Solovay functions and K-triviality

Laurent Bienvenu, Wolfgang Merkle, André Nies

Solovay functions and K-triviality *

Laurent Bienvenu[1], Wolfgang Merkle[2], and André Nies[3]

1 **LIAFA – CNRS & Université de Paris 7**
Case 7014, 75205 Paris Cedex 13, France
laurent.bienvenu@liafa.jussieu.fr
2 **Institut für Informatik, Universität Heidelberg**
Im Neuenheimer Feld 294, D-69120 Heidelberg, Germany
merkle@math.uni-heidelberg.de
3 **University of Auckland**
Private Bag 92019, Auckland, New Zealand
andre@cs.auckland.ac.nz

Abstract

As part of his groundbreaking work on algorithmic randomness, Solovay demonstrated in the 1970s the remarkable fact that there are computable upper bounds of prefix-free Kolmogorov complexity K that are tight on infinitely many values (up to an additive constant). Such computable upper bounds are called Solovay functions. Recent work of Bienvenu and Downey [STACS 2009, LIPIcs 3, pp 147-158] indicates that Solovay functions are deeply connected with central concepts of algorithmic randomness such as Ω numbers, K-triviality, and Martin-Löf randomess.

In what follows, among other results we answer two open problems posed by Bienvenu and Downey about the definition of K-triviality and about the Gács-Miller-Yu characterization of Martin-Löf randomess. The former defines a sequence A to be K-trivial if $K(A\upharpoonright_n) \leq^+ K(n)$, the latter asserts that a sequence A is Martin-Löf random iff $C(A\upharpoonright_n) \geq^+ n - K(n)$. So both involve the noncomputable function K. As our main results we show that in both cases $K(n)$ can be equivalently replaced by any Solovay function, and, what is more, that among all computable functions such a replacement is possible exactly for the Solovay functions. Moreover, similar statements hold for the larger class of all right-c.e. in place of the computable functions. These full characterizations, besides having significant theoretical interest on their own, will be useful as tools when working with K-trivial and Martin-Löf random sequences.

1998 ACM Subject Classification F.1.1, F.4.1

Keywords and phrases Algorithmic randomness, Kolmogorov complexity, K-triviality

Digital Object Identifier 10.4230/LIPIcs.STACS.2011.452

1 Introduction

1.1 Algorithmic randomness and Kolmogorov complexity

The goal of the theory of algorithmic randomness is to give a formal meaning to the notion of "random object". For finite discrete objects, such as finite binary sequences or strings, this was achieved by Solomonoff, Kolmogorov and Chaitin via the notion nowadays known as *Kolmogorov complexity*, where then a string is said to be random if it is incompressible in the sense of having roughly maximum Kolmogorov complexity. As usual, for a string w we

* Wolfgang Merkle was partially supported by the Deutsche Forschungsgemeinschaft (DFG) under grant ME 1806/3-1. André Nies was partially supported by the Marsden Fund of New Zealand under grant no. 08-UOA-187.

28th Symposium on Theoretical Aspects of Computer Science (STACS'11).
Editors: Thomas Schwentick, Christoph Dürr; pp. 452–463

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany


distinguish the plain Kolmogorov complexity $C(w)$ and its prefix-free variant $K(w)$. Both are defined as the length of the least description of w with respect to some fixed additively optimal Turing machine U. That is, we ask for the length of a shortest string p such that $U(p) = w$; however, for the prefix-free variant we restrict attention to Turing machines with prefix-free domain [5, 10].

For infinite objects, such as infinite binary sequences (or sequences, for short), various notions of randomness have been proposed and studied. This extensive study led to a consensus that the "best" notion of randomness is *Martin-Löf randomness*, mainly because in many respects, Martin-Löf randomness is well-behaved, in that the main properties of Martin-Löf random sequences do match our intuition of what random sequences should look like. Moreover, the concept of Martin-Löf randomness is robust in the sense that it admits various equivalent definitions that are all natural and intuitively meaningful. For example, Martin-Löf random sequences can be characterized as the sequences that are unpredictable in the sense that certain effectively approximable betting games cannot win on these sequence. In the early 1970s, another important characterization in terms of incompressibility was found by Schnorr — and independently, in a slightly different form, by Levin — which asserts that for any sequence A,

$$A \text{ is Martin-Löf random} \Leftrightarrow K(A\upharpoonright_n) \geq^+ n \tag{1}$$

where K is the prefix-free Kolmogorov complexity and $A\upharpoonright_n$ denotes the prefix of A of length n (here the notation $\geq^+$ means "greater or equal up to a constant additive term that does not depend on the variable n", for formal definitions and more detailed explanations of this and other notation see Section 1.3).

Solovay functions

Bienvenu and Merkle [2] observed that the incompressibility characterization (1) of Martin-Löf randomness remains valid in case the function K is replaced by a suitable computable function f:

$$A \text{ is Martin-Löf random} \Leftrightarrow \left[f(A\upharpoonright_n) \geq^+ n\right], \tag{2}$$

where it is easy to see that in addition the function f can be chosen to be an upper bound for K. Continuing this line of research, Bienvenu and Downey [1] considered "good" upper bounds for K, namely those that are computable, and are tight on infinitely many values. They called such bounds *Solovay functions*, as Solovay [14] was the first to show that such a function exists.

▶ **Definition 1.** A function $g\colon \mathbb{N} \to \mathbb{N}$ is an *upper bound for K (up to an additive constant)* if $K(n) \leq^+ g(n)$, and such a bound is *i.o. tight (up to an additive constant)* if for infinitely many n, $g(n) \leq^+ K(n)$. An i.o. tight upper bound g for K is a *Solovay function* in case g is computable. It is a *weak Solovay function* in case g is right-c.e.

Thus, K itself is a weak Solovay function. Among other results to be discussed below, Bienvenu and Downey demonstrated that any computable function g for which the equivalence (2) holds true must be a Solovay function.

▶ **Theorem 2** (Bienvenu-Downey)**.** *Any computable upper bound f of K which satisfies the equivalence*

$$A \textit{ is Martin-Löf random} \Leftrightarrow \left[f(A\upharpoonright_n) \geq^+ n\right]$$

is a Solovay function.

Left-c.e. reals and Ω numbers

Infinite binary sequences can be identified with binary expansions of reals in the unit interval in the canonical way. Then, a sequence is called *left-c.e.* if the corresponding real is the limit of an effectively given nondescending sequence of rational numbers. Martin-Löf random sequences that are left-c.e. exist, and have very interesting properties. For example, a left-c.e. real is Martin-Löf random if and only if it is Solovay complete, i.e., has only effective approximations from below that are as slow as any other effective approximation from below to any other left-c.e. real, up to a constant factor [5, 9]. Furthermore, Martin-Löf random left-c.e. reals can be characterized as the measures of the domains of universal Turing machines [5]. Letting

$$\Omega_g = \sum_{n\in\mathbb{N}} 2^{-g(n)}$$

one obtains as a variant of the latter result that a left-c.e. real is Martin-Löf random if and only if the real can be written in the form $\Omega_{\widetilde{K}}$ for some variant $\widetilde{K}$ of K obtained by using an alternate universal prefix-free Turing machine. In particular Ω_K is Martin-Löf random [5]. A full characterization of the computable functions g such that Ω_g is Martin-Löf random as the functions that are i.o. tight upper bounds for K was obtained by Bienvenu and Downey [1], and was extended to the class of right-c.e. functions by Hölzl et al. [8].

▶ **Theorem 3** (Bienvenu-Downey, Hölzl-Kräling-Merkle). *Let $g\colon \mathbb{N}\to\mathbb{N}$ be a right-c.e. function. Then g is a weak Solovay function if and only if Ω_g is a Martin-Löf random real. In particular, a computable function g is a Solovay function if and only if Ω_g is a Martin-Löf random real.*

Observe in this connection that by easy standard arguments, first, for any right-c.e. function g, the real Ω_g is finite if and only if g is an upper bound for K and second, a real is left-c.e. if and only if it can be written in the form Ω_g for some right-c.e. function g such that Ω_g is finite (where one exploits that left-c.e. reals have effective approximations from below by dyadic rationals, i.e., rationals of the form $p/2^q$ where $p, q \in \mathbb{N}$). Together, exactly the left-c.e. reals can be written in the form Ω_g for some right-c.e. upper bound g of K, and Theorem 3 states that the upper bound can be chosen to be i.o. tight if and only if the real is Martin-Löf random.

K-trivial sequences

From their incompressibility characterization, it can be seen that the Martin-Löf random sequences are those which have initial segments of roughly maximal Kolmogorov complexity. It is natural to ask which sequences have initial segments of *minimal* Kolmogorov complexity. It is immediate that any computable sequence has minimal Kolgomorov complexity because for such a sequence the prefix of any given length n will have the same Kolmorogov complexity as n itself, up to a fixed additive constant, which is then minimal since any code for the prefix can also be used as a code for n. Indeed, Chaitin [4] showed that the sequences A such that $C(A\upharpoonright_n) \leq^+ C(n)$ are exactly the computable ones. On the other hand, this is not true any longer for the class of sequences A such that $K(A\upharpoonright_n) \leq^+ K(n)$. While Chaitin [4] proved that any such sequence is computable from the halting problem, Solovay [14] was able to construct such a sequence that is noncomputable and computably enumerable. The class of such sequences was further studied by Downey, Hirschfeldt, Nies and Stephan [6, 12], who called these sequences *K-trivial*.

The K-trivial sequences turned out to have remarkable properties. Perhaps the most striking ones are that they can be characterized as the sequences that are low for Martin-Löf

randomness, or, alternatively, as the sequences that are low for prefix-free Kolmogorov complexity. In other words, a sequence A is K-trivial if and only if Martin-Löf randomness relativized to A coincides with Martin-Löf randomness, if and only if the prefix-free Kolmogorov complexity relativized to A is within an additive constant of the unrelativized one. There are many more interesting results about K-trivial sequences. We refer the reader to the books by Downey and Hirschfeldt [5], and by Nies [13].

In Section 2 we will argue that in the definition of the notion of K-trivial, the upper bound $K(n)$ can be equivalently replaced by any weak Solovay function, and that in fact the ability to do so characterizes the Solovay functions and the weak Solovay functions. A preliminary result in this direction was obtained by Bienvenu and Downey [1], who showed that K-triviality can be characterized via some particular Solovay function.

▶ **Theorem 4** (Bienvenu-Downey). *There exists a Solovay function g such that for all A,*

$$A \text{ is } K\text{-trivial} \Leftrightarrow \left[K(A\upharpoonright_n) \leq^+ g(n)\right]$$

The Gács-Miller-Yu Theorem

In view of the incompressibility characterization of Martin-Löf randomness in terms of prefix-free Kolmogorov complexity, it is suggestive to ask whether a similar characterization in terms of plain Kolmogorov complexity is possible. A first result in this direction was obtained by Gács [7] using conditional plain Kolmogorov complexity. He showed that for any sequence A,

$$A \text{ is Martin-Löf random} \Leftrightarrow \left[C(A\upharpoonright_n | n) \geq^+ n - K(n)\right].$$

Much later, Miller and Yu [11] were able to show that this equivalence remains true when conditional plain Kolmogorov complexity is replaced by its unconditional counterpart. That is, for any sequence A,

$$A \text{ is Martin-Löf random} \Leftrightarrow \left[C(A\upharpoonright_n) \geq^+ n - K(n)\right].$$

At the same time Miller and Yu showed that in addition the equivalence remains valid in case the term $K(n)$ is replaced by a suitable computable function g (a variation of the original Solovay function built by Solovay), which yields their celebrated characterization of Martin-Löf randomness based solely on plain Kolmogorov complexity: for some computable function g and for any sequence A,

$$A \text{ is Martin-Löf random} \Longleftrightarrow \left[C(A\upharpoonright_n) \geq^+ n - g(n)\right].$$

For a simplified proof of their result see Bienvenu et al. [3].

1.2 Overview

By results discussed above, and by many other results not mentioned here, prefix-free Kolmogorov complexity is one of the most central notions in algorithmic randomness, and is indeed closely related to many other fundamental concepts in this area. In particular, as discussed above, the following assertions all become true in case we let g be equal to prefix-free Kolmogorov complexity K.

(i) The real Ω_g is Martin-Löf random.
(ii) A sequence A is K-trivial if and only if $K(A\upharpoonright_n) \leq^+ g(n)$.

(iii) A sequence A is Martin-Löf random if and only if $C(A\upharpoonright_n) \geq^+ n - g(n)$.

However, known results suggest that these close relations might not just hold for prefix-free Kolmogorov complexity but also for Solovay functions and weak Solovay functions in general. As stated above, the first assertion is true for a right c.e. function g if and only if g is a weak Solovay function [8]. Hence, as a special case, the first assertion is true for a computable function g if and only if g is a Solovay function [1]. For the two other assertions, on the other hand, it is only known that the second assertion is true for some Solovay function g [1], and, by the aforementioned result of Miller and Yu, that the third assertion holds true for some computable function g, while any function of the latter type must be a Solovay function [1].

In the present paper, we will investigate the question of which functions g make the second and third assertion true. Similar to the first assertion, we obtain a full characterization in the sense that the second as well as the third assertion is true for a right-c.e. function g if and only if g is a weak Solovay function, hence, is true for a computable function g if and only if g is a Solovay function.

▶ Remark. The result of Bienvenu and Downey that any computable upper bound f of K which satisfies the equivalence

$$A \text{ is Martin-Löf random} \Leftrightarrow \left[f(A\upharpoonright_n) \geq^+ n\right]$$

must be a Solovay function does not extend to a characterization of Solovay functions, i.e., there are Solovay functions for which this equivalence is wrong. Indeed one can easily construct a Solovay function which is tight only on highly compressible sequences: take a Solovay function g, and define f by $f(0^n) = g(n)$ for all n, and $f(\sigma) = 3|\sigma|$ for all the other strings σ. It is clear that f is a Solovay function, but for $A = 10000\ldots$, one has $f(A\upharpoonright_n) =^+ 3n \geq^+ n$, hence f does not characterize Martin-Löf randomness.

1.3 Notation

Here we gather some notation that will be used throughout the paper. A (binary) string is a finite sequence over the alphabet $\{0,1\}$. The set of all strings is denote by $\{0,1\}^*$, while $\{0,1\}^n$ and $\{0,1\}^{\leq n}$ denote the set of strings of length n and of length at most n, respectively. Strings are identified with natural numbers via the order isomorphism that takes the length-lexicographical order on strings to the usual order on $\mathbb{N} = \{0,1,\ldots\}$, for example, the empty string λ is identified with the natural number 0. Sequence refers to an infinite binary sequence, unless explicitly stated otherwise, and the set of sequences is denoted by $\{0,1\}^\omega$. For a sequence A, we write $A = A(0)A(1)\ldots$ and the prefix of A of length i is denoted by $A\upharpoonright_i = A(0)\ldots A(i-1)$.

For a string σ, the *cylinder* $[\sigma]$ is the set of sequences A such that σ is a prefix of A. If S is a *set* of strings, we write $[S]$ for the set of sequences having some prefix in S, i.e. $S = \bigcup_{\sigma\in S}[\sigma]$. When we talk about *measure* on the space $\{0,1\}^\omega$ of sequences, we mean Lebesgue measure μ, which is the probability measure one gets when each bit of a sequence is chosen at random with probability $(1/2, 1/2)$ independently of all the other bits.

For functions f and g defined on some domain D such as the set of all strings or all natural numbers, the notation $f(n) \leq^+ g(n)$ means that there is some constant c such that for all $n \in D$ we have $f(n) \leq f(n) + c$, and $f(n) \geq^+ g(n)$ and $f(n) =^+ g(n)$ are defined likewise. Observe that this notation comprises a universal quantifier that ranges over D, hence it is slight abuse of notation, though straightforward, to extend to statements to statements such as "$f(n) \leq^+ g(n)$ holds for all n in some subset D_0 of D".

Plain Kolmogorov complexity is denoted by C, and its prefix-free variant by K; for definitions and further explanations we refer to the literature [5, 13, 10]. Kolmogorov complexity (plain or prefix-free) is defined on the set of finite string, but as usual we also apply it to other objects (integers, rational numbers, pairs of strings, etc.) as long as they can be encoded into finite strings in a computable way.

A function $f : D \to \mathbb{R}$ is right-c.e. (a.k.a. *approximable or semi-computable from above*) if there exists a computable function $F : D \times \mathbb{N} \to \mathbb{Q}$ such that for all $x \in D$, the values $F(x,0), F(x,1), \ldots$ are nonincreasing and converge to $f(x)$ (the value $F(x,t)$ is called the approximation of $f(x)$ at stage t and is often denoted by $f_t(x)$ when the choice of a particular F is irrelevant in the argument). The plain and prefix-free variants of Kolmogorov complexity are examples of right-c.e. functions.

A bounded request set (a.k.a. Kraft-Chaitin set) is a computably enumerable set W of pairs (σ, n) of a string σ and a natural number n such that $\sum_{(\sigma,n)\in W} 2^{-n}$ is finite (enumerating a pair (σ, n) into a request set is often said to incur a *cost* of 2^{-n}; the request set being bounded if the total cost is finite). Having such a set, the *Kraft-Chaitin theorem* [5, 10, 13] asserts that for all $(\sigma, n) \in W$, one has $K(\sigma) \leq^+ n$.

2 K-triviality and Solovay functions

In this section, we prove that for any right-c.e. function g the equivalence

$$A \text{ is } K\text{-trivial} \iff [K(A\upharpoonright_n) \leq^+ g(n)] \tag{3}$$

holds if and only if g is a weak Solovay function, i.e., if and only if g is an i.o. tight upper bound for K. Hence, in particular, for computable g, the equivalence (3) holds if and only if g is a Solovay function. Note that any function g that satisfies equivalence (3) must already be an upper bound of K, since $K(A\upharpoonright_n)$ is always greater or equal to $K(n)$, up to an additive constant.

2.1 Solovay functions characterize K-triviality

We begin with the first part of the equivalence result, namely that K-triviality is characterized by weak Solovay functions and thus, in particular, by Solovay functions.

▶ **Theorem 5.** *Let g be a weak Solovay function. If $K(A\upharpoonright_n) \leq^+ g(n)$, then A is K-trivial.*

As mentioned earlier (Theorem 4), this was proven by Bienvenu and Downey for a particular Solovay function, actually the one originally built by Solovay, which we call g_S. Their proof involved the construction of a bounded request set (or Kraft-Chaitin set), a standard technique to ensure the K-triviality of a sequence. However, it relied on the particular properties of the function g_S. We now show that given any weak Solovay function h and a sequence A such that $K(A\upharpoonright_n) \leq^+ h(n)$, one can construct a bounded request set that ensures $K(A\upharpoonright_n) \leq^+ g_S(n)$, hence proving the K-triviality of A. This is achieved by the following technical proposition, which will guarantee that building a bounded request set to ensure $K(A\upharpoonright_n) \leq^+ g_S(n)$ does not "cost more" (in a specific sense to be explained below) than building a bounded request set to ensure $K(A\upharpoonright_n) \leq^+ h(n)$.

▶ **Lemma 6.** *Let g be a Solovay function, and h a weak Solovay function. There exists a positive constant c and a computable partition of $\mathbb{N}$ into intervals $(I_n)_{n\in\mathbb{N}}$ such that for all n*

$$2^{-g(n)} \leq 2^c \sum_{i\in I_n} 2^{-h(i)}$$

Proof. We design a procedure which uniformly in k tries to construct a partition $(I_n^{(k)})_{n\in\mathbb{N}}$ such that $2^{-g(n)} \leq 2^k \sum_{i\in I_n} 2^{-h(i)}$. The procedure goes as follows:

For n from 0 to ∞ do

(1) Let $s(k,n) \in \mathbb{N}$ be the first integer which does not belong to one of the previously constructed intervals $I_j^{(k)}$ for $j < n$.
(2) Wait until we find some t large enough to have

$$\sum_{i=s(k,n)}^{t} 2^{-h_t(i)} \geq 2^{-k}2^{-g(n)}$$

(3) When this happens, we define $I_n^{(k)}$ to be $[s(k,n), t]$.

It is possible that for some (k,n) the procedure of parameter k waits at step 2 forever while executing the n-loop. When this happens, we have by construction:

$$\sum_{i\geq s(k,n)} 2^{-h(i)} \leq 2^{-k}2^{-g(n)}$$

Hence by the Kraft-Chaitin theorem, for all $i \geq s(k,n)$:

$$K(i) \leq^+ K(k,n,s(k,n)) + h(i) - k - g(n)$$

Since the construction is effective, $s(k,n)$ can be described via the pair (k,n) alone, hence $K(s(k,n)) \leq^+ K(k,n) \leq^+ K(n) + 2\log k$. This, together with the above inequality and the fact that $K(n) \leq^+ g(n)$ (because g is a Solovay function) yields for all $i \geq s(k,n)$:

$$K(i) \leq^+ h(i) - k + 2\log k$$

Now, recall that h is a weak Solovay function so $K(i) \geq^+ h(i)$ for infinitely many i. Therefore the above situation can only happen for a finite number of k. In other words, for all k large enough, the procedure never waits forever at step 2 and hence produces effectively a partition $(I_n^{(k)})_{n\in\mathbb{N}}$ of $\mathbb{N}$ into intervals such that for all n, and each $I_n^{(k)} = [s,t]$ we obtain as wanted

$$2^{-k}2^{-g(n)} \leq \sum_{i=s}^{t} 2^{-h_t(i)} \leq \sum_{i=s}^{t} 2^{-h(i)}.$$

◀

▶ **Corollary 7.** *For every weak Solovay function h, there exists a Solovay function $\tilde{h}$ such that $h \leq \tilde{h}$.*

Proof. Let h be a weak Solovay function and let g be any Solovay function. By Lemma 6, there exists a constant c and a computable partition $(I_n)_{n\in\mathbb{N}}$ of $\mathbb{N}$ into intervals such that for all n

$$2^{-g(n)} \leq 2^c \sum_{i\in I_n} 2^{-h(i)}$$

Let $\tilde{h} : \mathbb{N} \to \mathbb{N}$ be the function defined as follows. For a given i, let I_n be the interval to which i belongs, and set

$$\tilde{h}(i) = h_t(i) \text{ where } t \text{ is the least integer s.t. } 2^{-g(n)} \leq 2^c \sum_{i\in I_n} 2^{-h_t(i)}$$

It is clear that $\tilde{h}$ is computable and is an upper bound of h. Moreover, the sum

$$\sum_i 2^{-\tilde{h}(i)} = \sum_n \sum_{i \in I_n} 2^{-\tilde{h}(i)}$$

is random. Indeed, by construction for all n, $\sum_{i \in I_n} 2^{-\tilde{h}(i)} \geq^{\times} 2^{-g(n)}$. Hence $\sum_n 2^{-g(n)}$ is Solovay reducible to $\sum_i 2^{-\tilde{h}(i)}$ (the former being random, the latter must be too by the Kučera-Slaman theorem [9]). Therefore $\tilde{h}$ is a Solovay function. ◀

We are now ready to prove Theorem 5. Let h be a weak Solovay function, d a constant and A a sequence such that $K(A\upharpoonright_n) \leq h(n)+d$ for all n. We want to prove that A is K-trivial. Since by Corollary 7 any weak Solovay function is dominated by a Solovay function, we only need to prove this theorem for h computable. We apply Lemma 6 to get a constant c and a computable partition of $\mathbb{N}$ into intervals $(I_n)_{n\in\mathbb{N}}$ such that for all n, $2^{-g_S(n)} \leq 2^c \sum_{i\in I_n} 2^{-h(i)}$. Without loss of generality, we also assume that for all n, $n < \min(I_n)$ (this can be ensured easily in the proof of Lemma 6).

We show that A is K-trivial by building a bounded request set. For all n and all strings σ of length n, we wait until we find an extension τ of σ whose length is $\max(I_n)$ and such that for all $i \in I_n$, some description of $\tau\upharpoonright_i$ of length at most $h(i)+d$ is in the domain of $\mathbb{U}$ (by "description" we mean a string p such that $\mathbb{U}(p) = \tau\upharpoonright_i$, where $\mathbb{U}$ is the universal prefix-free machine defining K). When (and if) this happens (we know when it does by computability of h), we enumerate a pair $(\sigma, g_S(n)+c+d)$ in our request set. The cost of this for us is $2^{-g_S(n)-c-d}$, which we can account against the cost for $\mathbb{U}$ to enumerate descriptions of $\tau\upharpoonright_i$ as above, which is at least $\sum_{i\in I_n} 2^{-h(i)-d}$, which in turn is at least $2^{-g_S(n)-c-d}$ by construction of the intervals I_n. Hence, we never spend more than $\mathbb{U}$ does, which ensures that our request set is bounded. Now, by assumption on A, for every n, for every $i \in I_n$, the universal machine must issue a description of $A\upharpoonright_i$ of length at most $h(i)+d$, hence some pair $(A\upharpoonright_n, g_S(n)+c+d)$ enters our bounded request set at some point. Therefore, for all n, $K(A\upharpoonright_n) \leq g_S(n)+c+d$. Applying Theorem 4, this shows that A is K-trivial.

2.2 K-triviality characterizes Solovay functions

We now prove that any right-c.e. function g that makes the equivalence

$$A \text{ is } K\text{-trivial} \iff [K(A\upharpoonright_n) \leq^+ g(n)] \tag{4}$$

true is a weak Solovay function, and hence is a Solovay function in case g is computable. In the proof of our result, we need only to consider the case where g is an upper bound for K because otherwise the class of sequences A that satisfy the right-hand side of equivalence (4) is empty. We then prove the stronger fact that in the case g is a right-c.e. upper bound for K but is not a weak Solovay function, there are uncountably many sequences A such that $K(A\upharpoonright_n) \leq^+ g(n)$. This is enough for our purposes, since there are only countably many K-trivial sequences (indeed, as we mentioned earlier, they are all computable in the halting problem).

▶ **Theorem 8.** *Let g be a right-c.e. function such that $K(n) \leq^+ g(n)$ but where g is not a weak Solovay function. Then the set $\{A \mid K(A\upharpoonright_n) \leq^+ g(n)\}$ is uncountable.*

Proof. We will build an increasing sequence $a_1 < a_2 < a_3 < \ldots$ of integers such that any subset A of $\{a_1, a_2, a_3, \ldots\}$ satisfies $K(A\upharpoonright_n) \leq^+ g(n)$.

The sequence is defined by induction (but not effectively), where we set $a_1 = 0$ and where we ensure by induction that for all k, for any subset B of the finite set $\{a_1, \dots, a_k\}$ and for all $n \geq a_k$, for some constant d that does neither depend on B nor on k we have that

$$K(B\upharpoonright_n) \leq g(n) - k + d. \tag{5}$$

This suffices to prove the desired result: let A be any subset of $\{a_1, a_2, a_3, \dots\}$, and let n be some position. Let k be such that $a_k \leq n < a_{k+1}$. Let $B = A \cap \{a_1, \dots, a_k\}$. Since $B\upharpoonright_n = A\upharpoonright_n$, one has by the above property $K(A\upharpoonright_n) \leq^+ g(n) - k \leq^+ g(n)$.

We now explain the inductive definition of the sequence a_k. Suppose we have already defined $a_1, \dots, a_k$ with the property (5). Let us choose c to be a very large integer, say $c > 2a_k + k + 1$. Consider the sum $\Omega_g = \sum_n 2^{-g(n)}$. By Theorem 3, this is not a random real as g is not a weak Solovay function. Hence, there exists a prefix σ of Ω_g such that $K(\sigma) \leq |\sigma| - c$. Let p be a shortest description for σ. Knowing p, one can effectively perform the following operations: first, retrieve $\sigma = \mathbb{U}(p)$; then, enumerate Ω_g from below and wait until it becomes larger than the real value $0.\sigma$ (treated as a real number written in binary) using the approximation of the values $g(n)$ from above; when this happens, let a_{k+1} be the least number m such that for all $i \geq m$, so far there has been no contribution to Ω_g by the value $g(i)$ (more precisely, via the approximation of these values from above). Since σ is a prefix of Ω_g, this means in particular that $\sum_{n \geq a_{k+1}} 2^{-g(i)}$ does not exceed $2^{-|\sigma|}$, so by the Kraft-Chaitin theorem, any integer $n \geq a_{k+1}$ can be described by p and some additional $g(n) - |\sigma|$ bits of information. Therefore, if $n \geq a_{k+1}$ and B is a subset of $\{a_1, \dots, a_{k+1}\}$, then $B\upharpoonright_n$ can be described in a prefix-free way by

- $B\upharpoonright_{a_k}$,
- p (from which a_{k+1} can be retrieved),
- the single bit $B(a_{k+1})$,
- some additional $g(n) - |\sigma|$ bits.

Thus $K(B\upharpoonright_n) \leq^+ 2a_k + |p| + 1 + g(n) - |\sigma| \leq^+ g(n) - (k+1)$ (using the fact that $c > 2a_k + k + 1$ and $|p| \leq |\sigma| - c$). This concludes the inductive step. ◀

3 Solovay functions and the Gács-Miller-Yu theorem

We now turn to the link between Solovay functions and the Gács-Miller-Yu theorem. Recall from the introduction that this theorem states that a sequence A is Martin-Löf random if and only if $C(A\upharpoonright_n) \geq^+ n - K(n)$, and that moreover there exists a computable upper bound f of K such that A is Martin-Löf random if and only if $C(A\upharpoonright_n) \geq^+ n - f(n)$. Bienvenu and Downey proved that any such function f must be a Solovay function. We now prove the converse, i.e. that *any* Solovay function makes this equivalence true, and the same is true for weak Solovay functions.

▶ **Theorem 9.** *Let g be a (weak) Solovay function. The following are equivalent.*
(i) $A \in \{0,1\}^\omega$ is Martin-Löf random.
(ii) $C(A\upharpoonright_n) \geq^+ n - g(n)$.

We begin our proof with a combinatorial lemma.

▶ **Lemma 10.** *Let σ be a string. Let $I = [s, t]$ be a finite interval of integers with $s \geq |\sigma|$. Let $(a_i)_{i \in I}$ be a finite set of integers such that*

$$\sum_{i \in I} a_i 2^{-i} \geq 2^{-|\sigma|+1}.$$

Then, there exists a subset J of I and a finite set of strings S such that

(i) $[S] = [\sigma]$
(ii) *for all* $\tau \in S$, $|\tau| \in J$
(iii) *for all* $j \in J$, $|S \cap \{0,1\}^{\leq j}| \leq a_j$
Moreover, J *and* S *can be constructed effectively given* σ, I *and* $(a_i)_{i\in I}$.

Proof. We construct J and S via the following procedure. We initialize J and S to $\emptyset$. Now the procedure is as follows:

For all i from s to t do
If $a_i \leq |S|$ do nothing. Otherwise:

(1) Put i into J
(2) Split $[\sigma] \setminus [S]$ into cylinders of measure 2^{-i}. Let T be the set of strings of length i generating those cylinders.
(3) Let T' be the set containing the $c_i = a_i - |S|$ first strings of T in the lexicographic order (if $c_i > |T|$ then let $T' = T$).
(4) Enumerate all strings of T' into S.

We now verify that this procedure works, i.e., that the algorithm is well-defined and that the set S we obtain after the t-loop is as wanted. First, notice that at the beginning of the i-loop, S contains only strings of length smaller than i, therefore $[S]$ can be split into cylinders of measure 2^{-i}. Since $|\sigma| \leq s \leq i$, this is also the case for $[\sigma]$, hence for $[\sigma] \setminus [S]$, so step (2) is well-defined. We also immediately see that the conditions (ii) and (iii) of the lemma are satisfied: indeed, we only enumerate strings of a given length i after enumerating i into J, and if we do so, we ensure that at the end of the i-loop, the cardinality of $S \cap \{0,1\}^{\leq i}$ is at most a_i. It remains to verify condition (i). First it is clear that $S \subseteq [\sigma]$ as we only enumerate cylinders that are contained in $[\sigma]$. Suppose that this inclusion is strict. Then, when running the above procedure, at step 3, we are never in the case where $c_i > |T|$, hence for all i, at the end of i-loop, we have $|S \cap \{0,1\}^{\leq i}| \geq a_i$, whether i is in J or not. Therefore, at the end of the procedure, we have:

$$\sum_{i=s}^{t} a_i 2^{-i} \leq \sum_{i=s}^{t} |S \cap \{0,1\}^{\leq i}| 2^{-i} \leq \sum_{i=s}^{t} \sum_{k=s}^{i} |S \cap \{0,1\}^k| 2^{-i} \leq \sum_{k=s}^{t} |S \cap \{0,1\}^k| \sum_{i=k}^{t} 2^{-i}$$
$$< \sum_{k=s}^{t} |S \cap \{0,1\}^k| 2^{-k+1} < 2\mu([S]) < 2\mu([\sigma]) < 2^{-|\sigma|+1}$$

and this contradicts the hypothesis of the lemma. ◀

Proof of Theorem 9. The part $(i) \to (ii)$ follows directly from the Gács-Miller-Yu theorem. We prove the converse. Let g be a weak Solovay function and $A \in \{0,1\}^\omega$ a sequence which is not Martin-Löf random. We shall prove that $C(A\upharpoonright_n) \leq n - g(n) - k$ for infinitely many n and arbitrarily large k. By Corollary 7, we can assume that g is computable. We further assume, for technical reasons which will become clear at the end of the proof, that for all i, either $g(i) \leq 2\log(i)$ or $g(i) = +\infty$. If it is not the case, replace g by the bigger function $\tilde{g}$ defined by $\tilde{g}(i) = g(i)$ if $g(i) \leq 2\log(i)$, and $\tilde{g}(i) = +\infty$ otherwise. Then we have:

$$\sum_i 2^{-\tilde{g}(i)} = \sum_i 2^{-g(i)} - \sum_{\substack{i \\ g(i) \geq 2\log i}} 2^{-g(i)}$$

the third sum is a computable real number as the i-th term is bounded by $1/i^2$. Thus $\sum_i 2^{-\tilde{g}(i)}$ is equal to a random real minus a computable real, hence is a random real and thus $\tilde{g}$ is still a Solovay function.

Now, let $(\mathcal{U}_k)_{k\in\mathbb{N}}$ be a Martin-Löf test covering A and such that $\mu(\mathcal{U}_k) \leq 2^{-2k-1}$ for all k. We design a procedure (P_k) which for all k tries to enumerate a set of strings S_k such that $[S_k] = \mathcal{U}_k$, with additional properties on the length of the strings it contains. We ensure that this procedure succeeds for almost all k by building an auxiliary test $\mathcal{V}_k$ which tests the randomness of $\sum_i 2^{-g(i)}$. The procedure (P_k) works as follows.

(1) Wait for a new cylinder $[\sigma]$ to be enumerated into $\mathcal{U}_k$.
(2) Choose a large integer s, say larger than 2^N with N larger than any integer mentioned so far in the construction (including k).
(3) Enumerate into $\mathcal{V}_k$ the real dyadic interval

$$\left[\sum_{i<s} 2^{-g(i)}, 2^{-|\sigma|+1+k} + \sum_{i<s} 2^{-g(i)}\right]$$

(4) Wait for a stage t such that

$$\sum_{i\leq t} 2^{-g(i)} > 2^{-|\sigma|+1+k} + \sum_{i<s} 2^{-g(i)}$$

(5) When this happens, we have $\sum_{i=s}^{t} 2^{-g(i)} > 2^{-|\sigma|+1+k}$. We then apply Lemma 10 with $a_i = 2^{i-g(i)-k}$ to get a finite set of strings S_k^σ and a finite set of integers J_k^σ such that $[S_k^\sigma] = [\sigma]$, for all $\tau \in S_k^\sigma$, $|\tau| \in J_k^\sigma$ and for all $j \in J_k^\sigma$, $|S^\sigma \cap \{0,1\}^{\leq j}| \leq a_j$. We then put all strings of S_k^σ into S_k and go back to step 1.

It is possible that for some k, (P_k) will at some point reach step 4 and wait there forever. We claim that this can only happen for finitely many k. Indeed, for a given k, we have $\mu(\mathcal{V}_k) \leq 2^{-k}$, because whenever a cylinder $[\sigma]$ enters $\mathcal{U}_k$ at step 1, an interval of length $2^{-|\sigma|+1+k}$ enters $\mathcal{V}_k$, hence $\mu(\mathcal{V}_k) \leq 2^{k+1}\mu(\mathcal{U}_k) \leq 2^{-k}$. Thus, $(\mathcal{V}_k)_{k\in\mathbb{N}}$ is a Martin-Löf test. Furthermore, if the procedure for S_k waits forever at some step 4, this precisely means that $\sum_i 2^{-g(i)}$ belongs to the dyadic interval which was put into $\mathcal{V}_k$ at step 3, and thus in that case $\sum_i 2^{-g(i)} \in \mathcal{V}_k$. Since $\sum_i 2^{-g(i)}$ is random, it can only belong to finitely many $\mathcal{V}_k$, hence for almost all k the procedure (P_k) never waits forever at step 4. In that case, the c.e. set S_k it builds does satisfy $[S_k] = \mathcal{U}_k$ by construction.

To finish the proof, let k be such that (P_k) succeeds. Since A is not Martin-Löf random, A belongs to $\mathcal{U}_k$, hence to $[S_k]$. This means that for some n, $A\upharpoonright_n$ belongs to S_k. To describe $A\upharpoonright_n$, it suffices to describe k (this can be done with $2\log k + O(1)$ bits), and its position inside S_k. For its position inside S_k, we simply describe the position of $A\upharpoonright_n$ inside the S_k^σ it belongs to, when the latter is sorted in the length-lexicographic order. By construction of S_k^σ, n must be in J_k^σ (otherwise S_k^σ would be empty), and there are at most $a_n = 2^{n-g(n)-k}$ strings of length less than or equal to n in S_k^σ, and therefore we can specify the position of $A\upharpoonright_n$ inside S_k^σ with $n - g(n) - k$ bits. Thus, our description of $A\upharpoonright_n$ has total length $n - g(n) - k + 2\log k + O(1)$. Since k can be taken as large as wanted, this will be enough to prove the theorem, but one last thing we need to check is that this description is enough to retrieve $A\upharpoonright_n$. Indeed, while we give the index of $A\upharpoonright_n$ inside the S_k^σ it belongs to, we do not describe σ explicitly. However, σ can be found as follows. The description of $A\upharpoonright_n$ we give has length $n - g(n) - k + 2\log k + O(1)$. By assumption, $g(n) \leq 2\log n$ and by construction of S_k^σ, $k \leq \log s \leq \log n$. Hence our description has length between $n - 3\log n + O(1)$ and $n + O(1)$. Hence the length of our description gives us n with logarithmic precision. This is

enough to find the string σ such that $A\restriction_n$ belongs S_k^σ because by construction of S_k, if l is the length of some string in $S_k^{\sigma'}$ with $\sigma' \neq \sigma$, then either $2^l < n$ or $2^n < l$, and hence either $l < n - 3\log n$ or $n < l - 3\log l$. ◀

References

1. Laurent Bienvenu and Rodney Downey. Kolmogorov complexity and Solovay functions. In *Symposium on Theoretical Aspects of Computer Science (STACS 2009)*, volume 09001 of *Dagstuhl Seminar Proceedings*, pages 147–158, http://drops.dagstuhl.de/opus/volltexte/2009/1810, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany.
2. Laurent Bienvenu and Wolfgang Merkle. Reconciling data compression and Kolmogorov complexity. In *International Colloquium on Automata, Languages and Programming (ICALP 2007)*, volume 4596 of *Lecture Notes in Computer Science*, pages 643–654. Springer, 2007.
3. Laurent Bienvenu, Wolfgang Merkle, and Alexander Shen. A simple proof of Miller-Yu theorem. *Fundamenta Informaticae*, 83(1-2):21–24, 2008.
4. Gregory Chaitin. A theory of program size formally identical to information theory. *Journal of the Association for Computing Machinery*, 22:329–340, 1975.
5. Rodney Downey and Denis Hirschfeldt. *Algorithmic randomness and complexity*. Springer, 2010.
6. Rodney Downey, Denis Hirschfeldt, André Nies, and Frank Stephan. Trivial reals. In *Proceedings of the 7th and 8th Asian Logic Conferences*, pages 103–131. Singapore University Press, 2003.
7. Peter Gács. Exact expressions for some randomness tests. *Z. Math. Log. Grdl. M.*, 26:385–394, 1980.
8. Rupert Hölzl, Thorsten Kräling, and Wolfgang Merkle. Time-bounded Kolmogorov complexity and Solovay functions. In *Mathematical Foundations of Computer Science (MFCS 2009)*, volume 5734 of *Lecture Notes in Computer Science*, pages 392–402, 2009.
9. Antonin Kučera and Ted Slaman. Randomness and recursive enumerability. *SIAM Journal on Computing*, 31:199–211, 2001.
10. Ming Li and Paul Vitányi. *An introduction to Kolmogorov complexity and its applications*. Texts in Computer Science. Springer-Verlag, New York, 3rd edition, 2008.
11. Joseph Miller and Liang Yu. On initial segment complexity and degrees of randomness. *Transactions of the American Mathematical Society*, 360(6):3193–3210, 2008.
12. André Nies. Lowness properties and randomness. *Advances in Mathematics*, 197(1):274–305, 2005.
13. André Nies. *Computability and randomness*. Oxford Logic Guides. Oxford University Press, 2009.
14. Robert Solovay. Draft of a paper (or series of papers) on Chaitin's work. Unpublished notes, 215 pages, 1975.