Reprinted from INFORMATION AND CONTROL
All Rights Reserved by Academic Press, New York and London
Vol. 70, No. 2/3, August/September 1986
Printed in Belgium

Every Sequence Is Reducible to a Random One

PÉTER GÁCS*

Computer Science Department, Boston University,
Boston, Massachusetts 02215

Every infinite sequence is Turing–reducible to an infinite sequence which is random in the sense of Martin–Löf.

INTRODUCTION

Charles Bennett asked whether every infinite binary sequence can be obtained from an "incompressible" one by a Turing machine. He proved that this is the case for arithmetical sequences. The question has some philosophical interest because it permits us to view even very pathological sequences as the result of the combination of two relatively well-understood processes: the completely chaotic outcome of coin-tossing, and a transducer algorithm. A related problem was stated much earlier in Proposition 3 of (Levin, 1976). An interesting generalization of Levin's problem is still open, but its statement would require more definitions.

1. STATEMENT OF THE RESULT

First we introduce some basic definitions and convenient notation. In these, we follow approximately the works (Martin–Löf, 1966; Shoenfield, 1967; Levin, 1973). Let N denote the set of natural numbers. The cardinality of a set H is denoted by $|H|$. The set $S = \{0, 1\}^*$ is the set of all finite binary strings, and $B = \{0, 1\}^N$ is the set of all infinite binary strings. The length of a binary string x is denoted by $|x|$. For any binary string x and subset H of $S \cup B$, let $H[x]$ denote the set of all extensions of x in H. Sets of the form $B[x]$ are called *intervals*. For a subset E of S, let

$$H[E] = \bigcup_{x \in E} H[x].$$

* This work was supported in part by the National Science Foundation Grants MCS 8110430, MCS 8104008, MCS 8302874, and DCR 8405270, and in part by the DARPA Grant N 00014–82–K 0193 monitored by the ONR. Part of the work was done while the author was at the University of Rochester, Rochester, N.Y.

0019-9958/86 $3.00

The *open sets* of B are the ones of the form $B[H]$ for some subset H of S. Every open set can be written as a disjoint union

$$\bigcup_{x \in H} B[x] \tag{1.1}$$

for an appropriate set H. If H can be chosen to be recursively enumerable, the set $B[H]$ is called a *constructive open* set. Constructive closed sets are the complements of constructive open sets. Let G be a disjoint union (1.1). The (Lebesgue-) *measure* $\lambda(G)$ of G is

$$\sum_{x \in H} 2^{-|x|}.$$

For any closed set F, we define $\lambda(F) = 1 - \lambda(B - F)$.

A function $F: S \to S \cup B$ is *monotonic* if the following holds: if x is a prefix of y then $F(x)$ is a prefix of $F(x)$. A monotonic function F can be extended to $S \cup B$ by

$$F(x) = \sup\{F(y): y \in S, y \text{ is a prefix of } x\}.$$

We say that F is a *monotonic operator* if the set

$$\{(x, y): x, y \in S, y \text{ is a prefix of } F(x)\}$$

is recursively enumerable. A string x is *nonterminal* for F if there is an extension y of x such tha $F(y) \neq F(x)$. A monotonic operator is a *process* if the set of pairs

$$\{(x, F(x)): x \text{ is nonterminal for } F\}$$

is recursively enumerable. Processes are the monotonic operators which can be implemented by a Turing machine with a read-only tape moving in, with the argument x written on it, a write-only tape moving out with the value $F(x)$ accumulating on it, and working tapes. Indeed, the additional property to make a monotonic operator a process is just the one needed for the Turing machine to know when to read the next character of the input.

It is easy to see that if a monotonic operator is a recursive function from S to S then it is a process. It is also easy to see that for any monotonic operator F there is a monotonic operator G which is a recursive function from S to S such that for all infinite sequences x we have $F(x) = G(x)$.

An infinite sequence x is *Turing-reducible* to a infinite sequence y if there is a monotonic operator F with $x = F(y)$.

Martin-Löf introduced the notion of a random sequence in 1966. His definition is widely accepted now, and for the sake of completeness, we will

recall it in a paragraph at the end of the present section. We will need only the last fact mentioned in that paragraph: that the set of random sequences contains a constructive closed set E with $\lambda(E) > 0$. The assertion claimed in the title of the paper is a consequence of the following theorem.

THEOREM. *Let E be a constructive closed set with $\lambda(E) > 0$. Then there exists a process F such that $F(E) = B$. Moreover, there is a constant c such that on every nonterminal string x of length n we have*

$$|F(x)| \geq n - 3\sqrt{n} \log n + c.$$

COROLLARY. *Every infinite sequence is Turing-reducible to a random sequence.*

The last property of F says that we need no more than $3\sqrt{n} \log n$ bits of redundant information in our uniform generation of arbitrary sequences from random ones.

The theorem is not true without the assumption that the set E be a *constructive* closed set; it is easy to construct a counterexample by diagonalization. However, notice that the corollary, which is the main assertion of the paper, does not speak of closed sets at all, only of randomness and Turing-reducibility.

RANDOM SEQUENCES. According to Martin-Löf, an infinite binary sequence is *not random* if it is contained in a *constructive nullset*. A constructive nullset is defined as $\bigcap_i U_i$ for some sequence $\{U_i\}$ of open sets with the following two properties. First, we have $\lambda(U_i) < 2^{-i}$. Second, there is recursive function $(i, j) \to u_{ij}$ such that $U_i = \bigcup_j B[u_{ij}]$. A sequence $\{U_i\}$ with the above two properties is called a *test*. Martin-Löf (1966) showed that the union of all possible constructive nullsets is also a constructive nullset. If a test $\{U_i\}$ gives rise to this biggest constructive nullset then it is called a *universal test*. Thus, the set of random sequences is $\bigcup_i (B - U_i)$ for some universal test $\{U_i\}$. The measure of the constructive closed set $B - U_i$ is at least $1 - 2^{-i}$.

2. PROOF

The complement of the constructively closed set E is the union of a recursive sequence of intervals $B[x_t]$ for $t = 1, 2, \ldots$. Let

$$E_t = B - \bigcup_{u=1}^{t} B[x_i].$$

Then $E = \bigcap_t E_t$. In what follows the index k runs over the nonnegative integers. Let

$$k_0 = \lceil 1/\lambda(E) \rceil.$$

For $k \geqslant k_0$, let

$$n_k = k^2 + k\lfloor 2 \log k \rfloor, \qquad m_k = (k-1)^2 - (k_0 - 1)^2.$$

Let these numbers be 0 for $0 \leqslant k < k_0$. It is easy to see that there is a constant c such that for $n < n_{k+1}$ we have

$$m_k \geqslant n - 3\sqrt{n} \log n + c. \tag{2.1}$$

Let

$$T_k = \{0,1\}^{n_k}, \qquad U_k = \{0,1\}^{m_k}, \qquad T = \bigcup_k T_k, \qquad U = \bigcup_k U_k.$$

We will define $F(x)$ as well as all auxiliary monotonic operators in this paper only for x in T. For all other strings, the operator F will be extended by monotonicity. On T_k, we will define F with values in U_k. Let e denote the empty string. Let

$$\Phi_t^k = \left\{ x \in T_k : \lambda(E_t \cap B[x]) \geqslant \frac{1}{k} 2^{-n_k} \right\}$$

for $k \geqslant k_0$ and $\{e\}$ for $0 \leqslant k < k_0$. Let

$$\Phi_t = \bigcup_k \Phi_t^k, \qquad \Phi = \bigcap_t \Phi_t.$$

The elements of the set Φ_t are the intervals in T which have a large enough intersection with E_t. Let us call these intervals "fresh" at time t. The image comes from the observation that Φ_t is monotonically decreasing with t; as t increases, more and more intervals lose their freshness. It can be checked immediately that Φ_t^k is nonempty for al $k \geqslant 0$. The following lemma says that, moreover, freshness is inherited to enough subintervals.

LEMMA 1. *For x in Φ_t^k we have*

$$|\Phi_t^{k+1}[x]| \geqslant 2^{m_{k+1} - m_k}.$$

Proof. The statement is trivial for $k < k_0$. Otherwise, the set $T_{k+1}[x]$ has $2^{n_{k+1} - n_k}$ elements. If r of these are in Φ_t then

$$\frac{1}{k} 2^{-n_k} \leqslant \lambda(E_t \cap B[x]) \leqslant r 2^{-n_{k+1}} + \frac{1}{k+1} 2^{-n_k}.$$

Using

$$n_{k+1} - n_k \geqslant 2(k + \log k)$$

we have

$$r \geqslant \frac{2^{n_{k+1}-n_k}}{k(k+1)} \geqslant \frac{2^{2k}k^2}{k(k+1)} \geqslant 2^{2k-1} = 2^{m_{k+1}-m_k}. \quad \blacksquare$$

We will define the process F as $\sup_t F_t(x)$, where $F_t(x)$ is recursive and monotonic in x and t. The functions F_t will have the following additional properties

(i) For each t, k, for each string x in T_k, we have $|F_t(x)| \leqslant m_k$, with equality for nonterminal strings x.

(ii) If x is nonterminal for F_t then $F_t(x) = F(x)$.

The latter implies that F is a process. For a nonterminal x of length n with $n_k \leqslant n < n_{k+1}$ we have (using (i) and (2.1))

$$|F(x)| = m_k \geqslant n - 3\sqrt{n}\log n + c.$$

Let M_t^k be the set of all strings in Φ_t^k for which $|F_t(x)| = m_k$. Let $M_t = \bigcup_k M_t^k$. The sets M_t are neither obviously increasing nor obviously decreasing with t. Indeed, the sets Φ_t are decreasing with t, while F_t is increasing with t. However, the sequence M_t is almost monotonic, since the following assertion holds.

LEMMA 2. *Suppose that x is in $M_t - M_{t+1}$. Then x does not belong to any M_i for $i > t$.*

Proof. Suppose that $x \in M_t^k$. Then $|F_t(x)| = m_k$. The monotonicity of F_t implies together with (ii) that then $|F_{t+1}(x)| = m_k$. But then $x \notin M_{t+1}$ implies that x is not in Φ_{t+1}. Since E_t is decreasing in t we have that x is not in Φ_i for any i greater than t. $\blacksquare$

The lemma implies that if x belongs to M_t for infinitely many t then it belongs to M_t for all but finitely many t. Let M denote the set of those x having this property. The set M can be considered the *limit* of the sequence M_t. Let $M^k = M \cap T_k$.

We will ensure the following property of F_t.

(iii) $F_t(e) = e$. For any k, any x in M_t^k, the function F_t is a bijection between $M_t^{k+1}[x]$ and $U_{k+1}[F_t(x)]$.

Lemma 2 implies from (iii) that: $F(e) = e$. For any x in M^k, the function F is a bijection between $M^{k+1}[x]$ and $U^{k+1}[x]$.

This latter fact implies $F(E) = B$. Indeed, the set V is a tree whose infinite paths are all infinitary binary sequences. Let x_0 be an element of M^{k_0}. We can define the tree V as

$$V^{k_0} = \{x_0\}, \qquad V^{k+1} = \bigcup_{x \in V^k} M^{k+1}[x], \qquad V = \bigcup_{k \geqslant k_0} V^k.$$

Then F is an isomorphic mapping of the tree V onto the tree U which thus maps the infinite paths of the tree V into those of U. Since for each x in M we also have $x \in \Phi$, the infinite paths of V are elements of the closed set E. Hence we have $F(E) = B$.

It remains to construct F_t with the desired properties. We define it recursively with the help of the auxiliary process $G_t(z, x, y)$. Here G_t is like F_t with the additional restriction that x is forced to map to y.

Let us have an $x \in \Phi_t^k$ for some k, further $y \in U_k$ and $z \in T[x]$. By Lemma 1, the set Φ_t^{k+1} has at least $r = 2^{m_{k+1} - m_k}$ elements. Let $x_1, \ldots, x_r$ be, say, the first r of them in lexicographical order, and let $y_1, \ldots, y_r$ be an enumeration of $U_{k+1}[y]$. We define G_t recursively as

$$G_t(x, x, y) = y.$$

If z is in $T[x_i]$ for some i in $\{1, \ldots, r\}$ then

$$G_t(z, x, y) = G_t(z, x_i, y_i).$$

For all other arguments, G_t is extended by monotonicity. It is clear from this definition that G_t is indeed a process.

We define $F_0(z) = G_0(z, e, e)$. This F_0 satisfies (iii). Suppose that F_t is defined, and satisfies (iii). We proceed to define F_{t+1}. Let x be an element of $M_t^k \cap \Phi_{t+1}$. Then we define $F_{t+1}(x) = F_t(x)$. We define F_{t+1} for continuations of x which are not in M_t^{k+1}. By (iii), the mapping F_t is one-to-one on the set $L = \Phi_{t+1} \cap M_t^{k+1}[x]$. Let

$$l = |L|, \qquad r = 2^{m_{k+1} - m_k}, \qquad s = r - l.$$

We can suppose that $s > 0$. By Lemma 1 we have $|\Phi_{t+1}^{k+1}[x]| \geqslant r$. Let $y_1, \ldots, y_s$ be an enumeration of $U^{k+1}[y] - F_t(L)$. Let $x_1, \ldots, x_s$ be the first s elements of $\Phi_{t+1}^{k+1} - L$. Let $F_{t+1}(z) = G_{t+1}(z, x_i, y_i)$ for i in $\{1, \ldots, s\}$ and z in $T[x_i]$. We extend F_{t+1} further by monotonicity. It is straightforward to check that F_t has properties (i)–(iii). ∎

RECEIVED August 3, 1984; ACCEPTED April 1, 1986

References

LEVIN, L. A. (1973), On the notion of a random sequence, *Soviet Math. Dokl.* **14** No. 5, 1413–1416.

LEVIN, L. A. (1976), On the principle of conservation of information in intuitionistic mathematics, *Soviet Math. Dokl.* **17**, No. 2, 601–605.

MARTIN-LÖF, P. (1966), The definition of random sequences, *Inform. Control* **9**, 602–619.

SCHOENFIELD, J. R. (1967), "Mathematical Logic," Chapt. 7.2, Addison–Wesley, Reading, Mass.

Printed by the St. Catherine Press Ltd., Tempelhof 41, Bruges, Belgium