The Agafonov and Schnorr-Stimm theorems for probabilistic automata

Laurent Bienvenu, Hugo Gimbert, Subin Pulari

February 10, 2025

Abstract

For a fixed alphabet A, an infinite sequence X is said to be normal if every word w over A appears in X with the same frequency as any other word of the same length. A classical result of Agafonov (1966) relates normality to finite automata as follows: a sequence X is normal if and only if any subsequence of X selected by a finite automaton is itself normal. Another theorem of Schnorr and Stimm (1972) gives an alternative characterization: a sequence X is normal if and only if no gambler can win large amounts of money by betting on the sequence X using a strategy that can be described by a finite automaton. Both of these theorems are established in the setting of deterministic finite automata. This raises the question as to whether they can be extended to the setting of probabilistic finite automata. In the case of the Agafonov theorem, this question was positively answered by Léchine et al. (2024) in a restricted case of probabilistic automata with rational transition probabilities.

In this paper, we settle the full conjecture by proving that both the Agafonov and the Schnorr-Stimm theorems hold true for arbitrary probabilistic automata. Specifically, we show that a sequence X is normal if and only if any probabilistic automaton selects a normal subsequence of X with probability 1. We also show that a sequence X is normal if and only if a probabilistic finite-state gambler fails to win on X with probability 1.

1 Introduction

Given a finite alphabet A of k letters, an infinite sequence X of letters is said to be *normal* if every word of A^* appear as sub-word of X with the same frequency as any other word of the same length, namely, $(1/k)^{|w|}$. The famous Champernowne sequence

$$0123456789101112131415161718192021222324252 6\ldots$$

can be shown to be normal over the alphabet $A = \{0, \ldots, 9\}$. The number π, for example, is conjectured to have a normal expansion in every base, though this very much remains an open question. Normal sequences are plentiful, and an easy way to generate a normal sequence X is to draw each letter $X(n)$ at random in the alphabet A (all letters having the same probability $1/|A|$) independently of the other chosen letters $X(m)$. The law of large numbers tells us that we obtain a normal sequence with probability 1.

Of course, there are also plenty of examples of non-normal sequences:

- Periodic, or ultimately periodic, sequences can never be normal: indeed if the period of X is k there are only k possible sub-words of X of length k hence most words of length k will not appear and their frequency will be 0.

- Sturmian sequences, which are sequences with only $k+1$ different sub-words of length k (such as the Fibonacci sequence 100101001001010010100100101001001... obtained by iterating the morphism $0 \mapsto 01$ and $1 \mapsto 0$) are not normal for the same reason.

- The Thue-Morse sequence 01101001100101101001011001101001... obtained by iterating the morphism $0 \mapsto 01$ and $1 \mapsto 10$ is not normal because 000 and 111 do not appear as sub-words.

- A sequence of 0 and 1's generated at random where each bit is chosen equal to the previous one with probability $2/3$ will have, with probability 1, all possible finite words as sub-words but will not (still with probability 1) be normal as for example the word 00 will appear with frequency $1/3$ instead of $1/4$.

It turns out that normality has a nice interpretation in terms of finite automata. Indeed, two classical results, one due to Agafonov and the other due to Schnorr and Stimm, assert that an infinite sequence is normal if and only if it cannot be predicted by a finite automaton with better-than-average accuracy. Of course, one needs to specify what 'predicted' means. We consider two prediction models.

(I) In the Agafonov model, an automaton reads the infinite sequence one letter at a time and updates its state in the usual way. Some of its states have a 'select' tag on them. When the current state has such a tag, the next letter will be selected and added to a subsequence Y. We consider the automaton successful at predicting X if the subsequence Y built in this process is infinite and some letter of A does not have asymptotic frequency $1/|A|$ in Y. This means that the automaton has exhibited a statistical anomaly in the sequence X and isolated this anomaly in the subsequence Y.

(II) In the Schnorr-Stimm model, the predictor is still an automaton but this time is viewed as a gambling strategy. The gambler starts with a capital of \$1. Each state q is labeled with a betting function $\gamma_q : A \to \mathbb{R}^{\geq 0}$. This function represents the amount by which the predictor would like her capital to be multiplied by depending on the value of the next bit. For example, suppose the player plays against a sequence $X \in \{a, b, c\}^\omega$. If her current capital is \$2 and the current state q is labelled by a betting function γ_q such that $\gamma_q(a) = 0.7$, $\gamma_q(b) = 1.1$ and $\gamma(c) = 1.2$, if the next letter is a, her new capital will be \$1.4, if it is b her new capital will be \$2.2 and if it is c her new capital will be \$2.4. For the game to be fair, each betting function γ_q must satisfy $\frac{1}{|A|} \sum_{a \in A} \gamma_q(a) = 1$. We say that the predictor wins if the capital of the player takes arbitrarily large values throughout the (infinite) game. That is, the predictor has spotted some type of statistical anomaly and is exploiting it to get rich!

Both of these models lead to the same conclusion: an infinite sequence X is normal if and only if it is unpredictable (by a finite automaton).

Theorem 1 (Agafonov [1]). *For $X \in A^\omega$, the following are equivalent.*

(i) X is normal.

(ii) For any automaton $\mathcal{A}$ that selects a subsequence Y of X as in model I, either Y is finite, or every letter of A appears in Y with asymptotic frequency $1/|A|$.

(iii) For any automaton $\mathcal{A}$ that selects a subsequence Y of X as in model I, either Y is finite, or Y is normal.

(see also Carton [3] and Seiller and Simonsen [10] for a modern account of this theorem).

Theorem 2 (Schnorr-Stimm [9]). *For $X \in A^\omega$, the following are equivalent.*

(i) X is normal.

(ii) Any automaton $\mathcal{A}$ betting on X according to model II does not win.

In both of these theorems, the finite automata used for prediction are assumed to be deterministic. Would the situation changed if one allowed probabilistic automata? In principle, one would not expect an unpredictable sequence to become predictable in the presence of a random source. Indeed, given a sequence X and a random source R it seems, informally speaking, that almost surely R will not 'know' anything about X and thus will not help predicting X. Surprisingly, this intuition is wrong in the setting where the predictors are not finite automata but Turing machines, as shown by Bienvenu et al. [2] who built a sequence that is unpredictable by deterministic Turing machines (in either prediction model of selection or gambling) and becomes predictable (in either model) if one allows probabilistic Turing machines. Nonetheless, finite automata are much weaker than Turing machines and Bienvenu et al.'s construction cannot work for such a memoryless model of computation. And indeed, recently, Léchine et al. showed that Agafonov's theorems holds for probabilistic automata in the restricted case where the transition probabilities are rational.

Theorem 3 (Léchine et al. [6]). *For $X \in A^\omega$, the following are equivalent.*

(i) X is normal.

(ii) For any probabilistic automaton $\mathcal{A}$ with rational probabilities, almost surely, $\mathcal{A}$ selects a subsequence Y of X that is either finite, or every letter of A appears in Y with asymptotic frequency $1/|A|$.

(iii) For any probabilistic automaton $\mathcal{A}$ with rational probabilities, almost surely, $\mathcal{A}$ selects a subsequence Y of X such that either Y is finite, or Y is normal.

This led them to conjecture that the probabilistic version of Agafonov's theorem holds in the general case. In this paper, we prove this conjecture and also prove the probabilistic version of the Schnorr-Stimm theorem. Additionally, we establish a probabilistic version of the Schnorr-Stimm dichotomy regarding the winning rates of probabilistic gamblers.

It is worth noting that Léchine et al.'s proof is a reduction of the rational probabilistic case to the deterministic case. Our proof is also a reduction to the deterministic case, but we will need an extension of the deterministic case to Bernoulli measures (an extension which was proved by Seiller and Simonsen [10]), which will be presented in the next section.

Notation and terminology

We finish this introduction by formalizing the concepts discussed so far and gathering the notation and terminology that will be used in the rest of the paper.

Given an alphabet A, we denote by A^* the set of finite words over A, by A^n the set of words of length n, by A^ω the set of infinite sequences of letters and by $A^{\leq\omega}$ the set $A^* \cup A^\omega$. For $X \in A^{\leq\omega}$, we denote by $X(i)$ the i-th letter of X (by convention there is a 0-th letter) and by $X[i,j]$ the word $X(i)X(i+1)\dots X(j)$. Let λ denote the empty string.

Given a word u of length k and a word w of length $n \geq k$, we denote by $\mathrm{NbOcc}(u,w)$ the number of occurrences of the word u in w, i.e.,

$$\mathrm{NbOcc}(u,w) = \#\{i : 0 \leq i \leq n-k,\ w[i,i+k-1] = u\}$$

and the frequency of occurrence $\mathrm{Freq}(u,w)$ of u in w is naturally defined by

$$\mathrm{Freq}(u,w) = \frac{\mathrm{NbOcc}(u,w)}{n-k+1}$$

When X is an infinite sequence, we define

$$\mathrm{Freq}^-(u,X) = \liminf_{n\to\infty} \mathrm{Freq}(u, X[0..n]) \quad \text{and} \quad \mathrm{Freq}^+(u,X) = \limsup_{n\to\infty} \mathrm{Freq}(u, X[0..n])$$

When $\mathrm{Freq}^-(u,X)$ and $\mathrm{Freq}^+(u,X)$ have the same value, we simply call this common value $\mathrm{Freq}(u,X)$.

Given $X \in A^\omega$, we say that X is *balanced* if all letters appear in X with the expected frequency, i.e., $\mathrm{Freq}(a,X) = 1/|A|$ for all $a \in A$. We say that X is *normal* if all words appear in X with the expected frequency, i.e., $\mathrm{Freq}(w,X) = |A|^{-|w|}$ for all $w \in A^*$.

A *deterministic automaton (DFA)* is a tuple (Q, A, q_I, δ) where Q is a finite set of states, A a finite alphabet, q_I the initial state and $\delta : Q \times A \to Q$ the transition function (in this paper, runs of automata are meant to be infinite hence there is no need for final states). We denote by δ^* the function from $Q \times A^*$ to Q defined inductively by $\delta^*(q,\epsilon) = q$ where ϵ is the empty word and for $w \in A^*$ and $a \in A$, $\delta^*(q, w \cdot a) = \delta(\delta^*(q,w), a)$, where $\cdot$ is the concatenation of words.

An *automatic selector* (or *selector* for short) is a tuple (Q, A, q_I, δ, S) where (Q, A, q_I, δ) is a DFA and S is a subset of Q, representing the selection states.

Given a selector $\mathcal{S} = (Q, A, q_I, \delta, S)$, we define the selection function from A^* to A^* inductively by:

$$\mathrm{Select}(\mathcal{S}, \epsilon) = \epsilon$$

and for $w \in A^*$ and $a \in A$:

$$\mathrm{Select}(\mathcal{S}, w \cdot a) = \begin{cases} \mathrm{Select}(\mathcal{S}, w) \cdot a & \text{if } \delta^*(q_I, w) \in S \\ \mathrm{Select}(\mathcal{S}, w) & \text{if } \delta^*(q_I, w) \notin S \end{cases}$$

If X is an infinite sequence in A^ω, the sequence of words $\mathrm{Select}(\mathcal{S}, X[0..n])$ is non-decreasing with respect to the prefix order and thus converges to a sequence $Y \in A^{\leq\omega}$ which we call *the selected subsequence of* X *selected by* $\mathcal{S}$ and denote by $\mathrm{Select}(\mathcal{S}, X)$.

An *automatic gambler* (or *gambler* for short) is a tuple $(Q, A, q_I, \delta, \gamma)$ where (Q, A, q_I, δ) is a DFA and γ is a function from $Q \times A$ to $\mathbb{R}^{\geq 0}$ such that for all q, $\frac{1}{|A|}\sum_{a\in A} \gamma(q,a) = 1$. As said above, the value of $\gamma(q,a)$ should be interpreted as the multiplier by which the gambler, being currently in state q, would like her capital to be multiplied by if the next read letter is a. The condition $\frac{1}{|A|}\sum_{a\in A} \gamma(q,a) = 1$ ensures that the game is fair.

Given a gambler $\mathcal{G} = (Q, A, q_I, \delta, \gamma)$ and $w \in A^*$, we define $\mathrm{Capital}(\mathcal{G}, w)$ inductively by

$$\mathrm{Capital}(\mathcal{G}, \epsilon) = 1$$

and for $w \in A^*$ and $a \in A$,

$$\mathrm{Capital}(\mathcal{G}, w \cdot a) = \mathrm{Capital}(\mathcal{G}, w) \cdot \gamma(\delta^*(q_I, w), a)$$

and we say that a gambler $\mathcal{G}$ *wins against* $X \in A^\omega$ if

$$\limsup_{n\to+\infty} \mathrm{Capital}(\mathcal{G}, X[0..n]) = +\infty$$

(otherwise we say that $\mathcal{G}$ loses).

2 Deterministic prediction for Bernoulli measures

In classical normality, all letters of the alphabet occur with the same frequency. We can however consider the extension of normality to Bernoulli measures. A Bernoulli measure over A^ω is a probability measure where letters of an infinite sequence X are drawn at random independently of one another but the distribution over the alphabet A is non-uniform.

Definition 1. *Let $\mu : A \to [0,1]$ be a distribution over the alphabet A (hence satisfies $\sum_{a\in A} \mu(a) = 1$). The Bernoulli measure induced my μ, which we also denote by μ by abuse of notation, is the unique probability measure such that for all i, k, for every word $w = a_0, \dots a_k \in A$,*

$$\Pr_{X\sim\mu}\Big[X[i, i+k] = w\Big] = \prod_{j=0}^{k} \mu(a_j)$$

We also denote by $\mu(w)$ the quantity $\prod_{j=0}^{k} \mu(a_j)$.

Normality generalizes very naturally to Bernoulli measures.

Definition 2. *Let μ be a Bernoulli measure. A sequence $X \in A^\omega$ is* μ-balanced *if* $\mathrm{Freq}(a, X) = \mu(a)$ *for all $a \in A$. It is* μ-normal *if for all words $w \in A^*$,* $\mathrm{Freq}(w, X) = \mu(w)$.

We say that a Bernoulli measure μ is *positive* when $\mu(a) > 0$ for every letter a. In the rest of the paper, all Bernoulli measures will be assumed to be positive, and we simply say 'Bernoulli measure' to mean 'positive Bernoulli measure'.

The Agafonov theorem can be extended to Bernoulli measures, as proven by Seiller and Simonsen [10]. It is this theorem that we will use in the next section to obtain a proof of the Agafonov theorem for probabilistic selectors.

Theorem 4 (Agafonov theorem for Bernoulli measures [10])**.** *For $X \in A^\omega$, the following are equivalent.*

(i) X is μ-normal.

(ii) For any selector $\mathcal{S}$ that selects a subsequence Y of X, either Y is finite or Y is μ-balanced.

(iii) For any selector $\mathcal{S}$ that selects a subsequence Y of X, either Y is finite or Y is μ-normal.

We can also easily generalize the notion of gambler to the setting of Bernoulli measures: it suffices to define a μ-gambler $\mathcal{G} = (Q, A, q_I, \delta, \gamma)$ as before but with the fairness condition on γ replaced by $\sum_{a\in A} \mu(a)\gamma(q, a) = 1$ for every q. The function Capital and the notion of success are defined as before.

We will now prove that the Schnorr-Stimm theorem, just like the Agafonov theorem, can also be extended to Bernoulli measures.

Theorem 5 (Schnorr-Stimm theorem for Bernoulli measures)**.** *For $X \in A^\omega$ and μ a Bernoulli measure, the following are equivalent.*

(i) X is μ-normal.

(ii) No μ-gambler $\mathcal{G}$ wins by betting on X.

Proof. $(i) \Rightarrow (ii)$. Suppose that $X \in A^\omega$ is μ-normal and consider a μ-gambler $\mathcal{G} = (Q, A, q_I, \delta, \gamma)$.

We can assume that the μ-gambler only has one state on which it places a non-trivial bet. Indeed, define for every $q \in Q$ the μ-gambler $\mathcal{G}^{[q]} = (Q, A, q_I, \delta, \gamma^{[q]})$ where $\gamma^{[q]}(q, a) = \gamma(q, a)$ for all a and $\gamma^{[q]}(q', a) = 1$ for $q' \neq q$. That is, $\mathcal{G}^{[q]}$ is the gambler $\mathcal{G}$ where all states but state q are neutralized (no bet is placed while on them). By the multiplicative nature of the function Capital, we have for all n:

$$\text{Capital}(\mathcal{G}, X[0..n]) = \prod_{q \in Q} \text{Capital}(\mathcal{G}^{[q]}, X[0..n])$$

Thus if we can show that all quantities $\text{Capital}(\mathcal{G}^{[q]}, X[0..n])$ are bounded, we are done. Let thus assume there is a state r that is the unique state on which $\mathcal{G}$ bets. If instead of $\mathcal{G}$ we consider the selector $\mathcal{S} = (Q, A, q_I, \delta, \{r\})$ with only r as selecting state, we know by Agafonov's theorem for Bernoulli measures (Theorem 4) that the subsequence Y of X selected by $\mathcal{S}$ is μ-normal, hence in particular μ-balanced. But this subsequence is precisely the values of X on which $\mathcal{G}$ bets!

We can further assume that Y is infinite, otherwise it means that the run of $\mathcal{G}$ on X passes by r finitely often, hence $\mathcal{G}$ certainly cannot win as other states are not betting states. Now, suppose that at stage n of the run on X the state r has been visited $k = k(n)$ times. We have

$$\text{Capital}(\mathcal{G}, X[0..n+1]) = \prod_{a \in A} \gamma(r, a)^{\text{NbOcc}(a, Y[0..k])}$$

But since Y is μ-normal we have, for every a, $\text{NbOcc}(a, Y[0..k]) = \mu(a)k + o(k)$. Thus,

$$\text{Capital}(\mathcal{G}, X[0..n+1]) = \prod_{a \in A} \gamma(r, a)^{\mu(a)k + o(k)}$$

or equivalently

$$\log \text{Capital}(\mathcal{G}, X[0..n+1]) = (k + o(k)) \cdot \sum_{a \in A} \mu(a) \cdot \log \gamma(r, a)$$

(here we assume that all values $\gamma(r, a)$ involved in the product are positive for if not then the capital falls to 0 and we are done). Since we have $\sum_{a \in A} \mu(a) = 1$ (μ being a distribution), we can use the strict concavity of the function log on $(0, +\infty)$ to apply Jensen's inequality and get

$$\sum_{a \in A} \mu(a) \cdot \log \gamma(r, a) \leq \log \left(\sum_{a \in A} \mu(a) \gamma(r, a) \right)$$

with strict inequality when not all $\gamma(q, a)$ are equal (which is the case where $\mathcal{G}$ makes non-trivial bets). But by the fairness condition, we have $\sum_{a \in A} \mu(a)\gamma(r, a) = 1$ hence we see that $\log \text{Capital}(\mathcal{G}, X[0..n+1])$ is either 0 or ultimately negative which either way means that $\mathcal{G}$ does not win.

$(ii) \Rightarrow (i)$ Assume that X is not μ-normal. This means that there is some word w such that $\text{Freq}(w, X[0..n])$ does not converge to $\mu(w)$. Let us assume that w is a minimal such word and write $w = ux$ with $u \in A^*$ and $x \in A$.

Consider the sequence of vectors f_n defined be

$$f_n = \left(\frac{\text{Freq}(ua, X[0..n])}{\sum_{b \in A} \text{Freq}(ub, X[0..n])} \right)_{a \in A}$$

or equivalently,

$$f_n = \left(\frac{\mathrm{NbOcc}(ua, X[0..n])}{\sum_{b \in A} \mathrm{NbOcc}(ub, X[0..n])} \right)_{a \in A}$$

All of these vectors belong to the set $\Gamma = \{f : A \to [0,1] : \sum_{a\in A} f(a) = 1\}$. This is a compact set, hence the sequence f_n must have cluster points. By definition of u and x, we know that f_n does not converge to μ because $f_n(x)$ does not converge to $\mu(x)$: Indeed, in the definition of $f_n(x)$, the denominator converges to $\mathrm{Freq}(u, X)$ which by minimality of w is defined and equal to $\mu(u)$, while the numerator is equal to $\mathrm{Freq}(w, X[0..n])$ which by definition of w does not converge to $\mu(w) = \mu(u)\mu(x)$.

Therefore, the sequence (f_n) must have at least one cluster point ν different from μ. Fix such a cluster point ν.

We now build our gambler $\mathcal{G} = (Q, A, q_I, \delta, \gamma)$. The idea is that the gambler will record the last $|u|$ bits it read and will only place bets when these exactly form the word u. Let us thus take $Q = \{q_v : v \in A^*, |v| \leq |u|\}$ initial state $q_I = q_\epsilon$ and define δ by

$$\delta(q_v, a) = \begin{cases} q_{va}, & \text{if } |v| < |u| \\ q_{v'a} & \text{if } |v| = |u| \text{ and } v = xv' \text{ with } x \in A \end{cases}$$

Now, define $\gamma(q_v, a) = 1$ whenever $v \neq u$ and $\gamma(q_u, a) = \nu(a)/\mu(a)$ for all $a \in A$

Observe that this is a valid μ-gambler as the fairness condition is satisfied: $\sum_{a\in A} \nu(a)/\mu(a) \cdot \mu(a) = \sum_a \nu(a) = 1$.

Suppose that after reading n letters of X the state q_u has been visited $k = k(n)$ times.

First, observe that $k(n)$ tends to $+\infty$. Indeed, the state q_u is visited whenever u is seen as a subword of X. But we assumed that u appears in X with frequency $\mu(u)$, by minimality of w.

Second, unfolding the definition of f_k, we have,

$$\mathrm{Capital}(\mathcal{G}, X[0, n+1]) = \prod_{a\in A} \left(\frac{\nu(a)}{\mu(a)} \right)^{k \cdot f_k(a)}$$

This gives

$$\log \mathrm{Capital}(\mathcal{G}, X[0, n+1]) = k \cdot \sum_{a \in A} f_k(a) \log \left(\frac{\nu(a)}{\mu(a)} \right)$$

Since ν is a cluster point of the sequence f_k, for any fixed $\varepsilon > 0$ there are infinitely many n (or k) such that

$$\sum_{a\in A} f_k(a) \log \left(\frac{\nu(a)}{\mu(a)} \right) \geq \sum_{a\in A} \nu(a) \log \left(\frac{\nu(a)}{\mu(a)} \right) - \varepsilon$$

But the term $\sum_{a\in A} \nu(a) \log \left(\frac{\nu(a)}{\mu(a)} \right)$ is the relative entropy from μ to ν (also known as Kullback-Liebler divergence, see for example [5]) $\mathrm{D}_{KL}(\nu || \mu)$. This quantity is nonnegative in general and is positive when $\nu \neq \mu$, which is the case here. We have thus established that for any fixed ε, there are arbitrarily large n and k such that

$$\log \mathrm{Capital}(\mathcal{G}, X[0, n+1]) \geq k \left(\mathrm{D}_{KL}(\nu || \mu) - \varepsilon \right)$$

Taking any $\varepsilon < \mathrm{D}_{KL}(\nu || \mu)$, this shows that

$$\limsup_{n \to +\infty} \mathrm{Capital}(\mathcal{G}, X[0, n]) = +\infty$$

□

Let us note that this last proof actually gives us a finer analysis of normality, in terms of the rate of failure or success of the gambler. This was already observed by Schnorr and Stimm in their seminal paper, where they proved the following.

Theorem 6 (Schnorr-Stimm dichotomy theorem [9])**.** *Let X be an infinite sequence in A^ω.*
(i) If X is normal and $\mathcal{G}$ is a gambler, then the capital of $\mathcal{G}$ throughout the game either is ultimately constant or decreases at an exponential rate.
(ii) If X is not normal, then there exists a gambler $\mathcal{G}$ which wins against X at an 'infinitely often' exponential rate (i.e., $\limsup_n \log(\text{Capital})/n > 0$).

As a byproduct of our proof of Theorem 5, we have the same dichotomy for positive Bernoulli measures (i.e., Bernoulli measures such that $\mu(a) > 0$ for every letter):

Theorem 7 (Schnorr-Stimm dichotomy theorem for Bernoulli measures)**.** *Let X be an infinite sequence in A^ω and μ a positive Bernoulli measure.*
(i) If X is μ-normal and $\mathcal{G}$ is a μ-gambler, then the capital of $\mathcal{G}$ throughout the game either is ultimately constant or decreases at an exponential rate.
(ii) If X is not μ-normal, then there exists a μ-gambler $\mathcal{G}$ which wins against X at an 'infinitely often' exponential rate.

Our proof of Theorem 5 almost establishes this, but we do need an additional technical lemma:

Lemma 1. *Let (Q, A, q_I, δ) be a finite-state automaton, μ be a positive Bernoulli measure and let $V_q(n, X)$ denote the number of times the state q is visited upon running the automaton using the first n bits of $X \in A^\omega$.*

Then, for every $q \in Q$, there exists a real number $\pi_q \geq 0$ such that, for every μ-normal sequence $X \in A^\omega$:

- *either $V_q(n, X)$ is ultimately constant (i.e., q is visited only finitely often during the run on X)*
- *or, $\lim_{n\to\infty} \frac{V_q(n,X)}{n} = \pi_q$ (i.e., the state q is visited with asymptotic frequency π_q) and this second case can only happen when $\pi_q > 0$.*

Proof. Running an automaton upon a normal sequence, starting from any state, a strongly connected component must be reached in finitely many steps. Similar to [10], let us consider the Markov chain corresponding to the $|Q| \times |Q|$ stochastic matrix $\mathbf{P}$ where,

$$\mathbf{P}_{ij} = \sum_{a \in A} \mu(a) \cdot 1_{\delta(i,a)=j}.$$

The proof follows using the Ergodic Theorem for Markov chains and the same steps in the proof of Lemma 4.5 from [9] by replacing the uniform measure with the positive Bernoulli measure induced by μ. We note that for this line of proof uses the notion of *μ-block normality* defined using the block-wise occurrences of a words within an infinite sequence instead of the notion of normality we used in this paper. However, the two notions are equivalent. See section 6 for a full proof. □

Proof of Theorem 7. In part $(i) \Rightarrow (ii)$ of our proof of Theorem 5, we showed that on a given state r, either r is not a betting state (γ is the constant 1 on this state) or it is and then the gambler loses money exponentially fast in the number of times this state is visited, the exponent being $\alpha_r = \sum_{a\in A} \mu(a) \cdot \log \gamma(r, a)$ which we proved to be negative. By Lemma 1, betting states are either visited finitely often or with positive asymptotic density. If they are all visited finitely often, the capital stabilizes after the last bet is made. Otherwise, if betting states r are visited with frequency π_r and at least one π_r is positive, then the gambler loses at an exponential rate, where the exponent is $\sum_r \pi_r \alpha_r$.

In part $(ii) \Rightarrow (i)$, under the assumption of non-μ-normality of X, we built a gambler $\mathcal{G}$ satisfying the following: for any fixed ε, there are arbitrarily large n and k such that $\log \mathrm{Capital}(\mathcal{G}, X[0, n+1]) \geq k\,(\mathrm{D}_{KL}(\nu||\mu) - \varepsilon)$. Here, k in the number of visits to a state q_u. In the proof of Theorem 5 part (ii), this state q_u is visited with frequency $\mu(u)$ for large enough n. Hence the gambler has an 'infinitely often' exponential rate of success with exponent $\mu(u) D_{KL}(\nu||\mu)$. □

3 The Agafonov theorem for PFAs

We now want to prove the extension of Theorem 4 to probabilistic automata/selectors. A *probabilistic finite automaton (PFA)* is a tuple (Q, A, q_I, δ) where Q is a finite set of states, A a finite alphabet, q_I the initial state and $\delta : Q \times A \to \Delta(Q)$ is a probabilistic transition function, that is, $\Delta(Q)$ is the set of probability distributions over Q. In this setting, we define inductively the random variables $\delta^*(q, w)$ by $\delta^*(q, \epsilon) = q$ and for $w \in A^*$ and $a \in A$, the event $\delta^*(q, w \cdot a) = q'$ is defined as the union

$$\bigcup_{r \in Q} \left[\delta^*(q, w) = r \wedge \delta_{|w|+1}(r, a) = q'\right]$$

where $\{\delta_n(r, b) : n \in \mathbb{N}, r \in Q, b \in A\}$ is a family of independent random variables such that for all (n, r, b), the distribution of $\delta_n(r, b)$ is $\delta(r, b)$.

Modulo this change of type of transition, probabilistic selectors are defined as before as well as Select. This makes $\mathrm{Select}(\mathcal{S}, X)$ a random variable for every given $X \in A^{\leq \omega}$.

In order to lift the deterministic Agafonov theorem for Bernoulli measures to the probabilistic case, we will need some preliminary lemmas about normality.

Given two alphabets A and B, and given two sequences $X \in A^\omega$ and $Y \in B^\omega$, we denote by $X \otimes Y$ the sequence Z over the alphabet $A \times B$ where $Z(n) = (X(n), Y(n))$. The product $v \otimes w$ for v and w two words of the same length over A and B respectively is defined in the same way. Likewise, if μ and ν are Bernoulli measures over A and B respectively, $\mu \otimes \nu$ is the Bernoulli measure ξ over $A \times B$ where $\xi((a, b)) = \mu(a)\nu(b)$ for all $(a, b) \in A \times B$. Finally, if Z is a sequence over a product alphabet $A \times B$, we denote by π_0 and π_1 its first and second projection respectively (in other words, $\pi_0(X \otimes Y) = X$ and $\pi_1(X \otimes Y) = Y$ for all X, Y).

Lemma 2. *Let A and B be two alphabets and μ, ν two Bernoulli measures over A and B respectively. If a sequence $X \in A^\omega$ is μ-normal and a sequence $Y \in B^\omega$ is drawn at random according to ν, then ν-almost surely, $X \otimes Y$ is $\mu \otimes \nu$-normal.*

Proof. Let $w = u \otimes v$ be a non-empty word of $(A \times B)^*$. Let N be a large integer. We split $X \otimes Y$ into blocks of length N :

$$X \otimes Y = (X_1 \otimes Y_1) \cdot (X_2 \otimes Y_2) \cdot \ldots$$

with $|X_i| = |Y_i| = N, i \geq 1$. Introduce the random variables $(B_i)_{i \geq 1}$ which count the number of occurrences of $u \otimes v$ in each block:

$$B_i = \mathrm{NbOcc}(u \otimes v, X_i \otimes Y_i), i \geq 1 \ .$$

For every integer $n \in \mathbb{N}$, within $(X \otimes Y)[0 \dots n]$ there are $\lfloor n/N \rfloor$ complete blocks. Some occurrences of $u \otimes v$ in $(X \otimes Y)[0 \dots n]$ do occur inside a block $X_i \otimes Y_i$ while some other do not because they overlap two contiguous blocks. There can be at most $|w|$ such overlapping occurrences between two given blocks. That observation leads to two ways to count occurrences of $u \otimes v$: the exact way C_n and the N-block way $C'_{n,N}$, two random variables satisfying:

$$\begin{aligned} C_n &= \mathrm{NbOcc}(u \otimes v, (X \otimes Y)[0 \dots n]) \\ C'_{n,N} &= \sum_{i \in 1 \dots \lfloor n/N \rfloor} B_i \\ C'_{n,N} &\le C_n \le C'_{n+N,N} + |w| \cdot \lfloor n/N \rfloor \ . \end{aligned} \tag{1}$$

Let us focus on $C'_{n,N}$ first. The variables B_i can be grouped into $|A|^N$ different buckets, with respect to the corresponding value of X_i. For every $x \in A^N$, set $I_n(x) = \{1 \le i \le \lfloor n/N \rfloor \mid X_i = x\}$, which is non-empty for n large enough, since X is normal. The random variables $(X_i \otimes Y_i)$ are mutually independent, and so are the random variables $(B_i)_{i \ge 1}$. Moreover, for a fixed $x \in A^N$, the random variables $(B_i)_{i \in I_\infty(x)}$ are distributed identically, denote by $\xi(x)$ the corresponding probability distribution on $0 \dots N - |w| + 1$. Then $\mathbb{E}[\xi(x)]$ measures the expected number of occurrences of $u \otimes v$ in a sequence where the left part is fixed, equal to x, and the right part is independently generated according to ν. Thus

$$\mathbb{E}[\xi_N(x)] = \mathrm{NbOcc}(u, x) \cdot \nu(v) \ .$$

According to the Law of Large Numbers, $\frac{1}{|I_n(x)|} \sum_{i \in I_n(x)} B_i \to_n \mathrm{NbOcc}(u, x) \cdot \nu(v)$. Since X is normal, every word $x \in A^N$ occurs with frequency $\mu(x)$ in $X_1, X_2, \dots$ thus, almost-surely,

$$\lim_n \frac{1}{n} C'_{n,N} = \left(\sum_{x \in A^N} \mu(x) \cdot \mathrm{NbOcc}(u, x) \right) \cdot \nu(v) \ . \tag{2}$$

Since X is normal, the right part in (2) converges to $\mu(u) \cdot \nu(v)$ when N grows large. Using (1), we get the desired result:

$$\frac{1}{n} C_n \to_n \lim_N \left(\lim_n C'_{n,N} \right) = (\mu \otimes \nu)(u \otimes v) \ .$$

□

Lemma 3. *If a sequence Z is $\mu \otimes \nu$-normal over $A \times B$, then $\pi_0(Z)$ and $\pi_1(Z)$ are μ- and ν-normal respectively.*

Proof. Suppose $Z \in (A \times B)^\omega$ is $\mu \otimes \nu$-normal. We only need to show that $\pi_0(Z)$ is μ-normal, the proof of the ν-normality of $\pi_1(Z)$ being the same by symmetry. Let $w \in A^n$. We have

$$\mathrm{Freq}^+(w, \pi_0(Z)) = \sum_{w' \in B^n} \mathrm{Freq}^+(w \otimes w', Z)$$

which, by $\mu \otimes \nu$-normality of Z, implies

$$\mathrm{Freq}^+(w, \pi_0(Z)) = \sum_{w' \in B^n} \mu \otimes \nu(w \otimes w') = \sum_{w' \in B^n} \mu(w)\nu(w') = \mu(w)$$

The same holds for $\mathrm{Freq}^-(w, \pi_0(Z))$, hence we have proven that $\mathrm{Freq}(w, \pi_0(Z)) = \mu(w)$, which is what we wanted.

□

We are now ready to prove Agafonov's theorem for PFAs.

Theorem 8 (Agafonov's theorem for PFA)**.** *Let $X \in A^\omega$, and μ a Bernoulli measure over A. The following are equivalent.*

(i) X is μ-normal.

(ii) For any probabilistic selector $\mathcal{S}$ that selects a subsequence Y of X, almost surely, either Y is finite, or Y is μ-normal.

(iii) For any probabilistic selector $\mathcal{S}$ that selects a subsequence Y of X, almost surely, either Y is finite, or Y is μ-normal.

Proof. Since deterministic selectors (for which we already have Agafonov's theorem) are a subset of the probabilistic ones, all is left to prove is $(i) \Rightarrow (iii)$.

Let $\mathcal{S} = (Q, A, q_I, \delta, S)$ be a probabilistic selector. Recall that each transition $\delta(q, a)$ (where $q \in Q$ and $a \in A$) is some probability distribution over Q. Consider the set $\mathcal{T}$ of all functions $Q \times A \to Q$. We can put a distribution τ over $\mathcal{T}$ such that if t is chosen according to τ, for every (q, a), the marginal distribution of $t(q, a)$ is $\delta(q, a)$. An easy way to do this is to take $\tau = \bigotimes_{(q,a)\in Q\times A} \delta(q, a)$.

This construction means that, for a fixed sequence $X \in A^\omega$, an equivalent way to simulate the probabilistic run of $\mathcal{S}$ on X is, every time we are in a state q and read a letter a, to pick t at random according to τ and move to state $t(q, a)$. But $\mathcal{T}$ is a finite set and τ a Bernoulli measure over it (this Bernoulli measure might not be positive. If it is not, we simply remove from $\mathcal{T}$ all functions whose τ-probability is 0). Reformulating slightly, yet another equivalent way to simulate the run of $\mathcal{S}$ over X is to do the following:

1. First choose $T \in \mathcal{T}^\omega$ at random with respect to the Bernoulli measure τ.
2. Then run on the sequence $X \otimes T$ the deterministic selector $\hat{\mathcal{S}}$ whose set of states is Q and transition $\hat{\delta}$ is defined by $\hat{\delta}(q, (a, t)) = t(q, a)$.

Now, the two following random variables have the same distribution:

- The subsequence Y of X selected by $\mathcal{S}$
- The sequence $\pi_0(\hat{Y})$, where $\hat{Y}$ is the subsequence of $X \otimes T$ selected by $\hat{\mathcal{S}}$ when T is chosen randomly according to τ.

Since X is μ-normal, by Lemma 2, $X \otimes T$ is $\mu \otimes \tau$-normal τ-almost surely. Thus, by Agafonov's theorem for deterministic selectors and Bernoulli measures (Theorem 4), almost surely, the subsequence $\hat{Y}$ selected by $\hat{\mathcal{S}}$ is $\mu \otimes \tau$-normal. Finally, by Lemma 3, this implies that almost surely, the subsequence $\pi_0(\hat{Y})$ of X is μ-normal. This concludes the proof.

□

4 The Schnorr-Stimm theorem for probabilistic gamblers

Finally, we establish the Schnorr-Stimm theorem for probabilistic finite-state automata. The definition of probabilistic (μ-)gambler is the same as the definition of deterministic (μ-)gambler except that it is based on PFAs instead of DFAs. In this setting, the quantities $\mathrm{Capital}(\mathcal{G}, w)$ become random variables.

Theorem 9 (Schnorr-Stimm theorem for PFAs)**.** *For $X \in A^\omega$ and μ a Bernoulli measure, the following are equivalent.*

(i) X is μ-normal.

(ii) Any probabilistic μ-gambler $\mathcal{G}$ betting on X loses almost surely.

Proof. We already have $(ii) \Rightarrow (i)$ by Theorem 5 (deterministic gamblers being a subset of probabilistic ones).

For $(i) \Rightarrow (ii)$, the idea is similar to the proof of Theorem 8. To simulate the run of a probabilistic μ-gambler $\mathcal{G} = (Q, A, q_I, \delta, \gamma)$ on a sequence X, one can equivalently run on the sequence $X \otimes T$, where $T \in \mathcal{T}^\omega$ is τ-random sequence ($\mathcal{T}$ and τ being defined in the same way as in Theorem 8) the deterministic gambler $\hat{\mathcal{G}} = (Q, A, q_I, \hat{\delta}, \hat{\gamma})$ where again $\hat{\delta}(q,(a,t)) = t(q,a)$ for all $(q,a,t) \in Q \times A \times \mathcal{T}$ and $\hat{\gamma}(q,(a,t)) = \gamma(q,a)$ (indeed the bet placed by a gambler at a given stage does not take into account which state will be reached next). Note that the fairness condition is respected since for every q,

$$\sum_{(a,t)\in A\times\mathcal{T}} (\mu\otimes\tau)(a,t)\cdot\hat{\gamma}(q,(a,t)) = \sum_{a\in A, t\in\mathcal{T}} \mu(a)\tau(t)\gamma(q,a) = \sum_{t\in\mathcal{T}} \tau(t) \sum_{a\in A} \mu(a)\gamma(q,a) = \sum_{t\in\mathcal{T}} \tau(t) = 1$$

(the second-to-last equality holds by fairness condition on γ and the last one because τ is a distribution).

In this way, the two following random variables will have the same distribution:

- Capital$(\mathcal{G}, X[0..n])$
- Capital$(\hat{\mathcal{G}}, (X \otimes T)[0..n])$ where $T \in \mathcal{T}^\omega$ is chosen at random according to τ.

Since X is μ-normal, by Lemma 2 again, $X \otimes T$ is $\mu \otimes \tau$-normal τ-almost surely. Thus, by Theorem 5, for τ-almost all T, Capital$(\hat{\mathcal{G}}, (X \otimes T)[0..n])$ is bounded by a constant C independent on n. By the equivalence of the two random variables above, this means that almost surely, there exists a constant C such that Capital$(\mathcal{G}, X[0..n]) < C$ for all n. In other words, almost surely, $\mathcal{G}$ loses on the sequence X.

□

We remark that the dichotomy theorem for Bernoulli measures yields the following dichotomy for probabilistic μ-gamblers.

Theorem 10 (Schnorr-Stimm dichotomy theorem for PFAs)**.** *Let X be an infinite sequence in A^ω and μ a Bernoulli measure.*
(i) If X is μ-normal and $\mathcal{G}$ is a μ-gambler, then almost surely the capital of $\mathcal{G}$ throughout the game either is ultimately constant or decreases at an exponential rate.
(ii) If X is not μ-normal, then there exists a μ-gambler $\mathcal{G}$ which wins against X at an 'infinitely often' exponential rate almost surely.

5 Alternate proof of Agafonov's theorem for PFA

For the curious reader, we provide another proof of Theorem 8, along the lines of Carton's techniques [3], with a different presentation based on the notion of *balanced input words*.

5.1 The strongly connected case

In the whole section we assume that

$$\mathcal{A} \text{ is strongly connected}$$

i.e. for all states $q, r \in Q$, there is a computation in $\mathcal{A}$ from q to r. We also assume that $\mathcal{A}$ has at least one selecting state i.e. $S \neq \emptyset$. We prove

Lemma 4 (The strongly connected probabilistic case)**.** *Let $\mathcal{A}$ be a PFA over an alphabet A, μ a Bernoulli measure over A and assume that the Markov chain $\mathcal{A}^\mu$ is irreducible. Let X be a μ-normal sequence. The word output by $\mathcal{A}$ is almost-surely finite or μ-normal.*

The proof of Lemma 4, given at the end of the section, relies on two ingredients:

- *balanced input words* (cf. Definition 3), do appear frequently when the input word of the PFA is generated with μ (cf. Lemma 5);
- and a generic result about Markov processes (cf. Lemma 6).

The notion of balanced input words isolates a frequent behaviour of the Markov chain $\mathcal{M}$ with state space $Q \cup Q \times A$, obtained by synchronizing the μ-generator of letters and the probabilistic automaton reading those letters: from a state $q \in Q$, a new letter $a \in A$ is selected according to μ and the new state is (q, a). Then a transition to Q occurs with probabilities $p_{\mathcal{A}}$. This homogeneous Markov process is fully described by the random variables $Q_0, A_1, Q_1, A_2, \ldots$. The computation at step k is

$$\pi_k = Q_0, A_1, Q_1, A_2, \ldots, Q_k \quad .$$

For $1 \leq i \leq j$, the output of the chain between dates i and j, both included, is the random variable $\mathrm{Out}_{i \ldots j}$ with values in A^* defined by:

$$\begin{aligned}\mathrm{Out}_{i\ldots j} = \ & \text{the subword of } A_i \ldots A_j \\ & \qquad \text{obtained by selecting letters at all dates } k \text{ such that } i \leq k \leq j \text{ and } Q_{k-1} \in S \quad .\end{aligned}$$

When running $\mathcal{M}$, the output in A^ω is generated according to μ, at an average output rate $0 < \lambda < 1$. The ergodic theorem describes what happens almost-surely in $\mathcal{M}$. Since $\mathcal{A}$ is strongly connected, every state of $\mathcal{M}$ belongs to the same recurrence class, and is visited at a fixed positive frequency. Every time one of the selecting states is visited, say $Q_k \in S$ at step k, then the letter A_{k+1} is output, it is picked up at random, independently of the rest of the computation π_k, including Q_k. If one fixes some output length $m \in \mathbb{N}$, all words in A^m are expected to appear in the output at the same frequency $\frac{1}{|A|^m}$. The output rate λ is equal to the frequency of visits to S: every N input letters, the Markov chain produces λN output letters on average. In the long run, the chain almost-surely generates an infinite output

$$\mathrm{Out}_{1,\infty} = \lim_N \mathrm{Out}_{1\ldots N} \in A^\omega \quad .$$

according to distribution μ.

When the automaton $\mathcal{A}$ runs on a fixed finite input word $z = a_1, a_2, \ldots, a_N$ then the output is in general not generated according to the distribution μ: due to the deterministic constraint on the input word, the corresponding Markov chain is not homogeneous anymore since the letters are picked deterministically, on a time-dependent basis. However, we show that if the input word is

μ-normal, the computation of $\mathcal{A}$ still resembles the computation of $\mathcal{M}$, at least when it comes to counting the occurrences of a (small) word $w \in A^m$ in the output.

We denote, for every non-empty word $w \in A^+$, and non-empty integer interval $1 \leq i \leq j$:

$$\begin{aligned}\mathrm{NbOcc}_{i\ldots j}(w) = \; & \text{the number of indices } k \in i\ldots j \\ & \qquad\qquad\qquad \text{such that } Q_{k-1} \in S \text{ and } w \text{ is a prefix of } \mathrm{Out}_{k\ldots j} \enspace .\end{aligned}$$

As already discussed, in $\mathcal{M}$ the frequency of appearance of some word w in the output is typically equal to $\mu(w)$. This is also approximately the case in $\mathcal{A}$, provided the input is *balanced.*

Definition 3 ((m, ϵ)-balanced input words)**.** *Let $m \in \mathbb{N}$ and $\epsilon > 0$. Denote λ the frequency of visits to F in $\mathcal{M}$. For every word $w \in A^+$, and $1 \leq i < j$ we define the event*

$$\mathrm{Balanced}_{i\ldots j}(w, \epsilon) = \left\{ \left| \frac{\mathrm{NbOcc}_{i\ldots j}(w)}{|\,\mathrm{Out}_{i\ldots j}\,|} - \mu(w) \right| \leq \epsilon \right\} \wedge \left\{ \left| \frac{|\,\mathrm{Out}_{i\ldots j}\,|}{j-i} - \lambda \right| \leq \epsilon \right\} \enspace ,$$

and for every $m \in \mathbb{N}$,

$$\mathrm{Balanced}_{i\ldots j}(m, \epsilon) = \bigwedge_{w \in A^m} \mathrm{Balanced}_{i\ldots j}(w, \epsilon) \enspace .$$

A finite word $a_1 \ldots a_N$ is said to be (m, ϵ)-balanced if for every initial state $q_0 \in Q$,

$$\mathbb{P}_{\mathcal{M}} \left(\mathrm{Balanced}_{1\ldots N}(m, \epsilon) \mid A_1 = a_1, \ldots, A_N = a_N, Q_0 = q_0 \right) \geq 1 - \epsilon \enspace . \tag{3}$$

In the specific case where $\mathcal{A}$ is deterministic, then there is a single possible computation on $a_1 \ldots, a_N$, hence the probability in (3) is either 0 or 1. In that case, provided $\epsilon < 1$ the property $\mathrm{Balanced}_{1\ldots N}(m, \epsilon)$ is guaranteed to hold for sure whenever the word $a_1 \ldots, a_N$ is balanced.

The following lemma shows that among input words that are long enough, the proportion of (m, ϵ)-balanced words is arbitrarily close to 1.

Lemma 5. *Let $m \in \mathbb{N}$ and $\epsilon > 0$. For every state $q_0 \in Q$,*

$$\mathbb{P}_{\mathcal{M}, q_0} \left(A_1 \ldots A_N \textit{ is } (m, \epsilon)\textit{-balanced } \right) \to_N 1 \enspace .$$

Proof. In this proof, $\mathbb{P}_{\mathcal{M}, q_0}$ is simply denoted as $\mathbb{P}$. We make use of

$$\mathrm{Balanced}(m, \epsilon, \infty) = \bigcup_{M \in \mathbb{N}} \bigcap_{M \leq N} \mathrm{Balanced}(m, \epsilon, N) \enspace .$$

By standard ergodic properties of Markov chains, $\mathbb{P}\left(\mathrm{Balanced}(m, \epsilon, \infty)\right) = 1$. The event $\mathrm{Balanced}(m, \epsilon, \infty)$ is a tail event, thus according to Lévy's law, $\mathbb{P}\left(\mathbb{P}\left(\mathrm{Balanced}(m, \epsilon, \infty) \mid A_1, \ldots, A_N\right) \to_N 1\right) = 1$. Choose any ϵ' such that $0 < \epsilon' < \epsilon$. Then, for N large enough,

$$\mathbb{P}\left(\mathbb{P}\left(\mathrm{Balanced}(m, \epsilon, N) \mid A_1, \ldots, A_N\right) \geq 1 - \epsilon\right) \geq 1 - \epsilon'.$$

Thus $\mathbb{P}_{q_0}\left(A_1 \ldots A_N \text{ is } (m, \epsilon)\text{-balanced }\right) \geq 1 - \epsilon'$. This holds for N large enough whatever ϵ', hence the conclusion. □

The second ingredient in the proof of Lemma 4 is a generic property of Markov processes:

Lemma 6. *Let $X_1, X_2, \dots$ be a Markov process, $1 = i_1 < i_2 < \dots$ a sequence of indices and $B_1, B_2 \dots$ a sequence of events and $\epsilon > 0$ such that for every $\ell \geq 1$,*

$$B_\ell \text{ is } (X_{i_\ell}, \dots, X_{i_{\ell+1}-1}) - \text{measurable} \tag{4}$$

$$\forall x \in X, \mathbb{P}(B_\ell \mid X_{i_\ell} = x) \geq (1 - \epsilon) \ . \tag{5}$$

For every $k \in \mathbb{N}$, denote $N_k = \sum_{\ell \in 1\dots k} 1_{B_\ell}$. Then

$$\mathbb{P}\left(\frac{N_k}{k} > (1 - 2\epsilon)\right) \to_k 1 \ .$$

Proof. For every $k \in \mathbb{N}$ set

$$G_k = N_k - (1 - \epsilon) \cdot k \ .$$

and let $\mathcal{F}_k$ the σ-algebra generated by $X_1, \dots, X_{i_{k+1}-1}$. Then $(G_k)_{k\in\mathbb{N}}$ is a submartingale for the filtration $(\mathcal{F}_k)_{k\in\mathbb{N}}$:

$$\begin{aligned}\mathbb{E}[G_{k+1} \mid \mathcal{F}_k] &= \mathbb{E}[N_{k+1} \mid \mathcal{F}_k] - (1-\epsilon) \cdot (k+1) \\ &\geq (1-\epsilon)(N_k + 1) + \epsilon N_k - (1-\epsilon) \cdot (k+1) \\ &= M_k\end{aligned}$$

Notice that $\forall k \in \mathbb{N}, |G_{k+1} - G_k| \leq \max\{\epsilon, (1-\epsilon)\}$ and by hypothesis $\epsilon \leq 1$ hence $|G_{k+1} - G_k| \leq 1$. We make use of:

Theorem 11 (Azuma's Inequality for Submartingales)**.** *Let $(X_n)_{n\geq 0}$ be a submartingale with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$, i.e., $\mathbb{E}[X_{i+1} \mid \mathcal{F}_i] \geq X_i$ for all $i \geq 0$. Assume there exist constants $c_1, c_2, \dots, c_n$ such that $|X_i - X_{i-1}| \leq c_i$ almost-surely for all $i \geq 1$. Then, for any $t > 0$, the following bound holds:*

$$\mathbb{P}(X_n - X_0 \leq -t) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right) .$$

Apply Azuma's inequality with $c_k = (1 - \epsilon))$ and $t = -k\epsilon$ and get

$$\mathbb{P}(N_k \leq (1 - 2\epsilon) \cdot k) = \mathbb{P}(G_k \leq -k\epsilon) \leq e^{-\frac{k^2\epsilon^2}{2k}} = e^{-k\frac{\epsilon^2}{2}} \to_k 0 \ .$$

□

Proof of Lemma 4. Take some normal input word z. Fix a word length $m > 0$ and a precision $\epsilon > 0$ and compute the corresponding N from Lemma 5. Cut z in chunks of length N, so that $z = u_1 u_2 \cdots$ where each $|u_i| = N$.

For every $k \geq 1$, we consider the concatenation of the first k chunks:

$$z_k = u_1 u_2 \cdots u_k \ ,$$

and we shall approximate the expected number of occurrences of $w \in A^m$ in the corresponding output $\text{Out}_{1\dots N\cdot k}$ on z_k, when k is large enough, as well as the length of this output.

Let us enumerate all indices where u_ℓ is normal as

$$\{\ell \geq 0 \mid u_\ell \text{ is normal}\} = \{i_0 < i_1 < i_2 \dots\} \ .$$

By choice of N, among all possible words in A^N, at most an ϵ-fraction are not (m,ϵ)-balanced. Since z is normal, in the sequence $u_1, u_2, \dots$ every possible word in A^N occurs with the same limit frequency $\frac{1}{|A|^N}$. Thus $\limsup_k \frac{i_k - k}{i_k} \leq \epsilon$ hence

$$\exists k_0, \forall k \geq k_0,\ (1-2\epsilon) i_k \leq k \leq i_k \tag{6}$$

For every $\ell \geq 0$, denote I_ℓ the interval of integers of length N defined by

$$I_\ell = 1 + (N \cdot i_\ell) \dots (N+1) \cdot i_\ell$$

and let us evaluate balancedness at step k as,

$$M_k = \mid \{\ell \in 0 \dots k \mid \text{Balanced}_{I_\ell}(m,\epsilon)\} \mid \quad .$$

By definition of balancedness, for every index $1 \leq \ell \leq k$,

$$\mathbb{P}(\text{Balanced}_{I_\ell}(m,\epsilon) \mid Q_0, A_1, Q_1, \dots, Q_{i_\ell \cdot N}) \geq 1 - \epsilon \quad .$$

According to Lemma 6:

$$\mathbb{P}\left(\frac{M_k}{k} \geq (1-2\epsilon)\right) \rightarrow_k 1 \quad ,$$

which with (6) implies:

$$\mathbb{P}\left(\frac{|\{\ell \in 0 \dots k \mid \text{Balanced}_{\ell * N \dots (\ell+1)*N-1}(m,\epsilon)\}|}{k} \geq (1-2\epsilon)^2\right) \rightarrow_k 1 \quad .$$

By definition of balancedness, for k large enough, for every $w \in A^m$,

$$\mathbb{P}_{\mathcal{A},z}\left(\text{NbOcc}_{1 \dots k \cdot N}(w) \geq (1-2\epsilon)^2 \left(\mu(w) - \epsilon\right) \cdot \lambda \cdot k \cdot N\right) \geq 1 - \epsilon \tag{7}$$

$$\mathbb{P}_{\mathcal{A},z}\left(\mid \text{Out}_{1 \dots k \cdot N} \mid \geq (1-2\epsilon)^2 \cdot (\lambda - \epsilon) \cdot k \cdot N\right) \geq 1 - \epsilon \quad . \tag{8}$$

And on every non balanced words the output size is at most N, hence

$$\mathbb{P}_{\mathcal{A},z}\left(\mid \text{Out}_{1 \dots k \cdot N} \mid \leq (1-2\epsilon)^2 \cdot (\lambda - \epsilon) \cdot k \cdot N + (1 - (1-2\epsilon)^2) \cdot k \cdot N\right) \tag{9}$$

Combining (7) – (9), for ϵ arbitrarily small (hence k arbitrarily large) we obtain:

$$\mathbb{P}_{\mathcal{A},z}\left(\frac{\mid \text{Out}_{1 \dots k \cdot N} \mid}{k \cdot N} \rightarrow_k \lambda\right) = 1 \tag{10}$$

$$\mathbb{P}_{\mathcal{A},z}\left(\liminf_k \frac{\text{NbOcc}_{1 \dots k \cdot N}(w)}{\mid \text{Out}_{1 \dots k \cdot N} \mid} \geq \mu(w)\right) = 1 \quad . \tag{11}$$

Those two limits are taken over $k \cdot N \in \{N, 2N, 3N, \dots\}$ but since $\frac{N}{k}$ is negligible for large k, the stronger limits over $m \in \mathbb{N}$ do hold as well:

$$\mathbb{P}_{\mathcal{A},z}\left(\liminf_n \frac{\text{NbOcc}_{1 \dots n}(w)}{\mid \text{Out}_{1 \dots n} \mid} \geq \mu(w)\right) = 1 \quad . \tag{12}$$

Now we are done, since $\mathbb{P}_{\mathcal{A},z}$-almost-surely,

$$\begin{aligned}\limsup_n \frac{\sum_{w\in A^k} \mathrm{NbOcc}_{1\dots n}(w)}{|\,\mathrm{Out}_{1\dots n}\,|} &\leq 1 && \text{(by def. of NbOcc and Out)}\\ &= \sum_{w\in A^k} \mu(w)\\ &\leq \sum_{w\in A^k} \liminf_n \frac{\mathrm{NbOcc}_{1\dots n}(w)}{|\,\mathrm{Out}_{1\dots n}\,|} && \text{(by (12))}\\ &\leq \liminf_n \sum_{w\in A^k} \frac{\mathrm{NbOcc}_{1\dots n}(w)}{|\,\mathrm{Out}_{1\dots n}\,|}\end{aligned}$$

hence those three inequalities are $\mathbb{P}_{\mathcal{A},z}$-almost-surely equalities. As a consequence

$$\mathbb{P}_{\mathcal{A},z}\left(\frac{\mathrm{NbOcc}_{1\dots n}(w)}{|\,\mathrm{Out}_{1\dots n}\,|} \to_n \mu(w)\right) = 1 \ . \tag{13}$$

This holds for every $w \in A^+$, hence the output is a.s. normal. □

5.2 The general case

Alternate proof of Theorem 8. We extend Lemma 4 to the general case where the automaton has a non empty set of transient states T. By definition from every state in T there is a word and a computation on this word which leaves T and enter a BSCC, from which there is no way back to T. We show

$$\mathbb{P}_{\mathcal{A},z}(\exists n, Q_n \notin T) = 1 \ .$$

From every state $q \in T$, there is at least one word of length $\leq |T|$ which exits T with nonzero probability, bounded from below by ${p_m}^{|T|}$ where p_m is the smallest non-zero probability appearing in p. We can concatenate $|T|$ such words in order to obtain a single word u which guarantees probability $\geq {p_m}^{|T|^2}$ to leave T, *whatever the initial state*. Since the input word z is normal, u appears infinitely often in z, thus T is eventually left almost-surely.

Once T is left, the suffix of the computation fits the hypotheses of Lemma 4. Normality is a tail property, hence any suffix of the input word is normal as well, hence almost-surely a suffix of the output word is normal as well, according to Lemma 4, hence the output is normal, since normality is a tail property. □

6 μ-block normality and μ-normality

The proof of Lemma 1 requires the block-wise characterization of μ-normality. Let $\mathrm{BFreq}(u, w)$ denote the frequency of block-wise occurrences of the word u in w. For $w \in A^*$, let $\mathrm{BFreq}^-(w, X)$, $\mathrm{BFreq}^+(w, X)$ and $\mathrm{BFreq}(w, X)$ denote the lower, upper and limit block frequency of w in X defined similarly as $\mathrm{Freq}^-(w, X)$, $\mathrm{Freq}^+(w, X)$ and $\mathrm{Freq}(w, X)$. A sequence X is *μ-block normal* if for all words $w \in A^*$, $\mathrm{BFreq}(w, X) = \mu(w)$. The following was shown in Simonsen and Seiller.

Lemma 7 ([10]). *If a sequence $X \in A^\omega$ is μ-normal then X is μ-block normal.*

We note that the converse implication also holds.

Lemma 8. *If a sequence $X \in A^\omega$ is μ-block normal then X is μ-normal.*

The above lemmas completes the proof of the following equivalence theorem between μ-normality and μ-block normality.

Theorem 12. *A sequence $X \in A^\omega$ is μ-normal if and only if X is μ-block normal.*

The proof of Lemma 8 uses a counting trick from the proof of Theorem 3.1 from [8] which in turn is based on the proof of the main theorem in [7].

Proof of Lemma 8. As in the proof of Theorem 3.1 from [8], for any finite length string $w = a_1a_2a_3\dots a_k \in A^k$ and large enough n,

$$\mathrm{Freq}(w, X[0..n]) = f_1(n) + f_2(n) + \dots + f_{(1+\lfloor \log_2 \frac{n}{k} \rfloor)}(n) + \frac{(k-1).O(\log n)}{n-k+1} \tag{14}$$

where $f_p(n)$ are defined as follows:

$$f_p(n) = \begin{cases} \frac{|\{i \ | \ X[ki,k(i+1)-1]=w \ , \ 0 \ \leq \ i \ \leq \ \lfloor n/k \rfloor\}|}{n-k+1}, & \text{if } \ p=1 \\ \sum_{j=1}^{k-1} \frac{|\{i|X[2^{p-1}ki,2^{p-1}k(i+1)-1] \in S_j, 0 \leq i \leq n/2^{p-1}k\}|}{n-k+1}, & \text{if } 1 < p \leq (1+\lfloor \log_2(n/k) \rfloor) \\ 0, \ \text{otherwise.} \end{cases}$$

In the above definition, S_j is the set of strings of the form, $u \ a_1a_2\dots a_k \ v$ where u is some string of length $2^{p-2}k - j$ and v is some string of length $2^{p-2}k - k + j$. Since X is μ-block normal,

$$\lim_{n\to\infty} f_1(n) = \lim_{n\to\infty} \frac{|\{i \ | \ X[ki,k(i+1)-1] = w \ , \ 0 \ \leq \ i \ \leq \ \lfloor n/k \rfloor\}|}{n-k+1} = \frac{\mu(w)}{k}.$$

Now, when $p \leq 1 + \lfloor \log_2(\frac{n}{k}) \rfloor$,

$$\lim_{n\to\infty} f_p(n) = \sum_{j=1}^{k-1} \lim_{n\to\infty} \frac{|\{i \mid X[2^{p-1}ki, 2^{p-1}k(i+1)-1] \in S_j, 0 \leq i \leq n/2^{p-1}k\}|}{n-k+1} = \frac{\mu(w)}{2^{p-1}k}.(k-1)$$

Since $\langle \sum_{i=1}^{m} f_i(n) \rangle_{m\in\mathbb{N}}$ is uniformly convergent, we have

$$\mathrm{Freq}(w, X[0..n]) = \lim_{n\to\infty} \sum_{i=1}^{\infty} f_i(n) = \sum_{i=1}^{\infty} \lim_{n\to\infty} f_i(n).$$

Therefore from (14),

$$\lim_{n\to\infty} \mathrm{Freq}(w, X[0..n]) = \frac{\mu(w)}{k} + (k-1)\mu(w)\Big[\sum_{i=1}^{\infty} \frac{1}{2^i k}\Big] = \mu(w).$$

Hence, X is μ-normal. □

7 Conclusion

The main contributions of this paper are a generalization of the Agafonov theorem for PFA with arbitrary transition probabilities which settles the open question posed by Léchine et al. [6], and an extension of Schnorr-Stimm theorem to probabilistic gamblers.

While we proved the probabilistic Agafonov theorem (Theorem 8) by reduction to the deterministic setting, it is also possible to follow with a more direct approach (i.e., without appealing to the Seiller-Simonsen result), similar to the one followed by Carton to generalize Agafonov's theorem for DFA [3]. This however makes the argument somewhat more complicated. For the curious reader, we provide this alternate proof in section 5.

An interesting direction for future research is to explore whether the 'uselessness of randomness' also holds for pushdown automata. These are a more powerful model of computation and indeed some normal sequences can be predicted by pushdown automata (some in a rather dramatic way, as proven by Carton and Perifel [4][1]). We can for example ask: If some probabilistic pushdown selector selects a biased subsequence from a sequence X, does there necessarily exist a deterministic pushdown selector which also selects a biased subsequence?. Similarly, if some probabilistic pushdown gambler wins against a sequence X, does there necessarily exist a deterministic pushdown gambler which wins against that same sequence X?

A related question concerns the speed of success in the gambling model. In the case of finite-state automata, Schnorr and Stimm proved that either a sequence X cannot be predicted or some gambler wins on it at an exponential rate. This dichotomy no longer holds for pushdown automata, but one may ask the following question: *If some probabilistic pushdown gambler wins at an exponential rate against a sequence X, does there necessarily exist a pushdown gambler which wins against that sequence X at an exponential rate?*

References

[1] V. N. Agafonov. Normal sequences and finite automata. *Soviet Mathematics Doklady*, 9:324–325, 1968.

[2] Laurent Bienvenu, Valentino Delle Rose, and Tomasz Steifer. Probabilistic vs deterministic gamblers. In *39th International Symposium on Theoretical Aspects of Computer Science, STACS 2022*, volume 219 of *LIPIcs*, pages 11:1–11:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[3] Olivier Carton. A direct proof of Agafonov's theorem and an extension to shift of finite type. *CoRR*, abs/2005.00255, 2020.

[4] Olivier Carton and Sylvain Perifel. Deterministic pushdown automata can compress some normal sequences. *Logical Methods in Computer Science*, Volume 20, Issue 3, Aug 2024.

[5] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley, 2nd edition edition, 2006.

[6] Ulysse Léchine, Thomas Seiller, and Jakob Grue Simonsen. Agafonov's theorem for probabilistic selectors. In *49th International Symposium on Mathematical Foundations of Computer Science, MFCS 2024*, volume 306 of *LIPIcs*, pages 67:1–67:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024.

[7] John E. Maxfield. A short proof of Pillai's theorem on normal numbers. *Pacific Journal of Mathematics*, 2(1):23–24, 1952.

[8] Satyadev Nandakumar, Subin Pulari, Prateek Vishnoi, and Gopal Viswanathan. An analogue of pillai's theorem for continued fraction normality and an application to subsequences. *Bulletin of the London Mathematical Society*, 53(5):1414–1428, 2021.

[1] The result proven by Carton and Perifel considers a slightly different paradigm, namely compression (which we did not discuss in this paper) instead of prediction.

[9] Claus Schnorr and Hermann Stimm. Endliche Automaten und Zufallsfolgen. *Acta Informatica*, 1(4):345–359, 1972.

[10] Thomas Seiller and Jakob Grue Simonsen. Agafonov's theorem for finite and infinite alphabets and probability distributions different from equidistribution. *CoRR*, abs/2011.08552, 2020.