INFORMATION AND CONTROL 38, 258–263 (1978)

Complexity Dips in Random Infinite Binary Sequences*

HOWARD P. KATSEFF

Computer Science Division, Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, California 94720

Given any function f with $\sum_{n=0}^{\infty} 2^{-f(n)}$ divergent, it is shown that every finite binary sequence x has an infinite number of initial segments, x^n, with $K(x^n) \leqslant n - f(n)$.

I. INTRODUCTION

Martin–Löf (1965) has shown that for all functions f with $\sum_{n=0}^{\infty} 2^{-f(n)}$ divergent and for all infinite binary sequences x, there are an infinite number of integers n where

$$K(x^n) \leqslant n - f(n).$$

In a later paper (Martin–Löf, 1971) a special case of this theorem is presented. This paper presents a proof of Martin–Löf's original theorem which is motivated by an easy proof of another special case of the theorem.

II. DEFINITIONS

Throughout this paper x and y denote infinite binary sequences and all other variables denotes denote either natural numbers or (finite) binary strings. Natural numbers and strings beginning with 1 are used interchangeably, making use of the usual binary representation of numbers. We also make use of the following definitions:

$|s|$ = the length of s;

$x^{n,m}$ = the nth through mth bits of x where $x^{1,1}$ is the first bit of the string x;

$x^m = x^{1,m}$;

$s\hat{\ }t$ = s catenated to t;

$l(n) = \lfloor \log_2(n) \rfloor$;

$[m, n] = \{i \mid m \leqslant i \leqslant n\}$;

* Research Supported by National Science Foundation Grant MCS75-21760-A01.

0019-9958/78/0383-0258$02.00/0

Card(S) = number of elements in set S;

$K(s)$ = the Kolmogorov complexity (Kolmogorov, 1968) of s: the length of a shortest program in an optimal programming language (Schnorr, 1974) to output s.

III. Complexity Dips

The proofs in this paper make use of a simple technique to encode strings. The first theorem illustrates its use.

Theorem 1. $(\exists c)(\forall x)(\exists^\infty n)[K(x^n) \leqslant n - l(n) + c]$.

Proof. To prove the theorem, we show that any sequence x has an infinite number of initial segments with short programs. The programs represent an algorithm together with a string s which is used by the algorithm. The algorithm simply outputs all but the first bit of $| s |$, followed by s itself. To see that x has an infinite number of initial segments which are output by a program of this type, note that for all $n \geqslant 1$, the initial segment $x^{n+(1\wedge x^n)}$ is output by the program containing the string $x^{n+1,n+(1\wedge x^n)}$.

By examining the decoding algorithm, one can see that $K(x^{n+l(n)}) \leqslant n + d$ for each x and infinitely many n, where d is the length of the description of the decoding algorithm. Define $r(n) = n + l(n)$, so this statement is equivalent to $K(x^{r(n)}) \leqslant r(n) - l(n) + d$. It is easy to see that $l(n) \geqslant l(n + l(n)) - 1$, so $K(x^{r(n)}) \leqslant r(n) - l(r(n)) + d + 1$ which proves the theorem. ∎

We now present a proof of Martin–Löf's original theorem which states that for all functions f with $\sum_{n=0}^{\infty} 2^{-f(n)}$ divergent, $K(x^n) \leqslant n - f(n)$ for infinitely many n. This proof is similar in spirit to the proof of Theorem 1. Again, we use a decoding algorithm which, on input s, outputs s preceded by some function of $| s |$, which we refer to as $g(| s |)$. Given any such f, we show that there is a function g satisfying both:

$$(\forall k)[| g(k)| = f(k)],$$

$$(\forall x)(\exists^\infty k)[g(k) \text{ is an initial segment of } x].$$

Following the example of Theorem 1, we could use a decoding algorithm which, on input s, outputs $g(| s |)^\frown s$. Note that g is defined so that each x has infinitely many n where $g(n)$ is an initial segment of x. Thus, given any such n, the algorithm, on input $x^{|g(n)|+1,|g(n)|+n}$, outputs $x^{|g(n)|+n}$ demonstrating that

$$(\exists c)(\exists^\infty n)[K(x^{n+f(n)}) \leqslant n + c].$$

This statement appears to be weaker than the theorem we wish to prove in two ways: The first is the presence of the constant c. We will see that this problem is easily taken care of. The second is more serious as this statement appears to be weaker than

$$(\exists c)(\exists^\infty n)[K(x^n) \leqslant n - f(n) + c]. \tag{1}$$

We use a different idea to solve this second problem. Suppose that f has the property that there are never two different n with the same value of $n - f(n)$. Then we use a decoding algorithm which, on input s, finds the n satisfying $n - f(n) = | s |$ and if such an n exists, outputs $g(n)^\frown s$. It is not hard to see that (1) can be proved by using this algorithm to provide the short programs.

In general, functions f do not have this property and thus the decoding algorithm chooses a particular n for each value of $n - f(n)$, and outputs $g(n)^\frown s$ for this chosen n. It remains to be shown that, for our particular method of choosing n, this algorithm generates infinitely many initial segments of every sequence.

We now proceed with a formal proof of the theorem, beginning with a number of lemmas concerning functions with $\sum_{n=0}^\infty 2^{-f(n)}$ divergent. We will find it easier to consider a slightly more general class of functions which we now define.

DEFINITION. An *extended natural number* is a natural number or the special symbol ω.

DEFINITION. $2^{-\omega} = 0$.

Thus, it appears that ω has properties similar to that of infinity. However, it is better to think of ω as meaning undefined.

DEFINITION. f is an *extended function* iff f maps the natural numbers to the extended natural numbers.

We first show that given any function f with $\sum_{n=0}^\infty 2^{-f(n)}$ divergent, there is a function, identical to f except for being undefined at some values, which has a number of useful properties.

LEMMA 1. *For any recursive extended function f with $\sum_{n=0}^\infty 2^{-f(n)}$ divergent, there is a recursive extended function f' satisfying*:

$$(\forall n)[f'(n) \neq \omega \Rightarrow f'(n) = f(n)], \tag{2}$$

$$\sum_{n=0}^{\infty} 2^{-f'(n)} \text{ diverges}, \tag{3}$$

$$(\forall n)[f'(n) \neq \omega \Rightarrow f'(n) \leqslant n], \tag{4}$$

$$(\forall k)[\mathrm{Card}\{n \mid n - f'(n) = k\} \leqslant 1]. \tag{5}$$

Remark. Lemma 1 states that given any function f, there is a function f', identical to f except that $f'(n) = \omega$ for some n, which has the properties that when $f'(n) \neq \omega$, $f'(n)$ never exceeds n and for each k, the function $m = f'(n)$ never intersects the line $m = n - k$ more than once. The construction of f' is illustrated in Fig. 1.

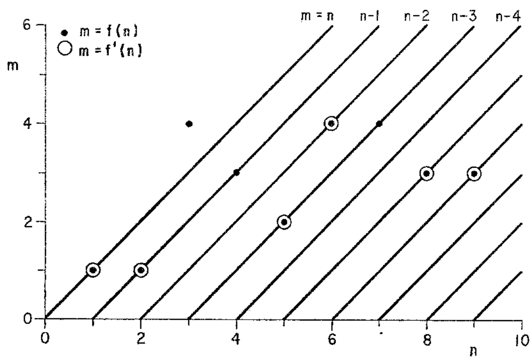

FIGURE 1

Note that f' cannot be set to ω on too many inputs since $\sum_{n=0}^{\infty} 2^{-f'(n)}$ is required to diverge.

Proof. Given f, define f' as follows:

$$f'(n) = \begin{cases} f(n) & \text{if } f(n) \leqslant n \quad \text{and} \quad (\forall m < n)[n - f(n) \neq m - f(m)] \\ \omega & \text{otherwise.} \end{cases}$$

It should be clear from its definition that f' satisfies (2), (4), and (5). The fact that $\sum_{n=0}^{\infty} 2^{-f'(n)}$ diverges may be verified with the following facts:

(i) The values of f where $f(n) \geqslant n$ contributes not more than 1 to $\sum_{n=0}^{\infty} 2^{-f(n)}$.

(ii) For each diagonal line in Fig. 1, consider all the points of f which lie on this line. We retain only the smallest for inclusion in f'. Thus, the points on the line not included in f' contribute no more to $\sum_{n=0}^{\infty} 2^{-f(n)}$ than the retained point. ▮

Lemma 2 states that, given a list of numbers with sufficient measure, there exist strings whose lengths are specified by these numbers such that at least one of these strings is an initial segment of every infinite sequence.

LEMMA 2. *Given a finite list* $N = \{n_1, n_2, \ldots, n_k\}$ *of numbers satisfying* $\sum_{n=1}^{k} 2^{-n} \geqslant 1$, *one can effectively find a set of* k *strings* $\{s_1, s_2, \ldots, s_k\}$ *satisfying*

$$(\forall i \leqslant k)[\,|s_i| = n_i], \tag{6}$$

$$(\forall x)(\exists i \leqslant k)\ [s_i \textit{ is an initial segment of } x]. \tag{7}$$

Proof. Sort N to obtain $\{n_{j_1}, n_{j_2}, \ldots, n_{j_k}\}$ with $n_{j_1} \leqslant n_{j_2} \leqslant \cdots \leqslant n_{j_k}$. The strings s_i are defined inductively starting with s_{j_1}:

s_{j_i} = "Find the set of strings, s, of length n_{j_i} such that the number of strings $s_{j_1}, s_{j_2}, \ldots, s_{j_{i-1}}$ which are initial segments of s is minimum. Let s_{j_i} be the lexicographically least of these strings."

The fact that (6) is satisfied is clear from the definition of s_i. The proof of (7) is by contradiction. Suppose that no s_i is an initial segment of x. Then, by the definition of the s_i, they form a finite prefix-free set which contains no initial segment of x, so $\sum_{i=1}^{k} 2^{-s_i} < 1$, a contradiction. ∎

Given a sufficiently slowly growing function, we now show that there is a list of strings whose lengths are determined by the function such that an infinite number of the strings are an initial segment of every infinite sequence.

LEMMA 3. *For any recursive extended function f with $\sum_{n=0}^{\infty} 2^{-f(n)}$ divergent, there exists a recursive g satisfying*

$$(\forall k \text{ with } f(k) \neq \omega)[|g(k)| = f(k)], \tag{8}$$

$$(\forall x)(\exists^{\infty} k)[g(k) \text{ is an initial segment of } x]. \tag{9}$$

Proof. The recursive function g is constructed in stages, starting with $m = 0$ at stage 0:

Stage i:

"Find the smallest number n satisfying $\sum_{k=m}^{n} 2^{f(k)} \geqslant 1$. Use Lemma 2 with the non-ω elements of $\{f(m), f(m+1), \ldots, f(n)\}$ to define values of $\{g(m), g(m+1), \ldots, g(n)\}$ which satisfy

$$(\forall k \in [m, n] \text{ with } f(k) \neq \omega)[|g(k)| = f(k)],$$

$$(\forall x)(\exists k \in [m, n])[g(k) \text{ is an initial segment of } x].$$

Set $m = n$.
Go to stage i + 1."

It should be clear that (8) is satisfied. To verify (9), note that for each x, some initial segment of x is assigned to the range of g in every stage. ∎

Lemma 4 shows that, given a sufficiently slowly growing function, we can obtain complexity dips whose magnitudes are specified by the function.

LEMMA 4. *($\forall$ recursive extended function f with $\sum_{n=0}^{\infty} 2^{-f(n)}$ divergent)*:

$$(\exists c)(\forall x)(\exists^{\infty} n \text{ with } f(n) \neq \omega)[K(x^n) \leqslant n - f(n) + c].$$

Proof. Let a recursive extended function f be given. Use Lemma 3 to find a recursive g satisfying (8) and (9) and assume without loss of generality that f satisfies (4) and (5) by Lemma 1. The algorithm to do the decoding, on input s, finds the n satisfying $n - f(n) = |s|$ and if such an n exists, outputs $g(n)^\frown s$.

To see that this algorithm encodes infinitely many initial segments of x, note that for each of the infinitely many k where $g(k)$ is an initial segment of x, $x^{|g(k)|+k-f(k)} = x^k$ is output by the algorithm with input $x^{|g(k)|+1,|g(k)|+k-f(k)} = x^{|g(k)|+1,k}$. Thus, we have shown that

$$(\exists c)(\exists^{\infty} n)[K(x^n) \leqslant n - f(n) + c]. \quad \blacksquare$$

We finally prove Martin–Löf's original theorem.

THEOREM 2. *($\forall$ recursive f with $\sum_{n=0}^{\infty} 2^{-f(n)}$ divergent)*:

$$(\forall x)(\exists^{\infty} n)[K(x^n) \leqslant n - f(n)].$$

Proof. Martin–Löf (1971) shows that for all recursive f with $\sum_{n=0}^{\infty} 2^{-f(n)}$ divergent, there is a recursive f' with $\sum_{n=0}^{\infty} 2^{-f'(n)}$ divergent and

$$(\forall c)(\forall^{\infty} n)[f'(n) \geqslant f(n) + c].$$

Apply Lemma 4 to f' to prove this theorem. $\blacksquare$

ACKNOWLEDGMENTS

I am grateful to C. P. Schnorr for suggesting the inclusion of Theorem 2 and for providing key ideas behind its proof. I also wish to thank Manuel Blum, William Sakoda, and Michael Sipser, who all contributed to the proof.

RECEIVED: February 25, 1977; REVISED: November 14, 1977

REFERENCES

KOLMOGOROV, A. (1968), Three Approaches for Defining the Concept of Information Quantity, *in* "Selected Translations in Mathematics, Statistics and Probability," AMS Publications, Providence.

MARTIN-LÖF, P. (1965), On the Oscillation of the Complexity of Infinite Binary Sequences, unpublished manuscript (in Russian).

MARTIN-LÖF, P. (1971), Complexity Oscillations in Infinite Binary Sequences, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **19**, 225–230.

SCHNORR, C. (1974), Optimal Enumerations and Optimal Godel Numberings, *Math. Systems Theory* **8**, 182–191.