Information and Computation 208 (2010) 292–308

Contents lists available at ScienceDirect

Information and Computation

journal homepage: www.elsevier.com/locate/ic

Algorithmically independent sequences☆

Cristian S. Calude [a,1], Marius Zimand [b,*,2]

[a] Department of Computer Science, University of Auckland, New Zealand
[b] Department of Computer and Information Sciences, Towson University, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 31 October 2008
Revised 19 May 2009
Available online 18 June 2009

Keywords:
Algorithmic information theory
Independence
Mutual information
Randomness

ABSTRACT

Two objects are independent if they do not affect each other. Independence is well-understood in classical information theory, but less in algorithmic information theory. Working in the framework of algorithmic information theory, the paper proposes two types of independence for arbitrary infinite binary sequences and studies their properties. Our two proposed notions of independence have some of the intuitive properties that one naturally expects. For example, for every sequence x, the set of sequences that are independent with x has measure one. For both notions of independence we investigate to what extent pairs of independent sequences, can be effectively constructed via Turing reductions (from one or more input sequences). In this respect, we prove several impossibility results. For example, it is shown that there is no effective way of producing from an arbitrary sequence with positive constructive Hausdorff dimension two sequences that are independent (even in the weaker type of independence) and have super-logarithmic complexity. Finally, a few conjectures and open questions are discussed.

1. Introduction

Intuitively, two objects are independent if they do not affect each other. The concept is well-understood in classical information theory. There, the objects are random variables, the information in a random variable is its Shannon entropy, and two random variables X and Y are declared to be independent if the information in the join (X, Y) is equal to the sum of the information in X and the information in Y. This is equivalent to saying that the information in X conditioned by Y is equal to the information in X, with the interpretation that, on average, knowing a particular value of Y does not affect the information in X.

The notion of independence has been defined in algorithmic information theory as well, but for finite strings [6]. The approach is very similar. This time the information in a string x is the complexity (plain or prefix-free) of x, and two strings x and y are independent if the information in the join string $\langle x, y\rangle$ is equal to the sum of the information in x and the information in y, up to logarithmic (or, in some cases, constant) precision.

☆ An extended abstract has appeared in M. Ito, M. Toyama (Eds.), Developments in Language Theory (DLT'08), Lectures Notes in Comput. Sci. 5257, Springer-Verlag, Berlin, 2008, pp. 183–195.

* Corresponding author.
E-mail address: mzimand@towson.edu (M. Zimand).
URLs: http://www.cs.auckland.ac.nz/~cristian (C.S. Calude), http://triton.towson.edu/~mzimand (M. Zimand).

[1] Calude was supported in part by UARC Grant 3607894/9343 and CS-PBRF Grant.
[2] Zimand has been partially supported by NSF Grant CCF 0634830. Part of this work was done while visiting the CDMTCS of the University of Auckland, New Zealand.

0890-5401/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ic.2009.05.004

The case of infinite sequences (in short, sequences) has been less studied. An inspection of the literature reveals that for this setting, independence has been considered to be synonymous with pairwise relative randomness, i.e., two sequences x and y are said to be independent if they are (Martin-Löf) random relative to each other (see [31,7]). The effect of this approach is that the notion of independence is confined to the situation where the sequences are random.

The main objective of this paper is to put forward a concept of independence that applies to *all* sequences, is natural, and is easy to use. One can envision various ways for doing this. One possibility is to use Levin's notion of mutual information for sequences [13] (see also the survey paper [10]) and declare two sequences to be independent if their mutual information is small.[3] We take another approach, which consists in extending in the natural way the notion of independence from finite strings to sequences. This leads us to two concepts: *independence* and *finitary-independence*. We say that (1) two sequences x and y are independent if, for all n, the complexity of $x\restriction n$ (the prefix of x of length n) and the complexity of $x\restriction n$ relativized with y are within $O(\log n)$ (and the same relation holds if we swap the roles of x and y), and (2) two sequences x and y are finitary-independent if, for all n and m, the complexity of $x\restriction n$ and the complexity of $x\restriction n$ given $y\restriction m$ are within $O(\log n + \log m)$ (and the same relation holds if we swap the roles of x and y). We have settled for the additive logarithmical term of precision (rather than some higher accuracy) since this provides robustness with respect to the type of complexity (plain or prefix-free) and other technical advantages.

We establish a series of basic facts regarding the proposed notions of independence. We show that independence is strictly stronger than finitary-independence. The two notions of independence apply to a larger category of sequences than the family of random sequences, as intended. However, they are too rough for being relevant for computable sequences. It is not hard to see that a computable sequence x is independent with any other sequence y, simply because the information in x can be obtained directly. In fact, this type of trivial independence holds for a larger family of sequences, namely for any H-trivial sequence, and trivial finitary-independence holds for any sequence x whose prefixes have logarithmic complexity. It seems that for this type of sequences (computable or with very low complexity) a more refined definition of independence is needed (perhaps, based on resource-bounded complexity). We show that the two proposed notions of independence have some of the intuitive properties that one naturally expects. For example, for every sequence x, the set of sequences that are independent with x has measure one.

We next investigate to what extent pairs of independent, or finitary-independent sequences, can be effectively constructed via Turing reductions. For example, is there a Turing reduction f that given oracle access to an arbitrary sequence x produces a sequence that is finitary-independent with x? Clearly, if we allow the output of f to be a computable sequence, then the answer is positive by the type of trivial finitary-independence that we have noted above. We show that if we insist that the output of f has super-logarithmic complexity whenever x has positive constructive Hausdorff dimension, then the answer is negative. In the same vein, it is shown that there is no effective way of producing from an arbitrary sequence x with positive constructive Hausdorff dimension two sequences that are finitary-independent and have super-logarithmic complexity.

Similar questions are considered for the situation when we are given two (finitary-) independent sequences. It is shown that there are (finitary-) independent sequences x and y and a Turing reduction g such that x and $g(y)$ are not (finitary-) independent. We consider that this is the only counter-intuitive effect of our definitions. Note that the notion of constructive Hausdorff dimension (or of partial randomness) suffers from the same problem. For example, it is not hard to see that there exist a sequence x with constructive Hausdorff dimension 1 and a computable function g (which can even be a computable permutation of the input bits) such that $g(x)$ has constructive Hausdorff dimension $1/2$. It seems that if one wants to extend the notion of independence to sequences that are not random (in particular to sequences that have arbitrary positive constructive Hausdorff dimension) such counter-intuitive effects cannot be avoided. On the other hand, for any independent sequences x and y and for any Turing reduction g, x and $g(y)$ are finitary-independent.

Our results show that partial random sequences can have complex structure: in particular, there are such sequences that cannot be obtained from random sequences by simple dilution operations (such as inserting a 0 between adjacent bits or doubling each bit).

We also raise the question on whether given as input several (finitary-) independent sequences x and y it is possible to effectively build a new sequence that is non-trivially (finitary-) independent with each sequence in the input. It is observed that the answer is positive if the sequences in the input are random, but for other types of sequences the question remains open. The same issue can be raised for finite strings and for this case a positive answer is obtained. Namely, it is shown that given three independent finite strings x, y and z with linear complexity, one can effectively construct a new string that is independent with each of x, y and z, has high complexity and length a constant fraction of the lengths of x, y and z.

1.1. Preliminaries

$\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}^+$ denote, respectively, the set of non-negative integers, the set of real numbers, and the set of positive real numbers; the size of a finite set A is denoted $||A||$. Unless stated otherwise, all numbers are in $\mathbb{N}$ and all logs are in base 2. We work over the binary alphabet $\{0,1\}$. A string is an element of $\{0,1\}^*$ and a sequence is an element of $\{0,1\}^\infty$. If x is a string, $|x|$ denotes its length; xy denotes the concatenation of the strings x and y. If x is a string or a sequence, $x(i)$ denotes the ith bit of x and $x\restriction n$ is the substring $x(1)x(2)\cdots x(n)$. For two sequences x and y, $x \oplus y$ denotes the sequence

[3] We note that Levin's definition is technically very complicated and some basic questions remain open. For example, it is not even known whether, in the setting of [13], every sequence (excluding the trivial cases) is dependent with itself (see Problems 8.2 and 8.3 in [22]).