mathbf{K}(A)$.*

Proof. Since $\mathbf{M} \overset{*}{>} \mathbf{m}(A)A\mathbf{M}$, $O(1)\mathbf{m}(A)\mathbf{h}(\alpha)$ is a lower semicomputable, relative to $\mathcal{H}$, $A\mathbf{M}$ test. So $\log 2^{-\mathbf{K}(A)}\mathbf{h}(\alpha) <^+ \log \mathbf{h}_A(\alpha) + \mathbf{K}(A)$. Putting this inequality and Corollary 1 together results in $\mathbf{I}(A\alpha;\mathcal{H}) <^+ \mathbf{I}(\alpha;\mathcal{H}) + 3\mathbf{K}(A)$.

New Bounds On Universal Partial Predicate Theorem

A partial predicate p is a finite or infinite set of pairs (x, b) consisting of indices $x \in \{0,1\}^*$ and bits $b \in \{0,1\}$. If $(x,b) \in p$, then we say $p(x) = b$, otherwise $p(x)$ is undefined. Let $\mathrm{Enc}(p) \in \{0,1\}^\infty$ be some standard encoding of partial predicate p. We use u to represent a universal partial recursive predicate, where for each partial recursive predicate p, there exists $z \in B^*$ such that $u(zx) = p(x)$, for all $x \in B^*$. The following theorem updates Theorem 1 from [Lev13], using the new information term $\mathbf{I}(\alpha;\mathcal{H})$, and achieves better bounds. The motivation for this theorem can also be found in [Lev13].

Theorem 7 *Let r be a partial predicate that on $\{0,1\}^n$ is a total extension of u.*
Then $\mathbf{I}(\mathrm{Enc}(r);\mathcal{H}) >^+ n$.

This theorem can be used to produce better bounds to Corollary 1 in [Lev13]. In a footnote of [Lev13] it is stated this bound can also be achieved by using the information term in [Lev84].

Corollary 2 *For randomized algorithm A, the probability that A computes on $\{0,1\}^n$ a total extension of u is at most $O(2^{-n})$.*

References

[Eps22a] S. Epstein. 22 examples of solution compression via derandomization. *CoRR*, abs/2208.11562, 2022.

[Eps22b] S. Epstein. The Kolmogorov Birthday Paradox. *CoRR*, abs/2208.11237, 2022.

[Eps22c] S. Epstein. The Outlier Theorem Revisited. *CoRR*, abs/2203.08733, 2022.

[Lee06] T. Lee. *Kolmogorov complexity and formula lower bounds*. PhD thesis, Amsterdam, 2006.

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

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

[Lev13] L. A. Levin. Forbidden information. *J. ACM*, 60(2), 2013.

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