Constant compression and random weights [1]

Wolfgang Merkle and Jason Teutsch

Ruprecht-Karls-Universität
Heidelberg, Germany

[1] presented at STACS 2012, full version will appear in *Computability*

Random sequences

Randomness of individual sequences

- Consider an infinite binary sequence

$$A(0)A(1)A(2)\ldots, \qquad A(i) \in \{0,1\}\,.$$

- When is such a sequence random?

  (Consider the symmetric case where 0 and 1 have the same probability.)
- From the point of view of probability theory, any given sequence is as random as any other.

In ALGORITHMIC RANDOMNESS one investigates into various notions of randomness for individual sequences and their relations to other concept from complexity or computability theory, e.g., one may study the computational power of random sequences.

Randomness via compressibility and via predictability

In order to obtain notions of randomness for individual sequences one considers effective COMPRESSIBILITY and PREDICTABILITY.

Randomness via compressibility

A sequence is random if the initial segments of the sequence cannot be "effectively compressed", i.e., for all m the initial segment of the sequence of length m has only codes of length at least $m - c$.

Randomness via predictability

A sequence is random if one cannot "effectively predict" bit $m + 1$ of the sequence after having seen the first m bits.

Coding by prefix-free Turing machines

Definition

A PREFIX-FREE MACHINE is a Turing machine with prefix-free domain. Given a prefix-free machine M, the PREFIX-FREE KOLMOGOROV COMPLEXITY of a string x with respect to M is

$$\mathrm{K}_M(x) = \min\{|p| : M(p) = x\}.$$

A prefix-free machine U is UNIVERSAL if for any other prefix-free machine M there is a constant c_M such that for all x it holds that

$$\mathrm{K}_U(x) \leq \mathrm{K}_M(x) + c_M.$$

There are universal prefix-free machines, e.g., given an effective listing $M_0, M_1, \ldots$ of all prefix-free machines, we obtain a universal prefix-free machine by letting

$$U(1^e 0 p) = M_e(p) \ .$$

Martin-Löf random sequences

Definition

We fix some some universal prefix-free machine U and let

$$\mathrm{K}(x) = \mathrm{K}_U(x)$$

be the PREFIX-FREE KOLMOGOROV COMPLEXITY OF **x**.

Definition

A sequence $A = A(0)A(1)\ldots$ is MARTIN-LÖF RANDOM if and only if for some constant c and all m it holds that

$$\mathrm{K}(A(0)\ldots A(m-1)) \geq m - c.$$

Why is Martin-Löf randomness defined in terms of prefix-free Kolmogorov complexity and not in terms of the more natural plain variant, where the restriction to prefix-free machines is dropped?

Martin-Löf random sequences

Randomness via prediction

Consider a betting game where one starts with finite capital and

- bets successively on the bits of an initially unknown sequence,

- payoff is fair in the sense that the stake is doubled or lost depending on whether the respective guess was correct,

- one succeeds on the given sequence if the gained capital is unbounded.

Consider further the model where given any initial segment $A(0) \dots A(i)$, the capital gained on this initial segment can be effectively approximated from below.

Theorem (Schnorr)

A sequence is Martin-Löf random if and only if one cannot succeed on the sequence by a betting strategy with a capital function that is approximable from below.

Left-r.e. Martin-Löf random sequences

Martin-Löf random sequences cannot exhibit effectively detectable patterns such as computable subsequences.

However, there are Martin-Löf random sequences that are effectively approximable in the sense that they are left-r.e.

Definition

A real α is LEFT-R.E. if there is a computable nondecreasing sequence of dyadic rationals that converges to α.

By identifying a sequence $A = A(0)A(1)\ldots$ with the real $\alpha = 0.A(0)A(1)\ldots$, the notion left-r.e. extends to sequences.

For a sequence, being left-r.e. amounts to an effective pointwise approximation starting from the all 0s sequence such that a bit 0 can always be switched to 1, but in order to switch a bit 1 to 0 there must be a corresponding switch from 1 to 0 farther left.

Chaitin's Omega numbers

Definition (Weight)

The WEIGHT of a string σ is $2^{-|\sigma|}$. The WEIGHT of a (not necessarily prefix-free) set A of strings is $\sum_{\sigma \in A} 2^{-|\sigma|}$.

Let Ω_M be the weight of the domain of the prefix-free machine M.

A real α is an OMEGA NUMBER if $\alpha = \Omega_U$ for some universal prefix-free machine.

Theorem (Kučera and Slaman)

A real in the interval between 0 and 1 is an Omega number if and only if the real is left-r.e. and Martin-Löf random.

More characterizations of Omega numbers are known, e.g., as

the Solovay complete left-r.e. reals, i.e., up to a constant factor, any effective approximation from below to any Omega number is slower than any such approximation to any other left-r.e real.

One- and two-sided Gamma sets

Definition (Gamma sets)

Let M be a prefix-free machine and let a and b be integers.

The SET OF **a**-COMPRESSIBLE STRINGS WITH RESPECT TO **M** is

$$\Gamma^a_M = \{\sigma \in \{0,1\}^* : (\exists \tau)\, M(\tau) = \sigma \text{ and } |\tau| \leq |\sigma| - a\}.$$

The SET OF **[a, b)**-COMPRESSIBLE STRINGS WITH RESPECT TO **M** is $\Gamma^{a\backslash b}_M = \Gamma^a_M - \Gamma^b_M$.

We will refer to a set of the form Γ^a_M or $\Gamma^{a\backslash b}_M$ as ONE-SIDED and TWO-SIDED GAMMA SET, respectively. Such a Gamma set is called UNIVERSAL in case M is a universal prefix-free machine.

A string σ is **a**-COMPRESSIBLE if $\mathrm{K}(\sigma) < |\sigma| - a$.

The number of compressible strings

Counting theorem (Chaitin)

For some positive constant d and all natural numbers a and n, it holds that $|\{\sigma \in \{0,1\}^n : \mathrm{K}(\sigma) \leq n + \mathrm{K}(n) - a\}| \leq 2^{n-a+d}$.

In particular, $|\{\sigma \in \{0,1\}^n : \mathrm{K}(\sigma) \leq n - a\}| \leq 2^{n-a-\mathrm{K}(n)+d}$.

Remark

Similarly, for any universal prefix-free machine U, we have

$$|\{0,1\}^n \cap \Gamma^a_U| \leq 2^{n-a-\mathrm{K}(n)+d'} .$$

Hence the ratio of a-compressible strings with respect to U among all strings of length n goes to 0 when n goes to infinity.

Rather tight lower and upper bounds for the number of strings of a given length in universal one-sided and two-sided Gamma sets can be obtained from the improved counting theorem by Miller and Yu.

The number of compressible strings

Improved counting theorem (Miller and Yu)

Let U be a universal prefix-free machine. There is a constant d such that for all natural numbers c and n it holds that

$$2^{n-c-\mathrm{K}(c|n^*)-d} \leq |\{\sigma \in \{0,1\}^n : \mathrm{K}_U(\sigma) \leq n + \mathrm{K}_U(n) - c\}| \leq 2^{n-c-\mathrm{K}(c|n^*)+d}.$$

Corollary

Let U be any universal prefix-free machine. There is a constant d such that for all natural numbers a and all n, as well as for all integers a and for all sufficiently large natural numbers n we have

$$2^{n-\mathrm{K}(n)-a-\mathrm{K}(a|n^*)-d} \leq |\Gamma^a_U \cap \{0,1\}^n| \leq 2^{n-\mathrm{K}(n)-a-\mathrm{K}(a|n^*)+d}.$$

One- and two-sided Theta numbers

Definition (Theta numbers)

For a prefix-free machine M and integers a and b let

$$\Theta_M^a = \sum_{\sigma \in \Gamma_M^a} 2^{-|\sigma|} \quad \text{and} \quad \Theta_M^{a\setminus b} = \sum_{\sigma \in \Gamma_M^{a\setminus b}} 2^{-|\sigma|}.$$

We will refer to a real of the form Θ_M^a or $\Theta_M^{a\setminus b}$ as ONE-SIDED and TWO-SIDED THETA NUMBER, respectively.

Theta numbers of the form Θ_M^a and $\Theta_M^{a\setminus b}$ are always finite since both can be at most as large as 2^{-a} times the weight $\Omega_M \leq 1$ of the domain of the prefix-free machine M.

By definition, $\Gamma_M^{a\setminus b} = \Gamma_M^a - \Gamma_M^b$, hence

- in case $b \leq a$, the set $\Gamma_M^{a\setminus b}$ is empty and $\Theta_M^{a\setminus b} = 0$,
- in case $b > a$, we have $\Gamma_M^b \subseteq \Gamma_M^a$ and $\Theta_M^{a\setminus b} = \Theta_M^a - \Theta_M^b$.

When are two-sided Theta numbers left-r.e.?

Tadaki observed that one-sided Theta numbers are always left-r.e., due to being equal to the weight of an r.e. one-sided Gamma set.

Proposition

Let a be any integer. There exists a prefix-free machine M such that for all $b > a$ the real $\Theta_M^{a\backslash b}$ is not left-r.e.

Proposition

Let U be a universal prefix-free machine, and let a be any integer. For any integer b, the set $\Gamma_U^{a\backslash b}$ does not contain an infinite r.e. set.

Theorem 1

Let U be a universal prefix-free machine, and let a be any integer. For all sufficiently large integers b, the real $\Theta_U^{a\backslash b}$ is left-r.e.

When are two-sided Theta numbers Martin-Löf random?

Theorem (Tadaki)

Let a be a natural number and let U be a universal prefix-free machine. Then the real Θ_U^a is Martin-Löf random.

Theorem 2

Let a be a natural number and let U be a universal prefix-free machine. Then for almost all natural numbers $b > a$, the real $\Theta_U^{a\backslash b}$ is Martin-Löf random.

Proposition

For all pairs of natural numbers a and b there is a universal prefix-free machine U such that $\Gamma_U^{a\backslash b}$ is empty, hence $\Theta_U^{a\backslash b} = 0$.

Which reals are Theta numbers?

Proposition

Let a and $b > a$ be natural numbers and let $\alpha < 2^{-a}$ be nonnegative and left-r.e.

Then there is a prefix-free machine M such that

$$\alpha = \Theta_M^{a\backslash b} = \Theta_M^a.$$

Theorem 3

Let a and $b > a$ be natural numbers and let $\alpha < 2^{-a}$ be an Omega number.

Then there are universal prefix-free machines V and V' such that

$$\alpha = \Theta_V^{a\backslash b} = \Theta_{V'}^a \ .$$

More characterizations of Omega numbers

Corollary

A real in the interval between 0 and 1 is an Omega number if and only if the real is a universal one-sided Theta number.

Corollary (proof not yet written up)

For any universal prefix-free machine U there is a natural number c_U such that the following equivalence holds.

A real α in the interval between 0 and 1 is an Omega number if and only if there are natural numbers a and b and a universal prefix-free machine U such that

$$\alpha = \Theta_U^{a \setminus b} \quad \text{and} \quad b - a \geq c_U.$$

The proof of Theorem 1

Lemma

Let U be a universal prefix-free machine and let a and b be any integers where $a < b$. Suppose that for each integer t an enumeration without repetitions of the set Γ_U^t is given uniformly effectively in t and let $\sigma_0, \sigma_1, \ldots$ and $\tau_0, \tau_1, \ldots$ be the corresponding enumerations of Γ_U^a and Γ_U^b, respectively.

Then for all sufficiently large b there is a strictly increasing recursive function g such that for all i,

(I) $\left|\sigma_{g(i)}\right| = |\tau_i|$,

(II) $\sigma_{g(i)} \neq \tau_j$ for $j = 0, \ldots, i$.

Sketch of proof: Letting $\gamma(i)$ be equal to the least string of length $|\tau_i|$ that differs from $\tau_0, \ldots, \tau_i, \sigma_0, \ldots, \sigma_{max\{g(0),\ldots,g(i-1)\}}$, it suffices to let $g(i)$ be equal to the index of $\gamma(i)$ in $\sigma_0, \sigma_1, \ldots$. That $\gamma(i)$ is indeed a-compressible follows for all large enough b by

$$\mathrm{K}_U(\gamma_i) \leq |a^* b^* \tau_i^*| + c \leq |\gamma(i)| - b + |a^* b^*| + c \leq |\gamma(i)| - a.$$

The proof of Theorem 1

Theorem 1

Let U be a universal prefix-free machine, and let a be any integer. For all sufficiently large integers b, the real $\Theta_U^{a\backslash b}$ is left-r.e.

Fix any b that is so large that there are enumerations $\sigma_0, \sigma_1, \ldots$ and $\tau_0, \tau_1, \ldots$ of Γ_U^a and Γ_U^b, respectively, and a recursive function g as in the lemma above. Recall that g is strictly increasing, hence is one-to-one and its range R is recursive. Now

$$\begin{aligned}
\Theta_U^{a\backslash b} &= \sum_{\sigma \in \Gamma_M^{a\backslash b}} 2^{-|\sigma|} = \sum_{\sigma \in \Gamma_M^a} 2^{-|\sigma|} - \sum_{\tau \in \Gamma_M^b} 2^{-|\tau|} \\
&= \sum_{k \in \mathbb{N}\backslash R} 2^{-|\sigma_k|} + \sum_{k \in \mathbb{N} \cap R} 2^{-|\sigma_k|} - \sum_{k \in \mathbb{N}} 2^{-|\tau_k|} \\
&= \sum_{k \in \mathbb{N}\backslash R} 2^{-|\sigma_k|} + \sum_{k \in \mathbb{N}} \underbrace{2^{-|\sigma_{g(k)}|} - 2^{-|\tau_k|}}_{=0},
\end{aligned}$$

hence $\Theta_U^{a\backslash b}$ is left-r.e.