On selection functions that do not preserve normality

Wolfgang Merkle Jan Reimann

Ruprecht-Karls-Universität Heidelberg
Institut für Informatik
Im Neuenheimer Feld 294
D-69120 Heidelberg, Germany
{merkle|reimann}@math.uni-heidelberg.de

Abstract. The sequence selected from a sequence $R(0)R(1)\dots$ by a language L is the subsequence of R that contains exactly the bits $R(n+1)$ such that the prefix $R(0)\dots R(n)$ is in L. By a result of Agafonoff, a sequence is normal if and only if any subsequence selected by a regular language is again normal. Kamae and Weiss and others have raised the question of how complex a language must be such that selecting according to the language does not preserve normality. We show that there are such languages that are only slightly more complicated than regular ones, namely, normality is preserved neither by deterministic one-counter languages nor by linear languages. In fact, for both types of languages it is possible to select a constant sequence from a normal one.

1 Introduction

It is one of the fundamental beliefs about chance experiments that we can expect that an infinite binary sequence obtained by independent tosses of a fair coin will contain, in the long run, any given finite sequence of length n with frequency 2^{-n}. Infinite binary sequences having this property are called *normal*. It is a basic result of probability theory that, with respect to the uniform Bernoulli measure, almost all sequences are normal. In what follows, we investigate into the following problem: If we select from a normal sequence an infinite subsequence, under what selection mechanisms is the thereby obtained sequence again normal, i.e., which restrictions on the class of admissible selection rules suffice to guarantee that normality is preserved.

This problem originated in the work of von Mises [14]. His aim was to base a mathematical theory of probability on the notion of a *"Kollektiv"*. In the case of kollektivs that are infinite binary sequences, he postulates, first, that the symbols 0 and 1 possess asymptotic frequencies, which allow in turn to assign probabilities and, second, that these limiting frequencies are preserved when a subsequence is selected from the original sequence. The latter postulate implies that one has to restrict the set of admissible selection rules, or place selection rules, as von Mises calls them, since otherwise, i.e., by using arbitrary selection rules, one may simply select all zeroes or all ones from any given sequence.

Von Mises did not specify which selection rules should be admissible, except for basic requirements such as the condition that the selection of a place must not depend on the value of the sequence at that place. There have been various attempts to clarify and rigorously define what an admissible selection rule and hence a kollektiv is. One approach allowed only rules that were in some sense effective, for instance, computable by a Turing machine. This effort was initiated by Church [8] and lead to the study of *stochastic sequences*, where a sequence is stochastic with respect to a given set of admissible selection rules if any infinite subsequence selected by an admissible selection rule is unbiased in the sense that in the limit the frequencies of 0's and 1's are both equal to 1/2 (see the survey by Uspensky, Semenov and Shen [19] for more on this). In this context it is natural to ask which types of sequences are stochastic with respect to what types of selection rules, where then a sequence is considered to be "more random" if the sequence is stochastic with respect to a more comprehensive set of selection rules. More specifically, one may ask with respect to which selection rules normal sequences are stochastic and, related and to some extent equivalent to this question, which selection rules preserve normality, i.e., map normal sequences to normal ones (see Remark 12).

From a purely measure theoretic point of view, normal sequences might appear as reasonable candidates for kollektivs, as they possess limit frequencies. Furthermore, they are random in the sense that their dynamic behavior is as complex as possible: They are generic points in $\{0,1\}^\infty$ with respect to a measure with highest possible entropy – the uniform Bernoulli measure (see Section 3 and Weiss [20]). On the other hand, from Champernowne's construction of a normal sequence that is highly regular (see Section 2 and Champernowne [7]), it is clear that normal numbers can only be stochastic with respect to admissible selection rules that are defined via rather restricted models of computation.

There are two kinds of selection rules that are commonly considered: oblivious ones, for which the decision of selecting a bit for the subsequence does not depend on the input sequence up to that bit (i.e., the places to be selected are fixed in advance), and selection rules that depend on the input sequence.

For oblivious selection rules, Kamae [9, 11, 20] gave a complete characterization of the selection rules that preserve normality. In the setting of input-dependent selection rules, Agafonoff [1] obtained the result that selection by regular languages preserves normality, i.e., any infinite sequence selected by a regular language from a normal sequence is again normal (for further details and references see Section 3). An easy consequence of Agafonoff's result is that a sequence is normal if and only if it is stochastic with respect to selection rules induced by regular languages.

Kamae and Weiss [10] asked whether Agafonoff's result can be extended to classes of languages that are more comprehensive than the class of regular languages, e.g., to the class of context-free languages (see also Li and Vitányi [12], p. 59, problem 1.9.7). In what follows, we give a negative answer to this question for two classes of languages that are minimum among all superclasses of the class of regular languages that are usually considered in the theory of formal languages. More precisely, Agafonoff's result can neither be extended to the class of languages that are recognized by deterministic pushdown automata with unary stack alphabet, also known as deterministic one-counter languages, nor to the class of linear languages. Recall that these two classes are incomparable and that the latter fact is witnessed, for example, by the languages used in the proofs of Propositions 14 and 16, i.e., the language of all words that contains as many 0's as 1's and the language of even-length palindromes (for background on formal language theory we refer to the survey by Autebert, Berstel, and Boasson [4]). While our results add some insight on normality preserving input-dependent selection rules, the general problem of an exact classification of such selection rules in the spirit of Kamae's result remains open.

The outline of the paper is as follows. In Section 2 we review the basic definitions related to normality and recap Champernowne's constructions of normal sequences. Section 3 discusses the two kinds of selection rules, oblivious and input-dependent ones. In Section 4, we show that normality is not preserved by selection rulesdefined by deterministic push-down automata with unary stack alphabet, while Section 5 is devoted to proving that normality is not preserved by linear languages.

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, 12, 13, 16]. Unless explicitly stated otherwise, sequences are always infinite and binary. A word is a finite sequence. For $i = 0, 1, \ldots$, we write $A(i)$ for bit i of a sequence A, hence $A = A(0)A(1)\ldots$, and we proceed similarly for words. A word w is a prefix of a sequence A if $A(i) = w(i)$ for $i = 0, \ldots, |w|-1$, where $|w|$ is the length of w. The prefix of a sequence A of length m is denoted by $A|m$. The concatenation of two words v and w is denoted by vw. A word u is a subword of a word w if $w = v_1 u v_2$ for appropriate words v_1 and v_2.

2 Normal sequences

For a start, we review the concept of a normal sequence and standard techniques for the construction of such sequences.

Definition 1. *(i) For given words u and w, let $\mathrm{occ}_u(w)$ be the number of times that u appears as a subword of w, and let $\mathrm{freq}_u(w) = \mathrm{occ}_u(w)/|w|$.*
(ii) A sequence N is normal *if and only if for any word u*

$$\lim_{m\to\infty} \mathrm{freq}_u(N|m) = \frac{1}{2^{|u|}}. \tag{1}$$

Remark 2. A sequence N is normal if for any word u and any $\varepsilon > 0$, we have for all sufficiently large m,

$$\mathrm{freq}_u(N|m) < \frac{1}{2^{|u|}} + \varepsilon. \tag{2}$$

For a proof, it suffices to observe that for any given $\varepsilon > 0$ and for all sufficiently large m, inequality (2) holds with u replaced by any word v that has the same length as u, while the sum of the relative frequencies $\mathrm{freq}_v(N|m)$ over these $2^{|u|}$ words differ from 1 by less than ε; hence by (2) all such m,

$$(1-\varepsilon) \le \sum_{\{v:|v|=|u|\}} \mathrm{freq}_v(N|m) \le \mathrm{freq}_u(N|m) + (2^{|u|}-1)(\tfrac{1}{2^{|u|}} + \varepsilon),$$

and by rearranging terms we obtain

$$\frac{1}{2^{|u|}} - 2^{|u|}\varepsilon < \mathrm{freq}_u(N|m)\ .$$

Together with (2) this implies that $\mathrm{freq}_u(N|m)$ converges to $2^{-|u|}$, because $\varepsilon > 0$ has been chosen arbitrarily.

Definition 3. *A set W of words is* normal in the limit *if and only if for any nonempty word u and any $\varepsilon > 0$ for all but finitely many words w in W,*

$$\frac{1}{2^{|u|}} - \varepsilon < \mathrm{freq}_u(w) < \frac{1}{2^{|u|}} + \varepsilon. \tag{3}$$

Definition 4. *For any n, let $\mathrm{v}_n = 0^n\ 0^{n-1}1\ 0^{n-2}10\ \ldots\ 1^n$ be the word that is obtained by concatenating all words of length n in lexicographic order.*

Proposition 5. *The set $\{\mathrm{v}_1, \mathrm{v}_2, \ldots\}$ is normal in the limit.*

Proof. By an argument similar to the one given in Remark 2, it suffices to show that for any word u and any given $\varepsilon > 0$ we have for almost all words v_i,

$$\mathrm{freq}_u(\mathrm{v}_i) < \frac{1}{2^{|u|}} + \varepsilon. \tag{4}$$

So fix u and $\varepsilon > 0$ and consider any index i such that $|u|/i < \varepsilon$. Recalling that v_i is the concatenation of all words of length i, call a subword of v_i undivided if it is actually a subword of one of these words of length i, and call all other subwords of v_i divided. It is easy to see that u can occur at most $2^i|u|$ many times as a divided subword of v_i. Furthermore, a symmetry argument shows that among the at most $|\mathrm{v}_i|$ many undivided subwords of v_i of length $|u|$, each of the $2^{|u|}$ words of length $|u|$ occurs exactly the same number of times. In summary, we have

$$\mathrm{occ}_u(\mathrm{v}_i) \leq \frac{|\mathrm{v}_i|}{2^{|u|}} + 2^i|u| = |\mathrm{v}_i| \left(\frac{1}{2^{|u|}} + \frac{2^i|u|}{|\mathrm{v}_i|} \right) < |\mathrm{v}_i| \left(\frac{1}{2^{|u|}} + \varepsilon \right) ,$$

where the last inequality follows by $|\mathrm{v}_i| = 2^i i$ and the choice of i. Inequality (4) is then immediate by definition of $\mathrm{freq}_u(\mathrm{v}_i)$. □

Lemma 6. *Let W be a set of words that is normal in the limit. Let $w_1, w_2, \ldots$ be a sequence of words in W such that*

$$\text{(i) for all } w \in W,\ \lim_{t\to\infty} \frac{|\{i \leq t \colon w_i = w\}|}{t} = 0, \quad \text{(ii)} \lim_{t\to\infty} \frac{|w_{t+1}|}{|w_1 \ldots w_t|} = 0.$$

Then the sequence $N = w_1 w_2 \ldots$ is normal.

Proof. By Remark 2 it suffices to show that (2) holds for any nonempty word u, any $\varepsilon > 0$, and all sufficiently large m. Fix any such u and ε and let $\varepsilon_0 = \varepsilon/4|u|$. By definition of the concept of normal in the limit, pick $n > 1/\varepsilon_0$ such that for all words w in W where $|w| > n$, we have

$$\mathrm{occ}_u(w) < |w|(\frac{1}{2^{|u|}} + \varepsilon_0). \tag{5}$$

For the scope of this proof, call the finitely many words in W of length at most n short words and call the remaining words in W long words. By assumption of the lemma, pick t_0 such that for all $t \geq t_0$,

$$\text{(i)} \quad \frac{|\{i \leq t \colon w_i \text{ short }\}|}{t} < \varepsilon_0, \qquad \text{(ii)} \quad \frac{|w_{t+1}|}{|w_1 \ldots w_t|} < \varepsilon_0. \tag{6}$$

Any prefix z of N that properly extends $w_1 \ldots w_{t_0}$ can be written in the form

$$z = w_1 \ldots w_t v$$

where $t \geq t_0$ and v is a nonempty proper prefix of w_{t+1}. Then, by definition of $\mathrm{freq}_u(N|m)$, in order to demonstrate (2) it suffices to show that we have (for numbers $j_1, \ldots, j_4$ to be specified in a minute),

$$\mathrm{occ}_u(z) \leq j_1 + j_2 + j_3 + j_4 < |z| \left(\frac{1}{2^{|u|}} + \varepsilon \right). \tag{7}$$

The j_i count the number of four different types of occurrences of u as a subword of z; more precisely,

j_1 counts the occurrences of u that are a subword of a single long w_i,
j_2 counts the occurrences of u that overlap with both of two adjacent long w_i,
j_3 counts the occurrences of u that overlap with a short subword w_i,
j_4 counts the occurrences of u that overlap with v.

Then the first inequality in (7) holds because each occurrence of u as a subword of z counts for at least one of these four numbers. In order to derive the second inequality, first observe that for all long w_i the number of occurrences of u as a subword of w_i is bounded according to (5), hence

$$j_1 < \sum_{\{i \leq t \colon w_i \text{ long}\}} |w_i| \left(\frac{1}{2^{|u|}} + \varepsilon_0 \right) \leq |z| \left(\frac{1}{2^{|u|}} + \varepsilon_0 \right).$$

Concerning j_2, recall that long words have length of at least $n \geq 1/\varepsilon_0$, hence there are at most $|z|\varepsilon_0$ many pairs of adjacent long words and at most $|z|\varepsilon_0|u|$ many subwords u that overlap with both words in such a pair. Concerning j_3, recall that by inequality (i) in (6), the fraction of short words is bounded by ε_0, while any short word is strictly shorter than any long word. Consequently, the number of positions of z that belong to a short word is at most $|z|\varepsilon_0$, and the number of subwords u that overlap with a short word is at most $|z|\varepsilon_0|u|$. Finally, by inequality (ii) in (6), we have $|v| < |w_{t+1}| \leq |z|\varepsilon_0$, hence the number of subwords u that overlap with v is again bounded by $|z|\varepsilon_0|u|$. In summary, by the preceding discussion and choice of ε_0, we obtain

$$j_1 + j_2 + j_3 + j_4 < |z| \left(\frac{1}{2^{|u|}} + 4\varepsilon_0|u| \right) = |z| \left(\frac{1}{2^{|u|}} + \varepsilon \right). \qquad \square$$

Remark 7. The sequence $\mathrm{v}_1\mathrm{v}_2\mathrm{v}_2\mathrm{v}_3\mathrm{v}_3\mathrm{v}_3\mathrm{v}_4\ldots$, which consists of i copies of v_i concatenated in length-increasing order, is normal. This assertion is immediate by definition of the sequence, Proposition 5, and Lemma 6.

The arguments and techniques in this section are essentially the same as the ones used by Champernowne [7], who considered normal sequences over the decimal alphabet $\{0, 1, \ldots, 9\}$ and proved that the decimal analogues of the sequences

$$N_1 = \mathrm{v}_1\mathrm{v}_2\mathrm{v}_2\mathrm{v}_3\mathrm{v}_3\mathrm{v}_3\mathrm{v}_4\ldots \qquad \text{and} \qquad N_2 = \mathrm{v}_1\mathrm{v}_2\mathrm{v}_3\mathrm{v}_4\ldots$$

are normal. In Remark 7, we have employed Lemma 6 and the fact that the set of all words v_i is normal in the limit in order to show that N_1 is normal. In order to demonstrate the normality of the decimal analogue of the sequence N_2, Champernowne [7, item (ii) on page 256] shows a fact about the decimal versions of the v_i that is stronger than just being normal in the limit, namely, for any word u and any constant k, we have for all sufficiently large i, and all $m \leq |\mathrm{v}_i|$,

$$\mathrm{freq}_u(\mathrm{v}_i(0)\ldots\mathrm{v}_i(m-1)) < \frac{m}{2^{|u|}} + \frac{|\mathrm{v}_i|}{k}. \tag{8}$$

This result is then used in connection with a variant of Lemma 6 where in place of assumption (ii), which asserts that the ratio $|w_{t+1}|$ and $|w_1 \ldots w_t|$ converge to 0, it is just required that this ratio is bounded.

3 Selecting subsequences

Oblivious selection rules are a restricted type of selection rule, where the places to be included in the selected sequence are fixed in advance, i.e., do not depend on the sequence from which they are selected.

Definition 8. *An* oblivious selection rule *is a sequence $S \in \{0,1\}^\infty$. The sequence B selected by an oblivious selection rule S from a sequence A is just the subsequence of A of all bits $A(i)$ with $S(i) = 1$, that is,*

$$B = A(i_1)A(i_2)\ldots \quad \textit{where } i_1 < i_2 < \ldots \textit{ and } \{i_1, i_2, \ldots\} = \{i : S(i) = 1\}.$$

Kamae [9] gave a complete characterization of the class of oblivious selection rules that preserve normality. Let T be the *shift map*, which transforms a sequence $S = S(0)S(1)S(2)\ldots$ into another sequence by cancelling the first bit, i.e., $T(S) = S(1)S(2)\ldots$. Given a sequence S, let δ_S denote the Dirac measure induced by S, that is, for any class $\boldsymbol{C}$ of sequences, $\delta_S(\boldsymbol{C}) = 1$ if $S \in \boldsymbol{C}$ and $\delta_S(\boldsymbol{C}) = 0$ otherwise. Kamae showed that for any oblivious selection rule S that has non-zero density in the sense that $\liminf_{n\to\infty} \mathrm{freq}_1(S|n) > 0$, the selection rule S preserves normality if and only if S is *completely deterministic*, that is, any cluster point (in the weak-$*$ topology) of the sequence of measures

$$\left(\frac{1}{n}\sum_{j=0}^{n}\delta_{T^j(S)}\right)_{n\in\mathbb{N}} \tag{9}$$

has entropy 0. Observe in this connection that for example a normal sequence S will not preserve normality, and that indeed for normal S any cluster point of the sequence of measures in (9) yields the uniform Bernoulli measure, which has entropy 1 (see Weiss [20] for further details).

In the remainder of this paper we consider selection rules that depend on the input sequence and we investigate which input-dependent selection rules preserve normality. For a start, we formally define how selection by an input-dependent selection rule works.

Definition 9. *Let A be a sequence and let L be a language. The sequence* selected from A by L *is the subsequence of A that contains exactly the bits $A(i)$ of A such that the prefix $A(0)\dots A(i-1)$ is in L.*

Along the lines of the work of von Mises and Church, one can define a stochasticity concept based on selection rules defined by regular languages.

Definition 10. *A sequence A is* stochastic *with respect to a given class of languages if for every language L from this class, it holds that if L selects from A an infinite sequence, then in the limit the frequency of* 1 *in the selected subsequence is* $1/2$. *A sequence is* regular stochastic *if it is stochastic with respect to the class of regular languages.*

By a celebrated result of Agafonoff [1, 10, 18], the class of normal sequences is closed under selection of subsequences by regular languages. An immediate consequence and, by Remark 12 below, in fact a reformulation of Agafonoff's result is that any normal sequence is regular stochastic; since also the reverse implication holds, we have that

$$A \text{ is regular stochastic} \quad \text{if and only if} \quad A \text{ is normal} \tag{10}$$

The forward implication in (10) has been attributed to Agafonoff by Schnorr and Stimm [18, Section 6], who also give a detailed proof of this result. The backwards implication in (10) has been observed first by Postnikova [17]; for ease of reference we briefly review the corresponding argument in Remark 11. In connection with Agafonoff's result and equivalence (10), see also O'Connor [15] and Broglio and Liardet [6].

Remark 11. Any regular stochastic sequence is normal.

For a proof by contraposition, fix a sequence A that is not normal and let u be a nonempty word of minimal length such that u does not occur in A in the limit with frequency $1/2^{|u|}$. In case u has length 1, we are done because by selecting all bits we obtain a subsequence that is biased, i.e., where in the limit the frequency of 0 and 1 differ from $1/2$. Otherwise, we have $w = vr$ for some nonempty word v and r in $\{0,1\}$. Then the selection rule that selects exactly the bits following an occurrence of v is induced by a regular language and selects from A a biased sequence. Assuming otherwise, in the limit half of the occurrences of v are followed by r, while by minimality of w the prefix v of w occurs in A in the limit with frequency $1/2^{|v|}$, hence by $|w| = |v| + 1$ also w occurs in the limit with its expected frequency $1/2^{|w|}$, thus contradicting the choice of w.

Remark 12. The assertion that any normal sequence is regular stochastic is essentially a reformulation of Agafonoff's result.

The fact that selection by regular languages preserve normality implies by definition of the involved concepts that any normal sequence is regular stochastic. But also the reverse implication is straightforward. If there was a regular language that selected from a normal sequence N a sequence N' that is not normal, then by Remark 11 in turn one could select a biased sequence N'' from N' by another regular language. But then the biased sequence N'' could be selected from the normal sequence N directly by another regular language, i.e., the sequence N would be normal but not regular stochastic. For a proof of the latter, observe that the class of selection rules that are induced by regular languages is closed under composition, as can be seen by "composing" the corresponding finite automata.

The classes of deterministic one-counter languages and of linear languages are both minimum among the superclasses of the regular languages that are usually studied in formal language theory. In the remainder of this article we construct two counter-examples that show that selection by neither of these two types of languages preserve normality. While this adds some insight on normality preserving input-dependent selection rules, the general problem of an exact classification of such selection rules in the spirit of Kamae's result remains open. Note in this connection that by Remark 13, there are arbitrary complex languages that preserve normality.

Remark 13. Every Turing degree contains a language L such that selection by L preserves normality.

Recall the usual identification of languages and sets of natural numbers via identifying the natural number i with the $(i+1)$st word in length-lexicographical ordering. Fix an appropriate computable infinite sequence of pairwise incomparable words $w_0, w_1, \ldots$, say, $w_i = 0^i 1$ and for any given set X of natural numbers, let

$$L_X = \{0,1\}^* \setminus \{w_i : i \notin X\} .$$

By construction, the language L_X is Turing-equivalent to X. Furthermore, since the set of all w_i is prefix-free, for any given sequence, the language L_X contains all prefixes of the sequence except for at most one, hence L_X selects all bits of the sequence except for at most one and consequently L_X preserves normality.

4 Normality is not preserved by deterministic one-counter languages

Proposition 14. *There is a normal sequence N and a deterministic one-counter language L such that the sequence selected from N by L is infinite and constant.*

Proof. Recall that $\text{occ}_r(w)$ is the number of occurrences of the symbol r in w; for any word w, let

$$d(w) = \text{occ}_0(w) - \text{occ}_1(w) .$$

Let L be the language of all words that have as many 0's as 1's, i.e.,

$$L = \{w \in \{0,1\}^* : d(w) = 0\} .$$

The language L can be recognized by a deterministic push-down automata with unary stack alphabet that for the already scanned prefix v of the input stores the sign and the absolute value of $d(v)$ by its state and by the number of stack symbols, respectively.

Recall from Section 2 that v_i is obtained by concatenating all words of length i and that by Remark 7, the sequence

$$N = \mathrm{v}_1 \mathrm{v}_2 \mathrm{v}_2 \mathrm{v}_3 \mathrm{v}_3 \mathrm{v}_3 \mathrm{v}_4 \ldots \tag{11}$$

is normal. For the scope of this proof, call the subwords v_i of N in (11) designated subwords. Furthermore, for all t, let z_t be the prefix of N that consists of the first t designated subwords. Every prefix of the form z_t of N is immediately followed by the $(t+1)$st designated subword, where each designated subword starts with 0. Hence the proposition follows, if we can show that among all prefixes w of N, exactly the z_t are in L, or equivalently, exactly the prefixes w that are equal to some z_t satisfy $d(w) = 0$.

Fix any prefix w of N. Choose t maximum such that z_t is a prefix of w and pick v such that $w = \mathrm{z}_t v$. By choice of t, the word v is a proper prefix of the $(t+1)$st designated subword and v is equal to the empty word if and only if w is equal to some z_i. By additivity of d, we have

$$d(w) = d(\mathrm{z}_t) + d(v) = d(\mathrm{v}_{i_1}) + \ldots + d(\mathrm{v}_{i_t}) + d(v) \tag{12}$$

for appropriate values of the indices i_j. Then in order to show that w is equal to some z_i if and only if $d(w) = 0$, it suffices to show that for all s,

$$\text{(i) } d(\mathrm{v}_s) = 0, \qquad \text{(ii) } d(u) > 0 \text{ for any nonempty proper prefix } u \text{ of } \mathrm{v}_s\,. \tag{13}$$

We proceed by induction on i. For $i = 0$ there is nothing to prove, so assume $i > 0$. Let v_i^0 and v_i^1 be the first and the second half of v_i, respectively. For $r = 0, 1$, the word v_i^r is obtained from v_{i-1} by inserting 2^{i-1} times r, where $d(\mathrm{v}_{i-1}) = 0$ by the induction hypothesis. Hence (i) follows because

$$d(\mathrm{v}_i) = d(\mathrm{v}_i^0) + d(\mathrm{v}_i^1) = d(\mathrm{v}_{i-1}) + 2^{i-1} + d(\mathrm{v}_{i-1}) - 2^{i-1} = 0\,.$$

In order to show (ii), fix any nonempty proper prefix u of v_i. First assume that u is a proper prefix of v_i^0. Then u can be obtained from a nonempty, proper prefix of v_{i-1} by inserting some 0's, hence we are done by the induction hypothesis. Next assume $u = \mathrm{v}_i^0 v$ for some proper prefix v of v_i^1. We have already argued that by the induction hypothesis $d(\mathrm{v}_i^0)$ is equal to 2^{i-1}. Furthermore, v can be obtained from a proper prefix v' of v_{i-1} by inserting at most 2^{i-1} many 1's, where by the induction hypothesis we have $d(v') > 0$. In summary, we have

$$d(u) = d(\mathrm{v}_i^0) + d(v) \geq 2^{i-1} + d(v') - 2^{i-1} > 0\,,$$

which finishes the proof of the proposition. □

Recalling that every regular language is a deterministic one-counter language, the following corollary is immediate by Proposition 14 and because Agafonoff's result implies that normal sequences are stochastic with respect to regular languages.

Corollary 15. *The sequences that are stochastic with respect to deterministic one-counter languages form a proper subclass of the class of sequences that are stochastic with respect to regular languages.*

5 Normality is not preserved by linear languages

Proposition 16. *There is a normal sequence N and a linear language L such that the subsequence selected from N by L is infinite and constant.*

Proof. For any word $w = w(0)\ldots w(n-1)$ of length n, let

$$w^{\mathrm{R}} = w(n-1)\ldots w(0)$$

be the mirror word of w and let

$$L = \{ww^{\mathrm{R}} : w \text{ is a word}\}$$

be the language of palindromes of even length. The language L is linear because it can be generated by a grammar with start symbol S and rules $\mathrm{S} \to 0\mathrm{S}0 \mid 1\mathrm{S}1 \mid \lambda$.

The sequence N is defined in stages $s = 0, 1, \ldots$ where during stage s we specify prefixes $\widetilde{\mathrm{z}}_s$ and z_s of N. At stage 0, let $\widetilde{\mathrm{z}}_0$ and z_0 both be equal to the empty word. At any stage $s > 0$, obtain $\widetilde{\mathrm{z}}_s$ by appending 2^{s-1} copies of v_s to z_{s-1} and obtain z_s by appending to $\widetilde{\mathrm{z}}_s$ its own mirror word $\widetilde{\mathrm{z}}_s^{\mathrm{R}}$, i.e.,

$$\widetilde{\mathrm{z}}_s = \mathrm{z}_{s-1}\mathrm{v}_s \ldots \mathrm{v}_s \quad (2^s \text{ copies of } \mathrm{v}_s), \quad \text{and} \qquad \mathrm{z}_s = \widetilde{\mathrm{z}}_s \widetilde{\mathrm{z}}_s^{\mathrm{R}}; \tag{14}$$

i.e., for example, we have $\widetilde{\mathrm{z}}_1 = \mathrm{v}_1$,

$$\begin{aligned}
\mathrm{z}_1 &= \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}, \\
\widetilde{\mathrm{z}}_2 &= \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}\ \mathrm{v}_2\mathrm{v}_2, \\
\mathrm{z}_2 &= \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}\ \mathrm{v}_2\mathrm{v}_2\ \ \mathrm{v}_2^{\mathrm{R}}\mathrm{v}_2^{\mathrm{R}}\ \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}, \\
\widetilde{\mathrm{z}}_3 &= \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}\ \mathrm{v}_2\mathrm{v}_2\ \ \mathrm{v}_2^{\mathrm{R}}\mathrm{v}_2^{\mathrm{R}}\ \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}\ \mathrm{v}_3\mathrm{v}_3\mathrm{v}_3\mathrm{v}_3, \\
\mathrm{z}_3 &= \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}\ \mathrm{v}_2\mathrm{v}_2\ \ \mathrm{v}_2^{\mathrm{R}}\mathrm{v}_2^{\mathrm{R}}\ \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}\ \mathrm{v}_3\mathrm{v}_3\mathrm{v}_3\mathrm{v}_3\ \ \mathrm{v}_3^{\mathrm{R}}\mathrm{v}_3^{\mathrm{R}}\mathrm{v}_3^{\mathrm{R}}\mathrm{v}_3^{\mathrm{R}}\ \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}\mathrm{v}_2\mathrm{v}_2\ \mathrm{v}_2^{\mathrm{R}}\mathrm{v}_2^{\mathrm{R}}\ \mathrm{v}_1\mathrm{v}_1^{\mathrm{R}}.
\end{aligned}$$

We show next that the set of prefixes of N that are in L coincides with the set $\{\mathrm{z}_s \colon s \geq 0\}$. From the latter, it is then immediate that L selects from N an infinite subsequence that consists only of 0's, since any prefix z_s of N is followed by the word v_{s+1}, where all these words start with 0.

By definition of the z_s, all words z_s are prefixes of N and are in L. In order to show that the z_s are the only prefixes of N contained in L, let

$$u_s = 01^s1^s0\ .$$

By induction on s, we show for all $s > 2$ that

(i) in $\widetilde{\mathrm{z}}_s$ occurs exactly one subword u_{s-1} and no subword u_s;
(ii) in z_s occur exactly two subwords u_{s-1} and one subword u_s;

Inspection shows that both assertion are true in case $s = 3$. In the induction step, consider some $s > 3$. Assertion (i) follows by $\widetilde{\mathrm{z}}_s = \mathrm{z}_{s-1}\mathrm{v}_s \ldots \mathrm{v}_s$, the induction hypothesis on z_{s-1}, and because by definition of v_s, the block of copies of v_s cannot overlap with a subword u_s. Assertion (ii) is then immediate by assertion (i), by $\mathrm{z}_s = \widetilde{\mathrm{z}}_s\widetilde{\mathrm{z}}_s^{\mathrm{R}}$, and because u_s is equal to ${u_s}^{\mathrm{R}}$ and 01^s is a suffix of $\widetilde{\mathrm{z}}_s$.

Now fix any prefix w of N and assume that w is in L, i.e., is a palindrome of even length. Let s be maximum such that z_s is a prefix of w. We can assume $s \geq 3$, because inspection reveals that w cannot be a prefix of z_3 unless w is equal to some z_i, where in the latter case we are done. By (ii), the words z_s and z_{s+1} contain u_s as a subword exactly once and twice, respectively, hence w contains u_s as a subword at least once and at most twice. When mirroring the palindrome w onto itself, the first occurrence of the palindrome u_s in w occurrence must either be mapped to itself or, if present at all, to the second occurrence of u_s in w, in which cases w must be equal to z_s and z_{s+1}, respectively. Since w was chosen as an arbitrary prefix of N in L, this shows that the z_s are the only prefixes of N in L.

It remains to show that N is normal. Let

$$W = \{\mathrm{v}_i \colon i \in \mathbb{N}\} \cup \{{\mathrm{v}_i}^{\mathrm{R}} \colon i \in \mathbb{N}\}$$

and write the sequence N in the form

$$N = w_1 w_2 \ldots \tag{15}$$

where the words w_i correspond in the natural way to the words in the set W that occur in the inductive definition of N (e.g., w_1, w_2, and w_3 are equal to v_1, $\mathrm{v}_1{}^{\mathrm{R}}$, and v_2). For the scope of this proof, we will call the subwords w_i of N in (15) the designated subwords of N.

We conclude the proof by showing that the assumptions of Lemma 6 are satisfied and that hence N is normal. By Proposition 5, the set of all words of the form v_i is normal in the limit, and the same holds, by literally the same proof, for the set of all words $\mathrm{v}_i{}^{\mathrm{R}}$; the union of these two sets, i.e., the set W, is then also normal in the limit because the class of sets that are normal in the limit is easily shown to be closed under union.

Next observe that among the designated subwords of any prefix z_s of N each of the $2s$ words $\mathrm{v}_1, \dots, \mathrm{v}_s$ and $\mathrm{v}_1{}^{\mathrm{R}}, \dots, \mathrm{v}_s{}^{\mathrm{R}}$ occurs exactly 2^{s-1} many times; in particular, z_s contains at least $s2^s$ designated subwords and has length of at least $2^{s-1}|\mathrm{v}_s|$. Now fix any $t > 0$ and let $z = w_1 \dots w_t$; let s be maximum such that z_s is a prefix of z. By the preceding discussion, we have for any w in W,

$$\frac{|\{i \leq t \colon w_i = w\}|}{t} < \frac{2^s}{s2^s} = \frac{1}{s}$$

and, furthermore,

$$\frac{|w_{t+1}|}{|w_1 \dots w_t|} < \frac{|\mathrm{v}_{s+1}|}{|\mathrm{z}_s|} \leq \frac{|\mathrm{v}_{s+1}|}{2^{s-1}|\mathrm{v}_s|} \leq \frac{1}{2^{s-3}}.$$

Since t was chosen arbitrarily and s goes to infinity when t does, this shows that assumptions (i) and (ii) of Lemma 6 are satisfied. □

Similar to the argument for deterministic one-counter languages in the previous section, we can conclude that stochasticity with respect to linear languages differs from regular stochasticity.

Corollary 17. *The sequences that are stochastic with respect to linear languages form a proper subclass of the class of sequences that are stochastic with respect to regular languages.*

6 Acknowledgements

We are grateful to Klaus Ambos-Spies, Frank Stephan, and Paul Vitányi for helpful discussions.

References

1. V. N. Agafonoff. Normal sequences and finite automata. *Soviet Mathematics Doklady*, 9:324–325, 1968.
2. K. Ambos-Spies and A. Kučera. Randomness in Computability Theory. In P. Cholak et al. (eds.), *Computability Theory: Current Trends and Open Problems*, Contemporary Mathematics 257:1–14, American Mathematical Society, 2000.
3. K. Ambos-Spies and E. Mayordomo. Resource-bounded balanced genericity, stochasticity and weak randomness. In *Complexity, Logic, and Recursion Theory*, A. Sorbi (ed.), Marcel Dekker Inc, 1997.
4. J.-M. Autebert, J. Berstel, and L. Boasson. Context-free languages and pushdown automata. In *Handbook of Formal Languages*, G. Rozenberg and A. Salomaa (eds.), Springer, 1997.
5. J.L. Balcázar, J. Díaz and J. Gabarró. *Structural Complexity*, Vol. I and II. Springer, 1995 and 1990.
6. A. Broglio and P. Liardet. Predictions with automata. *Symbolic dynamics and its applications, Proc. AMS Conf. in honor of R. L. Adler, New Haven/CT (USA) 1991*, Contemporary Mathematics 135:111–124, 1992.
7. D. G. Champernowne, The construction of decimals normal in the scale of ten. *Journal of the London Mathematical Society* 8:254–260, 1933.

8. A. Church. On the concept of a random number. *Bulletin of the AMS*, 46:130–135,1940.
Applications.
9. T. Kamae. Subsequences of normal sequences. *Israel Journal of Mathematics*, 16:121–149, 1973.
10. T. Kamae and B. Weiss. Normal numbers and selection rules. *Israel Journal of Mathematics*, 21(2-3):101–110, 1975.
11. M. van Lambalgen. *Random Sequences*, Doctoral dissertation, Universiteit van Amsterdam, Amsterdam, 1987.
12. M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*, second edition, Springer, 1997.
classification of
13. J. H. Lutz. The quantitative structure of exponential time. In Hemaspaandra, L. A. and A. L. Selman, editors, *Complexity theory retrospective* II, p. 225–260, Springer-Verlag, 1997.
14. R. von Mises. *Probability, Statistics and Truth.* Macmillan, 1957.
15. M. G. O'Connor. An unpredictability approach to finite-state randomness. *Journal of Computer and System Sciences*, 37(3):324–336, 1988.
16. P. Odifreddi. *Classical Recursion Theory. Vol. I.* North-Holland, Amsterdam, 1989.
17. L. P. Postnikova, On the connection between the concepts of collectives of Mises-Church and normal Bernoulli sequences of symbols. *Theory of Probability and its Applications* 6:211–213, 1961.
18. C. P. Schnorr and H. Stimm. Endliche Automaten und Zufallsfolgen (Finite automata and random sequences). *Acta Informatica* 1:345–359, 1972.
sequence
19. V. A. Uspensky, A. L. Semenov, and A. Kh. Shen'. Can an individual sequence of zeros and ones be random? *Russian Math. Surveys*, 45:121–189, 1990.
Wahrscheinlickkeitsrechnung.
20. B. Weiss. *Single Orbit Dynamics.* AMS, 2000.