Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Dimension 1 sequences are close to randoms ☆

Noam Greenberg [a], Joseph S. Miller [b], Alexander Shen [c], Linda Brown Westrick [d,*]

[a] School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, Wellington, New Zealand
[b] Department of Mathematics, University of Wisconsin–Madison, 480 Lincoln Dr., Madison, WI 53706, USA
[c] LIRMM, CNRS & University of Montpellier, 161 rue Ada, 34095, Montpellier, France
[d] Department of Mathematics, University of Connecticut, 341 Mansfield Rd U-1009, Storrs, CT 06269, USA

ARTICLE INFO

Article history:
Received 28 February 2017
Accepted 27 September 2017
Available online xxxx
Communicated by L.M. Kirousis

Keywords:
Algorithmic randomness
Effective dimension
Harper's theorem
Kolmogorov complexity
Shannon entropy

ABSTRACT

We show that a sequence has effective Hausdorff dimension 1 if and only if it is coarsely similar to a Martin-Löf random sequence. More generally, a sequence has effective dimension s if and only if it is coarsely similar to a weakly s-random sequence. Further, for any $s < t$, every sequence of effective dimension s can be changed on density at most $H^{-1}(t) - H^{-1}(s)$ of its bits to produce a sequence of effective dimension t, and this bound is optimal.

The theory of algorithmic randomness defines an individual object in a probability space to be *random* if it looks plausible as an output of a corresponding random process. The first and the most studied definition was given by Martin-Löf [16]: a random object is an object that satisfies all "effective" probability laws, i.e., does not belong to any effectively null set. (See [4,23,22] for details; we consider only the case of uniform Bernoulli measure on binary sequences, which corresponds to independent tossings of a fair coin.) It was shown by Schnorr and Levin (see [20,21,12]) that an equivalent definition can be given in terms of description complexity: a bit sequence $X \in 2^\omega$ is Martin-Löf (ML) random if and only if the prefix-free complexity of its n-bit prefix $X\restriction_n$ is at least $n - O(1)$. (See [14,23,22] for the definition of prefix-free complexity and for the proof of this equivalence; one may use also monotone or a priori complexity.) This robust class also has an equivalent characterization based on martingales that goes back to Schnorr [19].

The notion of randomness is in another way quite fragile: if we take a random sequence and change to zero, say, its 10th, 100th, 1000th, etc. bits, the resulting sequence is not random, and for a good reason: a cheater that cheats once in a while is still a cheater. To consider such sequences as "approximately random", one option is to relax the Levin–Schnorr definition by replacing the $O(1)$ term in the complexity characterisation of randomness by a bigger $o(n)$ term, thus requiring that

☆ The first and fourth authors were supported by a Rutherford Discovery Fellowship from the Royal Society of NZ. The second author was partially supported by grant #358043 from the Simons Foundation. The third author was partially supported by RaCAF ANR-15-CE40-0016-01 grant and RBFR 16-01-00362 A.

* Corresponding author.
E-mail addresses: greenberg@msor.vuw.ac.nz (N. Greenberg), jmiller@math.wisc.edu (J.S. Miller), alexander.shen@lirmm.fr (A. Shen), westrick@uconn.edu (L.B. Westrick).
URLs: http://homepages.mcs.vuw.ac.nz/~greenberg/ (N. Greenberg), http://www.math.wisc.edu/~jmiller/ (J.S. Miller), http://www.lirmm.fr/~ashen (A. Shen), http://www.math.uconn.edu/~westrick/ (L.B. Westrick).

https://doi.org/10.1016/j.tcs.2017.09.031
0304-3975/

$\lim_{n\to\infty} K(X\upharpoonright_n)/n = 1$. Such sequences coincide with the sequences of effective Hausdorff dimension 1. (Effective Hausdorff dimension was first explicitly introduced by Lutz [13]. It can be defined in several equivalent ways via complexity, via natural generalizations of effective null sets, and via natural generalizations of martingales; again, see [4,23,22] for more information.)

Another approach follows the above example more closely: we could say that a sequence is approximately random if it differs from a random sequence on a set of density 0. Our starting point is that this also characterizes the sequences of effective Hausdorff dimension 1.

To set notation, for $n \geqslant 1$, we let d be the normalised Hamming distance on $\{0,1\}^n$, the set of binary strings of length n:

$$d(\sigma, \tau) = \frac{\#\{k : \sigma(k) \neq \tau(k)\}}{n};$$

and we also denote by d the Besicovitch distance on Cantor space 2^ω (the space of infinite binary sequences), defined by

$$d(X, Y) = \limsup_{n\to\infty} d(X\upharpoonright_n, Y\upharpoonright_n),$$

where $Z\upharpoonright_n$ stands for the n-bit prefix of Z. If $d(X, Y) = 0$, then we say that X and Y are *coarsely equivalent*.[1]

Theorem 1.7. *A sequence has effective Hausdorff dimension 1 if and only if it is coarsely equivalent to a ML-random sequence.*

In Section 2, we generalize this result to sequences of effective dimension s in various ways. Because a sequence X having effective dimension s implies that the prefix-free complexity of its n-bit prefix $X\upharpoonright_n$ is at least $sn - o(n)$, it is natural to consider the weakly s-randoms, those sequences X such that $K(X\upharpoonright_n) \geqslant sn - O(1)$.

Theorem 2.5. *Every sequence of effective Hausdorff dimension s is coarsely equivalent to a weakly s-random.*

Along the way to proving this, we pass through the question of how to raise the effective dimension of a given sequence while keeping density of changes at a minimum. If $d(X, Y) = 0$, then $\dim(X) = \dim(Y)$; so sequences of effective Hausdorff dimension $s < 1$ cannot be coarsely equivalent to a ML random sequence. It is natural then to ask, what is the minimal distance required between any sequence and a random? By Theorem 2.5, it is equivalent to ask about distances between sequences of dimension s and dimension 1; and naturally generalising, to ask, for any $0 \leqslant s < t \leqslant 1$, about distances between sequences of dimension s and dimension t. We start with a naive bound. For any $X, Y \in 2^\omega$,

$$|\dim(Y) - \dim(X)| \leqslant H(d(X, Y)).$$

This is our Proposition 3.1. Here $H(p) = -(p\log p + (1-p)\log(1-p))$ is the binary entropy function defined on $[0,1]$. The binary entropy function is used to measure the size of Hamming balls. If $V(n,r) = \sum_{k\leqslant nr}\binom{n}{k}$ is the size of a Hamming ball of radius $r < 1/2$ in 2^n, then

$$H(r)n - O(\log n) \leqslant \log(V(n,r)) \leqslant H(r)n$$

(see [18, Cor. 9, p. 310]).

In Proposition 3.5, we will see that this bound is tight, in the sense that if $s < t$ then there are $X, Y \in 2^\omega$ with $\dim(X) = s$, $\dim(Y) = t$ and $d(X, Y) = H^{-1}(t - s)$. Note that for H^{-1} we take the branch which maps $[0,1]$ to $[0,1/2]$.

Bounding the distance from an arbitrary dimension s sequence to the nearest dimension t sequence requires more delicate analysis. For example, fix $0 < s < t \leqslant 1$. If X is Bernoulli $H^{-1}(s)$-random, then its dimension is s. But its density of 1s is $H^{-1}(s)$. If $\dim(Y) \geqslant t$ then the density of 1s in Y is at least $H^{-1}(t)$, so $d(X, Y) \geqslant H^{-1}(t) - H^{-1}(s)$. Note that $H^{-1}(t) - H^{-1}(s) \geqslant H^{-1}(t - s)$, so this is a sharper bound, and it is tight:

Theorem 4.1. *For every sequence X with $\dim(X) = s$, and every $t \in (s, 1]$, there is a Y with $\dim(Y) = t$ and $d(X, Y) \leqslant H^{-1}(t) - H^{-1}(s)$.*

In particular, for $t = 1$, in light of Theorem 1.7, we obtain

Theorem 2.1. *For every $X \in 2^\omega$ there is a ML-random sequence Y such that*

$$d(X, Y) \leqslant 1/2 - H^{-1}(\dim(X)).$$

(We however prove Theorem 2.1 first, and elaborate on its proof to obtain Theorem 2.5 and then Theorem 4.1.)

We can also ask, starting from an arbitrary random, how close is the nearest sequence of dimension s guaranteed to be? For example, a typical construction of a sequence of effective dimension 1/2 starts with a random and replaces all the even

[1] One place this is defined is in [11], where it is called "generic similarity".

Please cite this article in press as: N. Greenberg et al., Dimension 1 sequences are close to randoms, Theoret. Comput. Sci. (2017), https://doi.org/10.1016/j.tcs.2017.09.031