Electronic Colloquium on Computational Complexity, Report No. 161 (2013)

The Frequent Paucity of Trivial Strings

Jack H. Lutz
Department of Computer Science
Iowa State University
Ames, IA 50011, USA
lutz@cs.iastate.edu

Abstract

A 1976 theorem of Chaitin, strengthening a 1969 theorem of Meyer, says that infinitely many lengths n have a *paucity of trivial strings* (only a bounded number of strings of length n having trivially low plain Kolmogorov complexities). We use the probabilistic method to give a new proof of this fact. This proof is much simpler than previously published proofs. It also gives a tighter paucity bound, and it shows that the set of lengths n at which there is a paucity of trivial strings is not only infinite, but has positive Schnirelmann density.

1 Background

A string of binary data is *trivial* if, like a string of all zeros, it contains negligible information beyond that implicit in its length. This notion of triviality has been made precise in several different ways, and these have been useful in the foundations of Kolmogorov complexity [6], information-theoretic characterizations of decidability and polynomial-time decidability [2, 8], formal language theory [4], and the theory of K-trivial sequences [7, 3].

These applications share several common features. Each uses some version of Kolmogorov complexity to quantify the information content of a string. Each *parametrizes* its triviality notion by a nonnegative integer c, defining a string to be c-*trivial* if its information content is within c bits of a smallness criterion. Most crucially, the key to each of these applications is a *paucity theorem*, stating that there are infinitely many lengths n at which there is a paucity (at most a fixed multiple of 2^c) of c-trivial strings of length n.

The first such paucity theorem, reported in 1969, was proved by Meyer [6]. Chaitin [2] subsequently strengthened Meyer's proof, slightly relaxing his triviality notion and obtaining the following.

ISSN 1433-8092

Theorem 1 (paucity theorem)**.** *There is a constant $a \in \mathbb{N}$ such that, for every $c \in \mathbb{N}$, there exist infinitely many lengths $n \in \mathbb{N}$ for which at most 2^{c+a} strings $x \in \{0,1\}^n$ satisfy $C(x) \leq c + \log n$.*

The logarithm here is base-2, and $C(x)$ is the *plain Kolmogorov complexity* of x, the minimum number of bits required to program a fixed universal Turing machine to print the string x. (Thorough treatments of $C(x)$ appear in [5, 7, 3].)

The proofs of Theorem 1 and Meyer's earlier paucity theorem are somewhat involved. Part of this is because these early proofs were aimed at proving more, namely that

(I) for every $c \in \mathbb{N}$ there are at most 2^{c+a} *infinite* binary sequences that are *c-trivial* in the sense that *every* nonempty prefix x of such a sequence satisfies $C(x) \leq c + log|x|$; and

(II) every such c-trivial sequence is decidable.

It is clear that (I) follows immediately from Theorem 1, and it is now well understood that (II) follows directly from (I), because every isolated infinite branch of a decidable tree is decidable [3].

In the 1990s Li and Vitanyi [4] gave a proof of Theorem 1 that is simpler than the original proof, even when one discounts the parts of the original proof devoted to (I) and (II). However, even this simplified proof is nontrivial.

2 Result

The purpose of this note is to give a *very* simple proof of Theorem 1. As it turns out, this very simple proof also strengthens Theorem 1 in two small respects. First, while Theorem 1 asserts that the set L of lengths n at which there is a paucity of c-trivial strings is infinite, our simple proof shows that the (*Schnirelmann*) *density*

$$\sigma(L) = \inf \left\{ \frac{|L_{<m}|}{m} \;\middle|\; m \in \mathbb{Z}^+ \right\}$$

of L, where we write $L_{<m} = L \cap \{0, \ldots, m-1\}$, is strictly positive. That is, the condition $n \in L$ holds *frequently*, not just infinitely often [9]. Second, while earlier proofs of Theorem 1 require the constant a to be as large as the

number of bits required to encode a nontrivial Turing machine, our simple proof shows that it suffices to take $a = 1$. We thus have the following.

Theorem 2 (frequent paucity theorem)**.** *For every $c \in \mathbb{N}$ the set of nonnegative integers n for which at most 2^{c+1} strings $x \in \{0,1\}^n$ satisfy $C(x) \leq c + \log n$ has density at least $(2^{c+1}+1)^{-1}$.*

Proof. Let $c \in \mathbb{N}$, and let $d = 2^{c+1}$. For each $n \in \mathbb{N}$, let

$$B_n = \{x \in \{0,1\}^n \,|\, C(x) \leq c + \log n\},$$

noting that $B_0 = \emptyset$, and let

$$L = \left\{ n \in \mathbb{N} \,\middle|\, |B_n| \leq d \right\}.$$

Let $m \in \mathbb{Z}^+$, and let $l = |L_{<m}|$. It suffices to show that

$$l > \frac{m}{d+1}. \tag{$*$}$$

Consider the average

$$\mu = \frac{1}{m} \sum_{n=0}^{m-1} |B_n|.$$

We have

$$\begin{aligned} \mu &= \frac{1}{m} \left| \bigcup_{n=0}^{m-1} B_n \right| \\ &\leq \frac{1}{m} \left| \{0,1\}^{<c+\log m} \right| \\ &< \frac{1}{m} 2^{c+\log m+1} \\ &= d \end{aligned}$$

and

$$\mu \geq \frac{1}{m}(m-l)(d+1),$$

whence

$$md > (m-l)(d+1),$$

which is equivalent to ($*$). □

3 Conclusion

The simplicity of the above proof is the main contribution of this note. Its simplicity arises from its use of the first moment probabilistic method [1, 9]: Rather than deal with the cardinalities $|B_n|$ individually, it examines their average.

A brief remark on pedagogy: Li and Vitanyi's Kolmogorov complexity characterization of regular languages [4, 5] yields a simple and intuitive method for proving that languages are not regular. A possible obstacle to teaching this method in undergraduate theory courses has been that the characterization theorem relies on the (seemingly) difficult paucity theorem. The simple proof here removes that obstacle.

Acknowledgments

This research was supported in part by National Science Foundation Grant 0652569. Part of this work was done during a sabbatical at Caltech and the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge.

References

[1] N. Alon and J.H. Spencer. *The Probabilistic Method* (third edition). John Wiley & Sons, 2008.

[2] Gregory J. Chaitin. Information-theoretic characterizations of recursive infinite strings. *Theoretical Computer Science*, 2:45–48, 1976.

[3] Rodney G. Downey and Denis R. Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer, 2010.

[4] Ming Li and Paul M.B. Vitányi. A new approach to formal language theory by Kolmogorov complexity. *SIAM Journal on Computing*, 24:398–410, 1995.

[5] Ming Li and Paul M.B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications* (third edition). Springer, 2008.

[6] Donald W. Loveland. A variant of the Kolmogorov concept of complexity. *Information and Control*, 15:510–526, 1969.

[7] André Nies. *Computability and Randomness*. Oxford University Press, New York, NY, USA, 2009.

[8] Pekka Orponen, Ker-I Ko, Uwe Schöning, and Osamu Watanabe. Instance complexity. *Journal of the ACM*, 41:96–121, 1994.

[9] Terence Tao and Van Vu. *Additive Combinatorics*. Cambridge University Press, 2006.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de