Ergod. Th. & Dynam. Sys., (2025), **45**, 3506–3539 © The Author(s), 2025. Published by Cambridge University Press.
doi:10.1017/etds.2025.19

Agafonov's theorem for finite and infinite alphabets and probability distributions different from equidistribution

THOMAS SEILLER† and JAKOB GRUE SIMONSEN‡

† *CNRS, LIPN – UMR 7030 CNRS & Université Sorbonne Paris Nord, Villetaneuse 93439, France*
(e-mail: thomas.seiller@cnrs.fr)

‡ *IT University of Copenhagen, Rued Langgaards Vej 7, DK-2300 Copenhagen S, Denmark*
(e-mail: jags@itu.dk)

(*Received* 23 *June* 2023 *and accepted in revised form* 18 *February* 2025)

Abstract. An infinite sequence α over an alphabet Σ is μ-distributed with respect to a probability map μ if, for every finite string w, the limiting frequency of w in α exists and equals $\mu(w)$. We prove the following result for any finite or countably infinite alphabet Σ: every finite-state selector over Σ selects a μ-distributed sequence from every μ-distributed sequence *if and only if* μ is induced by a Bernoulli distribution on Σ, that is, a probability distribution on the alphabet extended to words by taking the product. The primary—and remarkable—consequence of our main result is a complete characterization of the set of probability maps, on finite and infinite alphabets, for which finite-state selection preserves μ-distributedness. As a consequence, the shift-invariant measures μ on Σ^ω, such that any finite-state selector preserves the property of genericity for μ, are exactly the positive Bernoulli measures.

Key words: equidistribution, symbolic dynamical systems, Agafonov's theorem
2020 Mathematics Subject Classification: 70H08 (Primary); 37J40 (Secondary)

1. *Introduction*

Let $\alpha = x_1x_2\cdots$ be an infinite sequence over a finite alphabet Σ. A string $w \in \Sigma^*$ is said to occur in α with limiting frequency f if $\lim_{N\to\infty}(\#_w(x_1\cdots x_N)/N) = f$, where $\#_w(x_1\cdots x_N)$ is the number of times that w occurs as a contiguous subsequence in $x_1\cdots x_N$. Here, α is said to be *normal* if every finite string of length n over Σ occurs with limiting frequency $|\Sigma|^{-n}$ in α [11]. By standard results, the fractional part of the base-b expansion of almost all real numbers is a normal sequence for $b \geq 2$, so for base 10, almost all real numbers have the digit '0' occurring 1-in-10 times in all sufficiently long finite prefixes of their digit expansion, have '11' occurring 1-in-100 times, '110' occurring 1-in-1000

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

times, and so on. Concrete examples of normal sequences include Champernowne's sequence $1234567891011\cdots$ [17], the Copeland–Erdös sequence $235711131719\cdots$ consisting of concatenating the prime numbers [21], and for any polynomial f with positive integer coefficients the sequence $f(1)f(2)f(3)\cdots$ [22].

A *finite-state selector* is a deterministic finite automaton (DFA) that selects those symbols x_m from α such that $x_1\cdots x_{m-1}$ is accepted by the DFA. The sequence of selected symbols may thus be finite or infinite. *Agafonov's theorem* states that a sequence α is normal if and only if any DFA that selects an infinite sequence from α, selects a normal sequence. Colloquially, Agafonov's theorem can be stated as: 'any constant-space algorithm must preserve normality'.

The purpose of this paper is twofold: (i) we study whether analogues of Agafonov's theorem hold if the distribution of finite strings is *different* from equidistribution, that is, whether distributions where finite strings s are allowed to occur with frequency distinct from $|\Sigma|^{-|w|}$; and (ii) we study extensions of Agafonov's theorem to infinite alphabets (which in the traditional setup in Agafonov's theorem is meaningless as there is no equidistributed probability distribution on a countably infinite set).

As an example, consider the (non-normal) sequence $\alpha = 010101\cdots$. Clearly, every finite bit string occurs in α with some well-defined frequency (the simplest way to see this is that for each $n > 0$, there are exactly two distinct substrings of length n in α: one starting with 0 and one starting with 1), and the frequencies thus induce a probability distribution on $\{0, 1\}^n$ for each n. In particular, 0 and 1 each occur with limiting frequency $1/2$, but any DFA that selects symbols at even positions will select the sequence $111\cdots$, and thus the probability distribution on $\{0, 1\}$ is *not* preserved, showing that Agafonov's theorem in general fails to hold.

The motivation for studying infinite alphabets is that the study of normality is closely tied to the study of symbolic dynamics and (information-theoretic) coding theory [8, 9, 30, 31], and that both areas have witnessed recent advances using infinite alphabets [10, 12, 26, 33, 49]; in particular, the techniques of Madritsch and Mance [33] have allowed construction of Champernowne-like sequences for various distributions over infinite alphabets.

1.1. *Contribution.* The formal statement of the main theorem can be found in Theorem 3.1 below. In plain language, we prove the following.

Let Σ be a non-empty finite or countably infinite alphabet, and let $\mu : \Sigma^* \longrightarrow [0, 1]$ be a probability map (that is, for all $n \geq 0$, $1 = \sum_{w\in\Sigma^n} \mu(w)$) such that there exists at least one $\alpha \in \Sigma^\omega$ that is μ-distributed. Then, the following are equivalent.

(1) *μ is induced by a positive Bernoulli probability distribution p on Σ, that is, for every $a_1, \ldots, a_n \in \Sigma$, $\mu(a_1\cdots a_n) = \prod_{i=1}^n p(a)$, and for every $a \in \Sigma$, $p(a) > 0$.*

(2) *For every DFA A over Σ and every μ-distributed sequence $\alpha \in \Sigma^\omega$, if A selects an infinite sequence from α, then the selected sequence is μ-distributed.*

The above result completely characterizes the probability maps preserved by selection by DFAs, both for finite and infinite alphabets, and Agafonov's theorem follows immediately as a corollary. We briefly review the roadmap and techniques used for the proof of the main result in §1.3.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

As the study of distributions associated to limiting frequencies of finite strings in (right-)infinite strings is cryptomorphic to the study of shift-invariant probability measures on the shift space (Σ^ω, s) equipped with the σ-algebra induced by the basis of cylinder sets on Σ, we obtain a complete characterization of the shift-invariant probability measures ν for which any finite-state selector preserves genericity for ν, see §6.

1.2. *Related work.* Agafonov's theorem [57] was one of the end results of multiple efforts grappling with the two notions of: (i) *kollektiv*—roughly, $\alpha \in \{0, 1\}^\omega$ is a kollektiv with respect to a set $\mathcal{S}$ of selection strategies if the limiting frequency of 1 is unchanged after applying any strategy in $\mathcal{S}$ to α (we note that the exact definition of kollektiv differs subtly across different authors, compare e.g., [18, 40, 54]; the original notion of kollektiv introduced by von Mises [54] had no constraints on the set $\mathcal{S}$, but this turned out to be essentially fruitless [20, 28, 41, 50]); and (ii) *admissible sequence* and its relation to the notion of normal sequence [19, 41, 42, 59, 60]. Agafonov's theorem itself had a virtually unknown precursor in a beautiful result by Postnikova [40] that showed, with the terminology of the present paper, that $\alpha \in \{0, 1\}^\omega$ is normal if and only if the distribution of 1s is preserved by selection strategies depending only on a finite word (see Definition 2.5 for the formal definition of *Postnikova strategies*).

Both Postnikova [40] and Agafonov [58] considered selection functions on sequences in $\{0, 1\}^\omega$, where the limiting distribution of 1 was $0 < p < 1$ (that is, considered a Bernoulli distribution on $\{0, 1\}$), but considering Bernoulli distributions instead of the special case of equidistributions seems to have disappeared almost completely from all later work. One possible reason for this is that only the short version (without proofs or explanation of techniques) of Agafonov's result [57] appeared in English as [1]; in contrast, the original longer paper in Russian [58] was published in a more obscure journal and was never translated. We have provided a (very) embellished account of the arguments in [58] on the preprint server *arXiv* (https://arxiv.org/abs/2007.03249), where we expand Agafonov's terse use of existing results of the time in much more detail, including using more basic arguments with modern methods (e.g., using concentration bounds directly instead of appealing to the law of large numbers) and add further embellishments to Agafonov's original arguments. Many of the results in the present paper exist due to insights obtained due to this embellishment, rather than the original proof itself.

For equidistribution, the earliest extension to arbitrary alphabets seems to be by Broglio and Liardet [13], and a number of authors have since re-proved Agafonov's theorem in the special case of equidistribution using a variety of methods; for example, using predictors defined from finite automata (for $\Sigma = \{0, 1\}$) [38], using compressibility arguments [5, 6, 46], and a combination of automata-theoretic and probabilistic methods similar to Agafonov's original reasoning [14].

Agafonov's theorem itself has been generalized to treat selectors that are not necessarily (induced by prefix selection by) finite automata [2, 5, 15, 53], and some generalizations consider selectors based on relaxed finiteness criteria of the syntactic monoid of a language selecting prefixes of infinite sequences [27, 50]; conversely, results by Merkle and Reimann show that adding just slight computational power to the selection strategies beyond finite automata—e.g., using a Pushdown automaton with unary stack alphabet instead of a

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

DFA renders Agafonov's theorem invalid [35]. (In fact, one of the strategies considered by Merkle and Riemann, which consists in computing the language $\{ww^{\mathrm{R}} \mid w \in \Sigma^*\}$, where w^{R} is the *reverse* of w, can be computed by an arguably less expressive model of computation, namely two-way automata with two heads [25].) Similarly, selection by finite automata has been extended, and analogues for Agafonov's theorem been proved, in other settings than selection from elements of the set Σ^ω, e.g., for shifts of finite type [14]. All of these results only consider normality rather than more general classes of distributions on finite strings.

Conversely, *construction* of normal sequences (as opposed to selecting normal sequences from other normal ones) has been investigated thoroughly for more than a hundred years [17, 34, 36, 39, 48, 51], including explicit construction of real numbers with normal expansion for any integer base $b \geq 2$ [4, 29, 43] and real numbers with normal expansion in non-integer bases [32, 52]. Among this work, the result of most use to the present paper is the construction by Madritsch and Mance of generic sequences for any shift-invariant probability measure μ [33]—these are essentially sequences that are μ-distributed using the terminology of the present paper (see Definition 2.4).

In very recent work, Carton [14] proves that, for any Markov measure μ on Σ^ω induced by a pair (P, π) of a stochastic $|\Sigma| \times |\Sigma|$ matrix and a stationary distribution π for P, any sequence selected from a μ-distributed sequence by a finite-state selector from a particular subset of μ-*compatible* selectors will be μ-distributed. Roughly, a finite-state selector is compatible if it can only read consecutive symbols of Σ with non-zero transition probability in P and every state has only incoming transitions of at most one symbol from Σ. In contrast, we consider the full set of finite-state selectors. Moreover, Carton's results are restricted to the case of finite alphabets.

1.3. *Overview of techniques and the proof of the main theorem.* The main result has two directions: (i) proving that if μ-distributedness is preserved by selection by any DFA, then μ is necessarily induced by a Bernoulli distribution; and (ii) any μ induced by a Bernoulli distribution is preserved across selection by any DFA.

For direction (i), we prove the more general result that if μ is not induced by a Bernoulli distribution on Σ, selection by a particular *Postnikova strategy* (roughly, a Postnikova strategy selects an element of the sequence if and only if it follows a fixed finite word) will select a non-μ-distributed infinite sequence from a—bespoke—μ-distributed sequence. The Postnikova strategy contains prefixes in the form $u \cdot w \in \Sigma^*$ for a *fixed* w chosen such that $w \cdot a \in \Sigma^{|w|+1}$ is a minimal witness string such that $\mu(w \cdot a) \neq \mu(w) \cdot \mu(a)$. Using basic constructions, we can then prove that the Postnikova strategy can be implemented by a DFA that simulates a sliding fixed-width window.

For direction (ii), most of the modern methods of proving Agafonov's theorem (e.g., [5, 6, 46]) are not immediately adaptable because they use methods that are particular to equidistributions on finite alphabets (e.g., lossless finite-state compressors [6] or automatic Kolmogorov complexity [46])—and we consider both Bernoulli distributions and infinite alphabets. Instead, we work along the general lines of Agafonov's original proof [57] that more heavily uses probabilistic reasoning.

The key insights in Agafonov's original proof was: (i) that any *strongly connected* finite automaton (containing at least one accepting state) applied to a normal sequence must

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

select (always, not just with probability 1) more than a constant fraction of elements from any sufficiently long finite substring of its input; and (ii) that selecting more than a constant fraction of sufficiently long substrings entails that each element of Σ must be selected with approximately equal probability, by the law of large numbers. In Agafonov's original approach (for $\Sigma = \{0, 1\}$), an appeal to the strong law of large numbers was used in conjunction with the product measure on the product topology on $\{0, 1\}^\omega$, thus required reasoning about cylinder sets centered on sets A of finite strings; and to avoid 'double-counting' the probabilities, these sets had to be prefix-free. We avoid this difficulty by using concentration bounds to tally the occurrences of elements $a \in \Sigma$ in block decompositions of finite prefixes of α.

The proof that any DFA selects a μ-distributed infinite sequence from a μ-distributed infinite sequence then follows by observing that: (i) any run of a DFA on an infinite sequence eventually reaches a strongly connected component C of the DFA that is recurrent (that is, the run can never exit C); and (ii) that any such component induces an irreducible Markov chain, whence we can apply the ergodic theorem for Markov chains to conclude that accepting states are reached infinitely often and with appropriate frequency.

The extension to infinite alphabets is surprisingly straightforward in most proofs: essentially, instead of using combinatorial estimates for finite sets, we have to ensure that series taken over infinite alphabets converge properly, but almost all instances involve series that: (i) have non-negative elements; and (ii) are bounded above, whence the usual reasoning about absolutely convergent series can be employed. Similarly, the classic results for finite automata that we use need to be re-stated and re-proved in the case of infinite alphabets, but this in general turns out to be doable without too much leg-work (e.g., Lemma A.1). One caveat is that several important ancillary results have standard proofs that use combinatorial arguments on finite sets and we thus need to provide alternative proofs using different methods.

2. *Preliminaries*

Definition 2.1. We assume a non-empty, possibly (countably) infinite, alphabet Σ and denote by λ the empty string; the sets of finite and right-infinite sequences of elements of Σ are denoted by Σ^* and Σ^ω, respectively. We denote by Σ^+ the set of finite non-empty words, that is, $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$. Elements of Σ^* are ranged over by $v, w, \ldots$ and elements of Σ^ω by $\alpha, \beta, \ldots$. If $\alpha = a_1 a_2 \cdots \in \Sigma^\omega$ and N is a positive integer, we denote by $\alpha|_{\leq N}$ the finite string $a_1 a_2 \cdots a_N$.

Given $v \in \Sigma^*$ and $u \in \Sigma^* \cup \Sigma^\omega$, we write $v \cdot u$ for the element of $\Sigma^* \cup \Sigma^\omega$ obtained by concatenation. For words $v \in \Sigma^*$ and $w \in \Sigma^* \cup \Sigma^\omega$, v is said to be a *prefix* of w, written $v \preceq w$, if there exists $u \in \Sigma^* \cup \Sigma^\omega$ such that $w = v \cdot u$. If $v \preceq w$ and $v \neq w$, v is said to be a *proper prefix* of w, written $v \prec w$. For any $w \in \Sigma^*$, the *cylinder set* of w, denoted $[w]$, is the subset of Σ^ω defined by $[w] = \{\alpha \in \Sigma^\omega : \alpha = w \cdot \beta, \beta \in \Sigma^\omega\}$, that is, the set of right-infinite sequences that have w as prefix.

Definition 2.2. Let Σ be a non-empty, possibly (countably) infinite, alphabet. A *probability map* (over Σ) is a map $\mu : \Sigma^+ \longrightarrow [0, 1]$ such that, for all positive integers n, the series

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

$$\sum_{a_1 \cdots a_n \in \Sigma^n} \mu(a_1 \cdots a_n)$$

is convergent with limit 1. Note that convergence implies absolute convergence here.

A probability map μ is said to be:

- *induced by a Bernoulli distribution* $p : \Sigma \longrightarrow [0, 1]$ if, for all positive integers n, and all $a_1, \ldots, a_n \in \Sigma$, $\mu(a_1 \cdots a_n) = \prod_{i=1}^{n} \mu(a_i) = \prod_{i=1}^{n} p(a_i)$;
- *invariant* if, for all $w \in \Sigma^*$, the series $\sum_{a\in\Sigma} \mu(w \cdot a)$ and $\sum_{a\in\Sigma} \mu(a \cdot w)$ are convergent with limit $\mu(w)$;
- (when Σ is finite) *equidistributed* if, for any $w \in \Sigma^n$, $\mu(w) = |\Sigma|^{-n}$.

Observe that an equidistributed μ is also Bernoulli. For alphabets $|\Sigma| > 1$, any map $p : \Sigma \longrightarrow [0, 1]$ such that the series $\sum_{a\in\Sigma} p(a)$ converges to 1 induces a probability map μ_p by setting $\mu_p(a_1 \cdots a_n) = \prod_{j=1}^{n} p(a_j)$. For finite alphabets Σ, this map is equidistributed if and only if $p(a) = |\Sigma|^{-1}$ for every $a \in \Sigma$.

The expression 'induced by a Bernoulli distribution' is justified by the fact that Bernoulli probability maps correspond directly to the measure of cylinders in Bernoulli shifts [47]. (In the literature on normal numbers, the word Bernoulli is sometimes used slightly differently, for example, Schnorr and Stimm [44] use the term 'Bernoulli sequence' for sequences that are equidistributed in our terminology. We also note that μ-distributed sequences (defined later) with respect to Bernoulli distributions were first introduced by Postnikov and I. I. Piatetski-Shapiro under the name 'Bernoulli normal sequences' [60].)

PROPOSITION 2.1. *A probability map μ induced by a Bernoulli distribution is invariant.*

Proof. For any $w \in \Sigma^*$, $\sum_{a\in\Sigma} \mu(aw) = \sum_{a\in\Sigma} \mu(a)\mu(w) = \sum_{a\in\Sigma} \mu(w)\mu(a) = \sum_{a\in\Sigma} \mu(wa)$. And $\sum_{a\in\Sigma} \mu(w)\mu(a) = \mu(w) \sum_{a\in\Sigma} \mu(a) = \mu(w)$. □

We shall need probability maps to act as 'measures' on (possibly infinite) sets of finite strings.

Definition 2.3. Let Σ be a non-empty alphabet, let $W \subseteq \Sigma^*$, and let μ be a probability map over Σ. If $W = \emptyset$, we define $\mu(W) = 0$. If $\sum_{w\in W} \mu(w)$ converges, we define $\mu(W) = \sum_{w\in W} \mu(w)$.

Observe that as $\mu(w) \geq 0$ for all $w \in \Sigma^*$, if $\sum_{w\in W} \mu(w)$ converges, it is absolutely convergent (hence, we do not need to specify an ordering of W).

We are interested in the probability maps whose values can be realized as the limiting frequencies of finite words in right-infinite sequences over Σ.

Definition 2.4. Let $v = v_1 \cdots v_N$ and $w = w_1 \cdots w_n$ be finite words over Σ. We denote by $\#_w(v)$ the number of occurrences of w in v, that is, the quantity

$$|\{j \leq N + 1 - n : v_j v_{j+1} \cdots v_{j+n-1} = w_1 w_2 \cdots w_n\}|$$

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

Let μ be a probability map over Σ and let α be a right-infinite sequence over Σ. If the limit

$$\lim_{N\to\infty} \frac{\#_w(\alpha|_{\leq N})}{N}$$

exists and is equal to some real number f, we say that w occurs in α *with limiting frequency* f. If every $w \in \Sigma^+$ occurs in α with limiting frequency $\mu(w)$, we say that α is μ-distributed.

PROPOSITION 2.2. *Let μ be a probability map over Σ. If there exists a μ-distributed sequence, then μ is invariant.*

Proof. Let μ be a probability map over Σ and $\alpha = a_1a_2\ldots$ a μ-distributed sequence. We consider $w = w_1w_2\ldots w_k \in \Sigma^k$ and note that for all $N > 0$,

$$\left|\sum_{a\in\Sigma} \#_{aw}(\alpha|_{\leq N}) - \#_w(\alpha|_{\leq N})\right| \leq 1.$$

Indeed, every occurrence of w as $a_ia_{i+1}\ldots a_{i+k}$ such that $i > 1$ is also an occurrence of $b \cdot w$ for a (unique) $b \in \Sigma$, so the expressions $\#_w(\alpha|_{\leq N})$ and $\sum_{a\in\Sigma} \#_{aw}(\alpha|_{\leq N})$ are equal if and only if $a_1a_2\ldots a_k \neq w$ and their difference is equal to 1 otherwise.

Thus,

$$\left|\frac{\sum_{a\in\Sigma} \#_{aw}(\alpha|_{\leq N})}{N} - \frac{\#_w(\alpha|_{\leq N})}{N}\right| \leq \frac{1}{N}.$$

We therefore obtain that

$$\left|\frac{\sum_{a\in\Sigma} \#_{aw}(\alpha|_{\leq N})}{N} - \mu(w)\right| \leq \left|\frac{\sum_{a\in\Sigma} \#_{aw}(\alpha|_{\leq N})}{N} - \frac{\#_w(\alpha|_{\leq N})}{N}\right| + \left|\frac{\#_w(\alpha|_{\leq N})}{N} - \mu(w)\right|.$$

Since both expressions on the right converge to 0, the left-hand side converges to zero, showing that $\mu(w) = \lim_{n\to\infty}(\sum_{a\in\Sigma} \#_{aw}(\alpha|_{\leq n})/n) = \sum_{a\in\Sigma} \lim_{n\to\infty}(\#_{aw}(\alpha|_{\leq n})/n) = \sum_{a\in\Sigma} \mu(aw)$.

Similarly, for all $N > 0$,

$$\left|\sum_{a\in\Sigma} \#_{wa}(\alpha|_{\leq N}) - \#_w(\alpha|_{\leq N})\right| \leq 1,$$

by a similar argument as the one used above, noting that the number of occurrences is different if and only if $a_{N-k+1}a_{N-k+2}\cdots a_N = w$. We then conclude that $\mu(w) = \sum_{a\in\Sigma} \mu(wa)$ in the same way. □

Observe that an infinite sequence α is normal in the usual sense if and only if it is μ-distributed for (the unique) equidistributed probability map μ over Σ.

Example 2.3. An example of a probability map that is *not* Bernoulli, but such that there is at least one μ-distributed right-infinite sequence, is the map μ over $\Sigma = \{0, 1\}$ defined by $\mu(w) = 1/2$ if w does not contain any of the strings 00 or 11 (note that for each positive integer n, there are exactly two such strings of length n, namely $0101010\cdots$ and

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

$101010\cdots$), and $\mu(w) = 0$ otherwise. Observe that the right-infinite sequence $010101\cdots$ is μ-distributed.

In contrast to all previous work on Agafonov's theorem, we allow countably infinite alphabets Σ. Alphabets of larger cardinality do not in general have probability measures realizable by considering limiting frequencies of elements of Σ^ω—simply because most elements of Σ cannot occur at all in a single element of Σ^ω.

One reason why previous generalizations of Agafonov's theorem have not considered infinite alphabets is that there can be no equidistribution on a countably infinite set. However, there are Bernoulli measures μ on countably infinite alphabets Σ and μ-distributed infinite sequences over Σ.

Example 2.4. An example of a countably infinite alphabet with a Bernoulli measure is $\Sigma = \mathbb{N}$ and $p(n) = 6/(\pi n)^2$ (note that we have $\sum_{n\in\Sigma} p(n) = 1$). In general, any convergent series $\sum_{n=1}^{\infty} a_n$, where every a_n is non-negative, induces a Bernoulli distribution on $\mathbb{N}$ by setting $p(n) = a_n / \sum_{n=1}^{\infty} a_n$. Each such Bernoulli distribution p induces an invariant probability map μ_p, and by a result of Madritsch and Mance [33], there exists a μ_p-distributed sequence.

Remark 2.5. As we consider possibly infinite alphabets, we often have to consider infinite series instead of finite sums in the proofs. In most cases, these series will have elements that are known to be non-negative and the sum of all partial sums will be bounded above, whence the series will be absolutely convergent and the order of summation can thus be changed freely. A trivial example of use is to consider some $B \subseteq \Sigma$ and note that $\sum_{a\in B} p(a) = 1 - \sum_{a\in\Sigma\setminus B} p(a)$ (as $\sum_{a\in B} p(a) \leq 1$, $\sum_{a\in\Sigma\setminus B} p(a) \leq 1$, and $p(a) \geq 0$ the two series are absolutely convergent, and $\sum_{a\in B} p(a) + \sum_{a\in\Sigma\setminus B} = \sum_{a\in\Sigma} p(a) = 1$).

2.1. *Strategies.*

Definition 2.5. Let Σ be an alphabet. A *strategy* S over Σ is a subset $S \subseteq \Sigma^*$.

Given a strategy S and $\alpha \in \Sigma^\omega$, we define the sequence *selected* by S, denoted $S(\alpha)$, as follows: if $i_1, i_2, \ldots, i_k, \ldots$ is the (increasing) sequence of indices i_j such that $\alpha|_{<i_j} \in S$, then $S(\alpha)_j = \alpha_{i_j}$. When $w \in \Sigma^*$ is a finite word, we define $S(w)$ *mutatis mutandis*.

A strategy S is a *Postnikova strategy* if there is $w \in \Sigma^*$ such that $S = \Sigma^* w$.

Thus, $S(\alpha)$ is simply the subsequence of symbols from α that are 'picked out' by applying S to prefixes of α. Note also that if $w \in S$, then in any word on the form $w \cdot b \cdot v$, S must pick b. Thus, S cannot be made to, for instance, only pick out a single symbol from Σ—it must select 'the next symbol' after any $w \in S$. This precludes, for example, constant-memory strategies from selecting only 0s from a normal binary sequence.

Our primary object of study is the case where S is a regular language, described next.

2.2. *Finite-state selectors and selection by DFAs.*

As we treat both finite and (countably) infinite alphabets, we must consider automata over possibly infinite alphabets. Every automaton has a finite number of states as usual, but as the alphabet is infinite and a deterministic automaton has transitions on all symbols from every state, the underlying

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

graph of the automaton will be infinitely branching. To keep notation simple, we refer to a deterministic automaton with a finite number of states as a 'DFA' as usual, even if the underlying alphabet is infinite.

Definition 2.6. A *finite-state selector* over Σ is a DFA $A = (Q, \delta, q_s, Q_F)$, where Q is a finite set of states, q_s is the unique start state, Q_F is the set of accepting states, and $\delta : Q \times \Sigma \longrightarrow Q$ is the transition relation.

A DFA is strongly connected if its underlying directed graph (states are nodes, transitions are edges) is strongly connected.

Denote by $L(A)$ the language accepted by the automaton. If $\alpha = a_1 a_2 \cdots$ is a finite or right-infinite sequence over Σ, the subsequence *selected by A* is the (possibly empty) sequence of letters a_n such that the prefix $a_1 \cdots a_{n-1} \in L(A)$, that is, the automaton when started on the finite word $a_1 \cdots a_{n-1}$ in state q_s ends in an accepting state after having read the entire word. The *run* of A on input α is the sequence of states visited when A is applied to α from the starting state. For $(q, w) = (q, w_1 \cdots w_n) \in Q \times \Sigma^*$, we use the notation $\delta^*(q, w)$ to denote the state $\delta(\cdots \delta(\delta(q, w_1), w_2) \cdots w_n)$, that is, the state reached by starting from q and following the (unique) path induced by w.

Observe that a DFA may select an empty, finite, or infinite sequence when run on a right-infinite word.

Definition 2.7. Let A be a DFA. A strongly connected component C in (the underlying directed graph of) A is said to be *recurrent* if, for every state p in C and every $a \in \Sigma$, $\delta(p, a)$ is a state in C (that is, once a run of A on some infinite word reaches a state in C, the run cannot leave C).

Definition 2.8. Let $A = (Q, \Sigma, \delta, q_0, Q_F)$ be a connected DFA. For all $q \in Q$, we denote by A_q the automaton $(Q, \Sigma, \delta, q, Q_F)$, that is, where the state q is chosen as the initial state.

Definition 2.9. Let $A = (Q, \Sigma, \delta, q_0, Q_F)$ be a connected DFA and let $q \in Q$. Let α be a right-infinite sequence over Σ. We denote by $A_q[\alpha]$ the subsequence $\bar{\alpha}$ of α *picked out* by A_q, that is, $w_i \in \bar{w}$ if and only if $\delta^*(q, \alpha_{<i})$ is an accepting state.

Let us stress that $A_q[\alpha]$ differs in general from $S(\alpha)$, where $S \in \Sigma^*$ is the strategy defined as the language accepted by A. Indeed, these are equal only when q is chosen to be the initial state of A.

We shall use the following fundamental result in automata theory (the result in [44] is stated for finite alphabets, but the proof method carries through for infinite alphabets as well. We provide a proof in Appendix A).

LEMMA 2.6. [44, Lemma 2.6] *For every DFA $A = (Q, \delta, q_s, Q_F)$ over the (possibly infinite) alphabet Σ, there is a word $w \in \Sigma^*$ such that, for every $q \in Q$, there is a strongly connected recurrent component C (of the underlying directed graph of A) such that $\delta^*(q, w) \in C$.*

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

COROLLARY 2.7. *Let μ be a probability map induced by a positive Bernoulli distribution on Σ, let A be a DFA over Σ, and let $\alpha \in \Sigma^\omega$ be μ-distributed. Then, the run of A on α eventually reaches a strongly connected recurrent component of A.*

Proof. Let w be the word obtained from Lemma 2.6. As α is p-distributed, w appears in α, so write $\alpha = vw\alpha'$, and let q be the state of A reached after $|v|$ transitions in the run of A on α. Then, after at most a further $|w|$ transitions, the run reaches a state in a strongly connected component of (the underlying directed graph of) A. □

Corollary 2.7 ensures that we can assume without loss of generality that the finite-state selectors we treat are strongly connected. Note that the corollary does not imply that the strongly connected recurrent component contains an accepting state (indeed, the automata may have an empty set of accepting states). Thus, some automata do not always select infinite sequences, and additional assumptions are needed if this is desirable (this is discussed in Remark 3.2 below). However, this is not an issue for our main result which states that the output of a selector applied to a normal sequence is again normal *as long as it is infinite*.

3. *Main result*

THEOREM 3.1. *Let Σ be a non-empty (finite or infinite) alphabet and μ be a probability map such that there exists at least one $\alpha \in \Sigma^\omega$ that is μ-distributed. Then, the following statements are equivalent:*

(1) *μ is induced by a positive Bernoulli distribution p on Σ, that is, for every $a_1 \cdots a_n \in \Sigma$, $\mu(a_1 \cdots a_n) = \prod_{i=1}^n p(a)$ and $p(a) > 0$ for all $a \in \Sigma$;*

(2) *(Postnikova property) for every finite word $w \in \Sigma^*$ and μ-distributed sequence $\alpha \in \Sigma^\omega$, if the sequence selected from α by the Postnikova strategy $\mathcal{S}_w = \{u \in \Sigma^* \mid$ there exists v such that $u = vw\}$ is infinite, then it is μ-distributed;*

(3) *(Agafonov property) for every DFA A over Σ and every μ-distributed sequence $\alpha \in \Sigma^\omega$, if the sequence selected from α by A is infinite, then it is μ-distributed.*

Proof. For the implication (1) ⇒ (3), Theorem 2.7 yields that any run of a finite-state selector on a μ-distributed sequence eventually reaches a strongly connected recurrent component; the restriction of any DFA to the state set of one of its recurrent component is also a DFA, and the result now follows by Theorem 5.9. The implication (3) ⇒ (2) is clear from the definitions since the considered strategies are computed by finite automata (Theorem 4.3). Lastly, Theorems 4.1 and 4.2 prove that (2) ⇒ (1). □

Remark 3.2. Theorem 3.1 addresses the case where a DFA or Postnikova strategy selects an infinite sequence from a μ-distributed sequence. If one wants to restrict attention to automata that *always* select an infinite subsequence from any μ-distributed sequence, extra conditions sometimes occur in the literature, e.g., that every cycle in the (underlying graph of the) DFA contains an accepting state [**6**] ensuring that an infinite subsequence is selected from any (not just μ-distributed sequence). Another condition that ensures that an infinite subsequence is selected from any μ-distributed sequence is to consider only DFAs such that every strongly connected recurrent component contains at least one accepting state.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

In this case, Corollary 2.7 ensures that any run on the automaton on a μ-distributed sequence will reach a strongly recurrent component, and Lemma 5.4 below then ensures that the DFA accepts an infinite subsequence from α.

4. *Non-preservation of μ-distributedness for non-Bernoulli measures*

We first prove that if μ is a probability map such that any DFA selects a μ-distributed right-infinite sequence from any μ-distributed right-infinite sequence, then μ must be Bernoulli. This is an immediate consequence of a stronger property proved in Lemma 4.1 below.

The idea of the proof is that if μ is *not* Bernoulli, there exists a word $a_1 \cdots a_k$ such that $\mu(a_1 \cdots a_{k-1}) = \prod_{j=1}^{k-1} \mu(a_j)$, but $\mu(a_1 \cdots a_{k-1}a_k) \neq \mu(a_1 \cdots a_{k-1}) \cdot \mu(a_k)$. One can then construct a finite-state selector that acts like a 'sliding window' of size $k-1$, that is, remembers the last $k-1$ letters scanned and accepts if these are $a_1 \cdots a_k$. This selector will select every letter following $a_1 \cdots a_{k-1}$; after a prefix of length N of a right-infinite sequence has been scanned, approximately $N \cdot \mu(a_1 \cdots a_{k-1})$ have been selected and approximately $N \cdot \mu(a_1 \cdots a_{k-1}a_k)$ of these will be the symbol a_k. However, then the limiting frequency of a_k in the sequence selected will be $\mu(a_1 \cdots a_{k-1}a_k)/\mu(a_1 \cdots a_{k-1}) \neq \mu(a_k)$ and the result follows.

For completeness, we give a fully formal proof after the lemma, but the entirety of the reasoning is essentially as we just described.

LEMMA 4.1. *Let $\mu : \Sigma^* \longrightarrow [0,1]$ be a probability map. If μ is not induced by a Bernoulli distribution on Σ, there exists a finite word $w \in \Sigma^*$ such that if $\alpha \in \Sigma^\omega$ is μ-distributed, then the Postnikova strategy $\mathcal{S}_w = \{u \in \Sigma^* \mid$ there exists v such that $u = vw\}$ selects from α an infinite sequence $\beta \in \Sigma^\omega$ that is not μ-distributed.*

Proof. If no element of Σ^ω is μ-distributed, the lemma is vacuously true. Hence, assume that there is at least one $\alpha \in \Sigma^\omega$ that is μ-distributed. If $|\Sigma| = 1$, then there is exactly one probability map on Σ^*, namely the one that assigns probability 1 to the unique element of Σ^k for every $k \geq 0$, and this probability map is clearly Bernoulli and the lemma is thus vacuously true. Hence, in the remainder of the proof, assume that $|\Sigma| \geq 2$.

Assume that μ is not induced by a Bernoulli distribution on Σ. Then, there are k and a word $a_1 \cdots_{k-1} a_k \in \Sigma^k$ such that $\mu(a_1 \cdots a_{k-1}a_k) \neq \prod_{j=1}^{k} \mu(a_j)$. Observe that $k = 1$ is impossible and thus we must have $k \geq 2$. Assume without loss of generality that k is minimal among such k, and hence that $\mu(a_1 \cdots a_{k-1}) = \prod_{j=1}^{k-1} \mu(a_j)$, and note that this implies $\mu(a_1 \cdots a_{k-1}a_k) \neq \mu(a_1 \cdots a_{k-1}) \cdot \mu(a_k)$.

Assume for contradiction that $\mu(a_1 \cdots a_{k-1}) = 0$. Then, $\mu(a_i) = 0$ for at least one a_i. We also must have $\mu(a_1 \cdots a_{k-1}a_k) = 0$ because otherwise, the fact that there is at least one μ-distributed right-infinite sequence entails that $\mu(a_1 \cdots a_{k-1}a_k) > 0$, which in turn implies that $\mu(a_i) \geq \mu(a_1 \cdots a_{k-1}a_k) > 0$ for each i. However, this is a contradiction, because then we would have that $\mu(a_1 \cdots a_{k-1}a_k) = 0 = \prod_{j=1}^{k} \mu(a_j)$. Thus, $\mu(a_1 \cdots a_{k-1}) > 0$.

As $\mu(a_1 \cdots a_{k-1}a_k) \neq \mu(a_1 \cdots a_{k-1}) \cdot \mu(a_k)$, $\mu(a_1 \cdots a_{k-1}) > 0$, and we have $\mu(a_1 \cdots a_{k-1}a_k) \leq \mu(a_1 \cdots a_{k-1})$ (because μ is invariant by Proposition 2.2), there

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

is a real number γ with $0 < \gamma < 1$ such that:

$$\left| \frac{\mu(a_1 \cdots a_{k-1}a_k)}{\mu(a_1 \cdots a_{k-1})} - \mu(a_k) \right| > \gamma.$$

We now consider the Postnikova strategy $\mathcal{S}_w$ with $w = a_1 \cdots a_{k-1}$, that is, the strategy that selects exactly the symbols following the occurrences of $a_1 \cdots a_{k-1}$ in α.

Let $\alpha \in \Sigma^\omega$ be μ-distributed. Then, for every $\epsilon > 0$, there is an $N_\epsilon > 0$ such that for all $n > N_\epsilon$, we have

$$\left| \frac{\#_{a_1 \cdots a_{k-1}}(\alpha|_{\leq n})}{n} - \mu(a_1 \cdots a_{k-1}) \right| \leq \epsilon$$

and

$$\left| \frac{\#_{a_1 \cdots a_{k-1}a_k}(\alpha|_{\leq n})}{n} - \mu(a_1 \cdots a_{k-1}a_k) \right| \leq \epsilon.$$

Hence,

$$n\mu(a_1 \cdots a_{k-1}) - n\epsilon \leq \#_{a_1 \cdots a_{k-1}}(\alpha|_{\leq n}) \leq n\mu(a_1 \cdots a_{k-1}) + n\epsilon \tag{4.1}$$

and

$$n\mu(a_1 \cdots a_{k-1}a_k) - n\epsilon \leq \#_{a_1 \cdots a_{k-1}a_k}(\alpha|_{\leq n}) \leq n\mu(a_1 \cdots a_{k-1}a_k) + n\epsilon. \tag{4.2}$$

As $\mu(a_1 \cdots a_{k-1}) > 0$ and $\mathcal{S}_w$ selects the symbol after each occurrence of $a_1 \cdots a_{k-1}$, $\mathcal{S}_w$ selects an infinite sequence β from α. Let $\beta^{(n)} \in \Sigma^*$ be the finite sequence selected by $\mathcal{S}_w$ from $\alpha|_{\leq n}$. Observe that we have $|\beta^{(n)}| = \#_{a_1 \cdots a_{k-1}}(\alpha|_{<n})$ and $\#_{a_k}(\beta^{(n)}) = \#_{a_1 \cdots a_{k-1}a_k}(\alpha|_{<n})$. The fraction of occurrences $\#_{a_k}(\beta^{(n)})/|\beta^{(n)}|$ of a_k in $\beta^{(n)}$ thus satisfies

$$\frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} = \frac{\#_{a_1 \cdots a_{k-1}a_k}(\alpha|_{<n})}{\#_{a_1 \cdots a_{k-1}}(\alpha|_{<n})} = \frac{\#_{a_1 \cdots a_{k-1}a_k}(\alpha|_{<n})}{n} \cdot \frac{n}{\#_{a_1 \cdots a_{k-1}}(\alpha|_{<n})}$$

and hence, by equations (4.1) and (4.2), for all $n > N$,

$$\frac{\mu(a_1 \cdots a_{k-1}a_k) - \epsilon}{\mu(a_1 \cdots a_{k-1}) + \epsilon} \leq \frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} \leq \frac{\mu(a_1 \cdots a_{k-1}a_k) + \epsilon}{\mu(a_1 \cdots a_{k-1}) - \epsilon}. \tag{4.3}$$

Consider an arbitrary δ with $0 < \delta < \gamma/2$. By equation (4.3), for all sufficiently small ϵ, we have

$$\left| \frac{\mu(a_1 \cdots a_{k-1}a_k)}{\mu(a_1 \cdots a_{k-1})} - \frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} \right| < \delta$$

and thus for all $n > N_\epsilon$,

$$\begin{aligned}
\gamma &< \left| \frac{\mu(a_1 \cdots a_{k-1}a_k)}{\mu(a_1 \cdots a_{k-1})} - \mu(a_k) \right| \\
&\leq \left| \frac{\mu(a_1 \cdots a_{k-1}a_k)}{\mu(a_1 \cdots a_{k-1})} - \frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} \right| + \left| \frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} - \mu(a_k) \right| \\
&< \delta + \left| \frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} - \mu(a_k) \right| < \frac{\gamma}{2} + \left| \frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} - \mu(a_k) \right|
\end{aligned}$$

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

whence

$$\left| \frac{\#_{a_k}(\beta^{(n)})}{|\beta^{(n)}|} - \mu(a_k) \right| > \frac{\gamma}{2}$$

and as the sequence $(\beta^{(n)})_{n\in\mathbb{N}}$ consists of prefixes of the sequence $\mathcal{S}_w[\alpha]$ selected by $\mathcal{S}_w$ from α, and is eventually increasing, the frequency of occurrences of a_k differs infinitely often from $\mu(a_k)$ by at least $\gamma/2$, so $\mathcal{S}_w[\alpha]$ cannot be μ-distributed. □

Lemma 4.1 shows that if a probability map is not induced by a Bernoulli distribution on Σ, some Postnikova strategy will select a non-μ-distributed sequence from *any* μ-distributed sequence. In the case where μ *is* induced by a Bernoulli distribution, but not a positive Bernoulli distribution, we can show the weaker result that there will be a Postnikova strategy that selects a non-μ-distributed sequence from *some* μ-distributed sequences (and this is sufficient for our main theorem).

LEMMA 4.2. *Let* $\mu : \Sigma^* \longrightarrow [0, 1]$ *be a probability map induced by a Bernoulli distribution on* Σ *that is not positive. Then, there exists a finite word* $w \in \Sigma^*$ *and* μ*-distributed* $\alpha \in \Sigma^\omega$ *such that the Postnikova strategy* $\mathcal{S}_w = \{u \in \Sigma^* \mid$ *there exists* v such that $u = vw\}$ *selects from* α *an infinite sequence* $\beta \in \Sigma^\omega$ *that is not* μ*-distributed.*

Proof. As μ is not positive, pick $b \in \Sigma$ such that $\mu(b) = 0$, and let $\Gamma \subseteq \Sigma$ be a maximal subset such that the restriction of μ to Γ is a positive Bernoulli distribution (observe that Γ is non-empty because μ is a probability map and $1 = \sum_{a\in\Sigma} \mu(a)$ thus implies $\mu(a) > 0$ for some $a \in \Sigma$). By [33], there exists a μ-distributed infinite sequence $\beta \in \Gamma^\omega$; notice that β can be assumed without loss of generality to not contain any occurrences of b. Let $\alpha \in \Sigma^\omega$ be obtained by inserting the string bb at positions 2, 4, 8, 16, ... Then, α is μ-distributed because: (i) every $v \in \Gamma^*$ occurs with the same limiting frequency as in β (the key observation here is that since the 'bb's are inserted at exponentially increasing positions, the frequency of occurrence of all other strings is decreased by a very small (and quickly decaying) factor); and (ii) every $v \in \Sigma^*$ that contains an element of $\Sigma \setminus \Gamma$ occurs in α with limiting frequency 0. Set $w = b$; then the Postnikova strategy $\mathcal{S}_w = \{u \in \Sigma^* |$ there exists $v$such that$u = vw\}$ selects from α a sequence $\beta = \mathcal{S}_w[\alpha]$ such that, for every $n > 0$, $\#_b(\beta|_{\leq n}) \geq n/2 - 1$. Thus, the limiting frequency of b in β is not 0, and hence is not $\mu(b)$, proving that β is not μ-distributed. □

LEMMA 4.3. *Let* $w \in \Sigma^*$. *The Postnikova strategy* $\{u \in \Sigma^* \mid$ *there exists* v such that $u = vw\}$ *is computable by a strongly connected DFA over* Σ.

Proof. Note that the alphabet can possibly be infinite in the following proof. In the trivial case $|\Sigma| = 1$, the desideratum follows since there is only one infinite sequence α and every Postnikova strategy extracts α from α. If $|w| = 0$, the single-state DFA that accepts all input and with transitions on all symbols to the state itself clearly computes the Postnikova strategy. Hence, we may assume in the remainder of the proof that $|\Sigma| > 1$ and $|w| > 0$.

Write $w = w_1, w_2, \ldots, w_m$, where $w_1, \ldots, w_m \in \Sigma$. We construct a finite state selector M_w with exactly $m + 1$ states which will select a symbol from the input if and only if it is preceded by the word w. Let $M_w = (\mathbf{m}, \delta, q_s, Q_F)$ be defined as follows:

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

- $\mathbf{m} = \{0, 1, \ldots, m\}$; state $i \in \{0, \ldots, m\}$ represents the case where the *longest* prefix of m read from the last i symbols has length i (observe that there may be several shorter prefixes present and that a prefix of w may contain a shorter prefix of w as a suffix);
- $q_s = 0$;
- $Q_F = \{m\}$;
- δ is defined by

$$\delta(i, a) = \begin{cases} i + 1 & \text{if } i < m \text{ and } a = w_{i+1}, \\ j & \text{if } (i = m \text{ or } a \neq w_{i+1}) \text{ and where } j \text{ is the longest prefix of} \\ & w \text{ such that } w_{i-j+1} \cdots w_i a = w_1 \cdots w_j. \end{cases}$$

Observe that the above cases exhaust all possibilities and that δ thus is well-defined. Note in particular that $\delta(i, a) = 0$ when there is no j such that $w_{i-j+1} \cdots w_i a = w_1 \cdots w_j$.

As an example, consider the case where $\Sigma = \{a, b\}$, and $w = aabaabb$. If M_w is in state 6 (corresponding to the 6 most recently seen symbols being $aabaab$ and M_w reads the symbol a, then the longest prefix of w most recently read is $aaba$, corresponding to the suffix aab of $aabaab$ followed by a). Hence, M_w enters state 4.

Observe that M_w implements the Postnikova strategy as M_w accepts if and only if the last sequence of m symbols read is w.

We now prove that the underlying directed graph of M_w is strongly connected. To prove the claim, observe first that by construction, there is a directed edge from every state $i \in \{0, \ldots, m-1\}$ to state $i + 1$ (namely the transition on symbol w_{i+1} corresponding to the first case in the definition of δ). As $|\Sigma| > 1$, there is a symbol $a \in \Sigma$ with $a \neq w_1$, and hence there is an edge from state 0 to itself on symbol a. We claim that, for any state $i \in \{1, \ldots, m\}$, there is a path in (the underlying directed graph of) M_w to some state $j < i$; it then follows by induction that there is a path from i to 0 and, by the above remarks, there is a path from state 0 to any state in M_w, whence M_w is strongly connected.

To prove the claim, let $i \in \{1, \ldots, m\}$. Split on cases as follows.

- If $i = m$, choose (as $|\Sigma| > 1$) $a \in \Sigma$ such that $a \neq w_m$. Then, by construction, there is a transition from state m on symbol a to some state $j < m$ because the final symbol of $w = w_1 \cdots w_m$ is distinct from a, the longest prefix of w such that $w_{i-j+2} \cdots w_i a = w_1 \cdots w_j$ must satisfy $j < m$.
- If $i < m$, choose (as $|\Sigma|$) $a \in \Sigma$ such that $a \neq w_{i+1}$ and, hence, $w_1 \cdots w_i a$ is not a prefix of w, whence the longest prefix $w_1 \cdots w_j$ of w such that $w_{i-j+2} \cdots w_i a = w_1 \cdots w_j$ must satisfy $j \leq i$. If $j < i$, the claim is proved. If $j = i$, we have $w_1 \cdots w_i = w_2 \cdots w_i a$ and, hence, $w_1 = w_2 = \cdots = w_{i-1} = w_i = a$. However, then $w_1 \cdots w_i w_{i+1} = a^i w_{i+1}$ with $w_{i+1} \neq a$.
  - If $i + 1 = m$, there is a directed edge from state i to state m on symbol w_m, and as $w = a^{m-1} w_{i+1} = a^{m-1} w_m$ with $w_m \neq a$, there is a directed edge from state m to state 0 on symbol w_m because $w_m w_m$ does not occur as a substring of w and $w_m \neq a = w_1$, concluding the proof of the claim.
  - If $i + 1 < m$, let k be the greatest integer such that $k \leq m$ and $w_{i+1}, w_{i+2}, \ldots, w_k \neq a$. Then, there is a directed edge from state $i + k$ to state 0 on symbol w_{i+1} because $w = a^i w_{i+1} \cdots w_k c$ for some (possibly empty) string c and, thus, no

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

non-empty suffix of $a^i w_{i+1} \cdots w_k w_{i+1}$ can be a prefix of w. By construction, there is a directed path from state i to state $k > i$. Hence, there is a directed path from state i to state 0, as desired. □

5. *Finite-state selectors preserve μ-distributedness for Bernoulli measures*

The sequence of auxiliary results of this section follows the general lines of Agafonov's original proof in Russian for the case $\Sigma = \{0, 1\}$ [57], but with multiple proofs needing more careful analysis and adapted techniques.

5.1. *Ancillary definitions and results.*

Definition 5.1. Let Σ be an alphabet, $\alpha = x_1 x_2 \cdots x_n \cdots \in \Sigma^\omega$, and let n be a positive integer. The *n-block decomposition of α* is the sequence $(\alpha_{(n,r)})_{r \geq 1}$, where $\alpha_{(n,r)} = x_{(r-1)n+1} \cdots x_{rn} \in \Sigma^n$.

Thus, $\alpha_{(n,1)}$ is the string of the first n symbols of α, $\alpha_{(n,2)}$ is the string of the next n symbols, and so forth.

Definition 5.2. Let μ be a probability map over Σ and $\alpha = x_1 x_2 \cdots x_n \cdots \in \Sigma^\omega$. We say that α is *μ-block-distributed* if, for each $n \geq 1$ and every $w \in \Sigma^n$, the n-block decomposition $(\alpha_{(n,r)})_{r \geq 1}$ of α satisfies

$$\lim_{k \to \infty} \frac{|i \leq k : \alpha_{(n,i)} = w|}{k} = \mu(w).$$

For finite alphabets and the special case of p being an equidistribution on Σ, it is straightforward to prove that the properties of being μ_p-distributed and μ_p-block-distributed are equivalent [16, 37, 61]. For the present paper, we only use that μ_p-distributedness implies μ_p-block-distributedness. This result can be proved by tedious, but standard, counting arguments on sufficiently large finite prefixes of $\alpha|_{\leq N}$ using the same reasoning as the original proof by Niven and Zuckerman for finite alphabets and normality [37]. However, we are indebted to an anonymous referee for suggesting the—much more elegant and succinct—proof of Proposition 5.1 below.

PROPOSITION 5.1. *Let μ_p be a probability map induced by a Bernoulli distribution p on the alphabet Σ. If $\alpha \in \Sigma^\omega$ is μ_p-distributed, it is μ_p-block distributed.*

Proof. Let I be an arbitrary set of cylinder sets $[v]$ with $v \in \Sigma^*$ and let m be a positive integer (the 'block size'). For any non-negative integer n, let $\#_I^0(n)$ be the number of occurrences of an element of I in the prefix of α of length n at any position j with $j = km + 1$ for some non-negative integer k (that is, the element of I occurs in the beginning of a size-m block in the block decomposition of α), and let $\#_I(n)$ be the number of occurrences of an element of I in any position in the prefix of α of length n.

Let $W \subseteq \Sigma^*$ be a set such that no element of W is a prefix of any other element. Hence, for arbitrary $v, w \in W$ with $v \neq w$, the cylinder sets $[v]$ and $[w]$ are disjoint.

Then,

$$\limsup_{n \to \infty} \frac{\#_W^0(n)}{n/m} \leq \limsup_{n \to \infty} \frac{\#_W(n)}{n/m} \leq m \limsup_{n \to \infty} \frac{\#_W(n)}{n}.$$

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

However, as α is μ_p-distributed, for every $w \in W$, we have $\lim_{n\to\infty} \#_w(n)/n = \mu_p(w)$, and as μ_p is a probability map and no element of W is a prefix of any another, $\sum_{w\in W} \mu_p(w)$ exists by the dominated convergence theorem. By the bounded convergence theorem, we then have $\lim_{n\to\infty}(\#_W(n)/n) = \sum_{w\in W} \mu_p(w)$ and, thus,

$$m \limsup_{n\to\infty} \frac{\#_W(n)}{n} = m \lim_{n\to\infty} \frac{\#_W(n)}{n} = m \sum_{w\in W} \mu_p(w),$$

whence in particular $\limsup_{n\to\infty}(\#^0_W(n)/n/m) \le m \sum_{w\in W} \mu_p(w)$.

Any union of a countable set I of cylinder sets is also the union of a countable set $W \subseteq I$ of cylinder sets $[v]$ such that no string v is a prefix of any other string $w \neq v$ with $[w] \in I$ (simply winnow out all strings that are proper suffixes of elements already in I), and hence, by the above inequalities,

$$\limsup_{n\to\infty} \frac{\#^0_I(n)}{n/m} \le \limsup_{n\to\infty} \frac{\#^0_W(n)}{n/m} \le m \sum_{w\in W} \mu_p(w) \le m \sum_{w\in I} \mu_p(w),$$

where $\sum_{w\in W} \mu_p(w)$ always exists.

It then follows immediately by [3, Theorem 2] that for any cylinder set $[w]$ with $w \in \Sigma^m$, we have

$$\lim_{n\to\infty} \frac{\#^0_W(n)}{n/m} = \sum_{w\in W} \mu_p(w) = \mu_p(w).$$

The number of occurrences of w as a size-m block in the first n symbols of the m-block decomposition of $\alpha|_{\le n}$ is exactly $\#^0_w(n)$, and the result follows. □

We now prove that finite-state selectors can be composed appropriately; this will later be a key ingredient in reducing the problem of selecting finite strings $w \in \Sigma^*$ with frequency $\mu_p(w)$ to the problem of selecting single symbols $a \in \Sigma$ with frequency $p(a)$.

PROPOSITION 5.2. (Finite-state selectors are compositional) *Let A and B be DFAs over the same alphabet. Then, there is a DFA C such that, for each sequence w, $C[w] = B[A[w]]$. If A and B are both strongly connected, and A contains at least one accepting state, C can be chosen to be strongly connected.*

Proof. Let $A = (Q^A, \Sigma, \delta^A, q_0^A, Q_F^A)$ and $B = (Q^B, \Sigma, \delta^B, q_0^B, Q_F^B)$. Define $Q^C = Q^A \times Q^B$, and set $q_0^C = (q_0^A, q_0^B)$ and $Q_F^C = Q_F^A \times Q_F^B$. For each $q^B \in Q^B$, define the set $D_{q^B} = \{(q, q^B) : q \in Q^A\} \subseteq Q^C$. Observe that $Q^C = \bigcup_{q^B\in Q^B} D_{q^B}$ and that for $q^B, r^B \in Q^B$ with $q^B \neq r^B$, we have $D_{q^B} \cap D_{r^B} = \emptyset$, and thus $\{D_{q^B} : q^B \in Q^B\}$ is a partitioning of Q^C. Hence, the transition relation, δ^C, of C may be defined by defining it separately on each subset D_{q^B}:

$$\delta^C((q, q^B), a) = \begin{cases} (r, q^B) & \text{if } q \notin Q_F^A \text{ and } \delta^A(q, a) = r, \\ (r, r^B) & \text{if } q \in Q_F^A \text{ and } \delta^A(q, a) = r \text{ and } \delta^B(q^B, a) = r^B. \end{cases}$$

Thus, when C processes its input, it freezes the current state q^B of B (the freezing is represented by staying within D_{q^B}) and simulates A until an accepting state of A is reached

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

(that is, just before A would select the next symbol); on the next transition, C unfreezes the current state of B and moves to the next state r^B of B, and then freezes it and continues with a simulation of A.

Observe that a symbol is picked out by C if and only if the state is an element of $Q_F^C = Q_F^A \times Q_F^B$ if and only if the symbol is the next symbol read after simulation of A reaches an accepting state of A when the current frozen state of B is an accepting state of B.

By construction, C is strongly connected if both A and B are: for any pair of states (q_1^A, q_1^B) and (q_2^A, q_B^2) in Q^C, strong connectivity of B implies that there is a directed path from q_1^B to q_2^B in B. Let $q_1^B, q_{1,2}^B, q_{1,3}^B, \ldots, q_{1,k}^B$ be the states along this path. Strong connectivity of A and the assumption that there is some $q_1^F \in Q_F^A$ imply that there is a directed path from (q_1^A, q_1^B) to (q_1^F, q_1^B) in C, and by definition of δ^C, there is a transition in C from (q_1^F, q_1^B) to $(q_1^B, q_{1,2}^B)$. A straightforward induction on k now completes the proof. □

The following shows that to prove that the property of being μ_p-distributed is preserved under finite-state selection, it suffices to prove that the limiting frequency of each $a \in \Sigma$ exists and equals $p(a)$.

LEMMA 5.3. *Let μ_p be a probability map induced by a Bernoulli distribution p on Σ, and let $\alpha \in \Sigma^\omega$ be μ_p-distributed. The following are equivalent.*

- *For all strongly connected DFAs A, if $A[\alpha]$ is infinite, then $A[\alpha]$ is μ_p-distributed.*
- *For all strongly connected DFAs A and all $a \in \Sigma$, if $A[\alpha]$ is infinite, then the limiting frequency of a in $A[\alpha]$ exists and equals $p(a)$.*

Proof. If, for all A such that $A[\alpha]$ is infinite, $A[\alpha]$ is μ_p-distributed, then in particular the limiting frequency of a in $A[\alpha]$ exists and is equal to $p(a)$ for all A.

Conversely, suppose that, for all strongly connected DFAs A and all $a \in \Sigma$, if $A[\alpha]$ is infinite, then the limiting frequency of a in $A[\alpha]$ exists and equals $p(a)$. If $|\Sigma| = 1$, it follows immediately that $A[\alpha]$ is μ_p-distributed; hence, in the remainder of the proof, assume that $|\Sigma| \geq 2$.

Let A be a strongly connected DFA such that $A[\alpha]$ is infinite. If A has no accepting states, there is nothing to prove, so assume that A has at least one accepting state.

We will prove by induction on $k \geq 0$ that the limiting frequency of every $v_1 \cdots v_k v_{k+1} \in \Sigma^{k+1}$ exists and equals $\mu_p(v_1 \cdots v_k v_{k+1})$.

- $k = 0$. This is the supposition.
- $k \geq 1$. Suppose that the result has been proved for $k-1$. Let $v = v_1 \cdots v_k \in \Sigma^k$; by the induction hypothesis, the limiting frequency of $v_1 \cdots v_k$ in $A[w]$ is $\mu_p(v_1 \cdots v_k)$. We claim that there is a strongly connected DFA B that, from any sequence α, selects the symbol after each occurrence of $v_1 \cdots v_k$, and *only* those symbols, except for at most k symbols at the start of α. We construct such a DFA as follows. Set $B = (\{q_0, \ldots, q_k\}, \delta, q_0, \{q_k\})$, with δ to be defined below. For i with $0 \leq i \leq k$, the state q_i represents a situation where the last i symbols read by M_w is a length-i prefix of v among the last k symbols read, and i is maximal (that is, there is no j with $i < j \leq k$ such that the last j symbol read by B is also a prefix of v); observe that v can overlap with itself, e.g., $wv = 000$, so when B has read the string 100, $i = 2$, but the rightmost 0 in 100 is also a prefix of v.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

We define δ as follows, for any q_i and $a \in \Sigma$: $\delta(q_i, a) = q_j$, where j is the largest j with $0 < j \leq k$ such that $v_{k-j+1} \cdots v_i \cdot a$ is a prefix of v. Note in particular that if $v_1 \cdots v_i a = v_1 \cdots v_i v_{i+1}$, then $\delta(q_i, a) = q_{i+1}$. If no such j exists (that is, no prefix of v overlaps with the last k symbols read), we define $\delta(q_i, a) = q_0$. Observe in particular, that q_0 has transitions to itself on all symbols a such that $a \neq v_1$.

To see that M_w is strongly connected, we prove the stronger property that between any (not necessarily distinct) states q_i, q_j, there is a path containing the state q_k. Observe that the state q_i represents the situation where the prefix $v_1 \cdots v_i$ has been read by B, and that there is a path $(q_i, v_{i+1}) \cdots (q_{k-1}, v_k)$ to q_k. As $|\Sigma| \geq 2$, there is at least one symbol $a \in \Sigma$ such that no suffix of $v_1 \cdots v_k \cdot a$ is a non-empty prefix of v, and hence, there is at least one transition from q_k to q_0; this proves strong connectivity.

By Theorem 5.2, there is a strongly connected DFA C such that $C[w] = B[A[w]]$ for all $w \in \Sigma^*$.

For any $a \in \Sigma$ and any sufficiently large positive integer N, we have

$$\frac{\#_a(C[\alpha|_{\leq N}])}{|C[\alpha|_{\leq N}]|} = \frac{\#_a(B[A[\alpha|_{\leq N}]])}{|B[A[\alpha|_{\leq N}]]|} = \frac{\#_{v_1 \cdots v_k a}(A[\alpha|_{\leq N}])}{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])}.$$

By the induction hypothesis, for every $\epsilon > 0$, we have, for all sufficiently large N, that $|\#_a(C[\alpha|_{\leq N}])/|C[\alpha|_{\leq N}]| - p(a)| < \epsilon$, and hence,

$$\left| \frac{\#_{v_1 \cdots v_k a}(A[\alpha|_{\leq N}])}{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])} - p(a) \right| < \epsilon. \tag{5.1}$$

However, for all sufficiently large N, the induction hypothesis also furnishes that

$$\left| \frac{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])}{|A[\alpha|_{\leq N}]|} - \mu_p(v_1 \cdots v_k) \right| < \epsilon. \tag{5.2}$$

Nevertheless, as

$$\frac{\#_{v_1 \cdots v_k a}(A[\alpha|_{\leq N}])}{|A[\alpha|_{\leq N}]|} = \frac{\#_{v_1 \cdots v_k a}(A[\alpha|_{\leq N}])}{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])} \cdot \frac{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])}{|A[\alpha|_{\leq N}]|},$$

equations (5.1) and (5.2) thus yield

$$\begin{aligned}
&\left| \frac{\#_{v_1 \cdots v_k a}(A[\alpha|_{\leq N}])}{|A[\alpha|_{\leq N}]|} - \mu_p(v_1 \cdots v_k a) \right| \\
&= \left| \frac{\#_{v_1 \cdots v_k a}(A[\alpha|_{\leq N}])}{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])} \cdot \frac{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])}{|A[\alpha|_{\leq N}]|} - \mu_p(v_1 \cdots v_k) p(a) \right| \\
&< \epsilon^2 + \epsilon \left(\frac{\#_{v_1 \cdots v_k a}(A[\alpha|_{\leq N}])}{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])} + \frac{\#_{v_1 \cdots v_k}(A[\alpha|_{\leq N}])}{|A[\alpha|_{\leq N}]|} \right) \\
&\leq \epsilon^2 + 2\epsilon.
\end{aligned}$$

Hence, for all $a \in \Sigma$, the limiting frequency of $v_1 \cdots v_k a$ in $A[\alpha]$ exists and equals $\mu_p(v_1 \cdots v_k a)$, as desired. □

5.2. *Preservation of Bernoulli μ_p-distributedness under finite-state selection.* By Lemma 5.3, we may restrict our attention to proving that the frequency of single symbols from Σ are preserved under selection by DFAs. The strategy will be to consider an

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

arbitrary strongly connected DFA A, split the set of finite words Σ^* into multiple classes that depend on the selection behavior of A, and use a combination of concentration bounds and basic Markov chain theory applied to these classes to obtain upper and lower bounds on the frequency with which A selects each symbol from A.

Definition 5.3. Let $A = (Q, \Sigma, \delta, q_0, Q_F)$ be a strongly connected DFA. For any probability distribution $p : \Sigma \longrightarrow [0, 1]$, any $b \in [0, 1]$, $n \in \mathbb{N}$, and any $\epsilon > 0$, we define sets $D_n^p(b, \epsilon)$, $E_n(b)$, and $G_n(b, \epsilon)$ as follows:

$$
\begin{aligned}
D_n^p(b, \epsilon, q) &= \left\{ w \in \Sigma^n \ : \ |A_q[w]| > bn \text{ and } \sup_{a \in \Sigma} \left| \frac{\#_a(A_q[w])}{|A_q[w]|} - p(a) \right| < \epsilon \right\}, \\
D_n^p(b, \epsilon) &= \bigcap_{q \in Q} D_n^p(b, \epsilon, q), \\
E_n(b, q) &= \{ w \in \Sigma^n : |A_q[w]| \leq bn \}, \\
E_n(b) &= \bigcup_{q \in Q} E_n(b, q), \\
G_n(b, \epsilon, q) &= \left\{ w \in \Sigma^n : |A_q[w]| > bn \text{ and } \sup_{a \in \Sigma} \left| \frac{\#_a(A_q[w])}{|A_q[w]|} - p(a) \right| \geq \epsilon \right\}, \\
G_n(b, \epsilon) &= \bigcup_{q \in Q} G_n(b, \epsilon, q).
\end{aligned}
$$

The definitions of the above sets are necessarily quite technical, but we will try to provide some rough intuitive explanations. The set $D_n(b, \epsilon)$ consists of words w of length n such that the automata A, *given any state q as initial state*, will pick out from w a subsequence of length at least bn and in which is well-distributed up to ϵ. The set $E_n(b)$ denotes the set of words w of length n such that there exists at least one state q for which A, with initial state q, will pick out from w a word of length smaller that bn. Lastly, the set $G_n(b, \epsilon)$ is the set of words w of length n such that there exists at least one state q for which A, with initial state q, will pick out from w a word of length at least bn, but in which the proportion of at least one symbol a will differ from $p(a)$ by more than ϵ.

Observe that, for all b, n, ϵ,

$$\Sigma^n = E_n(b) \cup D_n^p(b, \epsilon) \cup G_n(b, \epsilon)$$

(and also note that $E_n(b)$ and $G_n(b, \epsilon)$ are not necessarily disjoint).

We will show in the following that as n grows, the measure of the set $D_n(b, \epsilon)$ (for a well-chosen value b) goes to 1. This is done by showing that the measure of the two other sets goes to 0. We will then be able to use this result to prove that the sequence picked out from A will be well distributed. A key observation here is the intersection over all states in the definition of $D_n(b, \epsilon)$ as it ensures that one can apply the result on any block in the block decomposition of the input sequence even if the state the automata A will have reached through the computation over preceding blocks is unknown.

LEMMA 5.4. *Let p be a positive Bernoulli distribution on Σ, let $S = (Q, \Sigma, \delta, q_s, Q_F)$ be a strongly connected finite automaton with $Q_F \neq \emptyset$, and let n be a positive integer.*

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

Then, there exists a real number $c > 0$ such that for all real numbers $\epsilon > 0$, we have $\lim_{n\to\infty} \mu_p(E_n(c-\epsilon)) = 0$.

Proof. Here, S induces a stochastic $|Q| \times |Q|$ matrix $\mathbf{P}$ by setting

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

Observe that if Σ is infinite, the fact that (i) $p(a) \cdot 1_{\delta(i,a)=j} \geq 0$, (ii) $p(a) \cdot 1_{\delta(i,a)=j} \leq p(a)$, and (iii) $\sum_{a\in\Sigma} p(a) = 1$ entails that the series $\sum_{a\in\Sigma} p(a) \cdot 1_{\delta(i,a)=j}$ is absolutely convergent.

Note also that $\mathbf{P}_{ij} = 0$ if and only if there are no transitions from i to j in Q on a symbol $a \in \Sigma$ with $p(a) > 0$. As S is strongly connected, there exists a path from state i to state j for each $i, j \in Q$. Let v be the word along this path; as $p(a) > 0$ for all $a \in \Sigma$, we have $\mu_p(v) > 0$, whence for each i, j, there is an integer n_{ij} such that $\mathbf{P}_{ij}^{n_{ij}} > 0$, that is, $\mathbf{P}$ (and its associated Markov chain) is irreducible. As all states of a finite Markov chain with irreducible transition matrix are positive recurrent, standard results (see, e.g., [45, Theorem 54]) yield that there is a unique positive stationary distribution $\pi : Q \longrightarrow [0, 1]$ (such that, for all $i \in Q$, we have $\pi(i) > 0$ and $\pi(i) = \sum_{j\in Q} \pi(j)\mathbf{P}_{ij}$). Furthermore, the expected return time M_i to state i satisfies $M_i = 1/\pi(i)$.

Let $(X_n)_{n\geq 0} = (X_0, X_1, X_2, \ldots)$ be the Markov chain with transition matrix $\mathbf{P}$ and some initial distribution λ on the states. Consider, for each $i \in Q$, the stochastic variable V_i, where

$$V_i(n) = \sum_{k=0}^{n-1} 1_{X_k=i},$$

that is, $V_i(n)$ is the number of times state i is visited in the first n elements of the Markov chain. As $\mathbf{P}$ is irreducible, the ergodic theorem for Markov chains (see, e.g., [45, Theorem 75]) yields that, independently of λ, we have for arbitrary $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(\left|\frac{V_i(n)}{n} - \pi(i)\right| \geq \epsilon\right) = \lim_{n\to\infty} \Pr\left(\left|\frac{V_i(n)}{n} - \frac{1}{M_i}\right| \geq \epsilon\right) = 0. \tag{5.3}$$

Let n be a positive integer, let $w = w_1 \cdots w_n \in \Sigma^n$, and let $\mathbf{q}_{S_j}(w) = q_0^w q_1^w \cdots q_n^w$ be the sequence of states visited in the run of S_j on w (that is, $q_0^w = j$). The probability of observing a state sequence $q_0 \cdots q_n$ in the Markov chain is (when the initial distribution λ has $\lambda(q_0) = \lambda(j) = 1$)

$$\begin{aligned}\Pr(q_0 \cdots q_n) &= \prod_{i=0}^{n-1} \sum_{a\in\Sigma} p(a) 1_{\delta(q_i,a)=q_{i+1}} \\ &= \sum_{a_1,\ldots,a_n\in\Sigma} p(a_1) 1_{\delta(q_0,a_1)=q_1} \cdots p(a_n) 1_{\delta(q_{n-1},a_n)=q_n},\end{aligned}$$

where we have used the fact that the Cauchy product of two absolutely convergent series is convergent.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

As for all integers i with $0 \leq i \leq n$, we have $\delta(q_{i-1}^w, w_i) = q_i^w$, we obtain

$$\sum_{a_1,\ldots,a_n \in \Sigma} p(a_1)1_{\delta(q_0,a_1)=q_1} \cdots p(a_n)1_{\delta(q_{n-1},a_n)=q_n} = \mu_p(\{a_1 \cdots a_n : \mathbf{q}_{S_j}(a_1 \cdots a_n) = q_0 \cdots q_n\}),$$

and hence,

$$\Pr(q_0 \cdots q_n) = \mu_p(\{w : \mathbf{q}_{S_j}(w) = q_0 \cdots q_n\}). \tag{5.4}$$

Thus, as S is deterministic and every $w_1 \cdots w_n \in \Sigma^n$ occurs along exactly one path of states in S, we have

$$\begin{aligned}\Pr\left(\left|\frac{V_i(n)}{n} - \pi(i)\right| \geq \epsilon\right) &= \sum_{q_0q_1\cdots q_n \in Q^n} \Pr(q_0 \cdots q_n)1_{|V_i(n)/n-\pi(i)|\geq\epsilon} \\ &= \sum_{q_0q_1\cdots q_n \in Q^n} \mu_p(\{w_1 \cdots w_n : \mathbf{q}_{S_j}(w_1 \cdots w_n) = q_0 \cdots q_n\}) \\ &= \mu_p\left(w : \left|\frac{V_i(n)}{n} - \pi(i)\right| \geq \epsilon\right). \end{aligned} \tag{5.5}$$

Hence, by equations (5.3) and (5.5), we have

$$\lim_{n\to\infty} \mu_p\left(w : \left|\frac{V_i(n)}{n} - \pi(i)\right| \geq \epsilon\right) = 0. \tag{5.6}$$

If $\mathbf{q}_{S_j}(w) = q_0 \cdots q_n$ and $q_k \in Q_F$ for some k with $0 \leq k \leq n-1$, then S_j selects w_{k+1}. Set $c = \min_{q_i \in Q_F} \pi(i)$ (c is well defined as $Q_F \neq \emptyset$), and let $i \in Q_F$ be such that $\pi(i) = c$. Then, for all $j \in Q$,

$$\begin{aligned}&\mu_p(E_n(c-\epsilon, j)) \\ &= \mu_p\left(\{w \in \Sigma^n : |S_j[w]| \leq (c-\epsilon)n\}\right) \leq \mu_p\left(\{w \in \Sigma^n : V_i(n) \leq (c-\epsilon)n\}\right) \\ &= \mu_p\left(\left\{w \in \Sigma^n : \frac{V_i(n)}{n} - c \leq -\epsilon\right\}\right) \leq \mu_p\left(\left\{w \in \Sigma^n : \left|\frac{V_i(n)}{n} - c\right| \geq \epsilon\right\}\right).\end{aligned}$$

Hence, by equation (5.6), we have $\lim_{n\to\infty} \mu_p(E_n(c-\epsilon, j)) = 0$, and as $j \in Q$ was arbitrary, we obtain

$$\begin{aligned}\lim_{n\to\infty} \mu_p(E_n(c-\epsilon)) &= \lim_{n\to\infty} \mu_p(\bigcup_{j\in Q} E_n(c-\epsilon, j) \leq \lim_{n\to\infty} \sum_{j\in Q} \mu_p(E_n(c-\epsilon, j)) \\ &= \sum_{j\in Q} \lim_{n\to\infty} \mu_p(E_n(c-\epsilon, j)) = 0,\end{aligned}$$

as desired. □

LEMMA 5.5. *Let S be a strategy, $a \in \Sigma$, b, ϵ be real numbers with $0 < b \leq 1$ and $\epsilon > 0$, and $p : \Sigma \longrightarrow [0, 1]$ be a positive Bernoulli distribution. Define, for all positive integers n,*

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

$$H_n(b,\epsilon) = \left\{ w \in \Sigma^n : |S(w)| > bn \text{ and } \left| p(a) - \frac{\#_a(S(w))}{|S(w)|} \right| \geq \epsilon \right\}$$
$$= \bigcup_{bn<\ell\leq n} \left\{ w \in \Sigma^n : S(w) \in \Sigma^\ell \text{ and } \left| p(a) - \frac{\#_a(S(w))}{\ell} \right| \geq \epsilon \right\}.$$

Then,

$$\lim_{n\to\infty} \mu_p(H_n(b,\epsilon)) = 0.$$

Proof. Define

$$F_n(b,\epsilon) = \bigcup_{bn<\ell\leq n} \left\{ y \in \Sigma^\ell : \left| p(a) - \frac{\#_a(y)}{\ell} \right| \geq \epsilon \right\}.$$

Observe that $H_n(b,\epsilon) = \{w \in \Sigma^n : S(w) \in F_n(b,\epsilon)\}$. By definition of μ_p, if $v = v_1 \cdots v_n \in \Sigma^n$, then the set $T_v = \{w \in \Sigma^* : \text{there exists } u_0, \ldots, u_n \in \Sigma^*. w = u_0 v_1 u_1 \cdots v_n u_n\}$ satisfies $\mu_p(T_w) \leq \mu_p(v)$, and hence in particular if S is any strategy, then $\mu_p(\{w : S(w) = v\}) \leq \mu_p(v)$. Thus, $\mu_p(H_n(b,\epsilon)) \leq \mu_p(F_n(b,\epsilon))$ for all n, and it thus suffices to prove that $\lim_{n\to\infty} \mu_p(F_n(b,\epsilon)) = 0$.

Consider the stochastic variable X_a that is 1 when a is picked from Σ with probability $p(a)$, and 0 otherwise. Then, the mean of X_a is $p(a)$ and the variance of X_a is $p(a)(1 - p(a))$. Now consider performing $\ell \geq 1$ independent Bernoulli trials drawn according to X_a. Define $q : \{0,1\}^+ \longrightarrow [0,1]$ inductively by $q(1) = p(a)$, $q(0) = 1 - p(a)$, and $q(1c) = q(c)p(a)$ and $q(0c) = q(c)(1 - p(a))$ for $c \in \{0,1\}^+$, and observe that q induces a probability distribution $\bar{q}$ on Σ^ℓ by setting $\bar{q}(w) = q(w)$. Now, for any $v \in \{0,1\}^\ell$, $\bar{q}(v)$ is the probability of obtaining v by performing ℓ independent Bernoulli trials as above.

Define the stochastic variable $X_a^\ell = X_a + X_a + \cdots + X_a$ (ℓ times). Then, X^ℓ counts the number of occurrences of a by performing the ℓ repeated Bernoulli trials.

By the Chernoff bound, X_a^ℓ satisfies

$$\Pr\left(\left| p(a) - \frac{X_a^\ell}{\ell} \right| \geq \epsilon \right) \leq 2e^{-\ell\epsilon^2/3p(a)}. \tag{5.7}$$

Define the map $g : \Sigma \longrightarrow \{0,1\}$ by $g(a) = 1$ and $g(b) = 0$ for all $b \in \Sigma \setminus \{a\}$. Clearly, g extends homomorphically to a map $\tilde{g} : \Sigma^\ell \longrightarrow \{0,1\}^\ell$ by setting $\tilde{g}(c_1 c_2 \cdots c_\ell) = g(c_1)g(c_2)\cdots g(c_\ell)$.

CLAIM. *For any $u \in \{0,1\}^\ell$,*

$$\bar{q}(u) = \mu_p(\{y \in \Sigma^\ell : \tilde{g}(y) = u\}). \tag{5.8}$$

Proof of claim. By induction on ℓ:

- if $\ell = 1$, then if $u = 0$, we have $\{y \in \Sigma^\ell : \tilde{g}(y) = u\} = \Sigma \setminus \{a\}$ and thus,
$$\bar{q}(u) = \bar{q}(0) = q(0) = 1 - p(a) = \sum_{b\in\Sigma\setminus\{a\}} p(b) = \mu_p(\Sigma \setminus \{a\}).$$
Similarly, if $u = 1$, we have $\{y \in \Sigma^\ell : \tilde{g}(y) = u\} = \{a\}$ and thus, $\bar{q}(u) = \bar{q}(1) = q(1) = p(a) = \mu_p(\{a\})$, as desired;

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

- if $\ell > 1$, write $u = b_1 \cdots b_{\ell-1} b_\ell$; by the induction hypothesis,

$$\bar{q}(b_1 \cdots b_{\ell-1}) = \mu_p(\{y \in \Sigma^\ell : \tilde{g}(y') = b_1 \cdots b_{\ell-1}\}) = \sum_{\substack{y' \in \Sigma^{\ell-1} \\ \tilde{g}(y') = b_1 \cdots b_{\ell-1}}} \mu_p(y').$$

If $b_\ell = 0$, then

$$\begin{aligned}
\bar{q}(b_1 \cdots b_{\ell-1} b_\ell) &= \bar{q}(b_1 \cdots b_{\ell-1}) q(0) \\
&= \bar{q}(b_1 \cdots b_{\ell-1})(1 - p(a)) \\
&= \sum_{\substack{y' \in \Sigma^{\ell-1} \\ \tilde{g}(y') = b_1 \cdots b_{\ell-1}}} \mu_p(y')(1 - p(a)) \\
&= \sum_{\substack{y' \in \Sigma^{\ell-1} \\ \tilde{g}(y') = b_1 \cdots b_{\ell-1}}} \left(\mu_p(y') \sum_{c \in \Sigma \setminus \{a\}} p(c) \right) \\
&= \sum_{\substack{y' \in \Sigma^{\ell-1} \\ \tilde{g}(y') = b_1 \cdots b_{\ell-1}}} \sum_{c \in \Sigma \setminus \{a\}} \mu_p(y') p(c) \\
&\overset{(\dagger)}{=} \sum_{\substack{y' \in \Sigma^{\ell-1} \\ \tilde{g}(y') = b_1 \cdots b_{\ell-1} \\ c \in \Sigma \setminus \{a\}}} \mu_p(y'c) \\
&= \sum_{\substack{y \in \Sigma^{\ell} \\ \tilde{g}(y) = b_1 \cdots b_{\ell-1} b_\ell}} \mu_p(y) \\
&= \mu_p(\{y \in \Sigma^\ell : \tilde{g}(y) = b_1 \cdots b_{\ell-1} b_\ell\}),
\end{aligned}$$

where (†) follows as both series on the left- and right-hand sides of the equality are absolutely convergent. The proof for the case $b_\ell = 1$ is symmetric, *mutatis mutandis*.

□

Observe that, for any $y \in \Sigma^\ell$, we have

$$|p(a) - \#_1(\tilde{g}(y))/\ell| \geq \epsilon \quad \text{if and only if} \quad |p(a) - \#_a(y)/\ell| \geq \epsilon. \tag{5.9}$$

Hence, by equation (5.8), for any event $\mathcal{U} \subseteq \{0,1\}^\ell$, we have

$$\begin{aligned}
\Pr(\mathcal{U}) &= \sum_{u \in \mathcal{U}} \bar{q}(u) = \sum_{u \in \mathcal{U}} \mu_p(\{y \in \Sigma^\ell : \tilde{g}(y) = u\})) \\
&= \mu_p(\{y \in \Sigma^\ell : \tilde{g}(y) \in \mathcal{U}\}).
\end{aligned} \tag{5.10}$$

The event $|p(a) - X_a^\ell/\ell| \geq \epsilon$ is shorthand for the set

$$\left\{ u \in \{0,1\}^\ell : \left| p(a) - \frac{\sum_{j=1}^\ell u_j}{\ell} \right| \geq \epsilon \right\} = \left\{ u \in \{0,1\}^\ell : \left| p(a) - \frac{\#_1(u)}{\ell} \right| \geq \epsilon \right\}.$$

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

We thus obtain

$$\begin{aligned}\Pr\left(\left|p(a)-\frac{X_a^\ell}{\ell}\right|\geq\epsilon\right) &= \Pr\left(\left\{u\in\{0,1\}^\ell : \left|p(a)-\frac{\#_1(u)}{\ell}\right|\geq\epsilon\right\}\right)\\ &= \mu_p\left(\left\{y\in\Sigma^\ell : \left|p(a)-\frac{\#_1(\tilde{g}(y))}{\ell}\right|\geq\epsilon\right\}\right) \quad \text{by equation (5.10)}\\ &= \mu_p\left(\left\{y\in\Sigma^\ell : \left|p(a)-\frac{\#_a(y)}{\ell}\right|\geq\epsilon\right\}\right) \quad \text{by equation (5.9).}\end{aligned} \tag{5.11}$$

Observe that

$$\begin{aligned}\mu_p\left(F_n(b,\epsilon)\right) &= \mu_p\left(\bigcup_{bn<\ell\leq n}\left\{y\in\Sigma^\ell\cap F_n(b,\epsilon) : \left|p(a)-\frac{\#_a(y)}{\ell}\right|\geq\epsilon\right\}\right)\\ &= \sum_{bn<\ell\leq n}\mu_p\left(\left\{y\in\Sigma^\ell\cap F_n(b,\epsilon) : \left|p(a)-\frac{\#_a(y)}{\ell}\right|\geq\epsilon\right\}\right)\\ &\leq \sum_{bn<\ell\leq n}\mu_p\left(\left\{y\in\Sigma^\ell : \left|p(a)-\frac{\#_a(y)}{\ell}\right|\geq\epsilon\right\}\right)\\ &= \sum_{bn<\ell\leq n}\Pr\left(\left|p(a)-\frac{X_a^\ell}{\ell}\right|\geq\epsilon\right) \quad \text{by equation (5.11)}\\ &\leq \sum_{bn<\ell\leq n}2e^{-\ell\epsilon^2/3p(a)} \quad \text{by equation (5.7)}\\ &\leq (1-b)n2e^{-bn\epsilon^2/3p(a)}.\end{aligned}$$

Thus, $\lim_{n\to\infty}\mu_p(F_n(b,\epsilon))=0$, as desired. □

COROLLARY 5.6. *Let b,ϵ be real numbers with $0<b\leq 1$ and $\epsilon>0$. Then,*

$$\lim_{n\to\infty}\mu_p(G_n(b,\epsilon))=0.$$

Proof. By Lemma 5.5 with S the strategy defined by the automaton A_q, we obtain that

$$\lim_{n\to\infty}\mu_p(G_n(b,\epsilon,q))=0$$

and as $G_n(b,\epsilon)=\bigcup_{q\in Q}G_n(b,\epsilon,q)$, we have

$$\mu_p(G_n(b,\epsilon))\leq\sum_{q\in Q}\mu_p(G_n(b,\epsilon,q)).$$

As Q is finite, we hence obtain $\lim_{n\to\infty}\mu_p(G_n(b,\epsilon))=0$. □

LEMMA 5.7. *There is a real number b with $0<b\leq 1$ such that for all $\epsilon>0$,*

$$\lim_{n\to\infty}\mu_p(D_n^p(b,\epsilon))=1.$$

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

Proof. Observe that, for all b with $0 < b \leq 1$,

$$\begin{aligned}
&\Sigma^n \setminus D_n^p(b,\epsilon)) \\
&\quad = \{w \in \Sigma^n : \text{there exists } q \in Q . |A_q[w]| \leq bn\} \\
&\quad \cup \left\{ w \in \Sigma^n : \text{there exists } q \in Q . |A_q[w]| > bn \wedge \sup_{a\in\Sigma} \left| \frac{\#_a(A_q[w])}{|A_q[w]|} - p(a) \right| \geq \epsilon \right\} \\
&\quad = \left(\bigcup_{q\in Q} E_n(b,q) \right) \cup \left(\bigcup_{q\in Q} G_n(b,\epsilon,q) \right)
\end{aligned}$$

and thus,

$$\begin{aligned}
\mu_p(\Sigma^n \setminus D_n^p(b,\epsilon)) &\leq \mu_p\left(\bigcup_{q\in Q} E_n(b,q) \right) + \mu_p\left(\bigcup_{q\in Q} G_n(b,\epsilon,q) \right) \\
&= \mu_p(G_n(b,\epsilon)) + \mu_p(E_n(b)).
\end{aligned}$$

By Theorem 5.4, choose a real number $c > 0$ such that $\lim_{n\to\infty} \mu_p(E_n(c-\epsilon)) = 0$, and set $b = c - \epsilon$.

By Theorem 5.6, we obtain that $\lim_{n\to\infty} G_n(b,\epsilon) = 0$, and thus $\lim_{n\to\infty} \mu_p(\Sigma^n \setminus D_n^p(b,\epsilon)) = 0$. The result now follows by $\mu_p(D_n^p(b,\epsilon)) = 1 - \mu_p(\Sigma^n \setminus D_n^p(b,\epsilon))$. □

LEMMA 5.8. *Let $p : \Sigma \longrightarrow [0,1]$ be a probability distribution, let $\alpha \in \Sigma^\omega$ be μ_p-block-distributed, and A a strongly connected DFA over Σ. Then, for all $a \in \Sigma$, the limiting frequency of a in the sequence $\beta = A[\alpha]$ exists and equals $p(a)$.*

Proof. For each n, r, let $\beta_{(n,r)}$ be the sequence of symbols picked out from the block $\alpha_{(n,r)}$ when A is applied to α; note that each $\beta_{(n,r)}$ has length between 0 and n.

For each positive integer m, define

$$L_m = \sum_{i=1}^{m} |\beta_{(n,i)}|$$

and for each $a \in \Sigma$, define ρ_a^m by

$$\rho_a^m = \frac{\sum_{i=1}^{m} \#_a(\beta_{(n,i)})}{L_m}.$$

To prove the lemma, it suffices to show that, for any real number $\epsilon > 0$, then for all sufficiently large m, we have $|\rho_a^m - p(a)| < \epsilon$.

Define

$$I_m = \left\{ i \leq m : \beta_{(n,i)} \notin D_n^p\left(b, \frac{\epsilon}{2}\right) \right\},$$

where $b < 1$ is a constant to be fixed later in the proof.

Define

$$\ell_m = \sum_{i\in I_m} |\beta_{(n,i)}|$$

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

Now, define θ_a^m by

$$\theta_a^m = \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{\sum_{i\in\{1,\dots,m\}\setminus I_m} |\beta_{(n,i)}|} = \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m - \ell_m}.$$

That is, θ_a^m is the frequency of occurrences of a when the blocks $\beta_{(n,i)]}$ picked out from blocks $\alpha_{(i,r)} \in D_n^p(b, \epsilon/2)$ with $i \le m$ are all concatenated. Observe that, by definition of D_n^p, we have $|\theta_a^m - p(a)| < \epsilon/2$.

We have

$$\begin{aligned}
\rho_a^m - \theta_a^m &= \frac{\sum_{i=1}^m \#_a(\beta_{(n,i)})}{L_m} - \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m - \ell_m} \\
&= \left(\frac{\sum_{i\in I_m} \#_a(\beta_{(n,i)})}{L_m} + \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m}\right) - \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m - \ell_m} \\
&= \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m} - \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m - \ell_m} + \frac{\sum_{i\in I_m} \#_a(\beta_{(n,i)})}{L_m} \\
&\le \frac{\sum_{i\in I_m} \#_a(\beta_{(n,i)})}{L_m} \le \frac{\sum_{i\in I_m} |\beta_{(n,i)}|}{L_m} = \frac{\ell_m}{L_m},
\end{aligned} \tag{5.12}$$

where the penultimate inequalities in the last line above follows because $L_m \ge L_m - \ell_m$ implies $\sum_{i\in\{1,\dots,m\}\setminus I} \#_a(\beta_{(n,i)})/L - \sum_{i\in\{1,\dots,m\}\setminus I} \#_a(\beta_{[n,i]})/L - \ell \le 0$, and the final inequality follows because $\sum_{i\in I_m} \#_a(\beta_{(n,i)}) \le \sum_{i\in I} |\beta_{(n,i)}| = \ell_m$.

By basic algebra, we have

$$\frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m} - \frac{\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m - \ell_m} = \frac{-\ell_m \sum_{i\in\{1,\dots,m\}\setminus I} \#_a(\beta_{(n,i)})}{L_m(L_m - \ell_m)}$$

and as

$$\sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)}) \le \sum_{i\in\{1,\dots,m\}\setminus I_m} |\beta_{(n,i)}| = L_m - \ell_m,$$

we conclude that

$$\frac{-\ell_m \sum_{i\in\{1,\dots,m\}\setminus I_m} \#_a(\beta_{(n,i)})}{L_m(L_m - \ell_m)} \ge -\frac{\ell_m}{L_m}$$

and thus by the first line of equation (5.12) and the fact that $\ell_m \le L_m$, we have

$$\rho_a^m - \theta_a^m - \frac{\sum_{i\in I_m} \#_a(\beta_{(n,i)})}{L_m} \ge -\frac{\ell_m}{L_m},$$

whence $-\ell_m/L_m \le \rho_a^m - \theta_a^m$, which combined with equation (5.12) yields $|\rho_a^m - \theta_a^m| \le \ell_m/L_m$.

By Theorem 5.7, pick a b such that for all $\epsilon > 0$, we have $\lim_{n\to\infty} \mu_p(D_n^p(b, \epsilon)) = 1$. Choose $\delta > 0$ with $\delta < b\epsilon/8$, and pick $n \in \mathbb{N}$ such that $\mu_p(D_n^p(b, \epsilon)) > 1 - \delta$. Now, pick $\gamma < b\epsilon/8$. Because α is μ_p-block-distributed, there exists $M \in \mathbb{N}$ such that for all $k \ge M$ and all $B \subseteq \Sigma^n$, the prefix $\alpha|_{\le kn}$ satisfies

$$\left|\frac{|\{i \le k : \alpha_{(n,i)} \in B\}|}{k} - \mu_p(B)\right| < \gamma.$$

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

In the particular case $B = D_n^p(b, \epsilon/2)$, we thus have

$$\left| \frac{|\{i \leq k : \alpha_{(n,i)} \in D_n^p(b, \epsilon/2)\}|}{k} - \mu_p(D_n^p(b, \epsilon/2)) \right| < \gamma$$

and thus,

$$\begin{aligned} 1 - \delta - \frac{|\{i \leq k : \alpha_{(n,i)} \in D_n^p(b, \epsilon/2)\}|}{k} \\ \leq \mu_p(D_n^p(b, \epsilon/2)) - \frac{|\{i \leq k : \alpha_{(n,i)} \in D_n^p(b, \epsilon/2)\}|}{k} < \gamma, \end{aligned}$$

whence we conclude

$$\left| \left\{ i \leq k : \alpha_{(n,i)} \in D_n^p\left(b, \frac{\epsilon}{2}\right) \right\} \right| > k(1 - \delta - \gamma). \tag{5.13}$$

By definition of $D_n^p(b, \epsilon/2)$, as remarked above, every $\alpha_{(n,i)} \in D_n^p(b, \epsilon/2)$ satisfies $|A_{\delta^*(q_0, \alpha_{\leq ni})}[\alpha_{(n,i)}]| > bn$, whichever state $\delta^*(q_0, \alpha_{\leq ni})$ is reached at the beginning of the block. We will in the following simply write $|A[\alpha_{(n,i)}]|$ to simplify notation. We thus have

$$L_m = \sum_{i=1}^{m} |\beta_{(n,i)}| = \sum_{i=1}^{m} |A[\alpha_{(n,i)}]| \geq \left| \left\{ i \leq m : \alpha_{(n,i)} \in D_n^p\left(b, \frac{\epsilon}{2}\right) \right\} \right| bn > m(1 - \delta - \gamma)bn. \tag{5.14}$$

Furthermore, by the definition of I_m and equation (5.13),

$$\begin{aligned} |I_m| &= \left| \left\{ i \leq m : \alpha_{(n,i)} \notin D_n^p\left(b, \frac{\epsilon}{2}\right) \right\} \right| = m - \left| \left\{ i \leq m : \alpha_{(n,i)} \in D_n^p\left(b, \frac{\epsilon}{2}\right) \right\} \right| \\ &< m - m(1 - \delta - \gamma) = m(\delta + \gamma). \end{aligned}$$

However, then,

$$\ell_m = \sum_{i \in I_m} |\beta_{(i,n)}| \leq |I_m| n < mn(\delta + \gamma) \tag{5.15}$$

and thus by equations (5.14) and (5.15),

$$\frac{\ell_m}{L_m} < \frac{mn(\delta + \gamma)}{m(1 - \delta - \gamma)bn} = \frac{\delta + \gamma}{b(1 - \delta - \gamma)} < \frac{b\epsilon/8 + b\epsilon/8}{b(1 - b\epsilon/8 - b\epsilon/8)} < \frac{\epsilon/4}{1 - 1/4} < \frac{\epsilon}{2},$$

where we have used that $b\epsilon < 1$ in the penultimate inequality.

We now finally have

$$|\rho_a^m - p(a)| \leq |\rho_a^m - \theta_a^m| + |\theta_a^m - p(a)| < \frac{\ell_m}{L_m} + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

concluding the proof. □

LEMMA 5.9. *Let Σ be an alphabet, p a positive Bernoulli distribution on Σ, let $\alpha \in \Sigma^\omega$ be μ_p-distributed, and let A be a strongly connected DFA over Σ. Then, $A[\alpha]$ is μ_p-distributed.*

Proof. By Lemma 5.3, it suffices to show for every $a \in \Sigma$ and every strongly connected A that the limiting frequency of a in $A[\alpha]$ exists and equals $p(a)$. As α is μ_p-distributed, it

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

follows from Proposition 5.1 that it is μ_p-block-distributed, and the result then immediately follows by Lemma 5.8. □

6. *An application in symbolic dynamics: characterizing measures where genericity is preserved by DFAs*

We now show an application of the main result to the area of symbolic dynamical systems. The following section recalls basic facts about symbolic dynamical systems, including establishing the correspondence between probability maps on Σ^* and probability measures on full shifts.

6.1. *Shift spaces and genericity.*

We briefly introduce basic notions; full accounts can be found in standard textbooks, e.g., [30].

Definition 6.1. Let Σ be a non-empty alphabet. The (one-sided) *shift* $s : \Sigma^\omega \longrightarrow \Sigma^\omega$ is the map defined by $s(a_1a_2a_3\cdots) = a_2a_3\cdots$. A *shift space* is a pair (X, s), where $X \subseteq \Sigma^\omega$ is a closed (in the product topology on Σ^ω when Σ is endowed with the discrete topology) subset such that $s(X) = X$ (for one-sided shifts, some authors require only $s(X) \subseteq X$; we shall not do so here) and s is the restriction of the shift to X.

As usual, we consider the σ-algebra $\mathcal{C}$ on Σ^ω having the set of cylinders $\{[w] : w \in \Sigma^*\}$ as basis. All measures μ in the remainder of the paper are understood to be measures on $(\Sigma^\omega, \mathcal{C})$.

The standard example of probability measures on shift spaces is the set of Bernoulli measures [47].

Definition 6.2. A probability measure on the shift space (Σ^ω, s) is a probability measure on Σ^ω with the σ-algebra generated by the cylinder sets $\{[v] : v \in \Sigma^*\}$. A probability measure $\bar{\mu}$ on the full shift is a *Bernoulli measure* if there is a probability distribution $p : \Sigma \longrightarrow [0, 1]$ such that the measure of each cylinder satisfies $\bar{\mu}([a_1 \cdots a_n]) = \prod_{i=1}^{n} p(a_i)$. In this case, we say that $\bar{\mu}$ is *induced* by p.

Definition 6.3. Let (X, s) be a shift space. A probability measure $\bar{\mu}$ on X is said to be *shift invariant* if $\bar{\mu}(s^{-1}(A)) = \bar{\mu}(A)$ for all $A \subseteq X$. A finite word $w \in \Sigma^k$ is said to be *admissible* for μ if $\bar{\mu}([w]) > 0$.

A right-infinite sequence $\alpha \in \Sigma^\omega$ is said to be *generic* for $\bar{\mu}$ if, for all words w admissible for $\bar{\mu}$, we have

$$\lim_{n\to\infty} \frac{\#_w(\alpha|_{\leq n})}{n} = \bar{\mu}([w]).$$

That is, w occurs in α with limiting frequency $\bar{\mu}([w])$.

The study of probability measures on the full shift is cryptomorphic to the study of invariant probability maps; this folklore result is contained in the following two propositions (proofs can be found in Appendix A).

PROPOSITION 6.1. *Every invariant probability map* $\mu : \Sigma^* \longrightarrow [0, 1]$ *induces a shift-invariant probability measure* $\bar{\mu} : \Sigma^\omega \longrightarrow [0, 1]$ *by setting* $\bar{\mu}([w]) = \mu(w)$.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

Conversely, every probability measure $\nu : \Sigma^\omega \longrightarrow [0,1]$ *induces a probability map* $\underline{\nu} : \Sigma^* \longrightarrow [0,1]$ *by defining* $\underline{\nu}(w) = \nu([w])$*; if* ν *is shift-invariant, then* $\underline{\nu}$ *is invariant. Furthermore,* $\mu = \underline{(\bar{\mu})}$ *and* $\nu = \overline{(\underline{\nu})}$.

PROPOSITION 6.2. *Let* $\mu : \Sigma^* \longrightarrow [0,1]$ *be a probability map. The following are equivalent:*

(1) *there exists a* μ*-distributed* $\alpha \in \Sigma^\omega$*;*
(2) μ *is invariant;*
(3) *there exists a shift-invariant probability measure* ν *on* Σ^ω *such that* $\bar{\mu} = \nu$.

Conversely, let ν *be a probability measure on* Σ*. The following are equivalent:*

(1) *there exists* $\alpha \in \Sigma^\omega$ *that is generic for* ν*;*
(2) ν *is shift-invariant;*
(3) *there exists an invariant probability map* $\mu : \Sigma^* \longrightarrow [0,1]$ *such that* $\underline{\nu} = \mu$.

It follows that the shift-invariant probability measures ν on the full shift such that genericity is preserved by finite-state selection are exactly the Bernoulli measures.

THEOREM 6.3. *Let* Σ *be a non-empty alphabet and let* ν *be a shift-invariant measure on the full shift* (Σ^ω, s)*. Then, every finite-state selector preserves genericity if and only if* ν *is a Bernoulli measure such that all words in* Σ^* *are admissible.*

Proof. Observe that for a Bernoulli measure $\bar{\mu}$ on the full shift on Σ, all words are admissible if and only if $\bar{\mu}(a) > 0$ for all $a \in \Sigma$. The theorem now follows from Theorem 3.1 and Proposition 6.2. □

7. *Future work*

The most obvious extension of our main results is to attempt to relax the requirement that selection is done by a DFA by using methods similar to those of Kamae and Weiss [27], and Kamae and Wang [56], where reasoning using a combination of density arguments and relaxed finiteness conditions on the syntactic monoid of the strategy (using our terminology) have been used for normal sequences over binary alphabets. We conjecture that some of these techniques can be adapted to positive Bernoulli distributions on arbitrary finite alphabets. Notably, Auslander and Dowker proved some 45 years ago a general result on dynamical systems on compact metric spaces equipped with invariant Borel probability measures that, as a special case, identifies a very broad class of transformations (essentially selectors in the vernacular of the present paper) that preserve μ-distributed sequences if μ is induced by a Bernouilli distribution on Σ, generalizing the work of Wall [55], Furstenberg [23], and others on preservation of normality by various, more restrictive, classes of selectors. Also of interest would be to ascertain whether the 'if' in Auslander and Dowker's work (or, less ambitiously, for more restricted classes of selectors) can be extended to 'if and only if'.

A different possible thrust is to consider generalizations of Agafonov's theorem on domains different from infinite sequence over alphabets. However, some results in the—sparse—literature on selection from normal sequence-like objects in other contexts are negative; for example, normality is not preserved by arithmetic progressions (so,

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

probably not by finite-state selectors in any reasonable sense) for continued fraction expansions [24]. However, very recent work by Bergelson *et al* has succesfully adapted the classical techniques of Kamae and Weiss [27] to show that certain Følner sequences preserve (the appropriate analog of) normality in cancellative amenable semigroups [7].

Acknowledgement. T. Seiller was partially supported by the ANR ANR-22-CE48-0003-01 project DySCo.

A. *Appendix. Auxiliary proofs and definitions*

A.1. *Automata and selectors.*

The following is a proof of the extension of [44, Lemma 2.6]. The proof follows the original in most details.

Definition A.1. Let $G = (V, E)$ be a directed multigraph, and denote by $\sim \subseteq V \times V$ the equivalence relation such that $v \sim w$ if and only if there is a directed path from v to w and a directed path from w to v. For every $v \in V$, denote by $[v]_\sim$ the equivalence class containing v. Define the partial order $<$ on $V/\sim$ by $\mathcal{V} < \mathcal{W}$ if and only if there are $v \in \mathcal{V}$ and $w \in \mathcal{W}$ such that there is a directed path from w to v.

If G has a finite number of nodes, $<$ is clearly well founded. As $<$ is clearly also transitive, every $\mathcal{W} \in V/\sim$ satisfies $\mathcal{W} \geq \mathcal{V}$ for some $<$-minimal $\mathcal{V} \in V/\sim$.

Also observe that every $<$-minimal $\mathcal{V}$ is a recurrent strongly connected component, because (i) it is strongly connected by definition, and (ii) $<$-minimality implies that no directed path from any node in $\mathcal{V}$ can reach a node in a strongly connected component distinct from $\mathcal{V}$.

LEMMA A.1. *Let $S = (Q, \delta, q_s, Q_F)$ be a finite automaton over a (possibly infinite) alphabet Σ. Then, there is a word $w \in \Sigma^*$ such that, for all states $q \in Q$, $\delta^*(q, w)$ is a state in a $<$-minimal element of $Q/\sim$.*

Proof. Write $Q = \{q_1, \ldots, q_m\}$. We prove by induction on $i \leq m$ that there is a word $w_i \in \Sigma^*$ such that for all $j \leq i$, $\delta^*(q_j, w_i)$ is a state in a $<$-minimal element of $Q/\sim$.

$i = 1$. Let $\mathcal{V}$ be a $<$-minimal element of $Q/\sim$ such that $[q_1]_\sim > \mathcal{V}$. Choose $q \in Q$ such that $[q]_\sim = \mathcal{V}$. Then, there is a directed path from q_1 to q. Let w_1 be the word along that path and observe that $\delta^*(q_1, w_1) = q$.

$i > 1$. Let $\mathcal{V}$ be a $<$-minimal element of $Q/\sim$ such that $\delta^*(q_{i+1}, w_i) \in \mathcal{V}$, and let $q \in \mathcal{V}$, whence there is a directed path from $\delta^*(q_{i+1}, w_i)$ to q. Let $w' \in \Sigma^*$ be the word along that path, whence $\delta^*(\delta^*(q_{i+1}, w_i), w') = q$. Define $w_{i+1} = w_i \cdot w'$ and observe that $\delta^*(q_{i+1}, w_{i+1}) = q$.

For $j \leq i$, we claim that $\delta^*(q_j, w_{i+1})$ is a state in a $<$-minimal element of $Q/\sim$. For, by the induction hypothesis, $\delta(q_j, w_i)$ is in a $<$-minimal element $\mathcal{V}_j$ of $Q/\sim$, and as $<$-minimal elements are recurrent strongly connected components, no directed path from $\delta(q_j, w_i)$ can end in a state outside $\mathcal{V}_j$. □

Proof of Lemma 2.6. Any $<$-minimal element of $Q/\sim$ is recurrent. By Lemma A.1, there is a word w such that from any state $q \in Q$, $\delta^*(q, w)$ is a state in a recurrent strongly

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

connected component of the automaton. As $p(a) > 0$ for all $a \in \Sigma$, $\mu_p(w) > 0$, and as α is p-distributed, w thus occurs (infinitely often) in α. After the first occurrence of w, the run of A on α has entered a strongly recurrent connected component. □

A.2. *Symbolic dynamical systems.*

Proof of Proposition 6.1. The two identities $\mu = \underline{\bar{\mu}}$ and $\nu = \bar{\underline{\nu}}$ follow directly from the definitions. If μ is invariant, then for every cylinder $[w]$, we have $\bar{\mu}([w]) = \mu(w) = \sum_{a\in\Sigma} \mu(w \cdot a) = \sum_{a\in\Sigma} \bar{\mu}([w \cdot a])$; from this, and the observation that $1 = \mu(\lambda) = \bar{\mu}([\lambda]) = \bar{\mu}(\Sigma^\omega)$, $\bar{\mu}$ is a probability measure on Σ with the sigma algebra generated by the cylinder sets. In addition, as μ is invariant, we have for any cylinder $[w]$ that

$$\bar{\mu}(S^{-1}([w])) = \bar{\mu}\left(\bigcup_{a\in\Sigma}[a \cdot w]\right) = \sum_{a\in\Sigma} \bar{\mu}([a \cdot w]) = \sum_{a\in\Sigma} \mu(a \cdot w) = \mu(w) = \bar{\mu}([w]),$$

whence $\bar{\mu}$ is shift-invariant.

Conversely, if ν is a shift-invariant probability measure on Σ^ω, we have for any w that

$$\underline{\nu}(w) = \nu([w]) = \nu\left(\bigcup_{a\in\Sigma}[w \cdot a]\right) = \sum_{a\in\Sigma} \nu([w \cdot a]) = \sum_{a\in\Sigma} \underline{\nu}(w \cdot a)$$

and

$$\underline{\nu}(w) = \nu([w]) = \nu(S^{-1}([w])) = \nu\left(\bigcup_{a\in\Sigma}[a \cdot w]\right) = \sum_{a\in\Sigma} \nu([a \cdot w]) = \sum_{a\in\Sigma} \underline{\nu}(a \cdot w),$$

showing that $\underline{\nu}$ is invariant. □

Proof of Proposition 6.2. For the first part, we prove (1) ⇒ (2) ⇒ (3) ⇒ (1).

Assume that $\alpha \in \Sigma^\omega$ is μ-distributed. Observe that every occurrence of a word of the form $a \cdot w$ in α contains an occurrence of w, and hence, $\#_w(\alpha|_{\leq n}) \geq \sum_{a\in\Sigma} \#_{a\cdot w}(\alpha|_{\leq n})$. Conversely, for every occurrence of w starting at some position $i \geq 2$ in α, there is exactly one $a \in \Sigma$ such that the word $a \cdot w$ occurs at position $i - 1$, whence $\#_w(\alpha|_{\leq n}) \leq 1 + \sum_{a\in\Sigma} \#_{a\cdot w}(\alpha|_{\leq n})$.

Let $\preceq_\alpha$ be any total ordering on Σ. List the elements of Σ as $a_1 \preceq_\alpha a_2 \preceq_\alpha a_3 \cdots$. For any positive integer m and any $\epsilon > 0$, choose, by μ-distribution of α, n large enough that $|\#_w(\alpha|_{\leq n})/n - \mu(w)| < \epsilon/2$ and $|\#_{a_i\cdot w}(\alpha|_{\leq n})/n - \mu(a_i \cdot w)| < \epsilon/(2m)$ for all $i \in \{1, \ldots, m\}$. Then,

$$\begin{aligned}\mu(w) + \frac{\epsilon}{2} \geq \frac{\#_w(\alpha|_{\leq n})}{n} &\geq \sum_{a\in\Sigma} \frac{\#_{a\cdot w}(\alpha|_{\leq n})}{n} \geq \sum_{i=1}^{m} \frac{\#_{a_i\cdot w}(\alpha|_{\leq n})}{n} \\ &\geq \sum_{i=1}^{m} \left(\mu(a_i \cdot w) - \frac{\epsilon}{2m}\right)\end{aligned}$$

and hence $\mu(w) + \epsilon \geq \sum_{i=1}^{m} \mu(a_i \cdot w)$. As this inequality holds for all positive m, and all $\mu(a_i \cdot w)$ are positive, we have by the monotone convergence theorem that the limit $\sum_{i=1}^{\infty} \mu(a_i \cdot (w))$ exists, and thus that $\mu(w) + \epsilon \geq \sum_{i=1}^{\infty} \mu(a_i \cdot w)$. Conversely, for m, n,

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

and ϵ as above, we have

$$\mu(w) - \frac{\epsilon}{2} \leq \frac{\#_w(\alpha|_{\leq n})}{n} \leq \frac{1 + \sum_{a\in\Sigma} \#_{a\cdot w}(\alpha|_{\leq n})}{n} \leq \frac{1}{n} + \sum_{i=1}^{m} \left(\mu(a_i \cdot w) + \frac{\epsilon}{2m}\right)$$

and thus $\mu(w) - \epsilon \leq 1/n + \sum_{i=1}^{m} \mu(a_i \cdot w)$. As n can be taken arbitrarily large, we have $\mu(w) - \epsilon \leq \sum_{i=1}^{m} \mu(a_i \cdot w)$, and we obtain by the same reasoning as above that $\mu(w) - \epsilon \leq \sum_{i=1}^{\infty} \mu(a_i \cdot w)$. Thus, $|\mu(w) - \sum_{i=1}^{\infty} \mu(a_i \cdot w)| < \epsilon$, and as ϵ was arbitrary, we conclude that $\mu(w) = \sum_{i=1}^{\infty} \mu(a_i \cdot w)$. As all $\mu(a_i \cdot w)$ are non-negative, the series $\sum_{i=1}^{\infty} \mu(a_i \cdot w)$ is absolutely convergent whence the ordering of the terms does not affect convergence, and we obtain $\mu(w) = \sum_{a\in\Sigma} \mu(a \cdot w)$. By reasoning identical to the preceding, we obtain *mutatis mutandis* that $\mu(w) = \sum_{a\in\Sigma} \mu(w \cdot a)$, whence μ is invariant, as desired.

If μ is invariant, then by Proposition 6.1, $\bar{\mu}$ is a shift-invariant probability measure on Σ^ω. If ν is a shift-invariant probability measure on Σ^ω such that $\bar{\mu} = \nu$, then by [33, Main Theorem 2.1], there exists $\alpha \in \Sigma^\omega$ generic for $\bar{\mu}$, and thus for any admissible $w \in \Sigma^*$, $\lim_{n\to\infty}(\#_w(\alpha|_{\leq n})/n) = \nu(w) = \bar{\mu}([w]) = \mu(w)$. Observe that any inadmissible word $w = a_1 \cdots a_n$ has $\prod_{i=1}^{n} \mu(a_i) = \mu(w) = 0$, whence $\mu(a_i) = 0$ for some i, and hence $\lim_{n\to\infty}(\#_w(\alpha|_{\leq n})/n) \leq \lim_{n\to\infty}(\#_{a_i}(\alpha|_{\leq n})/n) = 0$. Hence, α is μ-distributed.

For the second part, we prove (1) $\Rightarrow$ (3) $\Rightarrow$ (2) $\Rightarrow$ (1). Assume that α is generic for ν. By construction, $\underline{\nu}$ is a probability map such that α is $\underline{\nu}$-distributed, and by the first part of the proposition, $\underline{\nu}$ is invariant, as desired. If $\underline{\nu}$ is an invariant probability map, then as any measurable A can be written as a disjoint union of cylinder sets, and as we for any cylinder $[w]$ have $S^{-1}([w]) = \bigcup_{a\in\Sigma}[a \cdot w]$, we obtain

$$\nu(S^{-1}([w])) = \underline{\nu}\left(\bigcup_{a\in\Sigma}[a \cdot w]\right) = \sum_{a\in\Sigma} \underline{\nu}(a \cdot w) = \underline{\nu}(w) = \nu([w]),$$

showing that ν is shift-invariant. Finally, if ν is shift-invariant, it follows from [33, Main Theorem 2.1] that there exists $\alpha \in \Sigma^\omega$ generic for ν, as desired. □

References

[1] V. N. Agafonov. Normal sequences and finite automata. *Sov. Math. Dokl.* **9** (1968), 324–325 (in English), originally published in Russian as [57].

[2] D. Airey and B. Mance. Normality preserving operations for Cantor series expansions and associated fractals, I. *Illinois J. Math.* **59**(3) (2015), 531–543.

[3] D. Airey and B. Mance. Hotspot lemmas for noncompact spaces. *Math. Notes* **108** (2020), 434–439.

[4] C. Aistleitner, V. Becher, A.-M. Scheerer and T. Slaman. On the construction of absolutely normal numbers. *Acta Arith.* **180** (2017), 333–346.

[5] V. Becher, O. Carton and P. A. Heiber. Normality and automata. *J. Comput. System Sci.* **81**(8) (2015), 1592–1613.

[6] V. Becher and P. A. Heiber. Normal numbers and finite automata. *Theoret. Comput. Sci.* **477** (2013), 109–116.

[7] V. Bergelson, T. Downarowicz and J. Vandehey. Deterministic functions on amenable semigroups and a generalization of the Kamae–Weiss theorem on normality preservation. *JAMA* **148** (2022), 213–286.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

[8] F. Blanchard. Non literal tranducers and some problems of normality. *J. Théor. Nombres Bordeaux* **5**(2) (1993), 303–321.
[9] F. Blanchard, J. M. Dumont and A. Thomas. Generic sequences, transducers and multiplication of normal numbers. *Israel J. Math.* **80** (1992), 257–287.
[10] D. Bontemps, S. Boucheron and E. Gassiat. About adaptive coding on countable alphabets. *IEEE Trans. Inform. Theory* **60**(2) (2014), 808–821.
[11] É. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo (2)* **27** (1909), 247–271.
[12] S. Boucheron, A. Garivier and E. Gassiat. Coding on countably infinite alphabets. *Trans. Inf* **55**(1) (2009), 358–373.
[13] A. Broglio and P. Liardet. Predictions with automata. Symbolic dynamics and its applications. *Contemp. Math.* **135** (1992), 111–124; also appeared in *Proceedings of the AMS Conference (New Haven, CT, USA, 1991)* in honor of R. L. Adler.
[14] O. Carton. A direct proof of Agafonov's theorem and an extension to shifts of finite type. *Preprint*, 2020, arXiv:2005.00255.
[15] O. Carton and J. Vandehey. Preservation of normality by non-oblivious group selection. *Theory Comput. Syst.* **65** (2020), 241–256.
[16] J. W. S. Cassels. On a paper of Niven and Zuckerman. *Pacific J. Math.* **2**(4) (1952), 555–557.
[17] D. G. Champernowne. The construction of decimals normal in the scale of ten. *J. Lond. Math. Soc. (2)* **s1-8**(4) (1933), 254–260.
[18] A. Church. On the concept of a random sequence. *Bull. Amer. Math. Soc. (N.S.)* **46**(2) (1940), 130–135.
[19] A. H. Copeland. Admissible numbers in the theory of probability. *Amer. J. Math.* **50**(4) (1928), 535–552.
[20] A. H. Copeland. Point set theory applied to the random selection of the digits of an admissible number. *Amer. J. Math.* **58**(1) (1936), 181–192.
[21] A. H. Copeland and P. Erdös. Note on normal numbers. *Bull. Amer. Math. Soc. (N.S.)* **52**(10) (1946), 857–860.
[22] H. Davenport and P. Erdös. Note on normal decimals. *Canad. J. Math.* **4** (1952), 58–63.
[23] H. Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation. *Math. Systems Theory* **1** (1967), 1–49.
[24] B. Heersink and J. Vandehey. Continued fraction normality is not preserved along arithmetic progressions. *Arch. Math. (Basel)* **106** (2015), 363–370.
[25] M. Holzer, M. Kutrib and A. Malcher. Multi-head finite automata: characterizations, concepts and open problems *Proc. Int. Workshop on the Complexity of Simple Programs (CSP 2008, Cork, Ireland, 6–7 December 2008) (Electronic Proceedings in Theoretical Computer Science, 1)*. Ed. T. Neary, D. Woods, A. K. Seda and N. Murphy. EPTCS, 2009, pp. 93–107.
[26] M. Hosseini and N. Santhanam. On redundancy of memoryless sources over countable alphabets. *Proc. 2014 Int. Symp. on Information Theory and its Applications (Victoria, BC, Canada, 26–29 October 2014)*. IEEE, Piscataway, NJ, 2014, pp. 299–303.
[27] T. Kamae and B. Weiss. Normal numbers and selection rules. *Israel J. Math.* **21** (1975), 101–110.
[28] E. Kamke. Über neuere begründungen der Wahrscheinlichkeitsrechnung. *Jahresber. Dtsch. Math.-Ver.* **42** (1933), 14–27.
[29] M. Levin. Absolutely normal numbers. *Moscow Univ. Math. Bull.* **34**(1) (1979), 32–39.
[30] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.
[31] M. Madritsch. Normal numbers and symbolic dynamics. *Sequences, Groups and Number Theory (Trends in Mathematics)*. Ed. V. Berthé and M. Rigo. Springer International Publishing, Cham, 2018, pp. 271–329.
[32] M. Madritsch, A.-M. Scheerer and R. Tichy. Computable absolutely Pisot normal numbers. *Acta Arith.* **184** (2018), 7–29.
[33] M. G. Madritsch and B. Mance, Construction of μ-normal sequences. *Monatsh. Math.* **179** (2016), 259–280.
[34] B. Mance. Cantor series constructions of sets of normal numbers. *Acta Arith.* **156** (2012), 223–245.
[35] W. Merkle and J. Reimann. Selection functions that do not preserve normality. *Theory Comput. Syst.* **39**(5) (2006), 685–697.
[36] Y. Nakai and I. Shiokawa. Discrepancy estimates for a class of normal numbers. *Acta Arith.* **62**(3) (1992), 271–284.
[37] I. Niven and H. S. Zuckerman. On the definition of normal numbers. *Pacific J. Math.* **1**(1) (1951), 103–109.
[38] M. G. O'Connor. An unpredictability approach to finite-state randomness. *J. Comput. System Sci.* **37**(3) (1988), 324–336.
[39] P. Pollack and J. Vandehey. Some normal numbers generated by arithmetic functions. *Canad. Math. Bull.* **58** (2013), 160–173.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press

[40] L. P. Postnikova. On the connection between the concepts of collectives of Mises–Church and normal Bernoulli sequences of symbols. *Theory Probab. Appl.* **6**(2) (1961), 211–213; translation of [61] by E. Nishiura.
[41] H. Reichenbach. Axiomatik der wahrscheinlichkeitsrechnung. *Math. Z.* **34**(1) (1932), 568–619.
[42] H. Reichenbach. Les fondements logiques du calcul des probabilités. *Ann. Inst. Henri Poincaré* **7**(1937), 267–348.
[43] A.-M. Scheerer. Computable absolutely normal numbers and discrepancies. *Math. Comp.* **86** (2015), 2911–2926.
[44] C.-P. Schnorr and H. Stimm. Endliche automaten und zufallsfolgen. *Acta Inform.* **1** (1972), 345–359.
[45] R. Serfozo. *Basics of Applied Stochastic Processes (Probability and Its Applications)*. Springer-Verlag, Berlin, 2009.
[46] A. Shen. Automatic Kolmogorov complexity and normality revisited. *Fundamentals of Computation Theory (Proc. 21st Int. Symp., FCT 2017, Bordeaux, France, 11–13 September 2017) (Lecture Notes in Computer Science, 10472)*. Ed. R. Klasing and M. Zeitoun. Springer, Berlin, 2017, pp. 418–430.
[47] P. Shields. *The Theory of Bernoulli Shifts*. University of Chicago Press, Chicago, 1973.
[48] W. Sierpinski. Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d'une tel nombre. *Bull. Soc. Math. France* **45** (1917), 125–132.
[49] J. F. Silva and P. Piantanida. Almost lossless variable-length source coding on countably infinite alphabets. *Proc. 2016 IEEE Int. Symp. on Information Theory (ISIT) (Barcelona, Spain, 2016)*. IEEE, Piscataway, NJ, 2016, pp. 1–5.
[50] E. Tornier. Wahrscheinlichkeitsrechnung und Zahlentheorie. Erste Mitteilung. *J. Reine Angew. Math.* **1929**(160) (1929), 177–198.
[51] J. Vandehey. The normality of digits in almost constant additive functions, *Monatsh. Math.* **171** (2012), 481–497.
[52] J. Vandehey. New normality constructions for continued fraction expansions. *J. Number Theory* **166** (2016), 424 – 451.
[53] J. Vandehey. Uncanny subsequence selections that generate normal numbers. *Unif. Distrib. Theory* **12** (2017), 65–75.
[54] R. Von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.* **5**(191) (1919), 52–99.
[55] D. D. Wall. Normal numbers. *PhD Thesis*, University of California, Berkeley, CA, 1949.
[56] X. Wang and T. Kamae. Selection rules preserving normality. *Israel J. Math.* **232** (2019), 427–442.
[57] В. Н. Агафонов. Нормальные последовательности и конечные автоматы. *Докл. АН СССР* **179**(2) (1968), 255–256.
[58] В. Н. Агафонов. Нормальные последовательности и конечные автоматы. *Наука, АН СССР* **20** (1968), 123–129.
[59] А. Г. Постников. Арифметическое моделирование случайных процессов. *Тр. МИАН СССР* **57** (1960), 3–84.
[60] А. Г. Постников and И. И. Пятецкий. Нормальные по Вернулли последовательности знаков. *Изв. АН СССР. Сер. матем.* **21**(4) (1957), 501–514.
[61] Л. П. Постникова. О связи понятий коллектива Мизеса-Черча и нормальной по Вернулли последовательности знаков. *Теория вероятн. и ее примен.* **6**(2) (1961), 232–234.

https://doi.org/10.1017/etds.2025.19 Published online by Cambridge University Press