A quest for algorithmically random infinite structures

Bakh Khoussainov

Computer Science Department, The University of Auckland, New Zealand

The talk is based on paper accepted to LICS 2014

June 12, 2014

Plan

- Motivation
- Randomness via strings
- Basics of algebra
- Computable tree lemma
- ML-randomness for algebras
- Generator Independence Theorem
- Random graphs, trees, and monoids

Recent work

- The modern history is fascinating; goes back to the works of Kolmogorov, Martin-Löf, Chaitin, Schnorr and Levin.
- The last 15 years has seen significant advances in the study of algorithmic randomness on infinite strings.
- Monographs by Downey and Hirschfeldt, and Nies.
- Many notions of randomness, various techniques, and ideas have been studied.
- There are connections to other fields; e.g. the recent work of V. Brattka, J. Miller and A. Nies.

Strings as infinite structures

Identify a binary string $\beta \in 2^\omega$ with the structure $\mathcal{A}_\beta = (\omega; S, P)$:

$$S(i) = i + 1 \text{ and } P(n) \iff \beta(n) = 1.$$

So, algorithmic randomness of β is identified with algorithmic randomness of specific infinite structures $\mathcal{A}_\beta$.

This does not answer the following question:

What is an algorithmically random infinite tree, graph, monoid, or generally, a universal algebra?

Strings as infinite structures

Identify a binary string $\beta \in 2^\omega$ with the structure $\mathcal{A}_\beta = (\omega; S, P)$:

$$S(i) = i + 1 \text{ and } P(n) \iff \beta(n) = 1.$$

So, algorithmic randomness of β is identified with algorithmic randomness of specific infinite structures $\mathcal{A}_\beta$.

This does not answer the following question:

What is an algorithmically random infinite tree, graph, monoid, or generally, a universal algebra?

The main task: we need a good measure

Martin-Löf tests constitute the central concept for algorithmic randomness in the setting of infinite strings. This concept is based on the natural measure on the Cantor space.

In the case of infinite structures, how does one introduce a meaningful measure for the class?

We answer this question by putting a very limited and yet natural finiteness conditions on structures.

The main task: we need a good measure

Martin-Löf tests constitute the central concept for algorithmic randomness in the setting of infinite strings. This concept is based on the natural measure on the Cantor space.

In the case of infinite structures, how does one introduce a meaningful measure for the class?

We answer this question by putting a very limited and yet natural finiteness conditions on structures.

What do we expect from an infinite algorithmically random structure?

- **Absoluteness**: Algorithmic randomness should be an isomorphism invariant property. In particular, we do not want algorithmically random structures to be isomorphic to computable structures.
- **Continuum**: Random structures should be in abundance, the continuum. This is a property of a collective, the idea that goes back to Von Mises.
- **Selection**: There should be no effective way to describe the isomorphism type or an infinite part of the structure.

Somewhat naive approach: String-randomness

Let $\mathcal{A} = (\omega; P_0^{n_0}, \ldots, P_k^{n_k})$ be a structure with $A = \omega$. Form the following string $\alpha_{\mathcal{A}}$:

$$P_0^{n_0} c_{n_0}(0) \ldots P_k^{n_k} c_{n_k}(0) P_0^{n_0} c_{n_0}(1) \ldots P_k^{n_k} c_{n_k}(1) \ldots$$

This string codes up the atomic diagram of the structure.

Definition

The structure $\mathcal{A}$ is *string-random* if the string $\alpha_{\mathcal{A}}$ is ML-random.

To avoid much notation, we now consider graphs.

String-random implies model-theoretic random

Theorem

If $\mathcal{G}$ is a string-random graph then $\mathcal{G}$ is random model theoretically.

Proof.

One needs to show that the following property, known as **extension axiom**, holds for $\mathcal{G}$:

For any finite set X of vertices and non-trivial partition Y_1, Y_2 of X there exists a vertex z such that $\{z, y_1\}$ is an edge for all $y_1 \in Y_1$ and $\{z, y_2\}$ is not an edge for all $y_2 \in Y_2$.

This is guaranteed by the fact that $\alpha_{\mathcal{G}}$ is ML-random. □

Consquences

Thus, we have the following:

- Any two string-random structures are isomorphic.
- String-random structures are isomorphic to computable structures.
- The isomorphism type of string random structure is axiomatised by extension axioms.

All of the above defy our intuition that we postulated for algorithmically random infinite structures.

Consquences

Thus, we have the following:

- Any two string-random structures are isomorphic.
- String-random structures are isomorphic to computable structures.
- The isomorphism type of string random structure is axiomatised by extension axioms.

All of the above defy our intuition that we postulated for algorithmically random infinite structures.

Basics of algebra

An *algebra* $\mathcal{A}$ is $(A; f_1, \ldots, f_n, c_1, \ldots, c_m)$, where:

- The set $A \neq \emptyset$ is the *domain*,
- Each $f_i : A^{k_i} \to A$ an *atomic operation*,
- Each c_j is a *distinguished element*.

Ground terms are defined by induction:

- Each c_j is a ground term,
- If $t_1, \ldots, t_{k_i}$ are ground terms, then so is $f_i(t_1, \ldots, t_{k_i})$.

The height, $h(t)$, of the term t is defined as follows:

- $h(c_j) = 0$,
- $h(f_i(t_1, \ldots, t_{k_i})) = max\{h(t_1), \ldots, h(t_n)\} + 1$.

Basics of algebra

An *algebra* $\mathcal{A}$ is $(A; f_1, \ldots, f_n, c_1, \ldots, c_m)$, where:

- The set $A \neq \emptyset$ is the *domain*,
- Each $f_i : A^{k_i} \to A$ an *atomic operation*,
- Each c_j is a *distinguished element*.

Ground terms are defined by induction:

- Each c_j is a ground term,
- If $t_1, \ldots, t_{k_i}$ are ground terms, then so is $f_i(t_1, \ldots, t_{k_i})$.

The height, $h(t)$, of the term t is defined as follows:

- $h(c_j) = 0$,
- $h(f_i(t_1, \ldots, t_{k_i})) = max\{h(t_1), \ldots, h(t_n)\} + 1$.

Basics of algebra

An *algebra* $\mathcal{A}$ is $(A; f_1, \ldots, f_n, c_1, \ldots, c_m)$, where:

- The set $A \neq \emptyset$ is the *domain*,
- Each $f_i : A^{k_i} \to A$ an *atomic operation*,
- Each c_j is a *distinguished element*.

Ground terms are defined by induction:

- Each c_j is a ground term,
- If $t_1, \ldots, t_{k_i}$ are ground terms, then so is $f_i(t_1, \ldots, t_{k_i})$.

The height, $h(t)$, of the term t is defined as follows:

- $h(c_j) = 0$,
- $h(f_i(t_1, \ldots, t_{k_i})) = max\{h(t_1), \ldots, h(t_n)\} + 1$.

$\bar{c}$-generated algebras

Definition

An algebra $\mathcal{A}$ is *c-generated* if every element of $\mathcal{A}$ is a value of some ground term.

Thus, if $\mathcal{A}$ is c-generated, then $\forall a \in A \exists t(t_{\mathcal{A}} = a)$. Call t a *representation* of a in $\mathcal{A}$. Set:

$$h(a) = min\{h(t) \mid t_{\mathcal{A}} = a\}.$$

The *height* of $\mathcal{A}$ is the supremum of all the heights of its elements.

Proper partial algebras

Let $\mathcal{A}$ be a c-generated. For each $n \in \omega$, consider

$$A[n] = \{a \in A \mid h(a) \leq n\}.$$

Each atomic operation f defines a *partial operation* f_n on $A[n]$ as follows. For all $a_1, \ldots, a_{k_i} \in A[n]$:

- $f_n(a_1, \ldots, a_{k_i})$ equals $f(a_1, \ldots, a_{k_i})$ if $h(a_i) < n$ for all i;
- $f_{i,n}(a_1, \ldots, a_{k_i})$ is undefined otherwise.

Call the partial algebra $\mathcal{A}[n]$, the *n-th slice* of $\mathcal{A}$. We refer to the isomorphism types of these algebras as *proper partial algebras*.

Preparatory Lemmas

Lemma

Two c-generated algebras $\mathcal{A}$ and $\mathcal{B}$ are isomorphic iff they agree at n for all n. □

Lemma

Let $\mathcal{A}$ be an infinite c-generated algebra. For each $n \geq 0$ there is a proper partial algebra $\mathcal{B}$ such that $\mathcal{A}$ and $\mathcal{B}$ agree at n. □

Lemma

If $\mathcal{B}$ is a proper partial of height n, then there is an infinite c-generated algebra $\mathcal{A}$ such that $\mathcal{A}$ and $\mathcal{B}$ agree at n. □

Definition of tree $\mathcal{T}_m$

1. The root is $\emptyset$. This is level -1.
2. The nodes of at level $n \geq 0$ are proper partial algebras of height n.
3. Let $\mathcal{B}$ be a proper partial algebra of height n. Its successor is any proper partial algebra $\mathcal{C}$ of height $n+1$ such that $\mathcal{B}$ and $\mathcal{C}$ agree at n.

The function $n \to r_m(n)$, where $r_m(n)$ is the number of proper partial algebras of height n, is computable.

Properties of the tree $\mathcal{T}_m$

Γ_m^ω denotes all $\bar{c}$-generated infinite algebras of signature Γ_m.

Lemma (Computable Tree Lemma)

1. *Given any node x of the three, we can effectively compute the proper partial algebra $\mathcal{B}_x$ associated with the node x.*
2. *Each x in $\mathcal{T}_m$ has an immediate successor. We can compute the number of immediate successors of x.*
3. *Each path $\eta = \mathcal{B}_0, \mathcal{B}_1, \ldots$ determines the algebra $\mathcal{B}_\eta = \cup_i \mathcal{B}_i \in \Gamma_m^\omega$.*
4. *The mapping $\eta \to \mathcal{B}_\eta$ is a bijection from $[\mathcal{T}_m]$ to Γ_m^ω.* □

Topology

Using $\mathcal{T}_m$ we can introduce the topology into the class Γ_m^ω.

Definition (Topology)

Let $\mathcal{B}$ be a proper partial algebra of height n. The **cone** of $\mathcal{B}$ is:

$$Cone(\mathcal{B}) = \{\mathcal{A} \mid \mathcal{A} \in \Gamma_m^\omega, \text{ and } \mathcal{A} \text{ and } \mathcal{B} \text{ agree at } n\}.$$

Declare the cones $Cone(\mathcal{B})$ to be the *base open sets* of the topology on Γ_m^ω. We refer to the proper partial algebra $\mathcal{B}$ as the *base of the cone*.

Measure

Definition (Measure)

- The measure of the cone based at the root is 1.
- Assume that the measure $\mu_m(Cone(\mathcal{B}_x))$ has been defined. Let e_x be the number of immediate successors of x. Then for any immediate successor y of x the measure of $Cone(\mathcal{B}_y)$ is

$$\mu_m(Cone(\mathcal{B}_y)) = \frac{\mu_m(Cone(\mathcal{B}_x))}{e_x}.$$

Metric

Definition (Metric)

For two c-generated algebras $\mathcal{A}$ and $\mathcal{B}$, let n be the maximal level at which $\mathcal{A}$ and $\mathcal{B}$ agree. Let $\mathcal{C}$ be the n-th slice of $\mathcal{A}$ (hence of $\mathcal{B}$). The distance $d(\mathcal{A}, \mathcal{B})$ between the algebras is then defined as follows: $d(\mathcal{A}, \mathcal{B}) = \mu_m(Cone(\mathcal{C}))$.

Lemma

The function d is a metric in the space Γ_m^ω . □

Basic properties of the space Γ_m^ω

Fact

The space $\mathcal{M} = (\Gamma_m^\star \cup \Gamma_m^\omega, d)$ has the following properties:

1. *$\mathcal{M}$ is compact.*
2. *The set $\Gamma_m^\star$ is countable and dense in $\mathcal{M}$.*
3. *Finite unions of cones form clo-open sets in the topology.*
4. *The set of all μ_m-measurable sets is a σ-algebra.* □

ML-random algebras

Definition

1. A *Martin-Löf test* is a uniformly c.e. sequence $\{G_n\}_{n\geq 1}$ of Σ^0_1-classes in Γ^{ω}_{m} such that $G_{n+1} \subset G_n$ and $\mu_m(G_n) < 1/r_m(n)$ for all $n \geq 1$.
2. A *c*-generated algebra $\mathcal{A}$ *fails* the *Martin-Löf test* $\{G_n\}_{n\geq 1}$ if $\mathcal{A}$ belongs to $\cap_n G_n$. Otherwise, we say that the algebra $\mathcal{A}$ *passes* the test.
3. A *c*-generated algebra $\mathcal{A}$ is *ML-random* if if it passes every Martin-Löf test.

Corollary

The number of ML-random algebras is continuum. □

Generator independence theorem

Theorem (Generator independence theorem)

ML-randomness for algebras is independent on the generators.

Proof (idea). Let $\bar{a} = a_1, \ldots, a_m$ and $\bar{b} = b_1, \ldots, b_k$ be generators of $\mathcal{A}$. Thus, $(\mathcal{A}, \bar{a}) \in \Gamma_m^\omega$ and $(\mathcal{A}, \bar{b}) \in \Gamma_k^\omega$.

Goal: $(\mathcal{A}, \bar{a})$ is ML-random if and only if $(\mathcal{A}, \bar{b})$ is ML-random.

Generator independence

There exist ground terms $t_1, \ldots, t_k$ and $q_1, \ldots, q_m$ such that

$$t_i(\bar{a}) = b_i \ \ and \ \ q_j(\bar{b}) = a_j,$$

with $i = \overline{1, k}$ and $j = \overline{1, m}$. Call these the *base equalities* B.

If $(\mathcal{D}, b_1, \ldots, b_k) \models B$ then $(\mathcal{D}, q_1(\bar{b}), \ldots, q_m(\bar{b})) \in \Gamma_m^\omega$.

Lemma

The partial mapping $\alpha : (\mathcal{D}, b_1, \ldots, b_k) \to (\mathcal{D}, q_1(\bar{b}), \ldots, q_m(\bar{b}))$ *preserves ML-tests.*

This proves the theorem.

Computable algebras

Let $\mathcal{A}$ be a c-generated infinite algebra and $h : \mathcal{T}_G \to \mathcal{A}$ be the onto homomorphism. The *word problem* of $\mathcal{A}$ is:

$$WP(\mathcal{A}) = \{(t, q) \mid t, w \in T_G \,\&\, h(t) = h(q)\}.$$

Fact

If $\mathcal{A}$ is a computable algebra then $\mathcal{A}$ is not ML-random. □

Computability in the Halting set

Denote the halting set by $\mathcal{H}$.

Definition

An algebra $\mathcal{A}$ is *$\mathcal{H}$-computable* if *WP*$(\mathcal{A})$ is computable in $\mathcal{H}$.

Theorem

ML-random $\mathcal{H}$-computable algebras exist.

Proof idea

Consider a universal ML-test: $\{U_n\}_{n\geq 1}$.
We build $\mathcal{A}$ so that $\mathcal{A} \notin U_1$.

Using $\mathcal{H}$, write U_1 as a disjoint union $C(\mathcal{B}_1) \cup \mathcal{C}(B_2) \cup \ldots$.

Using $\mathcal{H}$, construct $\mathcal{A}$ by stages s so that:

1. $\mathcal{A}_{s-1} \subset \mathcal{A}_s$.
2. The cone $C(\mathcal{A}_s)$ avoids all the cones $C(\mathcal{A}_i)$, $i = \overline{1, s}$.
3. The measure of $C(\mathcal{A}_s)$ is greater than the measure of the remaining cones.

A set up for graphs

We consider connected graphs of bounded degree $d > 2$. Let $\mathcal{G}$ be an infinite graph. Fix an initial vertex, say c.

For $n \in \omega$, let $D_{\mathcal{G},n}(c)$ be the collection of all the vertices in $\mathcal{G}$ that are at distance at most n form c.

We call the graphs $D_{\mathcal{G},n}(c)$ the *n-neighbourhoods* of c.

The tree of neighbourhoods

Define the following tree $\mathcal{T}$.

1. The root is $\emptyset$. This is level -1.
2. The nodes at level $n \geq 0$ are the isomorphism types of the n-neighbourhoods of c.
3. Let $\mathcal{G}$ be a graph at level n. Its successor is any $(n+1)$-neighbourhood $\mathcal{G}'$ such that $\mathcal{G} \subset \mathcal{G}'$.

Computable tree lemma

1. Given any node x, we can effectively compute the graph $\mathcal{G}_x$ associated with x.
2. For every x in $\mathcal{T}$, we can compute the number of immediate successors of x.
3. For each path η in $\mathcal{T}$, The union $\mathcal{G}_\eta = \cup_{\mathcal{G}_i \in \eta} \mathcal{G}_i \in \Gamma_m^\omega$ is a connected graph of bounded degree d.
4. The mapping $\eta \to \mathcal{G}_\eta$ is a bijection between $[T]$ and all infinite connected graphs of bounded degree d.

ML-random graphs

Just like in the case of finitely generated universal algebras, we have the following result:

Theorem

1. *The ML-randomness for graphs is on the constant c.*
2. *ML-randomness is an isomorphism invariant property.*
3. *There are continually many ML-random graphs.*
4. *ML-random $\mathcal{H}$-computable graphs exist.* □

Computably enumerable trees

Let E be an equivalence relation on ω.

Definition

Relation $Edge \subseteq \omega^2$ *respects* E if $\forall x_1, y_1, x_2, y_2 \in \omega$ we have

$$[(x_1, x_2) \in E \ \& \ (y_1, y_2) \in E] \rightarrow (Edge(x_1, y_1) \leftrightarrow Edge(x_2, y_2))].$$

If $Edge$ respects E then we can naturally define the structure

$$(\omega/E; Edge).$$

Definition

A graph $\mathcal{G}$ is *c.e.* if there is a c.e. equivalence relation E on ω and a binary relation $Edge$ that respects E such that the graph $\mathcal{G}$ is isomorphic to the graph $(\omega/E; Edge)$.

The computable tree $\mathcal{T}$ for trees

- Select a node *c* in a *d*-ary tree. It is the root.
- Define the heights of finite trees.
- Construct a computable tree $\mathcal{T}$ such that
  1. For any node *v* of $\mathcal{T}$, we can effectively compute the tree $\mathcal{X}_v$ associated with the node *v*.
  2. For every node *v* in $\mathcal{T}$, we can compute the number of immediate successors of *v*.
  3. For each path η the mapping $\eta \to X_\eta$ is a bijection between $[\mathcal{T}]$ and all infinite *d*-ary trees.

ML-random trees

Everything goes as in the case of algebras and graphs. However, the theorems about the existence of ML-random $\mathcal{H}$-computable algebras and graphs is strengthen significantly:

Theorem

ML-random computably enumerable d-ary trees exist.

Proof.

The proof uses an c.e. reduction process that shrinks finite trees without obstructing their tree structure.

□

The reduction process

Let $\mathcal{X} = (I/E; Edge)$ be a tree, with I finite.

The root of $\mathcal{X}$ is $[c]$.

Let $[x_1], [x_2], [x_3] \in I/E$, all distinct, such that $[x_3]$ is a leaf, $([x_2], [x_3]) \in Edge$ and $([x_1], [x_2]) \in Edge$.

Set E' = equivalence relation generated by E and (x_1, x_3).

The structure $\mathcal{X}' = (I/E'; Edge)$ is a tree.

Denote this by $\mathcal{X} \triangleright \mathcal{X}'$.

Lemma

For the tree $\mathcal{X}$ and any of its subtrees $\mathcal{Y}$ of height at least 2 there is a sequence of reductions $\mathcal{X}_1 \triangleright \mathcal{X}_2 \triangleright \ldots \mathcal{X}_{n-1} \triangleright \mathcal{X}_n$ such that $\mathcal{X}_1 = \mathcal{X}$ and $\mathcal{X}_n = \mathcal{Y}$. □

Monoids case

Proceed just like in the the previous cases. Build a tree $\mathcal{T}$ for the class of finitely generated monoids. However, we need to be a careful in building $\mathcal{T}$. We do not want the tree to collapse at some nodes. This is guaranteed by the following lemma:

Lemma

Let $\mathcal{M}$ be a proper partial monoid of height n. There are at least two non-isomorphic infinite monoids that extend $\mathcal{M}$. In particular $\mathcal{M}$ has at least two non-isomorphic proper partial monoid extensions of the same height. □

Monoids case

Proceed just like in the the previous cases. Build a tree $\mathcal{T}$ for the class of finitely generated monoids. However, we need to be a careful in building $\mathcal{T}$. We do not want the tree to collapse at some nodes. This is guaranteed by the following lemma:

Lemma

Let $\mathcal{M}$ be a proper partial monoid of height n. There are at least two non-isomorphic infinite monoids that extend $\mathcal{M}$. In particular $\mathcal{M}$ has at least two non-isomorphic proper partial monoid extensions of the same height. □

ML-random monoids

Just like in the case of finitely generated universal algebras and graphs, we have the following result:

Theorem

1. *The ML-randomness for monoids is independent on the generators c.*
2. *ML-randomness is an isomorphism invariant property.*
3. *There are continually many ML-random monoids.*
4. *ML-random $\mathcal{H}$-computable monoids exist.* □

Open questions

1. Are there ML-random c.e. universal algebras and graphs?
2. Is there a finitely presented yet random universal algebra?
3. Is there an effectively infinite ML-random universal algebra?
4. Build ML-random finitely generated groups and rings.