Agafonov's Theorem for Probabilistic Selectors

Ulysse Léchine
Université Sorbonne Paris Nord, France

Thomas Seiller
CNRS, Paris, France

Jakob Grue Simonsen
University of Copenhagen, Denmark

Abstract

A normal sequence over $\{0,1\}$ is an infinite sequence for which every word of length k appears with frequency 2^{-k}. Agafonov's eponymous theorem states that selection by a finite state selector preserves normality, i.e. if α is a normal sequence and A is a finite state selector, then the subsequence $A(\alpha)$ is either finite or a normal sequence.

In this work, we address the following question: does this result hold when considering probabilistic selectors? We provide a partial positive answer, in the case where the probabilities involved are rational. More formally, we prove that given a normal sequence α and a rational probabilistic selector P, the selected subsequence $P(\alpha)$ will be a normal sequence with probability 1.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory; Theory of computation → Probabilistic computation; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases Normal sequences, probabilistic automata, Agafonov's theorem

Digital Object Identifier 10.4230/LIPIcs.MFCS.2024.67

Funding *Thomas Seiller*: Partially supported by the ANR-22-CE48-0003-01 project DySCo.

1 Introduction

Let $\alpha = x_1 x_2 \cdots$ be an infinite sequence over $\{0,1\}$; α is said to be *normal* if every finite string of length n over Σ occurs with limiting frequency 2^{-n} in α [5]. By standard reasoning, almost all infinite sequences are normal when the set $\{0,1\}^\omega$ of infinite sequences is equipped with the usual Borel measure. Concrete examples of normal sequences include Champernowne's binary sequence $0100011011000001\cdots$ [11], and many more examples exist [7].

A *finite-state selector* is a deterministic finite automaton (DFA) that selects the nth symbol from α if the length $n-1$ prefix of α is accepted by the DFA. *Agafonov's Theorem* [15, 1] is the celebrated result that a sequence α is normal iff any DFA that selects an infinite sequence from α, selects a normal sequence. While alternative proofs, generalizations [20] – and counter-examples to generalizations [13] – abound, all results in the literature consider deterministic or non-deterministic DFAs, but none consider probabilistic computation.

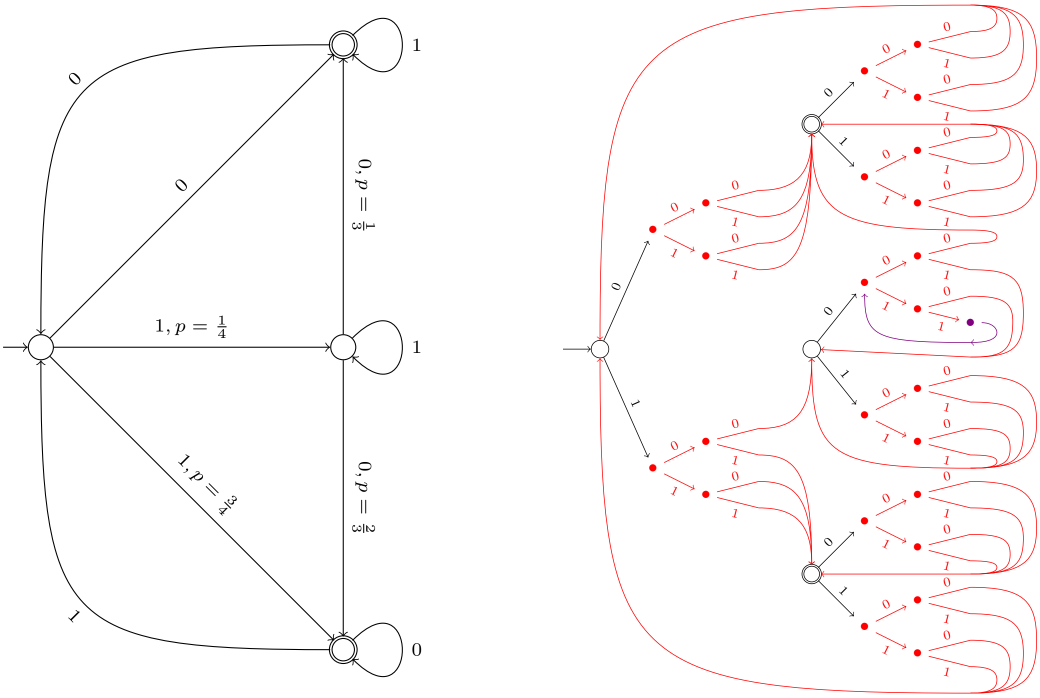
The extension to probabilistic selection is quite natural – not only are the underlying notions probabilistic in nature (i.e., normality of the transformed sequence), but the machinery of finite automata and similar computational devices itself has a 60-year history [18] of being extended to probabilistic devices.

In the present paper we study finite-state selectors equipped with probabilistic transitions from each state. As finite-state selectors can be viewed as devices sequentially processing successively larger prefixes of infinite sequences, we eschew the machinery of stochastic languages (where the initial state is a probability distribution on the states, and a string is accepted according to thresholding rules) – instead initial and accepting states are kept "as usual" in finite-state selectors. Probabilistic selection entails that normality may not be

49th International Symposium on Mathematical Foundations of Computer Science (MFCS 2024).
Editors: Rastislav Královič and Antonín Kučera; Article No. 67; pp. 67:1–67:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

preserved in all runs of an automaton: For example, consider an automaton with two states S_1 and S_2, only one of which is accepting, and transitions on 0 and 1 from S_i to S_i ($i \in \{1,2\}$) with probability $1/2$ and from S_i to $S_{i+1 \mod 2}$ with probability $1/2$; then for any normal sequence α, there is a run of the automaton on α that will select the sequence $0^\omega = 000\cdots$. The main result of the present paper is to show that the probability of having such runs is zero – in fact that for any probabilistic finite-state selector A with rational probabilities and any normal sequence α, *the probability that a run of* A on α *will select a normal sequence is 1.* The proof progresses by treating the relatively tame case of *dyadic* probabilities (i.e., of the form $a/2^k$ with a and k non-negative integers) first, and subsequently "simulating" finite-state selectors with arbitrary rational probabilities by "determinized" selectors with dyadic probabilities.

Figure 1 shows a probabilistic finite-state selector (Figure 1a) with two probabilistic transitions: one involves dyadic probabilities, the second one involves rational but non-dyadic probabilities. On the right-hand side (Figure 1b) is the determinization of this selector[1]. All unlabelled edges correspond to transitions valid for both 0 and 1. Determinizing the selector is done by introducing *gadgets* (shown in red) that simulate the probabilistic choices by drawing bits from a random advice sequence.



(a) A probabilistic selector.

(b) Determinization.

Figure 1 A probabilistic finite state selector and its determinization.
Unlabelled edges correspond to blind transitions, i.e. transitions valid for both 0 and 1.

[1] In fact, the determinisation as defined below would impose that all transition are represented as rationals of denominator 6 to ensure some regularity (this is discussed in Remark 20). For pedagogical purposes we however decided to show both a gadget for a dyadic transition and one for a rational but non-dyadic one.

Contributions. We prove that (the pertinent analogue of) Agafonov's Theorem holds in the setting of probabilistic selection, namely that a probabilistic finite state selector with rational probabilities preserves normality with probability 1. An added contribution is that the proof methods involved are novel, and may be of independent interest. As in Agafonov's original paper, and to keep complexity simple, all results are stated for binary alphabets. We fully expect all results to hold for arbitrary finite alphabets, *mutatis mutandis.*

Related work. Agafonov's Theorem has been generalized in multiple ways beyond finite automata (see, e.g.,[3, 2, 4, 10, 12, 20]). Conversely, it is known that when adding trifling computational expressivity to finite-state selectors, counterexamples to Agafonov's Theorem for the resulting selectors can be constructed [13]. While some existing work considers preservation of more general measures by finite automata, or similar selectors [9], and substantial work exists relating equidistribution to various types of automata [19, 4] no extant work considers stochastic *selection.* Agafonov's Theorem itself has been proved by a multitude of different techniques, e.g. [6, 8, 4]; it is conceivable that some of these can be adapted to alternative proofs, or extensions, of the results reported in the present paper.

2 Preliminaries and notation

Elements of $\{0,1\}^\omega$ are denoted by α, β, etc. Finite sequences (elements of $\{0,1\}^*$), or *words* are denoted by u, v, w, etc.

If $v, w \in \{0,1\}^*$, $v \cdot w$ denotes the concatenation of v and w; the definition extends to $v \cdot \alpha$ for $\alpha \in \{0,1\}^\omega$ *mutatis mutandis.*

The non-negative integers are denoted by $\mathbf{N}$, and the positive integers by $\mathbf{N}_{>0}$. If $N \in \mathbf{N}$ and $\alpha = a_1 a_2 \cdots \in \{0,1\}^\omega$, we denote by $\alpha|_{\leq N}$ the finite sequence $a_1 a_2 \cdots a_n$.

Given a set S we write $\mathrm{Dist}(S)$ the space of probability distributions on S. Given a probability distribution $\delta \in \mathrm{Dist}(S)$, we say that δ is *dyadic* (resp. *rational*) when for all $s \in S$, $\delta(s)$ is a dyadic number (resp. a rational number), that is a number of the form $\frac{p}{2^k}$ for integers p, k.

We consider the standard probability *measure* $\mathrm{Prob}_{\rho \in \{0,1\}^\omega}$ on $\{0,1\}^\omega$ equipped with the least Σ-algebra induced by the cylinder sets $C_w = \{\alpha \mid \exists \alpha' \in \{0,1\}^\omega, \alpha = w \cdot \alpha'\}$ and such that $\mathrm{Prob}_{\rho \in \{0,1\}^\omega} C_w = 2^{-|w|}$ for $w \in \{0,1\}^*$. Elements of $\{0,1\}^\omega$ drawn according to this measure are called *fair random infinite sequence.*

▶ **Definition 1.** *Let $a = a_1 \cdots a_m$ and $b = b_1 \cdots b_n$ be finite sequences such that $n < m$. An* occurrence *of b in a is an integer i with $1 \leq i \leq m$ such that $a_i = b_1, a_{i+1} = b_2, \ldots a_{i+n-1} = b_n$. If $\alpha = a_1 a_2 \cdots$ is an infinite sequence and $w = w_1 \cdots w_n$ is a word, we denote by $\sharp_N\{w\}(\alpha)$ the number of occurrences of w in $a_1 a_2 \cdots a_N$.*

A sequence $\alpha \in \{0,1\}^\omega$ is said to be normal *if, for any $m \in \mathbf{N}$ and any $w = w_1 \cdots w_m \in \{0,1\}^m$, the limit $\lim_{N \to \infty} \sharp_N\{w\}(\alpha)/N$ exists and equals 2^{-m}.*

▶ **Definition 2.** *Let $\alpha = a_1 a_2 \cdots$ be an infinite sequence, and i and n be positive integers. The ith* block of size n in α, *denoted $B_n^i(\alpha)$, is the finite sequence $a_{(i-1)n+1} a_{(i-1)n+2} \cdots a_{in}$. If w is a finite sequence of length k with $k = jn + r$ for appropriate j and $r < n$, the ith block of size n in a finite sequence of length $k \geq n$ is defined* mutatis mutandis *for any $i \leq j$.*

Given a word $w \in \{0,1\}^n$, we write $\sharp_N^{(n)}\{w\}(\alpha)$ for the number of blocks of size n that are equal to w in the prefix of size $N \times n$:

$$\sharp_N^{(n)}\{w\}(\alpha) = \mathrm{Card}\{i \in [0, N-1] \mid B_n^i(\alpha) = w\},$$

for $k \in \mathbf{N}$ and $w \in \{0,1\}^k$. We define $\mathrm{freq}(\alpha, w)$ *as the following limit, when it exists:*

$$\mathrm{freq}(\alpha, w) = \lim_{N\to\infty} \frac{\sharp_N^{(n)}\{w\}(\alpha)}{N}.$$

Let $k \in \mathbf{N}$, the sequence is length-k*-normal if for all words $w \in \Sigma^k$,* $\mathrm{freq}(\alpha, w)$ *is well defined and equal to 2^{-k}.*

The sequence α is said to be block normal *if it is* length-k*-normal for all k.*

By standard results, an infinite sequence α is normal iff it is block-normal [14, 16, 17].

We now define the crucial notion of *probabilistic finite state selector*. The usual notion of deterministic finite state selectors is a special case of this definition.

▶ **Definition 3.** *A probabilistic finite state selector $\mathcal{S}$ is a tuple (Q, t, ι, A) where Q is a finite set of states, $\iota \in Q$ is the initial state, $A \subset Q$ is the subset of accepting states, and $t : Q \times \{0,1\} \to \mathrm{Dist}(Q)$ is a probabilistic transition function.*

A rational (resp. dyadic, resp. deterministic) finite selector is a probabilistic finite state selector in which all distributions $t(q, a)$ (for $q \in Q$ and $a \in \{0,1\}$) are rational (resp. dyadic, resp. deterministic).

Given a probabilistic selector $\mathcal{S}$ and a sequence $\alpha \in \{0,1\}^\omega$, one can define a probability distribution over $\{0,1\}^\omega$ defined through *selection* of elements of α by $\mathcal{S}$.

Observe that if $\mathcal{S}$ is deterministic, the induced probability distribution assigns probability 1 to *the unique* selected subsequence $\mathcal{S}(\alpha)$ of α considered in the standard Agafonov theorem, i.e. the sequence of bits α_i in α such that $\mathcal{S}$ reaches an accepting state when given the prefix $\alpha_0 \dots \alpha_{i-1}$.

▶ **Definition 4.** *Given a probabilistic finite state selector $\mathcal{S}$ and an infinite sequence α, the* selection random variable *$S(\alpha)$ is the random variable over $\{0,1\}^\omega$ defined as follows on cylindrical sets C_w:*

$$S(\alpha)(C_w) = \sum_{i_1 < \dots < i_{|w|} \in \mathbf{N}, \alpha_{i_1}\alpha_{i_2}\dots\alpha_{i_{|w|}} = w} \mathrm{Prob}(\mathcal{S}, i_1 < \dots < i_{|w|})$$

where $\mathrm{Prob}(\mathcal{S}, i_1 < \dots < i_{|w|})$ denotes the probability that the first $|w|$ times the selector $\mathcal{S}$ reaches an accepting state on input α correspond to the indices $i_1 - 1, i_2 - 1, \dots, i_{|w|} - 1$.

We finally recall the standard Agafonov theorem.

▶ **Theorem 5.** *Let $\alpha \in \{0,1\}^\omega$ be a sequence, and $\mathcal{S}$ a deterministic finite selector. Then α is normal if and only if for all deterministic finite selector $\mathcal{S}$, the subsequence $\mathcal{S}(\alpha)$ is either finite or normal.*

2.1 Technical lemmas about normality

We will establish a few results on normal sequences that will be useful in later proofs. We first define notions that will be used in the proofs.

▶ **Definition 6.** *Let $\alpha \in \{0,1\}^\omega$ be a sequence, and $w \in \{0,1\}^*$ a word. We say that α is* w-normal *if $\lim_{N\to\infty} \frac{\sharp_N\{w\}(\alpha)}{N} = 2^{-|w|}$.*

Given $\epsilon \in \mathbf{R}$, we say that α is w-normal up to ϵ *if $\exists N_0, \forall N > N_0, \left|\frac{\sharp_N\{w\}(\alpha)}{N} - 2^{-|w|}\right| < \epsilon$.*

We now restate a weaker property for sequences than normality: being normal for words of a fixed length k. The main lemma associated to that notion will be that if a sequence is normal for words of length mk for a fixed integer k and all integers m, then it is normal (i.e. normal for words of arbitrary length).

▶ **Definition 7.** *Let $k \in \mathbf{N}$, the sequence is* length-k-normal *if for all words $w \in \{0,1\}^k$, $\mathrm{freq}(\alpha, w)$ is well defined and equal to 2^{-k}.*

▶ **Lemma 8.** *Let α be a sequence in $\{0,1\}^\omega$. The following are equivalent:*

- *α is normal*
- *there exists $m \in \mathbf{N}_{>0}$ such that α is* length-km *normal for all $k \in \mathbf{N}$.*

Proof. Consider a sequence α which is length-km normal for all $k \in \mathbf{N}$, and fix a word $w \in \{0,1\}^n$. We will use the length-mn normality of α. For this, we note that we can write:

$$\begin{aligned}
\sharp_N^{(n)}\{w\}(\alpha) &= \sum_{w_1,\dots,w_m \in \{0,1\}^n} \mathrm{Card}\{i \in \{1,\dots,m\} \mid w_i = w\}.\sharp_{N/m}^{(mn)}\{w_1 \cdot \dots \cdot w_m\}(\alpha) \\
&= \sum_{w_1,\dots,w_{m-1} \in \{0,1\}^n} \sum_{j=1}^{m} \sharp_{N/m}^{(mn)}\{w_1 \cdot \dots \cdot w_{j-1} \cdot w \cdot w_j \cdot w_{m-1}\}(\alpha) \\
&= \sum_{w_1,\dots,w_{m-1} \in \{0,1\}^n} m\sharp_{N/m}^{(mn)}\{w \cdot w_1 \cdot \dots \cdot w_{m-1}\}(\alpha)
\end{aligned}$$

As a consequence:

$$\begin{aligned}
\mathrm{freq}(\alpha, w) &= \lim_{N\to\infty} \frac{\sharp_N^{(n)}\{w\}(\alpha)}{N} \\
&= \lim_{N\to\infty} \sum_{w_1,\dots,w_m \in \{0,1\}^n} \frac{\mathrm{Card}\{i \in \{1,\dots,m\} \mid w_i = w\}.\sharp_{N/m}^{(mn)}\{w_1 \cdot \dots \cdot w_m\}(\alpha)}{N} \\
&= \lim_{N\to\infty} \sum_{w_1,\dots,w_{m-1} \in \{0,1\}^n} \frac{m.\sharp_{N/m}^{(mn)}\{w \cdot w_1 \cdot \dots \cdot w_{m-1}\}(\alpha)}{N} \\
&= \sum_{w_1,\dots,w_{m-1} \in \{0,1\}^n} \lim_{N\to\infty} \frac{\sharp_{N/m}^{(mn)}\{w \cdot w_1 \cdot \dots \cdot w_{m-1}\}(\alpha)}{N/m} \\
&= \sum_{w_1,\dots,w_{m-1} \in \{0,1\}^n} \mathrm{freq}(\alpha, w \cdot w_1 \cdot \dots \cdot w_{m-1}) \\
&= \sum_{w_1,\dots,w_{m-1} \in \{0,1\}^n} 2^{-mn} = 2^{n(m-1)}2^{-mn} = 2^{-n}
\end{aligned}$$ ◀

Now the following lemma states that the proportion of blocks equal to a fixed word w in a prefix of size N of a normal sequence asymptotically behaves as a linear function. The proof is quite straightforward.

▶ **Lemma 9.** *Let α be a normal sequence and $w \in \{0,1\}^n$. Then $\sharp_N^{(n)}\{w\}(\alpha) = 2^{-n}N + o(N)$.*

Proof. If it were not true, we would have that there exists some $\epsilon > 0$ and a sequence $(N_i)_{i\in\mathbf{N}}$ such that $\left|\sharp_{N_i}^{(n)}\{w\}(\alpha) - 2^{-n}N_i\right| > \epsilon N_i$ for all $i \in \mathbf{N}$. In other words,

$$\left|\frac{\sharp_{N_i}^{(n)}\{w\}(\alpha)}{N_i} - 2^{-n}\right| > \epsilon.$$

This contradicts the normality of α since it implies that $\lim_{i\to\infty}\left|\frac{\sharp_{N_i}^{(n)}\{w\}(\alpha)}{N_i} - 2^{-n}\right| = 0$. ◀

We will now define a partition of the set of blocks $(B_K^j(\alpha))_{j\in\mathbf{N}}$ into groups $(E_i)_{i\in\mathbf{N}}$ such that each $r \in \{0,1\}^K$ appears exactly once in each E_i. Those will be defined from a partition $(V_i^K(\alpha))_{i\in\mathbf{N}}$ of $\mathbf{N}$ such that the set $V_i^K(\alpha)$ contains the indices of the blocks of size K of α contained in E_i.

▶ **Definition 10.** *Let $\alpha \in \{0,1\}^\omega$ be a normal sequence, and $K \in \mathbf{N}$. We define $\theta_i(\alpha) : \{0,1\}^K \to \mathbf{N}$ as mapping a word w to the value j such that $B_K^j(\alpha)$ is exactly the i-th block of size K of α equal to w. (i.e. there are exactly $i-1$ indices $j_1 < j_2 < j_3 < \ldots < j_{i-1} < j$ such that $\forall k$, $B_K^{j_k}(\alpha) = w \wedge B_K^j(\alpha) = w$)*

The sets of indices $(V_i^K(\alpha))_{i\in\mathbf{N}} \subset \mathbf{N}$ are then defined as the image $\mathrm{Im}(\theta_i(\alpha))$.

The next lemma gives useful bounds on the V_i^K.

▶ **Lemma 11.** *Let α be a normal sequence, $K \in \mathbf{N}$. Consider the sets $V_i^K(\alpha)$ from Definition 10. We have that $\max_{i=1}^N \max V_i^K(\alpha) = N2^K + o(N)$, and $|[N]\setminus\bigcup_{i=1}^{N/2^K} V_i| = o(N)$.*

Proof. This comes from the fact that for any $w \in \{0,1\}^n$, $\sharp_N^{(n)}\{w\}(\alpha) = 2^{-|w|}N + o(N)$. ◀

The following probabilistic lemma is needed for the proof of Lemma 30.

▶ **Lemma 12.** *Let Ω_1 and Ω_2 be two sample spaces. Let $(X_i)_{i\in\mathbf{N}}$ be an iid family of r.v. which take value in $\Omega_1^{\mathbf{N}}$, $(Y_i)_{i\in\mathbf{N}}$ another iid family of r.v. which take value in $\Omega_2^{\mathbf{N}}$. Let $f(X,Y)$ be a function in $\Omega_1 \times \Omega_2 \mapsto \mathbf{R}$. Suppose $\forall x \in \Omega_1, \forall y \in \Omega_2, f(X_0, y)$ and $f(x, Y_0)$ have finite expected value and variance then*

$$\mathbb{P}_{Y_i}\left(\sum_x \left[\mathbb{P}(X = x)\sum_{i=1}^N f(x, Y_i)\right] = N \times \mathbb{E}_{X_0,Y_0}(f(X_0, Y_0)) + o(N)\right) = 1.$$

Proof. This is a direct application of the law of large numbers . ◀

3 Dyadic case

We first restrict to dyadic selectors. Given a dyadic selector $S = (Q, t, \iota, A)$, we define its dyadicity degree as the smallest integer D such that for all states $q, q' \in Q$ and element $a \in \{0,1\}$, the probability $t(q,a)(q')$ can be written as $\frac{m}{2^D}$.

We first define the determinisation of a dyadic selector.

▶ **Definition 13.** *Given a dyadic selector $\mathcal{S} = (Q, t, \iota, A)$ of dyadicity degree D, we define a* determinisation $\mathrm{Det}(\mathcal{S})$ *of S as the deterministic selector (Q', t', ι', A') where:*

- *$Q' = Q \cup Q \times \{0,1\} \times \{0,1\}^{\leq D-1}$;*
- *$\iota' = \iota$ and $A' = A$;*
- *the transition function t is defined as follows:*
  - *for all $q \in Q$, $t'(q,a) = (q, a, \epsilon)$ where ϵ is the empty word;*
  - *for all $((q,b,w)$ with $w \in \{0,1\}^{\leq D-2}$, $t'((q,b,w),a) = (q, b, w\cdot a)$;*
  - *for all (q,b,w) with $w \in \{0,1\}^{D-1}$, $t'((q,b,w),a) = q'$ where $q' = \phi(w\cdot a)$ for a chosen $\phi_{q,b} : 2^D \to Q$ such that the preimage of any $s \in Q$ has cardinality m_s where m_s is defined by $t(q,b)(s) = \frac{m_s}{2^D}$.*

Now, the principle is that the behaviour of a dyadic selector S on the sequence α can be simulated by the behaviour of a determinisation $\mathrm{Det}^D(\mathcal{S})$ computing on an interleaving of α and a random *advice* string ρ.

▶ **Definition 14.** *Let α, ρ be sequences in $\{0,1\}^\omega$, and $D \in \mathbf{N}$. The interwoven sequence $I^D(\alpha,\rho)$ is defined as the sequence:*

$$\alpha_0\rho_0 \dots \rho_{D-1}\alpha_1\rho_D \dots \rho_{2D-1} \dots .$$

Note that the interweaving of two normal sequences can be a non-normal sequence, e.g. the interweaving of α with itself $I^1(\alpha,\alpha)$ is not normal.

▶ **Lemma 15.** *Let $\mathcal{S}$ be a dyadic selector of dyadicity degree D. Then for all sequences $\alpha \in \{0,1\}^\omega$ the random variables $\mathcal{S}(\alpha)$ and $\mathrm{Det}(\mathcal{S})(I^D(\alpha,\rho))$ where ρ is drawn uniformly at random in $\{0,1\}^\omega$ have the same distribution.*

Proof. This is a special case of Lemma 23. ◀

Consider given a normal sequence α. We now prove that for almost all random advice sequence $\rho \in \{0,1\}^\omega$, the interwoven sequence $I^D(\alpha,\rho)$ is normal. This is the key lemma in the proof of Theorem 17.

▶ **Lemma 16.** *Let $\alpha \in \{0,1\}^\omega$ be a normal sequence. Then for all $D \in \mathbf{N}$,*

$$\mathrm{Prob}_{\rho\in\{0,1\}^\omega}[I^D(\alpha,\rho) \textit{is normal}] = 1.$$

Proof. In case $D = 0$, the interwoven sequence $I^D(\alpha,\rho)$ is equal to α. As a consequence, $\mathrm{Prob}_{\rho\in\{0,1\}^\omega}[I^D(\alpha,\rho)$ is normal$]$ is equal to 1.

We now suppose that $D \neq 0$. Given $m \in \mathbf{N}_{>0}$, we will show that $I^D(\alpha,\rho)$ is length-$(D+1)m$ normal with probability 1. This implies that for almost all $\rho \in \{0,1\}^\omega$, the sequence $I^D(\alpha,\rho)$ is length-$(D+1)m$ normal for every $m \in \mathbf{N}_{>0}$. By Lemma 8, this implies that for almost all $\rho \in \{0,1\}^\omega$, the sequence $I^D(\alpha,\rho)$ is normal.

We now fix $m \in \mathbf{N}_{>0}$, and $w \in \{0,1\}^{(D+1)m}$. We will consider the block decomposition of $I^D(\alpha,\rho)$ into blocks of size $(D+1)m$ and prove that:

$$\mathrm{Prob}_{\rho\in\{0,1\}^\omega}\left(\lim_{N\to\infty}\frac{\sharp_N^{((D+1)m)}\{w\}(I^D(\alpha,\rho))}{N} = 2^{-(D+1)m}\right) = 1.$$

We note that blocks of size $(D+1)m$ follow the pattern:

$$\alpha_i r_j \dots r_{j+D}\alpha_{i+1}r_{j+D}\dots r_{j+2D}\dots\alpha_{i+m-1}r_{j+(m-1)D}\dots r_{j+mD}.$$

We will consider $\lfloor w \rfloor_D = w_0 w_{D+1} w_{2(D+1)} \dots w_{(m-1)(D+1)}$ the subword of w corresponding to the positions of bits from α in this pattern.

We will consider the block decomposition of α into blocks of size m. Let $\mathrm{idx}(i) = j$ where j is the i-th block such that $B_j^m(\alpha) = \lfloor w \rfloor_D$. Note that this function is well defined because α is a normal sequence. Note that if a given block $B_i^{(D+1)m}(I^D(\alpha,\rho))$ is equal to w, then $\lfloor B_i^{(D+1)m}(I^D(\alpha,\rho))\rfloor_D$ should be equal to $\lfloor w \rfloor_D$. We write $\tilde{N} = \sharp_N^{(m)}\{\lfloor w \rfloor_D\}(\alpha)$, note that it is the maximal i such that $\mathrm{idx}(i) < N$. We also define $\bar{w}$ as the complementary subsequence of w:

$$\bar{w} = w_1 \dots w_D w_{D+2} \dots w_{2(D+1)-1} w_{2(D+1)+1} \dots w_{m(D+1)-1}.$$

We introduce a new notation: we will write $\sharp_{\mathrm{Im(idx)}<N}^{((D+1)m)}\{w\}(I^D(\alpha,\rho))$ to denote the number of blocks of size $(D+1)m$ equal to w within the blocks indexed by some $j < N$ in $\mathrm{Im(idx)}$.

$$
\begin{aligned}
P &= \mathrm{Prob}_{\rho\in\{0,1\}^\omega}\left(\lim_{N\to\infty}\frac{\sharp_N^{((D+1)m)}\{w\}(I^D(\alpha,\rho))}{N} = 2^{-(D+1)m}\right)\\
&= \mathrm{Prob}_{\rho\in\{0,1\}^\omega}\left(\lim_{N\to\infty}\frac{\sharp_{\mathrm{Im(idx)}<N}^{((D+1)m)}\{w\}(I^D(\alpha,\rho))}{N} = 2^{-(D+1)m}\right)
\end{aligned}
$$

Now by Lemma 9 we have that $\lim_{N\to\infty}\tilde{N} = N.2^{-m}$. Hence:

$$
\begin{aligned}
P &= \mathrm{Prob}_{\rho\in\{0,1\}^\omega}\left(\lim_{\tilde{N}\to\infty}\frac{\sharp_{\mathrm{Im(idx)}<N}^{((D+1)m)}\{w\}(I^D(\alpha,\rho))}{\tilde{N}.2^{-m}} = 2^{-(D+1)m}\right)\\
&= \mathrm{Prob}_{\rho\in\{0,1\}^\omega}\left(\lim_{\tilde{N}\to\infty}\frac{\sharp_{\mathrm{Im(idx)}<N}^{((D+1)m)}\{w\}(I^D(\alpha,\rho))}{\tilde{N}} = 2^{-Dm}\right)\\
&= \mathrm{Prob}_{\rho\in\{0,1\}^\omega}\left(\lim_{\tilde{N}\to\infty}\frac{\sharp_{\mathrm{Im(idx)}<N}^{(Dm)}\{\bar{w}\}(\rho)}{\tilde{N}} = 2^{-Dm}\right)
\end{aligned}
$$

By the law of large numbers, we have that

$$
\mathrm{Prob}_{\rho\in\{0,1\}^\omega}\left(\lim_{\tilde{N}\to\infty}\frac{\sharp_{\mathrm{Im(idx)}<N}^{(Dm)}\{\bar{w}\}(\rho)}{\tilde{N}} = 2^{-Dm}\right) = 1,
$$

which concludes the proof. ◀

This lemma then leads to the following theorem.

▶ **Theorem 17.** *Let $\alpha \in \{0,1\}^\omega$ be a sequence. Then α is normal if and only if for all dyadic finite selector $\mathcal{S}$ the probability that $\mathcal{S}(\alpha)$ is either finite or normal is equal to 1.*

Proof. The right to left implication is simply a consequence of Agafonov's theorem (Theorem 5) since if for all dyadic finite selector $\mathcal{S}$ the probability that $\mathcal{S}(\alpha)$ is either finite or normal is equal to 1, then for all deterministic finite selector $\mathcal{S}$ the selected subsequence $\mathcal{S}(\alpha)$ is either finite or normal.

Now, suppose that the above implication from left to right is false. Then by Lemma 15 there exists a subset $R \subset \{0,1\}^\omega$ of strictly positive measure such that $\mathrm{Det}(\mathcal{S})(I^D(\alpha,\rho))$ is infinite and not normal for all $\rho \in R$. Since almost for almost all $\rho \in \{0,1\}^\omega$ the interwoven sequence $I^D(\alpha,\rho)$ is normal, this implies that there exists a ρ such that $I^D(\alpha,\rho)$ is normal and $\mathrm{Det}(\mathcal{S})(I^D(\alpha,\rho))$ is infinite and not normal. But this contradict Agafonov's theorem (Theorem 5). ◀

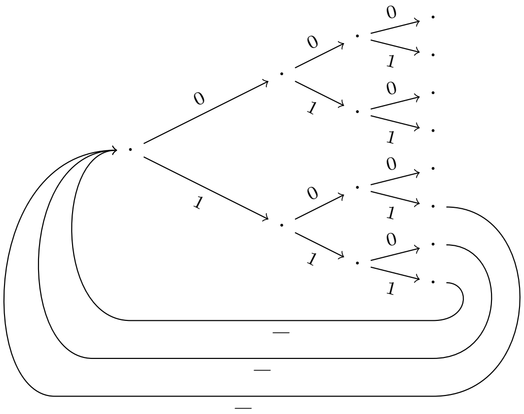
We will now consider the case of rational selectors. The difficulty in adapting the proof lies in the fact that the interwoven sequence has a less regular structure. In the above proof, each block of size $(D+1)m$ followed the same pattern. But in the case of rational selectors, the presence of feedback loops renders those pattern random, this makes the proof significantly harder. Indeed in the dyadic case the value of a block of size $(D+1)m$ was independent of the value of other blocks of size $(D+1)m$, in the rational case this is no longer true, thus we cannot apply the law of large numbers. Informally to make our proof work we divide $\mathcal{S}(\alpha)$ into non adjacent blocks whose values are independent, some bits are not contained in any blocks but we argue they are few of them and thus they don't prevent $\mathcal{S}(\alpha)$ from being normal.

4 Rational selector

4.1 Determinisation

The first step in extending the results to the rational case is to define the determinization. This will follow the same principle as for the dyadic case, but the parts of the determinized automaton that simulates probabilistic choices will contain feedback loops. Nonetheless, incorporating feedback loops is enough to represent any rational distribution, as shown in the next lemma.

▶ **Definition 18.** *A (k, f)-gadget is a regular binary tree of depth k, extended with blind transitions from the last F leaves to the root.*



▶ **Lemma 19.** *Any rational distribution* $\mathrm{Dist}(S)$ *is simulated by a gadget.*

Proof. Let $p_1, \dots, p_k$ be rationals with $\sum_i p_i = 1$, and suppose $p_i \leq p_{i+1}$ for all i. Consider M the smallest common multiple of all denominators of the elements p_i, and write $p_i = \frac{\tilde{p}_i}{M}$. We will denote by $q_i = \sum_{j=1}^{i} \tilde{p}_i$. Note that $q_0 = 0$ and $q_k = \sum_i \tilde{p}_i = M$. Now consider P the smallest natural number such that $2^P \geq M$. We build the regular automaton of depth P with feedback loops on $2^P - M$ leaves. We will show that the probability p of reaching a leaf within $[q_i + 1, q_{i+1}]$ is equal to p_i. One only need to compute:

$$p = \frac{\tilde{p}_i}{2^P} \sum_{m \geq 0} \left(\frac{2^P - M}{2^P} \right)^m = \frac{\tilde{p}_i}{2^P} \frac{1}{1 - \frac{2^P - M}{2^P}} = \frac{\tilde{p}_i}{2^P} \frac{2^P}{2^P - (2^P - M)} = \frac{\tilde{p}_i}{2^P} \frac{2^P}{M} = \frac{\tilde{p}_i}{M} = p_i \quad ◀$$

We will now define the determinisation of a rational selector in a similar way as for the dyadic case. First, since the selector is finite, one can write all rational numbers involved with a common denominator, say k. Given a rational selector $\mathcal{S}$, we will call k the *rationality degree* of $\mathcal{S}$. Then each transition will be simulated by a gadget as defined above.

▶ Remark 20. Note that since all rational distribution are represented with the same denominator, then all gadgets will have the same size. Indeed, let us define the *dyadicity degree* D of a rational selector $\mathcal{S}$ as the smallest integer such that $2^D \geq k$, where k is its rationality degree. Then the feedback edges of all gadgets corresponding to rational transitions correspond to the $2^D - k$ last edges in the gadget, and this does not depend on the specific transition considered.

The determinisation therefore has a quite *regular* structure which will be mirrored in the corresponding interwoven sequences.

▶ **Definition 21.** *Let $\mathcal{S} = (Q, \iota, t, A)$ be a rational selector of rationality degree k and dyadicity degree D. We define its* determinisation $\mathrm{Det}(\mathcal{S})$ *as the deterministic selector (Q', ι, t, A) where:*

- *$Q' = Q \cup Q \times \{0,1\} \times \left(\{0,1\}^{\leq k-1} \cup \{\mathrm{return}\}\right)$;*
- *$\iota' = \iota$ and $A' = A$;*
- *the transition function t is defined as follows:*
  - *for all $q \in Q$, $t'(q, a) = (q, a, \epsilon)$ where ϵ is the empty list;*
  - *for all $((q, b, w)$ with $w \in \{0,1\}^{\leq k-2}$, $t'((q, b, w), a) = (q, b, w \cdot a)$;*
  - *for all (q, b, w) with $w \in \{0,1\}^{k-1}$ and $a \in \{0,1\}$:*
    - *if $w \cdot a$ belongs to the $2^D - k$ last leaves (i.e. the $2^D - k$ largest elements of $\{0,1\}^D$ for the natural order), then $t'((q, b, w), a) = (q, b, \mathrm{return})$;*
    - *otherwise, $t'((q, b, w), a) = q'$ where $q' = \phi(w \cdot a)$ for a chosen $\phi_{q,b} : k \to Q$ such that the preimage of any $s \in Q$ has cardinality m_s where m_s is defined by $t(q, b)(s) = \frac{m_s}{k}$;*
  - *for all (q, b, return) and any $a \in \{0,1\}$, $t'(q, b, \mathrm{return}) = (q, b, \epsilon)$.*

▶ **Definition 22.** *Let $\mathcal{S}$ be a rational selector of rationality degree k and dyadicity degree D, $\alpha \in \{0,1\}^\omega$ an input sequence, and $\rho \in \{0,1\}^\omega$ an* advice *sequence. We define the interwoven sequence $I_k^D(\alpha, \rho)$ as:*

$$\alpha_1 \rho_1 \dots \rho_{i_1} \alpha_2 \rho_{i_1+1} \dots \rho_{i_2} \alpha_3 \dots,$$

where $i_1 < i_2 < \dots$ is the sequence of indices i_j such that $\rho_{i_j+1} \dots \rho_{i_{j+1}}$ is equal to $w_1 r_1 w_2 r_2 \dots r_m w_{m+1}$ where:

- *$w_1, w_2, \dots, w_m$ are among the $2^D - k$ greatest elements in $\{0,1\}^D$ (considered with the natural alphabetical order);*
- *w_{m+1} belongs to the k smallest elements in $\{0,1\}^D$;*
- *r_i are bits in $\{0,1\}$ which we will call* return bits, *corresponding to feedback loops.*

We note that this is a direct generalisation of the dyadic case, i.e. if the considered selector is dyadic, then the interwoven sequence $I_{2^D}^D$ just defined coincides with the definition from the previous section. Similarly, the determinisation of a dyadic selector is a special case of the determinisation of a rational selector. We can see here the difficulty in adapting the proof to the rational case arising: instead of interweaving one block of ρ of size D between each bit of α, we interweave a block of bits from ρ of variable length.

Note however that we carefully defined the determinisation so that the size of these blocks does not depend on the values α_i. Moreover, feedback loops introduce random *return* bits, allowing us to write the interwoven sequence as a sequence of blocks of the form $ar_1 \dots r_D$ where a is either a bit from α or a return bit from ρ and $r_1 \dots r_D$ are bits from ρ.

First, we check that the determinisation simulates the rational selector when given random advice strings.

▶ **Lemma 23.** *Let $\mathcal{S}$ be a rational selector of rationality degree k and dyadicity degree D. Then for all sequence $\alpha \in \{0,1\}^\omega$ the random variables $\mathcal{S}(\alpha)$ and $\mathrm{Det}(\mathcal{S})(I_k^D(\alpha, \rho))$ where ρ is an infinite fair random sequence, have the same distribution.*

Proof. Let ρ be an infinite fair random sequence. By construction of $\mathrm{Det}(\mathcal{S})(I_k^D(\alpha, \rho))$, for a any two state q and q' the probability of going from q to q' in $\mathrm{Det}(\mathcal{S})(I_k^D(\alpha, \rho))$ (ignoring the gadget states in between) is equal to the probability of going from q to q' in $\mathcal{S}(\alpha)$. ◀

4.2 Rational selectors preserve normality

In the following, α will be a infinite sequence, not considered normal unless explicitly stated. We will write w to denote a finite word. We denote by ρ and τ fair infinite random sequences, and by r a finite random sequence. Lastly, q will be the probability to loop back at the end of a gadget, equal to $1 - \frac{k}{2^D}$. We will denote by $A(N) \xrightarrow{N} B(N)(1 \pm \epsilon)$ the fact that

$$\exists N_0, \forall N > N_0, B(N)(1-\epsilon) < A(N) < B(N)(1+\epsilon).$$

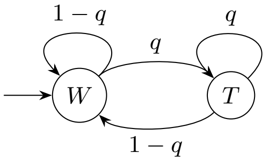
To prove that rational selectors preserve normality, we will prove in this section that $\mathbb{P}_\rho(I_k^D(\alpha,\rho)\text{is normal}) = 1$, that is the generalised version of Lemma 16. As in the dyadic case, this is the crux of the problem, and the proof of the main theorem will easily follow. In order to prove this technical lemma, we analyze a process we call $\mathcal{F}$ which takes a sequence α and inserts in between every bit of α a random amount of random bits. We will then show that if α is normal the sequence $\mathcal{F}(\alpha)$ obtained in this way is normal. Finally we will argue that normality of $I_k^D(\alpha,\rho)$ amounts to the normality of $\mathcal{F}(\alpha)$.

▶ **Definition 24** (Random process $\mathcal{F}_q$)**.** *Suppose given $K \in \mathbf{N}$, $w \in \{0,1\}^K$, $q \in [0;1[$, and $\tau \in \{0,1\}^\omega$. We define $\mathcal{F}_q(w,\tau) \in \{0,1\}^*$ as the random variable described in Figure 2 where we consume a bit of w when we get to state W and a bit of τ when we get to state T. The process stops when the state W is reached and there are no more bits of w to be consumed. The output is all the consumed bit in timely order.*

We denote by $\mathcal{F}_q(w)$ the random variable $\mathcal{F}_q(w,\tau)$ where τ is a fair random infinite sequence.

In the following, we may not specify q and just write $\mathcal{F}(w,\tau)$ when the context is clear.

▶ Remark 25. Note that τ needs to be infinite because we have no bound on how many bits of it we may consume.



Example: if $w = 0110$, $\tau = 10010...$, then

$$\mathcal{F}(w) = 010110010,$$

with the sequence of states

$$WTTWWTWTTW.$$

Figure 2 The random process $\mathcal{F}$.

For now $\mathcal{F}$ has only been defined on finite strings. We extend it to infinite strings in an intuitive way.

▶ **Definition 26.** *Suppose given $\alpha \in \{0,1\}^\omega$, $K \in \mathbf{N}$, and $q \in \mathbf{R}$. Let $(\mathcal{F}_i)_{i\in\mathbf{N}}$ be an iid family of random variables of law $\mathcal{F}$. The random variable $\mathcal{F}_q(\alpha)$ is the infinite sequence distributed as the concatenation of the $\mathcal{F}_i$ applied to the blocks $B_K^i(\alpha)$:*

$$\mathcal{F}_0(B_K^0(\alpha))\mathcal{F}_1(B_K^1(\alpha))\mathcal{F}_2(B_K^2(\alpha))\ldots.$$

Note that the value of K does not change the distribution of the random variable $\mathcal{F}(\alpha)$, hence the definition is unambiguous.

In the next lemma we analyze the length of $\mathcal{F}(w)$.

▶ **Lemma 27.** *Let $q \in [0;1[$ and $(\mathcal{F}_i)$ be an iid family of random variables of law $\mathcal{F}_q$. Then for all $K \in \mathbf{N}$ and for any family $(w_i)_{i\in\mathbf{N}} \in (\{0,1\}^K)^{\mathbf{N}}$,*

$$\mathbb{P}\left(\sum_{i\leq N} |\mathcal{F}_i(w_i)| = NKq^{-1} + o(NKq)\right) = 1.$$

Proof. By standard Markov chain analysis, the expected value of $|\mathcal{F}_i(w_i)|$ is Kq^{-1} and its variance is finite, furthermore the $\mathcal{F}_i(w_i)$ are independent the strong law of large number therefore applies and we get the desired result. ◀

In the next lemma we show that for any w we can approximate the number of w in $\mathcal{F}(\alpha) = \mathcal{F}_0(B_K^0(\alpha))\mathcal{F}_1(B_K^1(\alpha))\mathcal{F}_2(B_K^2(\alpha))\dots$ by adding up the number of w in each $\mathcal{F}_i(B_K^i(\alpha))$ separately. The larger K the more precise the approximation. What we gain from this separation is that the random variable $\sharp_{s_i-|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))$ are independent and we can apply the law of large numbers. In contrast in $\mathcal{F}(\alpha)$ where we concatenate the $\mathcal{F}_i(B_K^i(\alpha))$ we do not have independence because knowing that w appears at the end of $\mathcal{F}_1(B_K^1(\alpha))$ may influence that it appears at the beginning of $\mathcal{F}_2(B_K^2(\alpha))$.

▶ **Lemma 28.** *Let $\alpha \in \{0,1\}^\omega$ be a normal sequence, $w \in \{0,1\}^M$ be a word, and $(\mathcal{F}_i)_{i\in\mathbf{N}}$ be an iid family of random variables of law $\mathcal{F}$. For all $K \in \mathbf{N}$, we write $\beta = \mathcal{F}_0(B_K^0(\alpha))\mathcal{F}_1(B_K^1(\alpha))\mathcal{F}_2(B_K^2(\alpha))\dots$ and for all i we define $s_i = |\mathcal{F}_i(B_K^i(\alpha))|$ and $S_N = \sum_{i=0}^N s_i$. Then we have that*

$$\left[\sum_{i=0}^{N} \sharp_{s_i-|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))\right] - \sharp_{S_N}\{w\}(\beta) < MN.$$

Proof. First note that we count indices up to $s_i - |w|$ in $\sharp_{s_i-|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))$ because if w appears in $\mathcal{F}_i(B_K^i(\alpha))$ it must appear before the last $|w|$ bits. For this reason we also mention that $\sharp_{s_i-|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha))) = \sharp_{|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))$.

Then note that $\sharp_{S_N}\{w\}(\beta) \geq \left[\sum_{i=0}^N \sharp_{s_i-|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))\right]$ indeed if w appears somewhere in one of the $\mathcal{F}_i(B_K^i(\alpha))$ then it also appears in β.

Therefore every w is counted in $\sharp_{S_N}\{w\}(\beta)$ and not in $\left[\sum_{i=0}^N \sharp_{s_i-|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))\right]$ appears at an index in the $|w|$ last bits of an $F_i(B_K^i(\alpha))$. There are at most $|w| \times N = MN$ of those. ◀

We have that $\sum_{i=0}^N |\mathcal{F}_i(B_K^i(\alpha))|$ tends to NKq^{-1}, thus by taking large values of K the discrepancy MN of the number of w noticed in the previous theorem can be made negligible when compared to the size of the string.

In the next theorem we just prove that for a random ρ the proportion of w in $\mathcal{F}_i(B_K^i(\rho))$ is approximately $2^{-|w|}|\mathcal{F}_i(B_K^i(\rho))|$ on average.

▶ **Lemma 29.** *Suppose given $\rho \in \{0,1\}^\omega$ a fair random infinite sequence, $q \in [0;1[$, and $w \in \{0,1\}^M$. Let $(\mathcal{F}_i)_{i\in\mathbf{N}}$ be an iid family of random variables of law $\mathcal{F}_q$. For all i, we write $s_i = |\mathcal{F}_i(B_K^i(\rho))|$ and for all N, $S_N = \sum_{i=0}^N s_i$. Then for any $\epsilon > 0$, there exists $K \in \mathbf{N}$ such that:*

$$\left|\mathbb{E}(\sharp_{s_i-|w|}\{K\}(\mathcal{F}_i(B_w^i(\rho)))) - 2^{-|w|}Kq^{-1}\right| < \epsilon.$$

Proof. This result can be shown by standard analysis of fair random sequences of size Kq^{-1}. Indeed for a random ρ, $\mathcal{F}_i(B_K^i(\rho))$ is just a random sequence of expected size Kq^{-1}. Let $\epsilon \in \mathbf{R}$. If K is large enough, there exists some $\epsilon' \in \mathbf{R}$ such that a random sequence of size Kq^{-1} contains on average $2^{-|w|}Kq^{-1} + \epsilon'$ occurrences of w where $|\epsilon'| < \epsilon$. ◀

▶ **Lemma 30.** *Let α be a normal sequence, for any $w \in \{0,1\}^*$ and $q \in [0,1[$, $\mathcal{F}_q(\alpha)$ is w-normal with probability 1.*

Proof of Lemma 30. Let $(\mathcal{F}_i)_{i\in\mathbf{N}}$ be random independent processes $\mathcal{F}$. Let α be a normal sequence. Let $w \in \{0,1\}^*$, $\epsilon' \in \mathbf{R}$, and $q \in [0;1[$. Then

$$\mathbb{P}\left(\mathcal{F}_q(\alpha) \text{is } w\text{-normal up to } \epsilon'\right) = 1 \Leftrightarrow \mathbb{P}\left(\lim_N \frac{\sharp_N\{w\}(\mathcal{F}_q(\alpha))}{N} = 2^{-|w|} \pm \epsilon\right) = 1.$$

Using Lemma 28 by introducing independent random variables $\mathcal{F}_i$ of law $\mathcal{F}_q$, and writing $s_i = |\mathcal{F}_i(B^i_K(\alpha))|$, the above result is implied by:

$$\forall \epsilon, \exists K, \mathbb{P}\left(\sum_{i\leq N} \sharp_{s_i-|w|}\{w\}(\mathcal{F}_i(B^i_K(\alpha))) \xrightarrow{N} 2^{-|w|}KNq^{-1}(1\pm\epsilon)\right) = 1.$$

Take $V_i(\alpha)$ as defined in definition 10. Call $B_N = \{i \in [1; \frac{N}{2^K}] \mid \max(V_i(\alpha)) < N\}$. We group the indices of blocks $B^j_K(\alpha)$ into sets V_i of size 2^K and such that $|V_i| = 2^K$. We may also change $s_j - |w|$ to s_j as explained in the proof of Lemma 28. Then the above is equivalent to

$$\forall \epsilon, \exists K, \mathbb{P}\left(S_1 + S_2 \xrightarrow{N} 2^{-|w|}KNq^{-1}(1\pm\epsilon)\right) = 1,$$

where

$$S_1 = \sum_{i\in B_N} \sum_{j\in V_i(\alpha)} \sharp_{s_j}\{w\}(\mathcal{F}_j(B^j_K(\alpha))), \qquad S_2 = \sum_{j\in[N]\setminus\bigcup_{i\in B_N} V_i} \sharp_{s_j}\{w\}(\mathcal{F}_j(B^j_K(\alpha))).$$

By Lemma 11, we have that $|B_N| = \frac{N}{2^K} + g(N)$, where $g(N) = o(N)$. The equation can thus be further rewritten as:

$$\mathbb{P}\left(T_1 + T_2 + T_3 \xrightarrow{N} 2^{-|w|}KNq(1\pm\epsilon)\right) = 1,$$

where:

$$T_1 = \sum_{i\in[N/2^K]} \sum_{j\in V_i(\alpha)} \sharp_{s_j}\{w\}(\mathcal{F}_j(B^j_K(\alpha))),$$

$$T_2 = \sum_{i=1+N/2^K}^{\frac{N}{2^K}+g(N)} \sum_{j\in V_i(\alpha)} \sharp_{s_j}\{w\}(\mathcal{F}_j(B^j_K(\alpha))),$$

$$T_3 = \sum_{j\in[N]\setminus\bigcup_{i\in B_N} V_i} \sharp_{s_j}\{w\}(\mathcal{F}_j(B^j_K(\alpha))).$$

We now consider each term separately.

The term T_1. By construction of the V_i, as j ranges across all values in V_i, $B^j_K(\alpha)$ takes all values in $\{0,1\}^K$. By creating an appropriate bijection between j and (i,r), we can write

$$\sum_{i\in[\frac{N}{2^K}]} \sum_{j\in V_i(\alpha)} \sharp_{s_j}\{w\}(\mathcal{F}_j(B^j_K(\alpha))) = \sum_{i\in[\frac{N}{2^K}]} \sum_{r\in\{0,1\}^K} \sharp_{s_{i,r}}\{w\}(\mathcal{F}_{i,r}(r)).$$

We recognize a sum over expectations as in Lemma 12. By Lemma 29, we can take K big enough such that the expected value of $\sharp_{s_{i,r}}\{w\}(\mathcal{F}_{i,r}(r))$ is $Kq^{-1}2^{-|w|}(1\pm\epsilon)$. Thus:

$$\forall \epsilon, \exists K, \mathbb{P}\left(\sum_{i\in[\frac{N}{2^K}]} \sum_{j\in V_i(\alpha)} \sharp_{s_j}\{w\}(\mathcal{F}_j(B^j_K(\alpha))) \xrightarrow{N} 2^{-|w|}KNq^{-1}(1\pm\epsilon)\right) = 1.$$

The term T_2. We have that

$$\forall\epsilon, \forall K, \mathbb{P}\left(\left[\sum_{i=1+N/2^K}^{\frac{N}{2^K}+g(N)} \sum_{j\in V_i(\alpha)} \sharp_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha)))\right] = o(N)\right) = 1,$$

because $g(N) = o(N)$ and the random variable $\sharp_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha)))$ has finite expected value and variance (in particular constant in N).

The term T_3. We have that

$$\mathbb{P}\left(\sum_{j\in[N]\setminus\bigcup_{i\in B_N} V_i} \sharp_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha))) = o(n)\right) = 1.$$

The sum is over $o(N)$ terms by Lemma 11 and the random variable $\sharp_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha)))$ is of finite expected value and variance.
By combining the three results we can get that

$$\forall\epsilon' \in \mathbf{R}, \mathbb{P}(\mathcal{F}(\alpha)\text{is } w\text{-normal up to } \epsilon') = 1.$$

We define the sequence $(\epsilon_n)_{n\in\mathbf{N}} \in \mathbf{R}^{\mathbf{N}}$ as $\epsilon_n = 1/(n+1)$. We have that

$$\mathbb{P}(\forall n, \mathcal{F}(\alpha)\text{is } w\text{-normal up to } \epsilon_n) = 1$$

as an intersection of countably many events of probability 1. We then get, using the fact that all Cauchy sequences converge on $\mathbf{R}$, that $\mathbb{P}(\mathcal{F}(\alpha)\text{is } w\text{-normal}) = 1$. ◀

▶ **Lemma 31.** *Let α be a normal number then $\mathcal{F}(\alpha)$ is normal with probability 1.*

Proof. By lemma 30 $\forall w \in \{0,1\}^*$, $\mathcal{F}(\alpha)$ is w-normal with probability 1. $\mathbb{P}(\mathcal{F}(\alpha)\text{is normal}) = \mathbb{P}(\forall w \in \{0,1\}^*, \mathcal{F}(\alpha)\text{is } w\text{-normal})$, since this is an intersection of countably many event of probability 1, we have that $\mathcal{F}(\alpha)$ is normal with probability 1. ◀

▶ **Lemma 32.** *For any positive integer D, any $k \in [2^{D-1}; 2^D]$*

$$\mathbb{P}_\rho(I_k^D(\alpha,\rho)\textit{is normal}) = 1.$$

Proof. Let D be a positive integer, $k \in [2^{D-1}; 2^D]$. There are 3 kinds of bits in $I_k^D(\alpha,\rho)$: bits from α, bits from ρ appearing inside the gadgets (we call this sequence γ) and bits from ρ corresponding to return bits (we call this infinite sequence of bits τ). Note that γ and τ are both independent fair random infinite sequences.

Note that in $I_k^D(\alpha,\rho)$, we find every bit from α and τ at indices multiple of $D+1$. We define the infinite sequence y as such: $\forall i \in \mathbf{N}, y_i = I_k^D(\alpha,\rho)_{i(D+1)}$.

Notice that $I_k^D(\alpha,\rho) = I^D(y,\gamma)$ (where the second I is from defintion 14). Since γ is a fair infinite random sequence then by using theorem 16 if y is normal then so is $I_k^D(\alpha,\rho)$ with probability 1 over γ.

Thus now we only need to show that y is normal with probability 1. Notice that the distribution of y is the same as $\mathcal{F}_q(\alpha)$ with $q = 1 - \frac{k}{2^D}$. Therefore by theorem 31 y is normal with probability 1. ◀

This gives the main theorem. The proof follows the proof of Theorem 17, using Lemma 32 and and Lemma 23.

▶ **Theorem 33.** *Let $\alpha \in \{0,1\}^\omega$ be a sequence. Then α is normal if and only if for all rational selector $\mathcal{S}$ the probability that $\mathcal{S}(\alpha)$ is either finite or normal is equal to* 1*.*

While we think the equivalent statement to hold for general probabilistic selectors, we believe that establishing such a result would require a different proof method.

References

1 V. N. Agafonov. Normal sequences and finite automata. *Sov. Math., Dokl.*, 9:324–325, 1968. Originally published in Russian [15].

2 Dylan Airey and Bill Mance. Normality preserving operations for Cantor series expansions and associated fractals, i. *Illinois J. Math.*, 59(3):531–543, 2015.

3 J. Auslander and Y. N. Dowker. On disjointness of dynamical systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 85(3):477–491, 1979.

4 Veronica Becher, Olivia Carton, and Pablo Ariel Heiber. Normality and automata. *Journal of Computer and System Sciences*, 81(8):1592–1613, 2015.

5 Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Matem. Palermo*, 27:247–271, 1909.

6 A. Broglio and P. Liardet. Predictions with automata. symbolic dynamics and its applications. *Contemporary Mathematics*, 135:111–124, 1992. Also appeared in Proceedings of the AMS Conference in honor of R. L. Adler. New Haven CT – USA 1991.

7 Yann Bugeaud. *Distribution modulo one and Diophantine approximation*. Cambridge Tracts in Mathematics. Cambridge University Press, 2012.

8 Olivier Carton. A direct proof of agafonov's theorem and an extension to shift of finite type. *CoRR*, abs/2005.00255, 2020. `arXiv:2005.00255`.

9 Olivier Carton. A direct proof of Agafonov's theorem and an extension to shifts of finite type. *Preprint*, 2020.

10 Olivier Carton and Joseph Vandehey. Preservation of normality by non-oblivious group selection. *Theory of Computing Systems*, 2020.

11 D. G. Champernowne. The construction of decimals normal in the scale of ten. *Journal of the London Mathematical Society*, s1-8(4):254–260, 1933.

12 T. Kamae and B. Weiss. Normal numbers and selection rules. *Israel Journal of Mathematics*, pages 101–110, 1975.

13 Wolfgang Merkle and Jan Reimann. Selection functions that do not preserve normality. *Theory Comput. Syst.*, 39(5):685–697, 2006.

14 Ivan Niven and H. S. Zuckerman. On the definition of normal numbers. *Pacific J. Math.*, 1(1):103–109, 1951.

15 В. Н. Агафонов. Нормальные последовательности и конечные автоматы. Докл. АН СССР, 179(2):255–256, 1968.

16 Л. П. Постникова. О связи понятий коллектива Мизеса–Черча и нормальной по Бернулли последовательности знаков. Теория вероятн. и ее примен., 6(2):232–234, 1961.

17 LP Postnikova. On the connection between the concepts of collectives of Mises-Church and normal Bernoulli sequences of symbols. *Theory of Probability & Its Applications*, 6(2):211–213, 1961. translation of [16] by Eizo Nishiura.

18 Michael O. Rabin. Probabilistic automata. *Information and Control*, 6(3):230–245, 1963.

19 Claus-Peter Schnorr and H. Stimm. Endliche Automaten und Zufallsfolgen. *Acta Informatica*, 1:345–359, 1972.

20 Xiaowen Wang and Teturo Kamae. Selection rules preserving normality. *Israel Journal of Mathematics*, 232:427–442, 2019.