Enumerations including laconic enumerators[1]

Jason Teutsch

National University of Singapore

June 23, 2015

[1] Joint work with Sanjay Jain

Approximating complexity is impossible

Let $C(x)$ denote Kolmogorov complexity.

Definition

A binary string x is called *random* if $C(x) \geq |x|$.

Sample results

1. No algorithm enumerates more than finitely many random strings.
2. No unbounded, computable function is a lower bound for Kolmogorov complexity (Zvonkin, Levin).
3. Any algorithm mapping a string to a list of values containing its Kolmogorov complexity must, for all but finitely many lengths n, include in the list for some string of length n at least a fixed fraction of the lengths below $n + O(1)$ (Beigel, Buhrman, Fejer, Fortnow, Grabowski, Longpré, Muchnik, Stephan, Torenvliet).

Approximating complexity is impossible II

Definition

A set A is (m, k)-*recursive* if there exists a computable function f such that for any $x_1 < \ldots < x_k$, the k-bit vector $f(x_1, \ldots, x_k)$ agrees with the characteristic vector $A(x_1, \ldots, x_k)$ in at least m places.

Champions Lemma (Teutsch, Zimand)

Let M be a set of binary strings, let k be a positive integer. Suppose that M is $(1, k)$-recursive. Then for all non-empty finite sets A satisfying

$$|M \cap A| \geq \left(1 - \frac{1}{(k+1)!}\right) \cdot |A|,$$

there exists $x \in M \cap A$ with $C(x) \leq C(A) + 2 \log k + O(1)$.

Theorem (Teutsch, Zimand 2014)

The set of Kolmogorov random strings is not $(1, k)$-recursive for any k.

List approximations for shortest descriptions

Unlike Kolmogorov complexity, shortest strings admit computable list approximations.

Theorem (Bauwens, Makhlin, Vereshchagin, Zimand 2013)

There exists an algorithm which maps each binary string x to an $O(|x|^2)$-size list containing a length $C(x) + O(1)$ description for x.

Theorem (Teutsch 2013)

There exists a polynomial-time algorithm which maps each binary string x to a $\mathrm{poly}(|x|)$-size list containing a length $C(x) + O(1)$ description for x.

Theorem (Bauwens, Zimand 2014)

There exists a probabilistic algorithm which, for any positive $\delta < 1$, maps each binary string x to a $|x|$-size list which contains a length $C(x) + O[\log(|x|/\delta)]$ description for x with probability at least $1 - \delta$.

Enumerators

Can enumerations help us to achieve better list approximations?

Definition

An *enumerator* is an algorithm which takes an integer input and, over time, enumerates a list of integers.

- For any enumerator f, $f(e)$ denotes the set of all elements which f eventually enumerates on input e.
- Let C_A denote Kolmogorov complexity with respect to a machine A.

Definition

A machine U is called *universal* if for any machine M, there exists a constant d such that for all x, $C_U(x) \leq C_M(x) + d$.

Note that the definition of universal does not guarantee an effective translation from M to U.

Enumerating shortest descriptions for strings

A trivial enumeration construction beats the optimal quadratic bound for computable list approximations for shortest descriptions.

Proposition

For any universal machine U, there exists an enumerator f such that $f(x)$ enumerates a list of size at most $|x| + O(1)$ containing a U-shortest description for x.

Proof.

$f(x)$ enumerates the first string at each length less than or equal to $|x| + O(1)$ which computes x. □

Question

What if we want not just the shortest but the *lexicographically least* description for x?

Enumerating minimal indices for strings

Definition

Let $\min_U(x)$ denote the lexicographically least program p such that $U(p) = x$.

Theorem (Jain, Teutsch)

For every universal machine U, there exists an enumerator f such that for all strings x, $|f(x)| = O(|x|)$ and $\min_U(x) \in f(x)$.

Our algorithm for f is simple:

Enumerate (up to) constant many programs for x at each length.

How large must the constant be?

Proof of Theorem

For every universal machine U, there exists an enumerator f such that for all strings x, $|f(x)| = O(|x|)$ and $\min_U(x) \in f(x)$.

Let $T_{b,n}$ be the set of all x such that $U(q) = x$ for at least 2^b many different values q of length n.

- Cleverly define a machine witnessing that for some c and all b, $C_U(x) \leq n - b + 2\log b + c$ for all $x \in T_{b,n}$.
- Fix b such that $b > 2\log b + c$. Then for all $x \in T_{b,n}$, there exists a program of length less than n which computes x.

Let a be a constant such that for each string x there exists a program p of size at most $|x| + a$ such that $U(p) = x$. Define $f(x)$ as follows: for each length $n \leq |x| + a$, output the first 2^b U-programs of length n which compute x. It follows from the definition of $T_{b,n}$ that either:

- $f(x)$ enumerates all the U-programs of length n which compute x, or
- there exists a program of length less than n which computes x.

By induction, $\min_U(x) \in f(x)$ and $|f(x)| \leq 2^b \cdot (|x| + a) = O(|x|)$.

Proof of Theorem

For every universal machine U, there exists an enumerator f such that for all strings x, $|f(x)| = O(|x|)$ and $\min_U(x) \in f(x)$.

Let $T_{b,n}$ be the set of all x such that $U(q) = x$ for at least 2^b many different values q of length n.

- Cleverly define a machine witnessing that for some c and all b, $C_U(x) \leq n - b + 2\log b + c$ for all $x \in T_{b,n}$.
- Fix b such that $b > 2\log b + c$. Then for all $x \in T_{b,n}$, there exists a program of length less than n which computes x.

Let a be a constant such that for each string x there exists a program p of size at most $|x| + a$ such that $U(p) = x$. Define $f(x)$ as follows: for each length $n \leq |x| + a$, output the first 2^b U-programs of length n which compute x. It follows from the definition of $T_{b,n}$ that either:

- $f(x)$ enumerates all the U-programs of length n which compute x, or
- there exists a program of length less than n which computes x.

By induction, $\min_U(x) \in f(x)$ and $|f(x)| \leq 2^b \cdot (|x| + a) = O(|x|)$.

Linear size is optimal

The previous theorem is optimal via an observation by Bauwens, Mahklin, Vereshchagin, and Zimand.

Improved Gács's Theorem (Bauwens, Shen)

For any universal machine U, there exist infinitely many strings x such that $C_U[C_U(x) \mid x] \geq \log |x| - O(1)$.

Now for any f satisfying the theorem in question,

> *"For every universal machine U, there exists an enumerator f such that for all strings x, $|f(x)| = O(|x|)$ and $\min_U(x) \in f(x)$,"*

we have for infinitely many strings x,

$$\log |x| - O(1) \leq C_U[C_U(x) \mid x] \leq \log |f(x)| + O(1),$$

whence $|f(x)| = \Omega(|x|)$.

Programming langauges

We turn our attention from strings to functions.

Question

Can enumerators help us list-approximate minimal indices for partial-computable functions?

Definition

A *numbering* φ is an effective enumeration of partial-computable functions, denoted $\varphi_0, \varphi_1, \varphi_2, \ldots$.

- We say φ is a *Gödel* numbering if for every numbering ψ, there exists a computable function t such that $\varphi_{t(e)} = \psi_e$ for all e.
- If in addition t is bounded by a linear function, we say φ is a *Kolmogorov* numbering.
- The function t is called a *translator* from ψ to φ.

Indices from any programming language can be effectively translated into a Kolmogorov numbering with $O(1)$ blow-up in program size.

Computable lists for Kolmogorov numberings

Definition

For any numbering φ, let $\min_\varphi(e)$ denote the least index j such that $\varphi_j = \varphi_e$.

At CCR 2014, I posed the following problem:

Question

Does there exist a Kolmogorov numbering φ with a computable list that for every e contains $\min_\varphi(e)$ and has size $O(\log^2 e)$?

The answer is a resounding "no."

Theorem (Vereshchagin, 2014)

For every Kolmgorov numbering φ and any computable function f which maps each index e to a list containing $\min_\varphi(e)$, we have $|f(e)| = \Omega(e)$ for infinitely many e.

Enumerators for Kolmogorov numberings

We extend Vereshchagin's negative result in two ways:

- from computable list-approximations to enumerators, and
- from minimal to nearly-minimal indices.

Theorem (Jain, Teutsch)

For any Kolmogorov numbering φ, the following two statements hold.

1. *There exists a constant d such that for any enumerator f, $|f(e)| < e/d$ for all e implies $\min_\varphi(e) \notin f(e)$ for infinitely many e.*
2. *Let f be an enumerator f with $|f(e)| < \sqrt{e}$ for all e. Then for infinitely many indices e, any index $z \in f(e)$ such that $\varphi_z = \varphi_e$ satisfies $|z| - |\min_\varphi(e)| = \Omega(|e|)$.*

The hidden constant in 2 *depends on the numbering φ.*

Proof sketch (of theorem)

Our proof starts by constructing a single Kolmogorov numbering φ which is hostile towards enumerators (Properties 1 and 2 below).

Lemma (Jain, Teutsch)

Suppose that h is a non-decreasing, unbounded computable function such that $h(\ell) < \ell - 3$ for all $\ell > 1$. Then there exists a Kolmogorov numbering φ such that for any enumerator function f satisfying $|f(e)| < 2^{h(|e|)}$ and any Kolmogorov numbering ψ, there exists a linearly-bounded, computable translator t from φ to ψ and infinitely many indices e such that

1. *for all $z \in f(e)$, either $|z| \geq |e| - 1$ or $\varphi_z \neq \varphi_e$,*
2. $\min_\varphi(e) \leq 8 \cdot 2^{h(|e|)}$, *and*
3. $|t(e)| \geq |e| - 1$.

If some other Kolmogorov numbering ψ were friendly towards enumerators, then using 3 we reach a contradiction by effectively:

- translating a φ-index into a ψ index,
- enumerating a list of ψ-candidates for minimal indices, and then
- translating this list of ψ-indices back into φ.

Proof sketch (of lemma)

For the case $h(\ell) = \ell - 5$, we wish to construct the Kolmogorov numbering φ where for every enumerator f with $|f(e)| < h(|e|)$ there exists a *Kolmogorov* translator t satisfying the conclusions of the lemma:

1. for all $z \in f(e)$, either $|z| \geq |e| - 1$ or $\varphi_z \neq \varphi_e$,
2. $|\min_\varphi(e)| \leq |e| - 2$, and
3. $|t(e)| \geq |e| - 1$.

Construction idea

We assume that $f(\sigma)$ has already finished enumerating for a fraction of σ's with length $|e|$. If some $f(\sigma)$ ever gives a proper extension, which happens only finitely often, we invoke an "eraser" procedure and restart.

1. Maintain a $\varphi_\sigma = \varphi_\tau$ with $|\sigma| = |e|$ and $|\tau| = |e| - 2$ such that φ_τ differs from all functions in $f(e)$. τ exists because $|f(\sigma)| < |e| - 5$.
2. Diagonalize against potential translators t_r: force a collision $t_r(\sigma) = t_r(\sigma')$ for $\varphi_\sigma \neq \varphi_{\sigma'}$, if possible. Disregard a small number of remaining indices which do not satisfy 3 above.

for $j = 0$ to $2^p - 1$ **do**: (some notation changes from previous slide)

1. Let k be the least natural number less than $2^{g(p)}$ such that every $\xi \in f_{n,s}(\sigma_j)$ of length less than $p-1$ satisfies $\varphi_{\xi,t} \neq \varphi_{\tau_k}$ defined up to now. Such a k must exist because there are $2^{g(p)}$ distinct φ_{τ_i}'s while, by assumption, $|f_n(\sigma_j)| < 2^{g(p)}$. Define φ_{σ_j} to be φ_{τ_k} defined up to now.
2. If spoil$(r) =$ false and there exist σ_y and σ_z with $0 \leq y < z < j$ such that $v_{r,t}(\sigma_y)\downarrow = v_{r,t}(\sigma_z)$, then
   - make $\varphi_{\sigma_y,t}$ and $\varphi_{\sigma_z,t}$ different while keeping their domains finite and their ranges equal to the singleton $\{p\}$,
   - set spoil$(r) =$ true, and
   - go to the next iteration of the **for** loop.

   Otherwise proceed to Step 3.
3. Dovetail Steps 4 and 5 until one of them succeeds.
4. Search for a $t' > t$ and $\xi \in f_{n,s}(\sigma_j)$ of length less than $p-1$ such that $\varphi_{\xi,t'} = \varphi_{\tau_k}$ defined up to now. If found, proceed to the next iteration of the **for** loop with $t = t'$.
5. If spoil$(r) =$ false and $|v_r(\sigma_j)| < p-1$, then go to the next iteration of the **for** loop with $t = t + t'$, where t' is the number of steps needed to compute $v_r(\sigma_j)$.

end for

Gödel numberings

For Gödel numberings, we get a weaker lower bound on enumerator size.

Theorem (Jain, Teutsch)

Let φ be a Gödel numbering, and let $k \geq 1$ be a constant. There is no enumerator f satisfying $|f(e)| \leq k$ and $|\min_\varphi(e)| \in f(e)$ for all e.

Because one can encode functions in a Gödel number very sparsely, the size of the enumerator in this theorem is optimal (Teutsch, Zimand). Our proof of this theorem makes use of the following result.

Kummer Cardinality Theorem

Let A be a set of non-negative integers, and let k be a positive integer. Suppose that there exists an algorithm which, on any input $x_1, \ldots, x_k$ enumerates at most k integers among $\{0, 1, \ldots, k\}$ such that one of these integers equals $|A \cap \{x_1, \ldots, x_k\}|$. Then A is computable.

The Kolmogorov property

Definition

A numbering φ has the *Kolmogorov property* if for any other numbering ψ there exists a (not necessarily computable) linearly-bounded translator f such that for all e, $\varphi_{f(e)} = \psi_e$.

- The above definition is strictly weaker than Kolmogorov numbering because the Kolmogorov property does not require the translator f to be computable (Stephan).
- One can view the Kolmogorov property as a generalization of "universal for Kolmogorov complexity" for strings. Numberings with the Kolmogorov property and universal machines share the same translation bound.

All of our negative results on enumerators require computable translators.

Question (see Teutsch-Zimand for computable version)

Does there exist a numbering with the Kolmogorov property which is friendly towards enumerators for minimal indices?

Almost everywhere minimal indices

Definition

For two p.c. functions f and g, we say $f =^* g$ if f and g agree everywhere except on a finite set. Let $\min^*_\varphi(e)$ denote the least index j such that $\varphi_j =^* \varphi_e$.

Theorem (Jain, Teutsch)

Suppose h is a non-decreasing, unbounded computable function such that $h(\ell) < \ell - 3$ for all $\ell > 1$. Then there exists a Kolmogorov numbering φ such that for any enumerator function f satisfying $|f(e)| < 2^{h(|e|)}$, there exist infinitely many e such that

1. *for all $z \in f(e)$, $z \neq \min^*_\varphi(e)$, and*
2. $\min_\varphi(e) \leq 8 \cdot 2^{h(|e|)}$.

Question

Does there exist a Kolmogorov numbering which is *friendly* towards enumerators for $*$-minimal indices?

The MIN^* problem

We do not yet understand $*$-minimal indices well.

Definition (Case 1990)

For any numbering φ, let $\mathrm{MIN}^*_\varphi = \{e : (\forall j < e)\,[\varphi_j \neq^* \varphi_e]\}$.

Using the Kummer Cardinality Theorem, one can obtain the following result.

Theorem (Jain, Stephan, Teutsch 2011)

For any numbering φ with the Kolmogorov property, MIN^*_φ has Turing degree $\emptyset'''$. In particular, MIN^*_φ computes the halting set.

The following is still open.

Question (Schaefer 1998)

Does MIN^*_φ compute the halting set in every Gödel numbering φ?

Thank you.