Fundamenta Informaticae XXI (2007) 1–3
IOS Press

A simple proof of Miller–Yu theorem

Laurent Bienvenu
Laboratoire d'Informatique Fondamentale
CNRS & Université de Provence
Marseille, France
laurent.bienvenu@lif.univ-mrs.fr

Wolfgang Merkle
Insitut für Informatik
Ruprecht-Karls-Universität
Heidelberg, Germany
merkle@math.uni-heidelberg.de

Alexander Shen
Laboratoire d'Informatique Fondamentale
CNRS & Université de Provence
Marseille, France
alexander.shen@lif.univ-mrs.fr

Abstract. A few years ago a nice criterion of Martin-Löf randomness in terms of plain (neither prefix nor monotone) Kolmogorov complexity was found (among many other results, it is published in [4]). In fact Martin-Löf came rather close to the *formulation* of this criterion around 1970 (see [3] and [6], p. 98). We provide a simple proof of this criterion that uses only elementary arguments very close to the original proof of Levin–Schnorr criterion of randomness (1973) in terms of monotone complexity ([2, 5]).

Keywords: Martin-Löf randomness, Kolmogorov complexity

Address for correspondence: Address for correspondence goes here

Theorem 1. **A**. Let $f\colon \mathbb{N} \to \mathbb{N}$ be a total computable function such that $\sum 2^{-f(n)} < \infty$. Then for every random sequence ω there exists a constant c such that

$$C(\omega_1 \dots \omega_n | n) \geq n - f(n) - c$$

for all n.

B. There exists a total computable function $f\colon \mathbb{N} \to \mathbb{N}$ such that $\sum 2^{-f(n)} < \infty$ and for every non-random sequence ω and for every c there exists n such that

$$C(\omega_1 \dots \omega_n | n) < n - f(n) - c$$

(We consider binary sequences $\omega_1\omega_2 \dots$; the randomness means Martin-Löf randomness with respect to the uniform distribution on Cantor space Ω.)

Theorem 1 implies that for some computable function f (with $\sum 2^{-f(n)} < \infty$) the condition

$$C(\omega_1 \dots \omega_n | n) \geq n - f(n) - O(1)$$

is necessary and sufficient for ω being random.

Proof:

A. For a given c let us consider a set U_c of all strings x such that

$$C(x|n) < n - f(n) - c,$$

where n is the length of x (denoted by $l(x)$ in the sequel). It is enumerable. The total measure of all corresponding intervals Ω_x is less than $2^{-c}\sum 2^{-f(n)}$. (Here Ω_x stands for the set of all sequences that have prefix x.) Indeed, U_c contains at most $2^{n-f(n)-c}$ sequences of length n, and the total measure of corresponding intervals is $2^{-c}2^{-f(n)}$.

Therefore, if ω has prefix in every U_c, then ω is not random.

B. A universal randomness test is an algorithm that generates for every $c = 1, 2, 3, \dots$ a sequence of strings

$$x(c,0), x(c,1), x(c,2) \dots$$

such that for every c the total measure of all intervals $x(c,i)$ (for $i = 0, 1, 2 \dots$) [i.e., the sum $\sum_i 2^{-l(x(c,i))}$] does not exceed 2^{-2c}, and for every nonrandom sequence ω and every c one of the strings $x(c,i)$ is a prefix of ω.

Note that for technical reasons it is convenient to use bound 2^{-2c}. We may also assume without loss of generality that: (1) x is total, i.e., $x(c,i)$ is defined for all c and i; (2) the intervals are listed in non-decreasing length order, i.e., that $l(x(c,0)) \leq l(x(c,1)) \leq l(x(c,2)) \leq \dots$ for any c. Indeed, to achieve (1), we may add infinitely many "dummy" intervals with small total measure; to achieve (2), we may split any interval into many small intervals without changing the total measure or the subset of Ω that is covered.

Then for each c and n we have finitely many strings of length n in the sequence $x(c,\cdot)$, and there is an algorithm that produces the list of all these strings given c and n. Let $m(c,n)$ be the total measure of corresponding intervals (i.e., 2^{-n} times the number of strings). So we have

$$\sum_n m(c,n) \leq 2^{-2c}$$

for every c.

Now consider the function f defined by the equation

$$2^{-f(n)} = \sum_c 2^c m(c,n).$$

Since each $m(c,n)$ and even the sum $\sum_n m(c,n)$ does not exceed 2^{-2c}, the right hand side is a computably convergent computable series and f is a computable real-valued function. (In the statement we require f to be integer-valued, but this evidently does not matter, since we can replace f by its integer-valued approximation.) Let us check that f is the function we have looked for. First,

$$\sum 2^{-f(n)} = \sum_{n,c} 2^c m(c,n) \le \sum_c 2^{-2c} \le 1.$$

On the other hand, any string of length n in the sequence $x(c,\cdot)$ is uniquely and computably determined by c and the ordinal number of this string among $2^n m(c,n)$ of them. Therefore, its Kolmogorov complexity does not exceed

$$2\log c + \log(2^n m(c,n)) + O(1) \le 2\log c + n - f(n) - c + O(1).$$

(since the sum $\sum_c 2^c m(c,n)$ does not exceed $2^{-f(n)}$, the same is true for each term). Recall that every nonrandom sequence has prefix among those strings for every c (and some n); since $c - 2\log c$ can be arbitrarily large, we get the statement B.

□

Remark 1. One may wish to improve the statement B by replacing the conditional complexity by the unconditional one. (A similar replacement for A makes it weaker.) To get the same bound for unconditional complexity, we need to estimate the number strings of length *at most* n in the sequence $x(c,\cdot)$. The same bound would work if we knew that the number of strings of length less than n in this sequence does not exceed the number of strings of length n if the latter is not zero. This also can be easily achieved by splitting intervals (replacing some string u by all its continuations of a given greater length).

Remark 2. One can use similar ideas to get a characterization of randomness in terms of time-bounded Kolmogorov complexity, see [1] for details.

References

[1] Bienvenu, L., Merkle, W.: Reconciling data compression and Kolmogorov complexity, *International Colloquium on Automata, Languages and Programming (ICALP 2007)*, 4596, Springer, 2007.

[2] Levin, L.: The concept of random sequence, *Doklady Akademii Nauk SSSR*, **212**, 1973, 548–550.

[3] Martin-Löf, P.: Complex oscillations in infinite binary sequences, *Z. Wahrscheinlichkeittheorie verw. Geibiete*, **19**, 1971, 225–230.

[4] Miller, J. S., Yu, L.: On initial segment complexity and degrees of randomness, *Transaction of the American Mathematical Society*, to appear.

[5] Schnorr, C.: Process complexity and effective random tests,, *Journal of Computer and System Sciences*, **7**, 1973, 376–388.

[6] Zvonkin, A., Levin, L.: The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms, *Russian Mathematical Surveys*, **25**(6), 1970, 83–124.