



HAL
open science

Agafonov's theorem for probabilistic selectors

Ulysse Léchine, Thomas Seiller, Jakob Grue Simonsen

► **To cite this version:**

Ulysse Léchine, Thomas Seiller, Jakob Grue Simonsen. Agafonov's theorem for probabilistic selectors. 2024. hal-04563952

HAL Id: hal-04563952

<https://hal.science/hal-04563952>

Preprint submitted on 30 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Agafonov’s theorem for probabilistic selectors

2 Ulysse Léchine ✉ 

3 Université Sorbonne Paris Nord, France

4 Thomas Seiller ✉ 

5 CNRS, France

6 Jakob Grue Simonsen ✉ 

7 University of Copenhagen, Denmark

8 Abstract

9 A normal sequence over $\{0, 1\}$ is an infinite sequence for which every word of length k appears
10 with frequency 2^{-k} . Agafonov’s eponymous theorem states that selection by a finite state selector
11 preserves normality, i.e. if α is a normal sequence and A is a finite state selector, then the subsequence
12 $A(\alpha)$ is either finite or a normal sequence.

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

17 **2012 ACM Subject Classification** Theory of computation \rightarrow Formal languages and automata theory;
18 Theory of computation \rightarrow Probabilistic computation; Theory of computation \rightarrow Randomness,
19 geometry and discrete structures

20 **Keywords and phrases** Normal sequences, probabilistic automata, Agafonov’s theorem

21 **Digital Object Identifier** 10.4230/LIPIcs...

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

23 1 Introduction

24 Let $\alpha = x_1x_2 \dots$ be an infinite sequence over $\{0, 1\}$; α is said to be *normal* if every finite string
25 of length n over Σ occurs with limiting frequency 2^{-n} in α [5]. By standard reasoning, almost
26 all infinite sequences are normal when the set $\{0, 1\}^\omega$ of infinite sequences is equipped with
27 the usual Borel measure. Concrete examples of normal sequences include Champernowne’s
28 binary sequence 0100011011000001 \dots [9], and many more examples exist [6].

29 A *finite-state selector* is a deterministic finite automaton (DFA) that selects the n th
30 symbol from α if the length $n - 1$ prefix of α is accepted by the DFA. *Agafonov’s Theorem*
31 [1] is the celebrated result that a sequence α is normal iff any DFA that selects an infinite
32 sequence from α , selects a normal sequence. While alternative proofs, generalizations [16] –
33 and counter-examples to generalizations [11] – abound, all results in the literature consider
34 deterministic or non-deterministic DFAs, but none consider probabilistic computation.

35 The extension to probabilistic selection is quite natural – not only are the underlying
36 notions probabilistic in nature (i.e., normality of the transformed sequence), but the machinery
37 of finite automata and similar computational devices itself has a 60-year history [15] of begin
38 extended to probabilistic devices.

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



© Ulysse Léchine, Thomas Seiller, Jakob Grue Simonsen;
licensed under Creative Commons License CC-BY 4.0

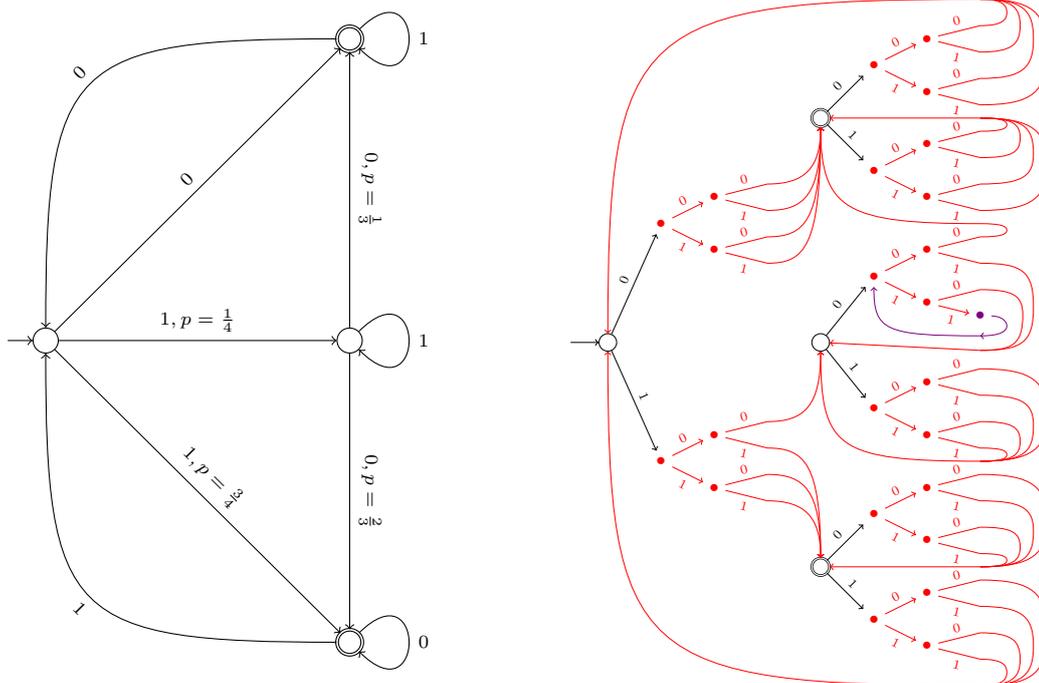
Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

XX:2 Agafonov's theorem for probabilistic selectors

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

56 Figure 1 shows a probabilistic finite-state selector (Figure 1a) with two probabilistic
 57 transitions: one involves dyadic probabilities, the second one involves rational but non-dyadic
 58 probabilities. On the right-hand side (Figure 1b) is the determinization of this selector¹.
 59 All unlabelled edges correspond to transitions valid for both 0 and 1. Determinizing the
 60 selector is done by introducing *gadgets* (shown in red) that simulate the probabilistic choices
 61 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.

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

68 **Related work:** Agafonov's Theorem has been generalized in multiple ways beyond finite
 69 automata (see, e.g., [3, 2, 4, 8, 10, 16]). Conversely, it is known that when adding trifling
 70 computational expressivity to finite-state selectors, counterexamples to Agafonov's Theorem
 71 for the resulting selectors can be constructed [11]. While some existing work considers
 72 preservation of more general measures by finite automata, or similar selectors [7], no extant
 73 work considers stochastic *selection*.

74 2 Preliminaries and notation

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

77 If $v, w \in \{0, 1\}^*$, $v \cdot w$ denotes the concatenation of v and w ; the definition extends to
 78 $v \cdot \alpha$ for $\alpha \in \{0, 1\}^\omega$ *mutatis mutandis*.

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

81 Given a set S we write $\text{Dist}(S)$ the space of probability distributions on S . Given a
 82 probability distribution $\delta \in \text{Dist}(S)$, we say that δ is *dyadic* (resp. *rational*) when for all
 83 $s \in S$, $\delta(s)$ is a dyadic number (resp. a rational number), that is a number of the form $\frac{p}{2^k}$
 84 for integers p, k .

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

89 **► Definition 1.** Let $a = a_1 \cdots a_m$ and $b = b_1 \cdots b_n$ be finite sequences such that $n < m$. An
 90 occurrence of b in a is an integer i with $1 \leq i \leq m$ such that $a_i = b_1, a_{i+1} = b_2, \dots, a_{i+n-1} =$
 91 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
 92 $\#_N\{w\}(\alpha)$ the number of occurrences of w in $a_1 a_2 \cdots a_N$.

93 $\alpha \in \{0, 1\}^\omega$ 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
 94 limit $\lim_{N \rightarrow \infty} \#_N\{w\}(\alpha)/N$ exists and equals 2^{-m} .

95 **► Definition 2.** Let $\alpha = a_1 a_2 \cdots$ be an infinite sequence, and i and n be positive integers.
 96 The i th block of size n in α , denoted $B_n^i(\alpha)$, is the finite sequence $a_{(i-1)n+1} a_{(i-1)n+2} \cdots a_{in}$.
 97 If w is a finite sequence of length k with $k = jn + r$ for appropriate j and $r < n$, the i th
 98 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 $\#_N^{(n)}\{w\}(\alpha)$ for the number of blocks of size n that
 are equal to w in the prefix of size $N \times n$:

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

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

$$\text{freq}(\alpha, w) = \lim_{N \rightarrow \infty} \frac{\#_N^{(n)}\{w\}(\alpha)}{N}.$$

XX:4 Agafonov's theorem for probabilistic selectors

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

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

102 By standard results, an infinite sequence α is normal iff it is block-normal [12, 14].

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

105 ► **Definition 3.** A *probabilistic finite state selector* \mathcal{S} is a tuple (Q, t, ι, A) where Q is a
106 finite set of states, $\iota \in Q$ is the initial state, $A \subset Q$ is the subset of accepting states, and
107 $t : Q \times \{0, 1\} \rightarrow \text{Dist}(Q)$ is a probabilistic transition function.

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

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

113 Observe that if \mathcal{S} is deterministic, the induced probability distribution assigns probability
114 1 to the *unique* selected subsequence $\mathcal{S}(\alpha)$ of α considered in the standard Agafonov theorem,
115 i.e. the sequence of bits α_i in α such that \mathcal{S} reaches an accepting state when given the prefix
116 $\alpha_0 \dots \alpha_{i-1}$.

► **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 Σ^ω defined as follows on
cylindrical sets C_w :

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

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

119 We finally recall the standard Agafonov theorem.

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

123 2.1 Technical lemmas about normality

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

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

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

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

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

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

136 ■ α is normal

137 ■ there exists $k \in \mathbb{N}_{>0}$ such that α is length- km normal for all $k \in \mathbb{N}$.

138 **Proof.** The proof is in the appendix. ◀

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

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

143 **Proof.** The proof is in the appendix. ◀

144 We will now define a partition of the set of blocks $(B_K^j(\alpha))_{j \in \mathbb{N}}$ into groups $(E_i)_{i \in \mathbb{N}}$ such
145 that each $r \in \{0, 1\}^K$ appears exactly once in each E_i . Those will be defined from a partition
146 $(V_i^K(\alpha))_{i \in \mathbb{N}}$ of \mathbb{N} such that the set $V_i^K(\alpha)$ contains the indices of the blocks of size K of α
147 contained in E_i .

148 ► **Definition 10.** *Let $\alpha \in \{0, 1\}^\omega$ be a normal sequence, and $K \in \mathbb{N}$. We define $\theta_i(\alpha) :$
149 $\{0, 1\}^K \rightarrow \mathbb{N}$ as mapping a word w to the value j such that $B_K^j(\alpha)$ is exactly the i -th block
150 of size K of α equal to w . (i.e. there are exactly $i - 1$ indices $j_1 < j_2 < j_3 < \dots < j_{i-1} < j$
151 such that $\forall k, B_K^{j_k}(\alpha) = w \wedge B_K^j(\alpha) = w$)*

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

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

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

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

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

158 ► **Lemma 12.** *Let Ω_1 and Ω_2 be two sample spaces. Let $(X_i)_{i \in \mathbb{N}}$ be an iid family of r.v. which
159 take value in $\Omega_1^{\mathbb{N}}$, $(Y_i)_{i \in \mathbb{N}}$ another iid family of r.v. which take value in $\Omega_2^{\mathbb{N}}$. Let $f(X, Y)$
160 be a function. Suppose $\forall x \in \Omega_1, \forall y \in \Omega_2, f(X, y)$ and $f(x, Y)$ have finite expected value and
161 variance then $\mathbb{P}_{Y_i} \left(\sum_x [\mathbb{P}(X = x) \sum_{i=1}^N f(x, Y_i)] = N \times \mathbb{E}_{X_0, Y_0}(f(X_0, Y_0)) + o(N) \right) = 1$*

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

163 3 Dyadic case

164 We first restrict to dyadic selectors. Given a dyadic selector $S = (Q, t, \iota, A)$, we define its
165 dyadicity degree as the smallest integer k such that for all states $q, q' \in Q$ and element
166 $a \in \{0, 1\}$, the probability $t(q, a)(q')$ can be written as $\frac{m}{2^k}$.

167 We first define the determinisation of a dyadic selector.

168 ► **Definition 13.** *Given a dyadic selector $S = (Q, t, \iota, A)$ of dyadicity degree d , we define a
169 determinisation $\text{Det}(S)$ of S as the deterministic selector (Q', t', ι', A') where:*

170 ■ $Q' = Q \cup Q \times \{0, 1\} \times \{0, 1\}^{\leq k-1}$;

171 ■ $\iota' = \iota$ and $A' = A$;

172 ■ the transition function t is defined as follows:

XX:6 Agafonov's theorem for probabilistic selectors

- 173 ■ for all $q \in Q$, $t'(q, a) = (q, a, \epsilon)$ where ϵ is the empty list;
- 174 ■ for all $((q, b, w))$ with $w \in \{0, 1\}^{\leq k-2}$, $t'((q, b, w), a) = (q, b, w \cdot a)$;
- 175 ■ for all (q, b, w) with $w \in \{0, 1\}^{k-1}$, $t'((q, b, w), a) = q'$ where $q' = \phi(w \cdot a)$ for a chosen
- 176 $\phi_{q,b} : 2^k \rightarrow Q$ such that the preimage of any $s \in Q$ has cardinality m_s where m_s is
- 177 defined by $t(q, b)(s) = \frac{m_s}{2^k}$.

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

181 ► **Definition 14.** Let α, ρ be sequences in $\{0, 1\}^\omega$, and $D \in \mathbf{N}$. The interwoven sequence
 182 $I^D(\alpha, \rho)$ is defined as the sequence:

$$183 \quad \alpha_0 \rho_0 \dots \rho_{d-1} \alpha_1 \rho_d \dots \rho_{2d-1} \dots$$

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

186 ► **Lemma 15.** Let S be a dyadic selector of dyadicity degree D . Then for all sequence
 187 $\alpha \in \{0, 1\}^\omega$ the random variables $\mathcal{S}(\alpha)$ and $\text{Prob}_{\rho \in \{0, 1\}^\omega}(\text{Det}(S)(I^D(\alpha, \rho)))$ are equal.

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

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

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

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

192 **Proof.** In case $D = 0$, the interwoven sequence $I^D(\alpha, \rho)$ is equal to α . As a consequence,
 193 $\text{Prob}_{\rho \in \{0, 1\}^\omega} [I^D(\alpha, \rho) \text{ is normal}]$ is equal to 1.

194 We now suppose that $D \neq 0$. Given $m \in \mathbf{N}_{>0}$, we will show that $I^D(\alpha, \rho)$ is length-
 195 $(d+1)m$ normal with probability 1. This implies that for almost all $\rho \in \{0, 1\}^\omega$, the sequence
 196 $I^D(\alpha, \rho)$ is length- $(D+1)m$ normal for every $m \in \mathbf{N}_{>0}$. By Lemma 8, this implies that for
 197 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:

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

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

$$199 \quad \alpha_i r_j \dots r_{j+D} \alpha_{i+1} r_{j+D} \dots r_{j+2D} \dots \alpha_{i+m-1} r_{j+(m-1)D} \dots r_{j+mD}.$$

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

202 We will consider the block decomposition of α into blocks of size m . Let $\text{idx}(i) = j$ where
 203 j is the i -th block such that $B_j^m(\alpha) = [w]_D$. Note that this function is well defined because
 204 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
 205 $[B_i^{(D+1)m}(I^D(\alpha, \rho))]_D$ should be equal to $[w]_D$. We write $\tilde{N} = \#_N^{(m)} \{[w]_D\}(\alpha)$, note that it

206 is the maximal i such that $\text{idx}(i) < N$. We also define \bar{w} as the complementary subsequence
 207 of w :

$$208 \quad \bar{w} = w_1 \dots w_D w_{D+2} \dots w_{2(D+1)-1} w_{2(D+1)+1} \dots w_{m(D+1)-1}.$$

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

$$211 \quad P = \text{Prob}_{\rho \in \{0,1\}^\omega} \left(\lim_{N \rightarrow \infty} \frac{\#_N^{((D+1)m)} \{w\}(I^D(\alpha, \rho))}{N} = 2^{-(D+1)m} \right)$$

$$212 \quad = \text{Prob}_{\rho \in \{0,1\}^\omega} \left(\lim_{N \rightarrow \infty} \frac{\#_{\text{Im}(\text{idx})}^{((D+1)m)} \{w\}(I^D(\alpha, \rho))}{N} = 2^{-(D+1)m} \right)$$

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

$$214 \quad P = \text{Prob}_{\rho \in \{0,1\}^\omega} \left(\lim_{\tilde{N} \rightarrow \infty} \frac{\#_{\text{Im}(\text{idx})}^{((D+1)m)} \{w\}(I^D(\alpha, \rho))}{\tilde{N} \cdot 2^{-m}} = 2^{-(D+1)m} \right)$$

$$215 \quad = \text{Prob}_{\rho \in \{0,1\}^\omega} \left(\lim_{\tilde{N} \rightarrow \infty} \frac{\#_{\text{Im}(\text{idx})}^{((D+1)m)} \{w\}(I^D(\alpha, \rho))}{\tilde{N}} = 2^{-Dm} \right)$$

$$216 \quad = \text{Prob}_{\rho \in \{0,1\}^\omega} \left(\lim_{\tilde{N} \rightarrow \infty} \frac{\#_{\text{Im}(\text{idx})}^{(Dm)} \{\bar{w}\}(\rho)}{\tilde{N}} = 2^{-Dm} \right)$$

217 By the law of large numbers, we have that

$$218 \quad \text{Prob}_{\rho \in \{0,1\}^\omega} \left(\lim_{\tilde{N} \rightarrow \infty} \frac{\#_{m^{-1}(N)}^{(Dm)} \{\bar{w}\}(\rho)}{\tilde{N}} = 2^{-Dm} \right) = 1,$$

219 which concludes the proof. ◀

220

221 This lemma then leads to the following theorem.

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

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

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

XX:8 Agafonov's theorem for probabilistic selectors

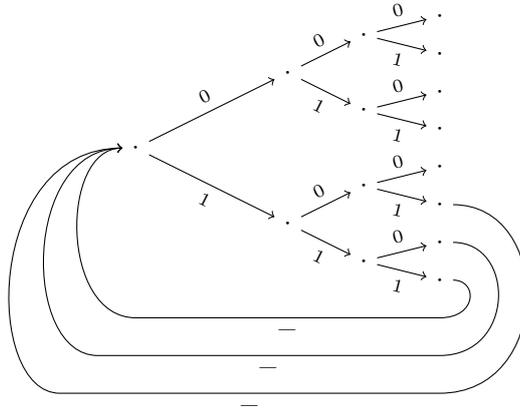
234 We will now consider the case of rational selectors. The difficulty in adapting the proof
 235 lies in the fact that the interwoven sequence has a less regular structure. In the above
 236 proof, each block of size $(d + 1)m$ followed the same pattern. But in the case of rational
 237 selectors,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

244 **4 Rational selector**

245 **4.1 Determinisation**

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

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



253

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

255 **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
 256 M the smallest common multiple of all denominators of the elements p_i , and write $p_i = \frac{\tilde{p}_i}{M}$.
 257 We will denote by $q_i = \sum_{j=1}^i \tilde{p}_j$. Note that $q_0 = 0$ and $q_k = \sum_i \tilde{p}_i = M$. Now consider P
 258 the smallest natural number such that $2^P \geq M$. We build the regular automaton of depth P
 259 with feedback loops on $2^P - M$ leaves. We will show that the probability p of reaching a leaf
 260 within $[q_i + 1, q_{i+1}]$ is equal to p_i . One only need to compute:

261
$$p = \frac{\tilde{p}_i}{2^P} \sum_{m \geq 0} \left(\frac{2^P - M}{2^P} \right)^m = \frac{\tilde{p}_i}{2^P} \frac{1}{1 - \frac{2^P - M}{2^P}} = \frac{\tilde{p}_i}{2^P} \frac{2^P}{2^P - (2^P - M)} = \frac{\tilde{p}_i}{2^P} \frac{2^P}{M} = \frac{\tilde{p}_i}{M} = p_i \blacktriangleleft$$

262 We will now define the determinisation of a rational selector in a similar way as for the
 263 dyadic case. First, since the selector is finite, one can write all rational numbers involved

264 with a common denominator, say k . Given a rational selector \mathcal{S} , we will call k the *rationality*
 265 *degree* of \mathcal{S} . Then each transition will be simulated by a gadget as defined above.

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

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

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

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

288 ► **Definition 22.** Let \mathcal{S} be a rational selector of rationality degree k and dyadicity degree D ,
 289 $\alpha \in \{0, 1\}^\omega$ an input sequence, and $\rho \in \{0, 1\}^\omega$ an advice sequence. We define the interwoven
 290 sequence $I_k^D(\alpha, \rho)$ as:

$$291 \quad \alpha_1 \rho_1 \dots \rho_{i_1} \alpha_2 \rho_{i_1+1} \dots \rho_{i_2} \alpha_3 \dots,$$

292 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
 293 $w_1 r_1 w_2 r_2 \dots r_m w_{m+1}$ where:

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

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

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

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

XX:10 Agafonov's theorem for probabilistic selectors

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

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

316 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

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

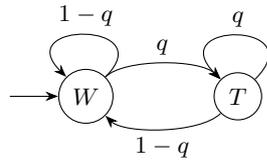
317 To prove that rational selectors preserve normality, we will prove in this section that
 318 $\mathbb{P}_\rho(I_k^D(\alpha, \rho) \text{ is normal}) = 1$, that is the generalised version of Lemma 16. As in the dyadic
 319 case, this is the crux of the problem, and the proof of the main theorem will easily follow. In
 320 order to prove this technical lemma, we analyze a process we call \mathcal{F} which takes a sequence
 321 α and inserts in between every bit of α a random amount of random bits. We will then show
 322 that if α is normal the sequence $\mathcal{F}(\alpha)$ obtained in this way is normal. Finally we will argue
 323 that normality of $I_k^D(\alpha, \rho)$ amounts to the normality of $\mathcal{F}(\alpha)$.

324 ► **Definition 24** (Random process \mathcal{F}_q). *Suppose given $K \in \mathbb{N}$, $w \in \{0, 1\}^K$, $q \in [0; 1[$, and
 325 $\tau \in \{0, 1\}^\omega$. We define $\mathcal{F}_q(w, \tau) \in \{0, 1\}^*$ as the random variable described in Figure 2 where
 326 we consume a bit of w when we get to state W and a bit of τ when we get to state T . The
 327 process stops when the state W is reached and there are no more bits of w to be consumed.
 328 The output is all the consumed bit in timely order.*

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

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

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



Example: if $w = 01110$, $\tau = 10010\dots$, then

$$\mathcal{F}(w) = 0101110010,$$

with the sequence of states

$$WTTWWTTTW.$$

■ **Figure 2** The random process \mathcal{F} .

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

► **Definition 26.** *Suppose given $\alpha \in \{0, 1\}^\omega$, $K \in \mathbb{N}$, and $q \in \mathbb{R}$. Let $(\mathcal{F}_i)_{i \in \mathbb{N}}$ 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 $B_K^i(\alpha)$:*

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

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

338 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 $(w_i)_{i \in \mathbf{N}} \in (\{0, 1\}^K)^{\mathbf{N}}$,*

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

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

342 In the next lemma we show that for any w we can approximate the number of w
 343 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
 344 $\mathcal{F}_i(B_K^i(\alpha))$ separately. The larger K the more precise the approximation. What we gain
 345 from this separation is that the random variable $\#_{s_i - |w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))$ are independent
 346 and we can apply the law of large numbers. In contrast in $\mathcal{F}(\alpha)$ where we concatenate the
 347 $\mathcal{F}_i(B_K^i(\alpha))$ we do not have independence because knowing that w appears at the end of
 348 $\mathcal{F}_1(B_K^1(\alpha))$ 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*

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

349 **Proof.** The proof is in the appendix. ◀

350 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
 351 discrepancy MN of the number of w noticed in the previous theorem can be made negligible
 352 when compared to the size of the string.

353 In the next theorem we just prove that for a random ρ the proportion of w in $\mathcal{F}_i(B_K^i(\rho))$
 354 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:*

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

355 **Proof.** This result can be shown by standard analysis of fair random sequences of size Kq^{-1} .
 356 Indeed for a random ρ , $\mathcal{F}_i(B_K^i(\rho))$ is just a random sequence of expected size Kq^{-1} . Let
 357 $\epsilon \in \mathbf{R}$. If K is large enough, there exists some $\epsilon' \in \mathbf{R}$ such that a random sequence of size
 358 Kq^{-1} contains on average $2^{-|w|}Kq^{-1} + \epsilon'$ occurrences of w where $|\epsilon'| < \epsilon$. ◀

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

XX:12 Agafonov's theorem for probabilistic selectors

361 **Proof.** The full proof is in the appendix. Here is a sketch. Using Lemma 28 and Lemma 11,
362 we have:

$$363 \quad \forall \epsilon', \mathbb{P}(\mathcal{F}_q(\alpha) \text{ is } w\text{-normal up to } \epsilon') = 1 \Leftrightarrow \forall \epsilon, \exists K, \mathbb{P}\left(T_1 + T_2 + T_3 \xrightarrow{N} 2^{-|w|} KNq(1 \pm \epsilon)\right) = 1,$$

364 where

$$365 \quad T_1 = \sum_{i \in [N/2^K]} \sum_{j \in V_i(\alpha)} \#_{s_j} \{w\}(\mathcal{F}_j(B_K^j(\alpha))),$$

$$366 \quad T_2 = \sum_{i=1+N/2^K}^{\frac{N}{2^K}+g(N)} \sum_{j \in V_i(\alpha)} \#_{s_j} \{w\}(\mathcal{F}_j(B_K^j(\alpha))),$$

$$367 \quad T_3 = \sum_{j \in [N] \setminus \bigcup_{i \in B_N} V_i} \#_{s_j} \{w\}(\mathcal{F}_j(B_K^j(\alpha))).$$

368 Then we show the right hand holds using Lemma 12 and Lemma 29 that T_1 goes to
369 $2^{-|w|}KNq(1 \pm \epsilon)$ with probability 1 and that T_2 and T_3 are $o(N)$ with probability 1.

370 This shows that $\forall \epsilon', \mathbb{P}(\mathcal{F}_q(\alpha) \text{ is } w\text{-normal up to } \epsilon') = 1$, which is enough to conclude. ◀

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

372 **Proof.** By lemma 30 $\forall w \in \{0, 1\}^*$, $\mathcal{F}(\alpha)$ is w -normal with probability 1. $\mathbb{P}(\mathcal{F}(\alpha) \text{ is normal}) =$
373 $\mathbb{P}(\forall w \in \{0, 1\}^*, \mathcal{F}(\alpha) \text{ is } w\text{-normal})$, since this is an intersection of countably many event of
374 probability 1, we have that $\mathcal{F}(\alpha)$ is normal with probability 1. ◀

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

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

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

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

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

384 Thus now we only need to show that y is normal with probability 1. Notice that the
385 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
386 with probability 1. ◀

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

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

391 While we believe the equivalent statement to hold for general probabilistic selector, we
392 believe establishing the result would require a different proof method.

393 — **References** —

- 394 **1** V. N. Agafonov. Normal sequences and finite automata. *Sov. Math., Dokl.*, 9:324–325, 1968.
395 Originally published in Russian (vol. 179:2, p. 255-266).
- 396 **2** Dylan Airey and Bill Mance. Normality preserving operations for Cantor series expansions
397 and associated fractals, i. *Illinois J. Math.*, 59(3):531–543, 2015.
- 398 **3** J. Auslander and Y. N. Dowker. On disjointness of dynamical systems. *Mathematical*
399 *Proceedings of the Cambridge Philosophical Society*, 85(3):477–491, 1979.
- 400 **4** Veronica Becher, Olivia Carton, and Pablo Ariel Heiber. Normality and automata. *Journal of*
401 *Computer and System Sciences*, 81(8):1592 – 1613, 2015.
- 402 **5** Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ.*
403 *Matem. Palermo*, 27:247–271, 1909.
- 404 **6** Yann Bugeaud. *Distribution modulo one and Diophantine approximation*. Cambridge Tracts
405 in Mathematics. Cambridge University Press, 2012.
- 406 **7** Olivier Carton. A direct proof of Agafonov’s theorem and an extension to shifts of finite type.
407 *Preprint*, 2020.
- 408 **8** Olivier Carton and Joseph Vandehey. Preservation of normality by non-oblivious group
409 selection. *Theory of Computing Systems*, 2020.
- 410 **9** D. G. Champernowne. The construction of decimals normal in the scale of ten. *Journal of the*
411 *London Mathematical Society*, s1-8(4):254–260, 1933.
- 412 **10** T. Kamae and B. Weiss. Normal numbers and selection rules. *Israel Journal of Mathematics*,
413 pages 101–110, 1975.
- 414 **11** Wolfgang Merkle and Jan Reimann. Selection functions that do not preserve normality. *Theory*
415 *Comput. Syst.*, 39(5):685–697, 2006.
- 416 **12** Ivan Niven and H. S. Zuckerman. On the definition of normal numbers. *Pacific J. Math.*,
417 1(1):103–109, 1951.
- 418 **13** Л. П. Постникова. О связи понятий коллектива Мизеса–Черча и нормальной по
419 Бернулли последовательности знаков. Теория вероятн. и ее примен., 6(2):232–234, 1961.
- 420 **14** LP Postnikova. On the connection between the concepts of collectives of Mises-Church and
421 normal Bernoulli sequences of symbols. *Theory of Probability & Its Applications*, 6(2):211–213,
422 1961. translation of [13] by Eizo Nishiura.
- 423 **15** Michael O. Rabin. Probabilistic automata. *Information and Control*, 6(3):230–245, 1963.
- 424 **16** Xiaowen Wang and Teturo Kamae. Selection rules preserving normality. *Israel Journal of*
425 *Mathematics*, 232:427–442, 2019.

426 **A Omitted proofs**

427 **Proof of Lemma 8.** Consider a sequence α which is length- km normal for all $k \in \mathbf{N}$, and
 428 fix a word $w \in \{0, 1\}^n$. We will use the length- mn normality of α . For this, we note that we
 429 can write:

$$\begin{aligned}
 430 \quad \#_N^{(n)}\{w\}(\alpha) &= \sum_{w_1, \dots, w_m \in \{0, 1\}^n} \text{Card}\{i \in \{1, \dots, m\} \mid w_i = w\} \cdot \#_{N/m}^{(mn)}\{w_1 \cdots w_m\}(\alpha) \\
 431 \quad &= \sum_{w_1, \dots, w_{m-1} \in \{0, 1\}^n} \sum_{j=1}^m \#_{N/m}^{(mn)}\{w_1 \cdots w_{j-1} \cdot w \cdot w_j \cdot w_{m-1}\}(\alpha) \\
 432 \quad &= \sum_{w_1, \dots, w_{m-1} \in \{0, 1\}^n} m \#_{N/m}^{(mn)}\{w \cdot w_1 \cdots w_{m-1}\}(\alpha)
 \end{aligned}$$

433 As a consequence:

$$\begin{aligned}
 434 \quad \text{freq}(\alpha, w) &= \lim_{N \rightarrow \infty} \frac{\#_N^{(n)}\{w\}(\alpha)}{N} \\
 435 \quad &= \lim_{N \rightarrow \infty} \sum_{w_1, \dots, w_m \in \{0, 1\}^n} \frac{\text{Card}\{i \in \{1, \dots, m\} \mid w_i = w\} \cdot \#_{N/m}^{(mn)}\{w_1 \cdots w_m\}(\alpha)}{N} \\
 436 \quad &= \lim_{N \rightarrow \infty} \sum_{w_1, \dots, w_{m-1} \in \{0, 1\}^n} \frac{m \cdot \#_{N/m}^{(mn)}\{w \cdot w_1 \cdots w_{m-1}\}(\alpha)}{N} \\
 437 \quad &= \sum_{w_1, \dots, w_{m-1} \in \{0, 1\}^n} \lim_{N \rightarrow \infty} \frac{\#_{N/m}^{(mn)}\{w \cdot w_1 \cdots w_{m-1}\}(\alpha)}{N/m} \\
 438 \quad &= \sum_{w_1, \dots, w_{m-1} \in \{0, 1\}^n} \text{freq}(\alpha, w \cdot w_1 \cdots w_{m-1}) \\
 439 \quad &= \sum_{w_1, \dots, w_{m-1} \in \{0, 1\}^n} 2^{-mn} \\
 440 \quad &= 2^{n(m-1)} 2^{-mn} \\
 441 \quad &= 2^{-n}
 \end{aligned}$$

442 ◀

Proof of Lemma 9. 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 $|\#_{N_i}^{(n)}\{w\}(\alpha) - 2^{-n}N_i| > \epsilon N_i$ for all $i \in \mathbf{N}$. In other words,

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

This contradicts the normality of α since it implies that

$$\lim_{i \rightarrow \infty} \left| \frac{\#_{N_i}^{(n)}\{w\}(\alpha)}{N_i} - 2^{-n} \right| = 0. \quad \blacktriangleleft$$

443 **Proof of Lemma 28.** First note that we count indices up to $s_i - |w|$ in $\#_{s_i - |w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))$
 444 because if w appears in $\mathcal{F}_i(B_K^i(\alpha))$ it must appear before the last $|w|$ bits. For this reason
 445 we also mention that $\#_{s_i - |w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha))) = \#_{|w|}\{w\}(\mathcal{F}_i(B_K^i(\alpha)))$.

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

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

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

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

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

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

451 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
 452 the indices of blocks $B_K^j(\alpha)$ into sets V_i of size 2^K and such that $|V_i| = 2^K$. We may also
 453 change $s_j - |w|$ to s_j as explained in the proof of Lemma 28. Then the above is equivalent to

$$454 \quad \forall \epsilon, \exists K, \mathbb{P}\left(S_1 + S_2 \xrightarrow{N} 2^{-|w|} KNq^{-1}(1 \pm \epsilon)\right) = 1,$$

455 where

$$456 \quad S_1 = \sum_{i \in B_N} \sum_{j \in V_i(\alpha)} \#_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha))),$$

$$457 \quad S_2 = \sum_{j \in [N] \setminus \bigcup_{i \in B_N} V_i} \#_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha))).$$

By Lemma 11, we have that $|B_N| = \frac{N}{2^K} + g(N)$, where $g(N) = o(N)$. This rewrites:

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

458 where

$$459 \quad T_1 = \sum_{i \in [N/2^K]} \sum_{j \in V_i(\alpha)} \#_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha))),$$

$$460 \quad T_2 = \sum_{i=1+N/2^K}^{\frac{N}{2^K}+g(N)} \sum_{j \in V_i(\alpha)} \#_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha))),$$

$$461 \quad T_3 = \sum_{j \in [N] \setminus \bigcup_{i \in B_N} V_i} \#_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha))).$$

462 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 , $B_K^j(\alpha)$ takes all values in $\{0, 1\}^K$. By creating an appropriate bijection between j and (i, r) , we can write

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



XX:16 Agafonov's theorem for probabilistic selectors

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:

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

■ **The term T_2 .** We have that

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

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

■ **The term T_3 .** We have that

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

465 The sum is over $o(N)$ terms by Lemma 11 and the random variable $\sharp_{s_j}\{w\}(\mathcal{F}_j(B_K^j(\alpha)))$
466 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

$$\mathbb{P}(\mathcal{F}(\alpha) \text{ is } w\text{-normal}) = 1.$$

