The complexity of stochastic sequences

Wolfgang Merkle

Ruprecht-Karls-Universität Heidelberg, Institut für Informatik, Im Neuenheimer Feld 294, D-69120 Heidelberg, Germany

Received 15 August 2003; received in revised form 31 March 2006
Available online 13 June 2007

Abstract

We review and slightly strengthen known results on the Kolmogorov complexity of prefixes of effectively random sequences. First, there are recursively random sequences such that for any computable, non-decreasing and unbounded function \( f \) and for almost all \( n \), the uniform complexity of the length \( n \) prefix of the sequence is bounded by \( f(n) \). Second, a similar result with bounds of the form \( f(n) \log n \) holds for partial-recursively random sequences.

Furthermore, we demonstrate that there is no Mises–Wald–Church stochastic sequence such that all non-empty prefixes of the sequence have Kolmogorov complexity \( O(\log n) \). This implies a sharp bound for the complexity of the prefixes of Mises–Wald–Church stochastic and of partial-recursively random sequences. As an immediate corollary to these results, we obtain the known separation of the classes of recursively random and of Mises–Wald–Church stochastic sequences.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Random sequences; Stochastic sequences; Kolmogorov complexity; Initial segment complexity

1. Introduction and overview

Theorem 3 below asserts that there are recursively and partial-recursively random sequences such that the uniform complexity of the prefixes of the sequences grows comparatively slow. First, there are recursively random sequences such that for any computable, non-decreasing and unbounded function \( f \) and for almost all \( n \), the uniform complexity of the length \( n \) prefix of the sequence is bounded by \( f(n) \). Second, a similar result with bounds of the form \( f(n) \log n \) holds for partial-recursively random sequences. Both assertions are slightly strengthened variants of basically equivalent known results, see Remark 6 below.

The assertion on recursively random sequences cannot be extended to a constant in place of \( f(n) \), because if there is a constant which bounds the uniform complexity of infinitely many prefixes of a sequence, then by definition of uniform complexity the sequence must be computable. The corresponding question for the assertion on partial-recursively random sequences has been reported as being open in the monograph by Li and Vitányi [9, Exercise 2.5.14] with partial-recursively random replaced by Mises–Wald–Church stochastic, i.e., by stochastic with respect to partially computable (monotonic) selection rules. Theorem 7, our main result, gives a negative answer to the question in both cases. As a consequence, we obtain a sharp bound for the complexity of the prefixes of partial-recursively random and of Mises–Wald–Church stochastic sequences. There are such sequences such that for any computable, non-decreasing
and unbounded function \( f \) almost all prefixes of the sequence have Kolmogorov complexity of at most \( f(n) \log n \); however, there are no such sequences such that all non-empty prefixes have Kolmogorov complexity \( O(\log n) \).

Furthermore, as an immediate corollary of the mentioned results we obtain the known separation of the classes of recursively random and of Mises–Wald–Church stochastic sequences.

2. Notation

Our notation is mostly standard, for unexplained terms and further details we refer to the textbooks and surveys cited in the bibliography [2,3,5,9,11,14].

All functions are meant to be total if not explicitly attributed as being partial; in particular, a computable function is a partially computable function that is total. We write \( \log n \) for the logarithm of \( n \) to base 2.

A word is a finite binary sequence. We write \(|w|\) for the length of a word \( w \); the unique word of length 0, the empty word, is denoted by \( \lambda \). An assignment is a function from a subset of \( \mathbb{N} \), the set of natural numbers, to \{0, 1\}. A word of length \( n \) is identified in the natural way with an assignment on \{0, ..., \( n - 1 \)\}. For an assignment \( \sigma \) with finite domain \( \{z_1 < \cdots < z_n\} \), the WORD ASSOCIATED WITH \( \sigma \) is \( \sigma(z_1) \ldots \sigma(z_n) \).

Unless explicitly stated otherwise, the term sequence refers to an infinite binary sequence. For a sequence \( S(0)S(1)\ldots \), we refer to \( S(i) \) as bit \( i \) of the sequence \( S \). A sequence \( S \) can be viewed as a function \( i \mapsto S(i) \), hence restricting a sequence \( S \) to a set \( I \) yields an assignment \( S|I \) with domain \( I \); in particular, restricting \( S \) to the set \{0, ..., \( n - 1 \)\} yields a word, the prefix of \( S \) of length \( n \).

3. Stochastic and random sequences

In this section, we briefly review notation and concepts related to stochastic sequences, random sequences, and Kolmogorov complexity; for more detailed accounts see Refs. [2,3,9,11].

Stochastic sequences are defined in terms of selection rules. Intuitively speaking, a (monotonic) selection rule defines a process that scans the bits of a given sequence \( A \) in the natural order where in addition the process has to determine for each place whether this place is to be selected. The decision whether a place \( n \) shall be selected has to be determined before the corresponding bit \( A(n) \) is scanned and depends solely on the previously scanned bits \( A(0) \ldots A(n - 1) \). Formally, a selection rule is a partial function that receives as input the word \( A(0)\ldots A(n - 1) \) and outputs a bit that indicates whether \( n \) is to be selected.

**Definition 1.** A SELECTION RULE is a not necessarily total function

\[
r : \{0, 1\}^* \rightarrow \{0, 1\},
\]

\[
w \mapsto r(w).
\]

The sequence that is SELECTED by a selection rule \( r \) from a sequence \( A \) is the subsequence of \( A \) that contains exactly the bits \( A(n) \) such that \( r(A(0)\ldots A(n - 1)) = 1 \).

For given sequence \( A \) and selection rule \( r \), the SEQUENCE OF SELECTED PLACES \( z_0, z_1, \ldots \) is the sequence of all natural numbers \( z \) where \( r(A(0)\ldots A(z - 1)) \) is equal to 1 in natural order (accordingly, the selected sequence is \( A(z_0)A(z_1)\ldots \)). Observe that the sequence of selected places and the sequence selected from \( A \) will both be finite in case \( r \) is undefined on some prefix of \( A \).

**Definition 2.** A sequence \( S \) is STOCHASTIC with respect to a given set of admissible selection rules if for every admissible selection rule the sequence of selected places \( z_0, z_1, \ldots \) either is finite or the frequencies of 1’s in the prefixes of the selected sequence converge to 1/2, i.e.,

\[
\lim_{n \to \infty} \frac{|\{i < n \mid S(z_i) = 1\}|}{n} = \frac{1}{2},
\]

(1)

A sequence is MISES–WALD–CHURCH STOCHASTIC if it is stochastic with respect to the set of all partially computable selection rules.
Stochastic sequences may lack certain statistical properties that are associated with the intuitive understanding of randomness, for example, there are such sequences such that every prefix of the sequence contains more zeroes than ones [16]. An attempt to overcome these deficiencies is to define concepts of random sequences in terms of betting games where a player bets on individual bits of an initially unknown sequence.

Formally, a player can be identified with a BETTING STRATEGY, i.e., a not necessarily total function that maps the information about the already scanned part of the unknown sequence to a bet on the place to be scanned next, where a bet is determined by a guess for the bit at this place, and a rational in the closed interval [0, 1] that is equal to the fraction of the current capital that shall be bet on this guess. Payoff is fair in the sense that after each individual bet the stake is lost in case the guess was wrong, and is doubled, otherwise. For given betting strategy \( b \) and initial capital, let \( d_b \) be the corresponding PAYOFF FUNCTION or MARTINGALE, i.e., \( d_b(\lambda) \) is the initial capital and \( d_b(w) \) is the capital that \( b \) has accumulated after the first \(|w|\) bets when betting against an unknown sequence that has the word \( w \) as a prefix. For a martingale \( d \), the fairness condition can then be written as

\[
d(w) = \frac{d(w0) + d(w1)}{2}.
\]

A martingale \( d \) SUCCEEDS on a sequence \( A \) if \( d \) is unbounded on the prefixes of \( A \), i.e., if

\[
\limsup_{n \to \infty} d(A(0) \ldots A(n)) = \infty.
\]

Furthermore, a sequence is RANDOM with respect to a given set of admissible martingales if no admissible martingale succeeds on the sequence. A sequence is RECURSIVELY RANDOM if it is random with respect to the class of all computable martingales and a sequence is PARTIAL-RECURSIVELY RANDOM if it is random with respect to all partially computable martingales.

For any given Turing machine \( M \), the Kolmogorov complexity \( C_M(w) \) of a word \( w \) with respect to \( M \) is the length of the shortest word \( x \) such that \( M \) on input \( x \) outputs \( w \). There are Turing machines \( U \) that yield optimal Kolmogorov complexity up to an additive constant, i.e., for any Turing machine \( M \) there is a constant \( c_M \) such that for all words \( w \), we have \( C_U(w) \leq C_M(w) + c_M \) [9, Section 2.1]. We fix such an additively optimal Turing machine \( U \) as reference machine and let the KOLMOGOROV COMPLEXITY \( C(w) \) of a word \( w \) be equal to \( C_U(w) \).

The uniform Kolmogorov complexity \( C_M(w; n) \) of a word \( w \) with respect to a Turing machine \( M \) is the length of the shortest word \( x \) such that for all \( i \leq n \), the machine \( M \) outputs on input \((x, i)\) the length \( i \) prefix of \( w \). Again, it can be shown that for the concept of uniform complexity there are additively optimal Turing machines; we fix such a Turing machine \( \bar{U} \) and let the UNIFORM COMPLEXITY \( C(w; n) \) of a word \( w \) be equal to \( C_{\bar{U}}(w; n) \) [9, 2.3.3].

4. The complexity of the prefixes of random sequences

Theorem 3 asserts that there are recursively and partial-recursively random sequences such that the uniform complexity of the prefixes of the sequences grows comparatively slow. Assertions (i) and (ii) of the theorem are slightly strengthened variants of basically equivalent results by Lathrop and Lutz [8] and by Muchnik [13], respectively, see Remark 6 for details and further references.

Theorem 3. Let \( F \) be the class of all computable functions from \( N \) to \( N \) that are non-decreasing and unbounded.

(i) There is a recursively random sequence \( R \) such that for all \( f \in F \) and almost all \( n \),

\[
C(R(0) \ldots R(n - 1); n) \leq f(n).
\]

(ii) There is a partial-recursively random sequence \( R \) such that for all \( f \in F \) and almost all \( n \),

\[
C(R(0) \ldots R(n - 1); n) \leq f(n) \log n.
\]

Proof. Let \( \varphi_0, \varphi_1, \ldots \) be a standard effective enumeration of the partial recursive functions from \( N \) to \( N \) and let \( d_0, d_1, \ldots \) be an appropriate effective enumeration of all partially computable martingales with initial capital 1.

Proof of (i). For \( s = 0, 1, \ldots \), let

\[
F_s = \{i < s: \varphi_i \in F\}, \quad D_s = \{i < s: d_i \text{ is total}\}.
\]
Furthermore, let $n_0 = 0$ and for all $s \geq 1$ let

$$n_s = \min\{n: n_{s-1} < n \text{ and } 3s \leq \phi_i(n) \text{ for all } i \in F_s\},$$

$$I_s = \{n \in \mathbb{N}: n_s \leq n < n_{s+1}\}.$$

Consider the following non-effective construction of a sequence $R$, which proceeds in stages $s = 0, 1, \ldots$. During stage $s$, the restriction of $R$ to the interval $I_s$ is determined by diagonalizing against an appropriate weighted sum $\tilde{d}_s$ of the martingales in $D_s$. More precisely, let

$$\tilde{d}_s = \sum_{i \in D_s} \alpha_i d_i,$$

where $\tilde{d}_s$ shall be constant with value 1 in case $D_s$ is empty. When defining $R(n)$ for an $n$ in $I_s$, we assume that the prefix $w = R(0) \ldots R(n-1)$ of $R$ is already determined and we let

$$R(n) = \begin{cases} 0 & \text{if } \tilde{d}_s(w0) \leq \tilde{d}_s(w), \\ 1 & \text{if } \tilde{d}_s(w1) < \tilde{d}_s(w). \end{cases}$$

Observe in connection with the definition of $R(n)$ that each $\tilde{d}_s$ is again a martingale and that consequently the fairness condition (2) implies that for any word $w$ the values $\tilde{d}_s(w0)$ and $\tilde{d}_s(w1)$ either both agree with $\tilde{d}_s(w)$ or exactly one of them is strictly smaller than $\tilde{d}_s(w)$. Next we show by induction on $n$ that we have

$$\tilde{d}_s(R(0) \ldots R(n)) < 2 - \frac{1}{2^s} \text{ for all } s \text{ and any } n \text{ in } I_s. \quad (3)$$

For $n$ in $I_0$ there is nothing to prove because $\tilde{d}_s$ is just 1. In the induction step we distinguish two cases. In case $n$ is not minimum in an interval $I_s$, the induction step is immediate because on each individual interval the martingale $\tilde{d}_s$ is non-increasing by construction of $R$. In case $n$ is equal to the minimum number $n_s$ of some interval $I_s$, we let $w = R(0) \ldots R(n-1)$ and obtain

$$\tilde{d}_s(w(R(n))) \leq \tilde{d}_s(w) \leq \tilde{d}_{s-1}(w) + \alpha_s d_s(w) \leq 2 \left(1 - \frac{1}{2^{s-1}}\right) = 2 - \frac{1}{2^s}.$$

The inequalities follow by construction of $R$, by definition of $\tilde{d}_s$, and by the induction hypothesis and choice of $\alpha_s$; in connection with the third inequality observe that $d_s(w)$ is at most $2^{n_s}$, because of $d(\lambda) = 1$ and $|w| = n_s$.

The sequence $R$ is recursively random. Otherwise, there would be an index $i$ such that $d_i$ is total and succeeds on $R$, which yields a contradiction because by construction of $\tilde{d}_s$ and by (3), we have for almost all $s$ and all $n$ in $I_s$,

$$\alpha_i d_i(R(0) \ldots R(n)) \leq \tilde{d}_s(R(0) \ldots R(n)) < 2.$$

In order to conclude the proof of (i), observe that the index $s$ and the sets $F_s$ and $D_s$ can be coded by a word $x_s$ of length $2s$ and that given access to these two sets, the construction up to and including stage $s$ can be simulated effectively; in fact, there is an effective procedure that given $x_s$ computes the restriction of $R$ to the union of the intervals $I_0$ through $I_s$. This procedure can be adjusted to output the length $i$ prefix of $R$ on input $(i, x_s)$ whenever $i$ is in one of the intervals $I_0$ through $I_s$. In particular, we have for some constant $c$, for all $s$, and any $n$ in $I_s$,

$$C(R(0) \ldots R(n-1); n) \leq 2s + c.$$

Assertion (i) follows because by choice of the $n_s$, for any $f$ in $F$ we have for almost all $s$ and all $n$ in $I_s$,

$$2s + c \leq 3s \leq f(n_s) \leq f(n).$$

Proof of (ii). The construction is similar to the one given for assertion (i) and we just indicate the necessary changes. The martingale against which we diagonalize during stage $s$ is now a convex sum over all $d_i$ with $i < s$, except that on input $w$ we omit all the $d_i$ where $d_i(v)$ is undefined for some prefix $v$ of $w$. In order to be able to effectively simulate the construction up to and including stage $s$, in worst case this requires the information about $s$ places at which one or more of the betting strategies $d(i)$ are not defined. So, in order to effectively simulate the construction up to the definition of $R(n)$ where $n$ is in interval $I_s$, it suffices to supply $s$ numbers less than or equal to $n$ plus the set $F_s$. Coding this information requires not more that $3s \log n$ bits, i.e., requires at most $f(n) \log n$ bits for all $f \in F$ and almost all $n$. □
Remark 4. Up to an additive constant, the Kolmogorov complexity $C(w)$ and the uniform complexity $C(w; n)$ of a word $w$ of length $n$ differ at most by $2 \log n$.

For a sketch of proof, consider the additively optimal reference Turing machines that have been used when defining Kolmogorov complexity and uniform complexity. Any code $x$ for a word $w$ with respect to one of these machines can be transformed into a code for $w$ with respect to the other machine by adding some finite information about how the information given by $x$ is to be used, plus, in case one transforms a code for the latter to a code for the former reference Turing machine, information about the length of $w$, e.g., in this case $x$ may be transformed into $1^{|z|}0^{|z|}x$ where $z$ is the binary expansion of the length of $w$.

Remark 4 shows that when considering bounds of the form $O(\log n)$ or $f(n) \log n$ for unbounded $f$, usually it will not be necessary to distinguish between Kolmogorov complexity and uniform complexity. Accordingly, in what follows we will only consider Kolmogorov complexity; for a start, we rephrase assertion (ii) in Theorem 3 in terms of Kolmogorov complexity.

Corollary 5. Let $F$ be the class of all computable functions from $\mathbb{N}$ to $\mathbb{N}$ that are non-decreasing and unbounded. There is a partial-recursively random sequence $R$ such that for all $f \in F$ and almost all $n$,

$$C(R(0) \ldots R(n-1)) \leq f(n) \log n. \quad (4)$$

Proof. Let $R$ be a sequence according to assertion (ii) in Theorem 3. Fix any $f$ in $F$. Then also the function

$$n \mapsto \max\{0, f(n) - 3\}$$

is in $F$, hence by choice of $R$ we have for almost all $n$,

$$C(R(0) \ldots R(n-1); n) \leq (f(n) - 3) \log n.$$ 

By Remark 4, the latter inequality implies (4) for all sufficiently large $n$. \qed

Remark 6. The assertions in Theorem 3 are variants of known results on stochastic and random sequences. Similar results on stochastic sequences are attributed to Loveland by Daley [6] (see also Li and Vitányi [9, Exercise 2.5.13 and comments]), where in Loveland’s results the notions of recursively and partial-recursively random are replaced by stochastic with respect to computable selection rules and by Mises–Wald–Church stochastic, respectively. Randomness implies the corresponding notion of stochasticity in the sense that for example any partial-recursively random sequence is Mises–Wald–Church stochastic [4,13], hence Loveland’s results are a special case of Theorem 3. However, the proof of Theorem 3 stated in terms of betting strategies is less involved than the original proofs of the more specific results on stochastic sequences, which rely on a combinatorial algorithm for constructing stochastic sequences, the LMS-algorithm [10].

Lathrop and Lutz [8] introduced ultracompressible sequences, i.e., sequences $X$ such that for every computable, non-decreasing and unbounded function $g$ and for almost all $n$ we have

$$K(X(0) \ldots X(n-1)) \leq K(n) + g(n),$$

where $K$ is the prefix-free version of Kolmogorov complexity [9]; they demonstrated that there is a recursively random ultracompressible sequence, which is roughly equivalent to and, in particular, is an easy consequence of assertion (i) in Theorem 3. A marginally weaker form of assertion (ii) in Theorem 3 has been demonstrated by Muchnik [13, Theorem 9.5], where the complexity of the prefixes of the constructed random sequence is bounded by $f(n) \log n$ for any given single computable, non-decreasing and unbounded function $f$, instead of being eventually bounded by any such function. The ideas and methods used by Lathrop and Lutz and by Muchnik for proving their results are essentially the same as in the proof of Theorem 3.

Can the factor $f(n)$ in Corollary 5 be improved to a constant, i.e., are there partial-recursively random sequences where for almost all $n$ or, equivalently, for all $n > 0$, the length $n$ prefix has complexity $O(\log n)$? The next theorem gives a negative answers to this question; sequences of such low complexity cannot even be found in the more comprising class of Mises–Wald–Church stochastic sequences.
Theorem 7. Let $X$ be a sequence such that for some natural number $c$ and almost all $n > 1$, 
\[ C(X(0) \ldots X(n - 1)) \leq c \log n. \] 
(5)

Then $X$ is not Mises–Wald–Church stochastic.

Proof. Let $k = 2c + 2$. Fix an appropriate computable sequence $m_0, m_1, \ldots$ where all the $m_s$ are multiples of $k + 1$ and such that $m_0 > 0$ and for all $s > 0$, 
\[ 10(m_0 + \cdots + m_{s-1}) < \frac{ms}{k+1}. \] 
(6)

Partition the natural numbers into consecutive, non-overlapping intervals $I_0, I_1, \ldots$ where $I_s$ has length $m_s$, i.e., $I_0$ contains the least $m_0$ natural numbers, $I_1$ the next $m_1$ natural numbers, and so on. Divide each interval $I_s$ into $k + 1$ consecutive, non-overlapping subintervals 
\[ J_s^{1}, \ldots, J_s^{k+1} \] 
(7)

of identical length $l_s = \frac{ms}{k+1}$ and let $w_s^e$ be the word that is associated to the restriction of $X$ to $J_s^e$.

Now assume that for some $t$, there is a procedure that given $s$ and the restriction of $X$ to the set of all numbers up to but not including the minimum element of $J_s^t$, enumerates a set $T_s^t$ of words such that

(i) $w_s^e$ is in $T_s^t$ for almost all $s$,

(ii) $|T_s^t| \leq 0.2l_s$ for infinitely many $s$.

Then one of the following selection rules $r_0$ and $r_1$ witness that $X$ is not Mises–Wald–Church stochastic. Intuitively speaking, $r_i$ tries to select places in intervals of the form $J_s^i$ where the corresponding bit is $i$. For all $s$, let $v_s^1, v_s^2, \ldots$ be the assumed enumeration of $T_s^i$, where we may suppose that the enumeration is without repetitions. Pick $s_0$ such that $w_s$ is in $T_s^t$ for all $s \geq s_0$. Both selection rules select only numbers in intervals of the form $J_s^i$ where $s \geq s_0$; on entering such an interval, $r_i$ lets $e = 1$ and starts scanning the numbers in the interval. Assuming that the restriction of $X$ to $I_s$ is given by $v_s^e$, the selection rule $r_i$ selects the number $n$ if and only if the corresponding bit of $v_s^e$ is $i$. This is done until either the end of the interval is reached or one of the scanned bits differs from the corresponding bit of $v_s^e$; in the latter case, the index $e$ is incremented and the procedure iterates by scanning the remaining bits. Observe that $v_s^e$ is always defined by choice of $s$ and because iteration $e$ is only reached in case the true word $w_s^e$ is not among $v_s^e$ through $v_s^{e-1}$.

By construction, for all $s \geq s_0$, every number in interval $J_s^i$ is selected by either $r_0$ or $r_1$. For the scope of this proof, say a number is selected correctly if it is selected by $r_i$ and the corresponding bit is indeed $i$. Then in each interval $J_s^i$, there are at most $|T_s^i| - 1$ numbers $n$ that are selected incorrectly. Hence by assumption, for infinitely many $s$, there at least $0.8l_s$ numbers in the interval $J_s^i$ that are selected correctly, and thus for some $i$ and infinitely many $s$, the selection rule $r_i$ selects among the numbers in $J_s^i$ at least $0.4l_s$ numbers correctly and at most $0.2l_s$ numbers incorrectly; moreover, by (6) and (7) there are at most $0.1l_s$ numbers that $r_i$ could have been selected before entering the interval. Hence up to and including each such interval $J_s^i$, the selection rule $r_i$ selects at least $0.4l_s$ numbers correctly and at most $0.3l_s$ numbers incorrectly, i.e., $r_0$ witnesses that $X$ is not Mises–Wald–Church stochastic.

It remains to argue that for some $t$, there is indeed a procedure as assumed above, i.e., which enumerates sets $T_s^t$ that satisfy (i) and (ii). For all $s$, let $w_s$ be the word that is associated to the restriction of $X$ to the interval $I_s$ and let 
\[ A_s = \{ w : |w| = m_s \text{ and } C(w) < k \log m_s \}. \]

Then $w_s$ is in $A_s$ for almost all $s$. For a proof, observe that for almost all $s$, due to (5) and $m_0 + \cdots + m_{s-1} < m_s$, we have
\[ C(X(0) \ldots X(m_0 + \cdots + m_s)) \leq c(\log(m_0 + \cdots + m_s)) \leq c(1 + \log(m_s)) \leq 2c \log(m_s), \]
i.e., there is a word $x$ of length at most $2c \log m_s$ from which the reference Turing machine computes the restriction of $X$ to the intervals $I_0$ through $I_s$. By prefixing $w$ with the word $0^t1$, which for all $s > 0$ has length of at most $\log m_s$,
we obtain a code from which some Turing machine that does not depend on \( s \) computes the restriction of \( X \) to \( I_s \). This implies that for almost all \( s \), up to an additive constant that does not depend on \( s \),

\[
C(X \mid I_s) \leq (2c + 1) \log m_s = (k - 1) \log m_s,
\]

hence by definition of \( A_s \), the word \( w_s \) is in \( A_s \) for almost all \( s \).

In order to obtain sets \( T^j_s \) as required, let for all \( s \geq 0 \) and for \( j = 1, \ldots, k + 1 \),

\[
T^{j+1}_s = \{ v \colon w^1_s \ldots w^k_s v \in A_s \},
\]

\[
T^j_s = \{ v \colon |v| = l_s \text{ and there are at least } (0.2l_s)^{k+1-j} \text{ words } u \text{ such that } w^1_s \ldots w^{j-1}_s vu \in A_s \}.
\]

There is a Turing machine that on input \( s \) enumerates \( A_s \), hence there is a Turing machine that given the indices \( j \) and \( s \) and the word \( w^1_s \ldots w^{j-1}_s \), enumerates the set \( T^j_s \); i.e., for any \( t \) the sets \( T^j_s \) satisfy the condition on enumerability and it suffices to show that (i) and (ii) are true for some \( t \).

For a start, observe that in case (ii) is not satisfied for some \( t > 0 \), then condition (i) is satisfied with \( t \) replaced by \( t - 1 \). Indeed, if (ii) is false, then for almost all \( s \) there are at least \( 0.2l_s \) words \( v \) in \( T^j_s \), where each of these words can be extended by at least \( (0.2l_s)^{k+1-j} \) words \( u \) to a word \( w^1_s \ldots w^{j-1}_s vu \in A_s \). Consequently, for each such \( s \) there are at least \( (0.2l_s)^{k+1-u-1} \) words \( uv \) that extend \( w^1_s \ldots w^{j-1}_s \) to a word in \( A_s \), i.e., for almost all \( s \), the word \( w^{j-1}_s \) is in \( T^j_s \).

Condition (i) is satisfied for \( t = k + 1 \), so if (ii) is satisfied, too, we are done by just letting \( t = k + 1 \). Otherwise, by the discussion in the preceding paragraph, condition (i) is satisfied for \( t = k \). Now we can iterate the argument; if (ii) is satisfied for \( t = k \), we are done by letting \( t = k + 1 \) while, otherwise, condition (i) holds for \( t = k - 1 \). This way we proceed inductively and it remains to argue that it cannot be that (ii) is false for \( t = k + 1, \ldots, 1 \). Assuming the latter, for almost all \( s \) there at least \( 0.2l_s \) many assignments on \( J^1_s \) that can be extended in \( (0.2l_s)^k \) ways to a word in \( A_s \), thus for all sufficiently large \( s \) and some constant \( \varepsilon > 0 \),

\[
|A_s| \geq (0.2l_s)^{k+1} \geq \left( \frac{m_s}{k+1} \right)^{k+1} = \varepsilon m_s^{k+1} > m_s^k.
\]

This contradicts the fact that the set \( A_s \) has by definition at most \( m_s^k \) members because for any \( n \), there are less than \( 2^n \) words \( w \) where \( C(w) < n \) since there are less than \( 2^n \) possible codes of length strictly less than \( n \). \( \square \)

By the first assertion in Theorem 3, there is a recursively random sequence \( R \) such that for almost all \( n \) the uniform complexity of the length \( n \) prefix of \( R \) is at most \( f(n) = \log \log n \), hence by Remark 4, the Kolmogorov complexity of the prefix is at most \( \log \log n + 2 \log n \), which is less that \( 3 \log n \) for almost all \( n \). The following known result [1] is then immediate by Theorem 7.

**Corollary 8.** The class of recursively random sequences is not contained in the class of Mises–Wald–Church stochastic sequences.

In fact it is known that neither of the two classes in the corollary is contained in the other; a sequence that is Mises–Wald–Church stochastic but not recursively random can be obtained by a probabilistic argument where the bits of the sequence are chosen by independent tosses of biased coins where the probabilities for 0 converge slowly enough to 1/2 [7,12,15].

**Acknowledgments**

We thank Klaus Ambos-Spies, Nicolai Vereshchagin, and Paul Vitányi for helpful discussions and Jack Lutz, Alexander Shen, and the anonymous referees for their comments.

**References**


