On the number of infinite sequences with trivial initial segment complexity

George Barmpalias

Joint work with my students from Amsterdam
Martijn Baarse and Tom Sterkenburg

ISCAS
Institute of Software
Chinese Academy of Sciences

July 2011, Ioannina

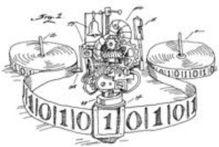
Kolmogorov complexity of strings

Kolmogorov, 1960s:

“The complexity of a binary string is the length of its shortest description.”

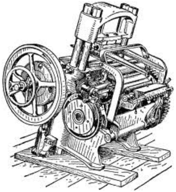
Descriptions given in an algorithmic way:

If M is a Turing machine and $M(\sigma) = \tau$ then σ is an M-description of τ.



Kolmogorov complexity of strings: Machines

Let $|M|$ be the **size** of the machine M.



Let $C_M(\sigma)$ be the complexity of σ w.r.t. M.

The complexity of σ is the least sum $|M| + C_M(\sigma)$ where M ranges over all machines.

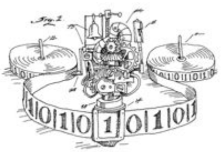
Let $C(\sigma)$ denote the complexity of σ.

By universality, Kolmogorov complexity is an effective concept.

Kolmogorov complexity of strings: Optimality

There is a machine that gives optimal descriptions:

On a given input, interpret the first few bits as a description of a machine. . .



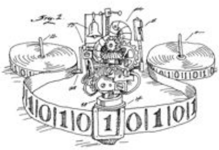
. . . and simulate that machine on the tail of the original input.

Kolmogorov complexity of strings: Optimality

There is a machine that gives optimal descriptions:

On a given input, interpret the first few bits as a description of a machine...

...and simulate that machine on the tail of the original input.



In other words... $U(0^e 1\sigma) = M_e(\sigma)$ *where M_e is the eth machine.*

Let $C(\sigma)$ denote the Kolmogorov complexity of σ.

Examples:

$C(\sigma) \leq |\sigma|$ and $C(n) \leq \log n$

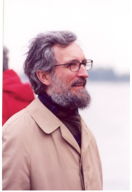
Randomness of infinite sequences

A string is *c*-incompressible if it has no description that is shorter than its length by at least *c* bits.

Kolmogorov sought to define:

“A binary stream is random if for some constant *c* all of its initial segments are *c*-incompressible.”

Unfortunately, Martin-Löf showed that

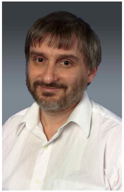


For every constant *c* and every stream *X* has a *c*-compressible initial segment.

Chaitin and Levin, 1970s

They observed that

Given a string, one can recover information from the bits of the string but also from its length.



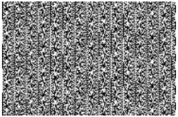
More faithful complexity measure via prefix-free machines:

Prefix-free machines cannot extract information from the length of a string.

Let $K(\sigma)$ be the prefix-free complexity of σ.

$K(\sigma) \leq |\sigma| + K(|\sigma|)$ and $K(n) \leq 2\log n$.

Algorithmic randomness



A stream X is random if there is a constant c such that $K(X \restriction_n) \geq n - c$ for all n.

This notion of randomness is robust:

- Coincides with other approaches (betting strategies, statistics)
- Random reals form a set of measure 1
- Meets laws of large numbers, normality etc.

Trivial sequences

A stream is random if it has high initial segment complexity.

> *To describe the first n bits of the sequence you need to use n bits (modulo a constant)*

On the other end of the spectrum:

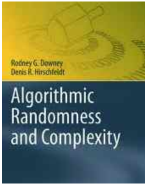
> *A stream is trivial if the complexity of its first n bits is as low as the complexity of* 0^n.

Solovay 1975

Chaitin asked if there are non-computable streams whose initial segment complexity is as low as a computable stream.

Solovay gave a positive answer.

Draft of a paper (or series of papers) on Chaitin's work. Unpublished notes, May 1975. 215 pages.



Solovay's work entirely included...

...in Downey/Hirschfeldt book

The world of K-trivial streams

$\rightsquigarrow$ Computable from the halting problem i.e. Δ^0_2 (Chaitin 70s)

$\rightsquigarrow$ Incomplete, and in fact low (Downey/Hirschfeldt/Nies/Stephan)

$\rightsquigarrow$ Provide a 'natural' solution to Post's problem.

The world of K-trivial streams

$\rightsquigarrow$ Computable from the halting problem i.e. Δ^0_2 (Chaitin 70s)

$\rightsquigarrow$ Incomplete, and in fact low (Downey/Hirschfeldt/Nies/Stephan)

$\rightsquigarrow$ Provide a 'natural' solution to Post's problem.

$$A = \{n \mid \exists e, s\, (\underbrace{W_{e,s} \cap A_s = \emptyset \ \wedge\ n > 2e \ \wedge n \in W_{e,s}}_{\text{Post's simple set}} \wedge \textstyle\sum_{n<j<s} 2^{-K_s(j)} < 2^{-e})\}$$

The world of K-trivial streams

$\rightsquigarrow$ Computable from the halting problem i.e. Δ^0_2 (Chaitin 70s)

$\rightsquigarrow$ Incomplete, and in fact low (Downey/Hirschfeldt/Nies/Stephan)

$\rightsquigarrow$ Provide a 'natural' solution to Post's problem.

$$A = \{n \mid \exists e, s\, (\underbrace{W_{e,s} \cap A_s = \emptyset \ \wedge \ n > 2e \ \wedge n \in W_{e,s}}_{\text{Post's simple set}} \wedge \textstyle\sum_{n<j<s} 2^{-K_s(j)} < 2^{-e})\}$$

$\rightsquigarrow$ Downward closed under oracle computation (Nies 2005)

$\rightsquigarrow$ Form an ideal in the Turing degrees.

Cumulative hierarchy of K-trivial streams

A stream X is K-trivial if $K(X \upharpoonright_n) \leq K(n) + c$ for all n, some c.

K-trivial streams are stratified in a hierarchy of length ω

. . . whose c-level contains the K-trivial streams with constant c.

Question by Downey, Miller, Nies, Yu

What is the arithmetical complexity of the map G which takes c to the number of K-trivial strings with constant c?

. . . or equivalently

How many quantifiers are needed in order to define G in arithmetic?

. . . or equivalently

How hard is to compute G?

Basic facts about G, by DMNY

- Computable from $\mathbf{0}^{(3)} \ldots$ i.e. Δ^0_4
- Not computable i.e. not Δ^0_1
- Not computable from the halting problem, i.e. not Δ^0_2

Is it computable from $\mathbf{0}^{(2)}$ i.e. is it Δ^0_3?

The classes of K_c-trivial streams

- They are uniformly Π^0_1 in the halting set
- The set of infinite paths through a $\mathbf{0}'$-computable tree.
- The width of these trees is computably bounded since

$$|\{\sigma \in 2^n \mid K(\sigma) \leq K(|\sigma|) + c\}| < 2^c$$

. . . by the coding theorem

Number of paths through trees of bounded width

- The number of infinite paths through a tree T with bounded width can be computed from T''.
- This is optimal!
- If a family of trees is computable from a low$_2$ oracle A then the number of paths is computable from $\mathbf{0}^{(2)}$.
- Oracle A is low$_2$ if A'' is computable from $\mathbf{0}^{(2)}$; $\Sigma^0_2(A) \subseteq \Delta^0_3$.

Representing the K_c-trivial classes with simpler trees

Theorem (B. and Tom Sterkenburg)

Given a Δ^0_2 tree T which only has K_c-trivial paths we can compute the index of another Δ^0_2 tree which is K-trivial and has the same infinite paths as the original tree.

The new trees have trivial initial segment complexity.

Fact: $\mathbf{0}^{(2)}$ can compute a low$_2$ index of a K_c-trivial stream given c and the Δ^0_2 index of the stream.

Computation of $G(c)$ from $\mathbf{0}^{(2)}$

- Get the index of the original Δ^0_2 tree representing the class K_c-trivial.
- Compute the index of the K-trivial tree representing this class.
- Use $\mathbf{0}^{(2)}$ to compute a low$_2$-ness index of the new tree.
- Use $\mathbf{0}^{(2)}$ again to compute the number of infinite paths through this tree.
- This is $G(c)$

A related class: low for K streams

If a computer is given access to a powerful oracle, it will achieve better compression for many strings.

X is called low for K if the prefix-free complexity relative to X is the same as the unrelativized prefix-free complexity.

> ......*if X cannot compress strings more than a constant number of bits.*

> ......*if as far as prefix-free complexity is concerned, it is not better than a computable oracle.*

This class was defined by Muchnik in 1999, who also exhibited non-computable elements in it.

Hierarchy of low for K and complexity

- Low for K streams are stratified in a cumulative hierarchy of finite classes.
- Hirschfeldt and Nies showed that they coincide with the K-trivial streams.
- Our methodology applies to this class, showing that

... the corresponding function giving the cardinality of the hierarchy classes is Δ^0_3.

Applications

A consequence of the main result is that $\mathbf{0}''$ can obtain the indices of the K_c-trivial strings.

This can be used to show that a number of K-related objects have lower complexity.

For example, gap functions for K-triviality.

Gap functions for K-triviality

These are non-decreasing unbounded functions f such that

$$\forall n\,[K(X \upharpoonright_n) \leq K(n) + f(n) + c] \Rightarrow X \text{ is } K\text{-trivial.}$$

- Constructed by Csima and Montalbán in 2006
- Used to obtain minimal pairs in the degrees of randomness
- Complexity: Δ^0_4
- Downey raised the question about their complexity

Complexity of gap functions

Theorem (B. and Martijn Baartse)

If f is Δ^0_2 unbounded and non-decreasing then there are uncountably many streams X such that

$$K(X \upharpoonright_n) \leq K(n) + f(n) \quad \textit{for all } n.$$

Theorem (B. and Martijn Baartse)

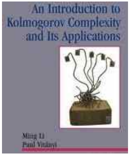
There is a Δ^0_3 gap function for K-triviality.

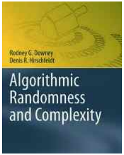
References

- Barmpalias/Sterkenburg, On the number of infinite sequences with trivial initial segment complexity
- Barmpalias/Baartse, On the gap between trivial and nontrivial initial segment prefix-free complexity

- Webpage: `http://www.barmpalias.net`

General references





- Li-Vitanyi, An introduction to Kolmogorov Complexity and its applications, Springer-Verlag.
- Nies, Computability and Randomness, Oxford Press.
- Downey and Hirschfeldt, Algorithmic randomness and complexity, Springer-Verlag.