EVERY 2-RANDOM REAL IS KOLMOGOROV RANDOM

JOSEPH S. MILLER

Abstract. We study reals with infinitely many incompressible prefixes. Call $A \in 2^\omega$ *Kolmogorov random* if $(\exists^\infty n)\ C(A \upharpoonright n) > n - \mathcal{O}(1)$, where C denotes plain Kolmogorov complexity. This property was suggested by Loveland and studied by Martin-Löf, Schnorr and Solovay. We prove that 2-random reals are Kolmogorov random.[1] Together with the converse—proved by Nies, Stephan and Terwijn [11]—this provides a natural characterization of 2-randomness in terms of plain complexity. We finish with a related characterization of 2-randomness.

§1. Introduction. This paper is part of an ongoing program to understand randomness for real numbers, which we take to be elements of 2^ω, by investigating the complexity of their initial segments. Solomonoff [13] and Kolmogorov [4] independently defined a measure of the information content of finite strings. Intuitively, a complex string should be difficult to compress. The *Kolmogorov complexity* of $x \in 2^{<\omega}$ relative to a partial function $M\colon 2^{<\omega} \to 2^{<\omega}$ is defined to be $C_M(x) = \min\{|y| \mid M(y) = x\}$. In words, $C_M(x)$ is the length of the shortest M-description of x. We are interested in Kolmogorov complexity relative to arbitrary partial computable functions, but it is enough to consider a single universal function. Call a partial computable function $V\colon 2^{<\omega} \to 2^{<\omega}$ *universal* if, for every partial computable M, there is an x such that $(\forall y \in 2^{<\omega})\ M(y) = V(xy)$. Fix a universal partial computable function V. We write C for C_V and call this (*plain*) *Kolmogorov complexity.* Note that, up to a constant, Kolmogorov complexity is independent of the choice of V.

A student of Kolmogorov, Martin-Löf [8] introduced the most successful notion of effective randomness for real numbers. He defined the 1-*random* reals to be those avoiding certain effective measure zero sets. We get stronger notions of randomness by increasing the class of null sets to be avoided. The n-*random* reals are defined in this way; see the next section for details. We are concerned with the connections between n-randomness and Kolmogorov complexity. In particular:

Question. To what extent can effective randomness for real numbers be characterized in terms of plain Kolmogorov complexity?

The author's research was supported by the Marsden Fund of New Zealand.

[1] This result has been proved independently by Nies, Stephan and Terwijn [11]. We remark that their proof differs substantially from the one given below. It should also be noted that the author was aware, prior to establishing the results reported here, that Nies, Stephan and Terwijn [11] had proved that every Kolmogorov random is 2-random.

One might naïvely define a real $A \in 2^\omega$ to be random if none of its initial segments can be compressed (beyond some constant)—that is, if $(\forall n)\ C(A \upharpoonright n) > n - \mathcal{O}(1)$. It is well known that no real number would satisfy this definition. The problem is that we can use the length of a string $y \in 2^{<\omega}$ to encode an extra $\ln |y|$ bits of information; every real $A \in 2^\omega$ has infinitely many prefixes $x \in 2^{<\omega}$ such that $C(x) \leq |x| - \ln |x| - \mathcal{O}(1)$. Both Levin and Chaitin addressed this problem by modifying Kolmogorov complexity so that the naïve definition would be equivalent to 1-randomness. Levin used *monotone* complexity [5] and Chaitin used *prefix-free* complexity [1]. These complexity measures have proved important to the study of randomness, but we are still interested in understanding the relationship between plain Kolmogorov complexity and random reals.

While it is impossible for every initial segment of a real to be incompressible with respect to C, most reals have infinitely many initial segments which are incompressible (up to a constant). Call $A \in 2^\omega$ *Kolmogorov random* if

$$(\exists^\infty n)\ C(A \upharpoonright n) > n - \mathcal{O}(1).$$

This definition was proposed by Loveland [7] as a natural notion of effective randomness.[2] What is the relationship between Kolmogorov randomness and n-randomness? Martin-Löf [9] proved that every Kolmogorov random real is 1-random, while Schnorr [12] refuted the converse. Nies, Stephan and Terwijn proved that every Kolmogorov random is actually 2-random [11]. In the other direction, Martin-Löf showed that almost every real is Kolmogorov random. More concretely, Yu, Ding and Downey [15] analyzed an argument of Solovay [14] to prove that all 3-randoms are Kolmogorov random. Putting these facts together, it is known that Kolmogorov randomness lies somewhere between 3-randomness and 2-randomness.

We prove that every 2-random real is Kolmogorov random. Therefore, Kolmogorov randomness is a natural characterization of 2-randomness in terms of plain Kolmogorov complexity. This gives a partial answer to our motivating question. Further progress has recently been made by the author and Liang Yu [10], who give two new characterizations of 1-randomness. They prove that A is 1-random iff $(\forall n)\ C(A \upharpoonright n) > n - K(n) - \mathcal{O}(1)$, where K denotes prefix-free Kolmogorov complexity (which is defined in the next section). This shows that we can determine if a real is 1-random by looking at the Kolmogorov complexity of its initial segments, but clearly falls short of being a true plain complexity characterization of 1-randomness. Such a characterization is given by another result from [10], closely related to the first: A is 1-random iff for every computable $g\colon \omega \to \omega$ for which $\sum_{n\in\omega} 2^{-g(n)}$ is finite, $(\forall n)\ C(A \upharpoonright n) > n - g(n) - \mathcal{O}(1)$.

The paper is organized as follows. The next section reviews the necessary definitions and notation. In Section 3, we prove that 2-randoms are Kolmogorov random, and even that the constant can be preserved (up to a constant). In the final section, we introduce a new complexity measure $\widehat{C}$, which is defined in terms of plain complexity. For $x \in 2^{<\omega}$, let $\widehat{C}(x) = |x| - \min_{z \succeq x}(|z| - C(z))$. We

[2] Loveland actually used *uniform* Kolmogorov complexity in his definition. However, it was known very early that the definition is robust enough that uniform, length conditional and ordinary Kolmogorov complexity all give the same class [7, 9, 2]. See [6] for information on *uniform* and *length conditional* Kolmogorov complexity.

prove that $A \in 2^\omega$ is 2-random iff $(\forall n)\ \widehat{C}(A \upharpoonright n) = n + \mathcal{O}(1)$, giving a second characterization of 2-randomness in terms of plain complexity.

§2. Preliminaries. We assume that the reader is familiar with certain basic notions from computability theory: the jump, relative computability and the arithmetical hierarchy. In this section, we give a brief introduction to algorithmic randomness. For a thorough introduction, see either Li and Vitanyi [6] or the upcoming monograph of Downey and Hirschfeldt [3]. Denote the standard measure on 2^ω by μ. For $x, y \in 2^{<\omega}$, we write $x \preceq y$ if x is a prefix of y. Similarly, $x \prec A$ means that x is a prefix of $A \in 2^\omega$. Let $[x] = \{A \in 2^\omega \mid x \prec A\}$; such sets form a clopen basis for the standard topology on Cantor space. To $V \subseteq 2^{<\omega}$ we associate the open set $[V] = \bigcup_{x \in V}[x]$.

Martin-Löf [8] defined random reals as those which avoid effectively presented null sets. A *Martin-Löf test* is a computable sequence $\{V_i\}_{i\in\omega}$ of computably enumerable subsets of $2^{<\omega}$ such that $\mu([V_i]) \leq 2^{-i}$. A real A *passes* a Martin-Löf test $\{V_i\}_{i\in\omega}$ if $A \notin \bigcap_{i\in\omega}[V_i]$ and a real which passes all Martin Löf tests is called *1-random* (or *Martin-Löf random*). By taking the sets $V_i \subseteq 2^{<\omega}$ to be Σ^0_n instead of Σ^0_1 in these definitions we get the relativized notions of Σ^0_n*-Martin-Löf tests* and *n-random* reals.

Chaitin [1] modified Kolmogorov complexity to define randomness for reals. Call $D \subseteq 2^{<\omega}$ *prefix-free* if for all $x, y \in D$ such that $x \neq y$, we have $x \not\preceq y$. A partial function $M\colon 2^{<\omega} \to 2^{<\omega}$ is *prefix-free* if its domain is a prefix-free set. We write K_M instead of C_M to emphasize the fact that M is prefix-free. There is a *universal* partial computable prefix-free function U. In other words, for every partial computable prefix-free M, there is a $c \in \omega$ such that $(\forall y \in 2^{<\omega})\ K_U(y) \leq K_M(y) + c$. We write K for K_U and call this *prefix-free Kolmogorov complexity.* Up to a constant, K is independent of the choice of U. Schnorr [12] proved that $A \in 2^\omega$ is 1-random iff $(\forall n)\ K(A \upharpoonright n) > n - \mathcal{O}(1)$.

Prefix-free complexity can also capture stronger notions of randomness. We can extend U to a partial oracle-computable function such that $U^X : 2^{<\omega} \to 2^{<\omega}$ is a universal prefix-free X-computable function for every $X \in 2^\omega$. Write K^X for K_{U^X}. Relativizing Schnorr's result, $A \in 2^\omega$ is $(n+1)$-random iff $(\forall n)\ K^{\emptyset^{(n)}}(A \upharpoonright n) > n - \mathcal{O}(1)$.

In the proof of Theorem 1, we will construct a prefix-free partial function $F\colon 2^{<\omega} \to 2^{<\omega}$ so that K_F meets certain requirements. Such constructions are simplified by using the Kraft-Chaitin Theorem [1], which allows us to define F purely in terms of K_F. Say that we want $K_F(x_i) \leq n_i$, for all $i \in \omega$, where $\{x_i\}_{i\in\omega}$ is a sequence of strings and $\{n_i\}_{i\in\omega}$ a sequence of natural numbers. The Kraft-Chaitin Theorem states that F exists, and can be computed uniformly from the two sequences, as long as $\sum_{i\in\omega} 2^{-n_i} \leq 1$. Therefore, to construct F it is enough to describe which strings it compresses and by how much, while being careful not to demand more than there is room in the domain of F to satisfy.

§3. The Main Theorem. Let $\mathcal{K}_s = \{A \in 2^\omega \mid (\forall n)\ K^{\emptyset'}(A \upharpoonright n) > n - s\}$ and $\mathcal{C}_s = \{A \in 2^\omega \mid (\exists^\infty n)\ C(A \upharpoonright n) > n - s\}$, for each $s \in \omega$. Note that $\bigcup_{s\in\omega}\mathcal{K}_s$ is the class of 2-randoms and $\bigcup_{s\in\omega}\mathcal{C}_s$ the class of Kolmogorov random reals.

The following theorem not only states that all 2-random reals are Kolmogorov random, but also that the constant is preserved (up to a constant).

Theorem 1. $(\forall s)\ \mathcal{K}_s \subseteq \mathcal{C}_{s+\mathcal{O}(1)}$.

Proof. The statement of the theorem is clearly independent of our choice of universal functions V and $U^{\emptyset'}$. So we can assume, without loss of generality, that if $x \in 2^{<\omega}$ and $n \geq C(x)$, there is a $y \in 2^n$ such that $V(y) = x$. For if $\widehat{V}$ is any universal partial computable function, then define $V(0^k 1y) = \widehat{V}(y)$, for all $y \in 2^{<\omega}$ and $k \in \omega$. This V is also universal and has the desired property.

Our goal is to construct a prefix-free partial function $F\colon 2^{<\omega} \to 2^{<\omega}$ computably in $\emptyset'$ such that

$$R_{t,m}: \text{ if } (\forall k \geq m)\ C(A \upharpoonright k) \leq k - t, \text{ then } (\exists k)\ K_F(A \upharpoonright k) \leq k - t,$$

for all $t, m \in \omega$. Using the Kraft-Chaitin theorem, we construct F as the limit of a $\emptyset'$-computable sequence $\{F_s\}_{s\in\omega}$ of finite partial functions.[3] At stage $s = \langle t, m\rangle + 1$ we ensure that $R_{t,m}$ is satisfied. Let

$$G_{n,s} = \{y \in 2^n \mid V(y)\downarrow= x \text{ and } (\exists z \preceq x)\ |z| - K_{F_s}(z) \geq |x| - |y|\}.$$

In words, these are the $y \in \operatorname{dom} V$ of length n such that F_s already compresses a prefix of $x = V(y)$ by at least as much as y compresses x. No further action must be taken on behalf of such strings. Let $B_{n,s} = 2^n \cap \operatorname{dom} V \setminus G_{n,s}$. For every stage $s \in \omega$, we must guarantee that

$$(\exists \varepsilon_s > 0)(\exists N_s)(\forall n \geq N_s)[\mu(\operatorname{dom} F_s) + 2^{-n}|B_{n,s}| < 1 - \varepsilon_s]. \tag{1}$$

This should be read as the assertion that there is room left in the domain of F_s to handle the elements in the domain of V which still require attention. Note that $\emptyset'$ can determine if (1) holds for given values of $\varepsilon_s > 0$ and $N_s \in \omega$. This is because the sets $\{B_{n,s}\}_{n\geq N_s}$ can be enumerated uniformly (given the finite partial function F_s).

The Construction.

Stage $s = 0$. Let $F_0 = \emptyset$. It is easy to see that condition (1) is satisfied because V simulates the empty machine, say with prefix $y \in 2^{<\omega}$. So, if $\varepsilon_0 < 2^{-|y|}$ and $N_0 > |y|$, then for any $n \geq N_0$ there are less than $(1 - \varepsilon_0)2^n$ strings of length n in the domain of V.

Stage $s = \langle t, m\rangle + 1$. For each $n \geq m$, define

$$\begin{aligned} X_n = \{x \in 2^n \mid &(\forall k \in [m, n])\ C(x \upharpoonright k) \leq k - t \\ &\text{but } (\forall k \leq n)\ K_{F_{s-1}}(x \upharpoonright k) > k - t\}. \end{aligned}$$

Let $\Gamma = \bigcap_{n\in\omega}[X_n]$; this is the set of reals which still require attention in order to satisfy $R_{t,m}$. Note that the intersection is nested, so $\mu(\Gamma) = \lim_{n\to\infty} \mu([X_n])$.

Let $\varepsilon_s = \varepsilon_{s-1}/2$. Using $\emptyset'$, search for an $N_s \geq \max\{N_{s-1}, m - t\}$ such that we can extend F_{s-1} to a finite partial function F_s which compresses the elements of X_{N_s+t} by t, and if so, such that (1) holds for this choice of ε_s, N_s and F_s. If such an N_s is found, then the stage completes successfully and $R_{t,m}$ is satisfied. We must verify that an appropriate N_s exists.

[3] We mean this in the strongest sense: that $\emptyset'$ can compute the sequence of canonical indices for these functions.

Let $N_s \geq \max\{N_{s-1}, m-t\}$ be large enough that $\mu([X_{N_s+t}]) - \mu([X_{n+t}]) \leq 2^{-t}\varepsilon_s$, for all $n \geq N_s$. We will prove that N_s satisfies the conditions of our search. If $n \geq N_s$ and $x \in X_{n+t}$, then $C(x) \leq n$. By our assumption on V, there is a $y \in 2^n$ such that $V(y) = x$. Note that $y \in B_{n,s-1}$ and that if we extend F_{s-1} to a finite partial function F_s which compresses the elements of X_{N_s+t} by t, then $y \notin B_{n,s}$. In this case, each element of X_{n+t} would correspond to an element of $B_{n,s-1} \smallsetminus B_{n,s}$. Hence, $|X_{n+t}| + |B_{n,s}| \leq |B_{n,s-1}|$. Compressing every element of X_{N_s+t} by t would add $2^t\mu([X_{N_s+t}])$ to the measure of the domain of F_{s-1}. So,

$$\begin{aligned}\mu(\operatorname{dom} F_s) + 2^{-n}|B_{n,s}| &\leq \mu(\operatorname{dom} F_{s-1}) + 2^t\mu([X_{N_s+t}]) + 2^{-n}|B_{n,s}| \\ &\leq \mu(\operatorname{dom} F_{s-1}) + 2^t\mu([X_{n+t}]) + 2^{-n}|B_{n,s}| + \varepsilon_s \\ &= \mu(\operatorname{dom} F_{s-1}) + 2^{-n}(|X_{n+t}| + |B_{n,s}|) + \varepsilon_s \\ &\leq \mu(\operatorname{dom} F_{s-1}) + 2^{-n}|B_{n,s-1}| + \varepsilon_s < 1 - \varepsilon_{s-1} + \varepsilon_s = 1 - \varepsilon_s,\end{aligned}$$

for every $n \geq N_s$. Therefore, F_s exists (because we are not adding too much measure to the domain of F_{s-1}) and (1) holds for this choice of ε_s, N_s and F_s. This proves that the search is successful and completes the construction of F.

Finally, take $c \in \omega$ such that $K^{\emptyset'}(x) \leq K_F(x) + c$, for all $x \in 2^{<\omega}$. If $A \notin \mathcal{C}_{s+c}$, then there is an $m \in \omega$ such that $(\forall k \geq m)\ C(A \upharpoonright k) \leq k - (s+c)$. By $R_{s+c,m}$, we know that $(\exists k)\ K_F(A \upharpoonright k) \leq k - (s+c)$. Therefore, $K^{\emptyset'}(A \upharpoonright k) \leq k - s$. In other words, $A \notin \mathcal{K}_s$. But A is an arbitrary real not contained in $\mathcal{C}_{s+c}$, so $\mathcal{K}_s \subseteq \mathcal{C}_{s+c}$. ⊣

As was mention above, Nies, Stephan and Terwijn [11] proved that every Kolmogorov random real is 2-random, and in fact, that $\mathcal{C}_s \subseteq \mathcal{K}_{s+\mathcal{O}(1)}$, for all $s \in \omega$. Therefore:

Corollary 2. *The Kolmogorov random and 2-random reals coincide.*

We finish this section by considering $\overline{\mathcal{C}_s}$, the closure of $\mathcal{C}_s$. Note that while $\mathcal{K}_s$ is a closed set by definition, $\mathcal{C}_s$ is only a G_δ set (i.e., a countable intersection of open sets). Even so, every real in $\overline{\mathcal{C}_s}$ is Kolmogorov random.

Corollary 3. $(\forall s)\ \overline{\mathcal{C}_s} \subseteq \mathcal{C}_{s+\mathcal{O}(1)}$.

Proof. Because $\mathcal{C}_s \subseteq \mathcal{K}_{s+\mathcal{O}(1)}$ and $\mathcal{K}_{s+\mathcal{O}(1)}$ is closed, we have $\overline{\mathcal{C}_s} \subseteq \mathcal{K}_{s+\mathcal{O}(1)}$, for all $s \in \omega$. Now apply Theorem 1. ⊣

§4. Another Characterization.

There is more to be said about the connection between 2-randomness and plain complexity. Recall that plain Kolmogorov complexity can "underestimate" the information content of a string $x \in 2^{<\omega}$ if $|x|$ is too well correlated to x. Intuitively, this effect can be nullified by considering extensions of x; complex strings should have incompressible extensions. This motivates the following complexity measure. Let

$$\widehat{C}(x) = |x| - \min_{z \succeq x}(|z| - C(z)),$$

for all $x \in 2^\omega$. This complexity measure behaves more like Levin's monotone complexity [5] than Chaitin's prefix-free complexity. In particular, it is not hard to see that $\widehat{C}(x) \leq |x| + \mathcal{O}(1)$, for all $x \in 2^{<\omega}$. We prove that $A \in 2^\omega$ is 2-random iff $(\forall n)\ \widehat{C}(A \upharpoonright n) = n + \mathcal{O}(1)$.

Define $\widehat{\mathcal{C}}_s = \{A \in 2^\omega \mid (\forall n)\ \widehat{C}(A \upharpoonright n) > n - s\}$, for all $s \in \omega$. We remark that the proof of the following proposition is essentially the same as the proof of Nies, Stephan and Terwijn that every Kolmogorov random real is 2-random [11] (which itself is a modification of a proof of Yu, Ding and Downey [15] that no Δ^0_2 real is Kolmogorov random).

Proposition 4. $(\forall s)\ \widehat{\mathcal{C}}_s \subseteq \mathcal{K}_{s+\mathcal{O}(1)}$.

Proof. We define a partial computable function $M\colon 2^{<\omega} \to 2^{<\omega}$ such that, if $U^{\emptyset'}$ compresses $x \in 2^{<\omega}$ by s, then M will compress all sufficiently long extensions of x by s. As usual, let $\emptyset'_t$ denote the finite subset of $\emptyset'$ enumerated by stage $t \in \omega$. For every $y \in 2^{<\omega}$, if $y = y_1y_2$ and $U^{\emptyset'_{|y|}}(y_1) \downarrow = x$, then let $M(y) = xy_2$. This is well-defined because $U^{\emptyset'_{|y|}}$ is prefix-free, so there is at most one $y_1 \preceq y$ such that $U^{\emptyset'_{|y|}}(y_1) \downarrow$.

Choose $c \in \omega$ so that $C(x) \leq C_M(x) + c$, for all $x \in 2^{<\omega}$. If $A \notin \mathcal{K}_{s+c}$, then there is an $n \in \omega$ such that $K^{\emptyset'}(A \upharpoonright n) \leq n - (s+c)$. Choose $m \in \omega$ large enough that $K^{\emptyset'_t}(A \upharpoonright n) \leq n - (s+c)$, for all $t \geq m$. Then $C(z) \leq C_M(z) + c \leq |z| - s$, for any $z \succeq A \upharpoonright n$ such that $|z| \geq m$. This implies that $\widehat{C}(A \upharpoonright m) \leq m - s$, so $A \notin \widehat{\mathcal{C}}_s$. Therefore, $\widehat{\mathcal{C}}_s \subseteq \mathcal{K}_{s+c}$. ⊣

It is clear that $\mathcal{C}_s \subseteq \widehat{\mathcal{C}}_s$. Combining this fact with the proposition and Theorem 1, we get another characterization of 2-randomness in terms of plain complexity.

Corollary 5. *$A \in 2^\omega$ is 2-random iff $(\forall n)\ \widehat{C}(A \upharpoonright n) = n + \mathcal{O}(1)$.*

Restating this corollary, $A \in 2^\omega$ is 2-random iff every $x \prec A$ can be extended to a $z \succeq x$ such that $C(z) > |z| - \mathcal{O}(1)$.

REFERENCES

[1] Gregory J. Chaitin, *A theory of program size formally identical to information theory*, ***J. Assoc. Comput. Mach.***, vol. 22 (1975), pp. 329–340.

[2] Robert P. Daley, *Complexity and randomness*, ***Computational complexity (Courant Comput. Sci. Sympos. 7, New York Univ., New York, 1971)***, Algorithmics Press, New York, 1973, pp. 113–122.

[3] R. Downey and D. Hirschfeldt, ***Algorithmic randomness and complexity***, Springer-Verlag, Berlin, to appear.

[4] A. N. Kolmogorov, *Three approaches to the definition of the concept "quantity of information"*, ***Problemy Peredači Informacii***, vol. 1 (1965), no. vyp. 1, pp. 3–11.

[5] L. A. Levin, *The concept of a random sequence*, ***Dokl. Akad. Nauk SSSR***, vol. 212 (1973), pp. 548–550.

[6] M. Li and P. Vitányi, ***An introduction to Kolmogorov complexity and its applications***, Texts and Monographs in Computer Science, Springer-Verlag, New York, 1993.

[7] Donald W. Loveland, *On minimal-program complexity measures*, ***ACM symposium on theory of computing (STOC '69)*** (New York), ACM Press, May 1969, pp. 61–78.

[8] Per Martin-Löf, *The definition of random sequences*, ***Information and Control***, vol. 9 (1966), pp. 602–619.

[9] Per Martin-Löf, *Complexity oscillations in infinite binary sequences*, ***Z. Wahrscheinlichkeitstheorie und Verw. Gebiete***, vol. 19 (1971), pp. 225–230.

[10] Joseph S. Miller and Liang Yu, *On initial segment complexity and degrees of randomness*, in preparation.

[11] André Nies, Frank Stephan, and Sebastiaan A. Terwijn, *Randomness, relativization and Turing degrees*, submitted.

[12] C. P. Schnorr, *A unified approach to the definition of random sequences*, ***Math. Systems Theory***, vol. 5 (1971), pp. 246–258.

[13] R. J. Solomonoff, *A formal theory of inductive inference I and II*, ***Information and Control***, vol. 7 (1964), pp. 1–22, 224–254.

[14] Robert M. Solovay, *Draft of paper (or series of papers) on Chaitin's work*, (1975), unpublished notes, 215 pages.

[15] Liang Yu, Ding Decheng, and Rod G. Downey, *The Kolmogorov complexity of the random reals*, submitted.

SCHOOL OF MATHEMATICAL AND COMPUTING SCIENCES
VICTORIA UNIVERSITY, P.O. BOX 600
WELLINGTON, NEW ZEALAND
E-mail: Joe.Miller@mcs.vuw.ac.nz