Efficient Constructions of Finite-State Independent Normal Pairs

Subin Pulari[1]

[1] National Research University Higher School of Economics Moscow

February 27, 2026

Abstract

Finite-state independence is a robust notion of algorithmic independence for infinite words. It was introduced for general infinite words by Becher, Carton, and Heiber via deterministic asynchronous two-tape finite automata. Álvarez, Becher, and Carton then studied the normal case and characterized finite-state independence in terms of deterministic finite-state *shufflers*. A shuffler is a finite automaton that reads from two input tapes $x, y \in \Sigma^\infty$ and, at each step, chooses one tape to read next, outputs the symbol read, and updates its state based only on that output symbol. In terms of this characterization, two normal sources are finite-state independent if *every* deterministic finite-state way of shuffling (interleaving) them still produces a normal sequence. Álvarez, Becher, and Carton also exhibited a doubly-exponential procedure to construct a finite-state independent normal pair and posed the following questions: (1) can one compute finite-state independent normal pairs efficiently, improving their doubly-exponential procedure; and (2) given a normal word x, can one effectively construct a normal word y that is finite-state independent from x?

We answer both questions by explicit deterministic constructions. First, we give a deterministic polynomial-time algorithm that, on input N, outputs the first N symbols of two normal words x and y such that for every shuffler S, the shuffled output $S(x, y)$ is normal; hence (x, y) is finite-state independent. The construction is driven by a computable potential that aggregates conditional failure probabilities of a sparse family of normality tests for the first N shufflers, together with uniform tail bounds that remain valid under conditioning on short fixed prefixes.

Second, we solve the one-sided companion problem effectively. Given any computable normal word $x \in \Sigma^\infty$, we give an explicit deterministic construction of a computable normal word $y \in \Sigma^\infty$ such that for every shuffler S, the shuffled output $S(x, y)$ is normal. In particular, x and y are finite-state independent by the shuffler characterization theorem.

1   Introduction

Finite-state independence is a notion of algorithmic independence for infinite words in which the observer is restricted to finite memory. It was introduced by Becher, Carton, and Heiber [1] via *conditional finite-state compression*: one measures how well a word $x \in \Sigma^\infty$ can be compressed by a one-to-one finite-state transducer when a second word $y \in \Sigma^\infty$ is available as an oracle input. Technically, this oracle-assisted compression is modeled by deterministic asynchronous transducers with two input tapes and one output tape, formalized as 2-*deterministic* 3-*automata* [1]. Two words are finite-state independent if neither helps to compress the other in this model. Under the uniform product measure on $\Sigma^\infty \times \Sigma^\infty$, almost every pair is finite-state independent [1].

Normality is the property that an infinite digit sequence has the expected uniform frequencies for all fixed finite blocks. A word $z \in \Sigma^\infty$ is *normal* (in the sense of Borel [7]) if, for each block length m,

every m-symbol pattern appears with the limiting frequency predicted by the uniform distribution. Borel proved that almost every infinite word is normal [7]. It is conjectured that the digit expansions of familiar constants such as $\sqrt{2}$, π, and e are normal, but this is unknown. There are nevertheless explicit constructions of normal sequences; the classical example is Champernowne's number [2], whose base-10 digit sequence is obtained by concatenating the successive integers,

$$C_{10} := 0.12345678910111213141516\cdots.$$

(See Kuipers–Niederreiter [3] for background on normal numbers and equivalent formulations.)

Álvarez, Becher, and Carton [6] investigated finite-state independence in the normal setting and gave several equivalent characterizations using deterministic asynchronous two-tape automata. One characterization is particularly convenient for constructions: when x and y are normal over the same alphabet, they are finite-state independent if and only if *every* deterministic finite-state *shuffler* preserves normality, i.e., $S(x, y)$ is normal for every shuffler S. A shuffler is a deterministic finite automaton that reads from two input tapes and outputs an interleaving of the two inputs, choosing at each step which tape to read next and updating its state based only on the symbol just output [6].

Based on this shuffler characterization, Álvarez, Becher, and Carton gave an explicit algorithm that constructs a finite-state independent normal pair, but its running time is doubly exponential in the output length [6]. They point out that the same type of bottleneck already appears in Turing's construction of absolutely normal numbers [5]: one enforces an expanding family of frequency constraints by repeatedly intersecting "good" sets at rapidly growing cutoff lengths, and the cost is dominated by checking these large constraint families [6]. Motivated by this, they asked two concrete questions [6]:

1. Can one compute finite-state independent normal pairs efficiently, improving the doubly-exponential procedure?

2. Given a normal word x, can one effectively construct a normal word y that is finite-state independent from x (in particular, for explicit examples such as Champernowne's word)?

These questions should be contrasted with the single-source setting: there are highly efficient algorithms for producing (absolutely) normal numbers/words [4], but finite-state independence demands uniform normality preservation under *all* finite-state shuffles.

Our contributions. We answer both questions by explicit deterministic constructions.

(1) Efficient construction of an independent normal pair. We give a deterministic polynomial-time algorithm that, on input N, outputs the first N symbols of two normal words x and y such that $S(x, y)$ is normal for every shuffler S; hence (x, y) is finite-state independent by the characterization theorem [6]. At a high level, our construction replaces nested-set enforcement by a *local* greedy extension strategy controlled by a computable potential: at each step we extend one symbol of one tape so as to keep an upper bound on the *total conditional failure mass* of a sparse family of shuffler normality tests. The main technical inputs are (a) uniform tail bounds for aligned block-frequency deviations that remain valid under conditioning on short fixed prefixes, and (b) a dynamic program that computes the relevant conditional probabilities in polynomial time.

(2) Effective one-sided construction: a computable companion for a given computable normal input. Given any computable normal word $x \in \Sigma^\infty$, we construct a computable normal word $y \in \Sigma^\infty$ such that $S(x, y)$ is normal for every shuffler S. In particular, this answers the question raised in [6] about constructing a finite-state independent companion for Champernowne's

normal word: since Champernowne's word is computable and normal [2], our procedure produces a computable normal y that is finite-state independent from it. The proof has two stages. First we prove an almost-sure statement: for any fixed shuffler S, if Y is i.i.d. uniform then $S(x, Y)$ is normal with probability 1. Second, we effectivize this by defining, for each finite family of shufflers and tests, a clopen set of high measure of "good" y's, and then computing a single computable y in the intersection of all these sets via a standard effective compactness/extraction argument.

Organization. Section 2 introduces the main definitions and notation used throughout, including finite-state independence, shufflers, and normality. Section 3 gives the polynomial-time construction of a finite-state independent normal pair. Section 4 gives the effective one-sided construction for a fixed computable normal input.

2 Preliminaries

Fix a finite alphabet Σ of size $k \geq 2$. We write Σ^n for the set of words of length exactly n, $\Sigma^{\leq n} := \bigcup_{j=0}^{n} \Sigma^j$, $\Sigma^* := \bigcup_{n \geq 0} \Sigma^n$ for the set of finite words, and Σ^∞ for the set of infinite words. The empty word is denoted by λ, and $|u|$ denotes the length of a finite word u. For $x \in \Sigma^\infty$ and $n \in \mathbb{N}$, we write $x[1..n] \in \Sigma^n$ for the length-n prefix of x.

We equip Σ^∞ with the usual product (Cantor) topology. For $u \in \Sigma^*$, the cylinder determined by u is

$$[u] := \{x \in \Sigma^\infty : x[1..|u|] = u\}.$$

Cylinder sets are clopen. Moreover, since Σ^∞ is compact and cylinders form a basis of the topology, every clopen set is a finite union of cylinders. Let ν be the uniform distribution on Σ, i.e. $\nu(a) = 1/k$ for all $a \in \Sigma$. We equip Σ^∞ with the product measure $\nu^{\otimes \mathbb{N}}$. On $\Sigma^\infty \times \Sigma^\infty$ we use the product measure $\mu := \nu^{\otimes \mathbb{N}} \otimes \nu^{\otimes \mathbb{N}}$. For $u, v \in \Sigma^*$ we write $[u] \times [v]$ for the corresponding cylinder rectangle, and we write $\mu(\cdot \mid [u] \times [v])$ for conditional probability with respect to this rectangle (which is well-defined since $\mu([u] \times [v]) = k^{-(|u|+|v|)} > 0$).

We use both overlapping and aligned block-occurrence counts. For a finite word $u \in \Sigma^*$ and a block $w \in \Sigma^m$, define the (overlapping) occurrence count

$$\mathrm{occ}^{\mathrm{in}}(u, w) := \#\{\, 1 \leq t \leq |u| - m + 1 : u[t..t + m - 1] = w \,\}.$$

Throughout the paper we write $\mathrm{occ}(u, w)$ for $\mathrm{occ}^{\mathrm{in}}(u, w)$. We also use aligned (non-overlapping) counts at a fixed block length: for $u \in \Sigma^*$ and $w \in \Sigma^m$, define

$$\mathrm{occ}^{\mathrm{al}}(u, w) := \#\left\{\, 0 \leq j < \left\lfloor \frac{|u|}{m} \right\rfloor : u[jm + 1 .. jm + m] = w \,\right\}.$$

Now we define normality in terms of overlapping block-occurrence counts.

Definition 1 (Normality [7])**.** *A word $z \in \Sigma^\infty$ is (Borel) normal if for every $m \geq 1$ and every $w \in \Sigma^m$,*

$$\lim_{n \to \infty} \frac{\mathrm{occ}(z[1..n], w)}{n} = k^{-m}.$$

Normality can equivalently be defined using aligned block counts $\mathrm{occ}^{\mathrm{al}}(\cdot, \cdot)$ (with normalization by $\lfloor n/m \rfloor$); see [3]. We will also use the standard computability notion for infinite words.

Definition 2 (Computable infinite word)**.** *A word $z \in \Sigma^\infty$ is computable if there exists an algorithm that, on input $n \in \mathbb{N}$ (in binary), outputs the prefix $z[1..n]$.*

Finite-state independence for arbitrary infinite words was introduced by Becher, Carton, and Heiber [1] via conditional finite-state compression, using deterministic asynchronous transducer models (formalized as 2-deterministic 3-automata). In the normal setting, Álvarez, Becher, and Carton [6] gave a characterization in terms of deterministic shufflers, which is the formulation used throughout this paper.

Definition 3 (Deterministic shuffler [6, Def. 12]). *A (deterministic) shuffler is a tuple* $S = (Q, q_0, \delta, \tau)$ *where* Q *is a finite set of states,* $q_0 \in Q$ *is a start state,* $\delta : Q \times \Sigma \to Q$ *is a transition function, and* $\tau : Q \to \{1, 2\}$ *is a tape-choice function. Given* $(x, y) \in \Sigma^\infty \times \Sigma^\infty$, *the output* $z = S(x, y) \in \Sigma^\infty$ *is produced by the following infinite process: maintain head positions* $a = b = 0$ *and a current state* q; *at each step let* $\ell = \tau(q)$, *read the next unused symbol from tape* ℓ *(from* x *if* $\ell = 1$, *from* y *if* $\ell = 2$*), output that symbol* α, *and update* $q \leftarrow \delta(q, \alpha)$.

In [6], shufflers are presented as 2-deterministic 3-automata whose transitions are of one of two copy-types (copy from tape 1 or copy from tape 2). Determinism implies that all transitions leaving a fixed state have the same copy-type, hence induce a well-defined tape-choice function τ, yielding Definition 3.

For normal words over a common alphabet, we will use the shuffler-based formulation as our working definition of finite-state independence.

Definition 4 (Finite-state independence for normal words). *Let* $x, y \in \Sigma^\infty$ *be normal. We say that* x *and* y *are* finite-state independent *if for every shuffler* S *over* Σ, *the shuffled output* $S(x, y)$ *is normal.*

The next theorem, proved by Álvarez, Becher, and Carton [6], shows that this formulation is equivalent to the original definition of finite-state independence introduced by Becher, Carton, and Heiber [1].

Theorem 1 (Shuffler characterization [6, Thm. 1]). *Let* $x, y \in \Sigma^\infty$ *be normal. Then* x *and* y *are finite-state independent in the sense of Becher, Carton, and Heiber [1] if and only if for every shuffler* S *over* Σ, *the output* $S(x, y)$ *is normal.*

3 Polynomial-time construction of a finite-state independent normal pair

In [6], Álvarez et al. asked whether one can compute, in polynomial time, a finite-state independent pair of normal sequences $x, y \in \Sigma^\infty$ (Question 3 from Section 6 of [6]). In the following theorem, we give a positive answer.

Theorem 2. *There is a deterministic algorithm that, on input* N, *outputs prefixes* $(x{\restriction}N, y{\restriction}N)$ *in time* $N^{O(1)}$ *such that the infinite limit sequences* $x, y \in \Sigma^\infty$ *satisfy:*

1. x *and* y *are normal;*

2. *for every shuffler* S, *the shuffled output* $S(x, y)$ *is normal.*

In particular, (x, y) *is a finite-state independent pair by the characterization theorem of [6].*

Proof organization. We build x and y incrementally, one symbol at a time, by maintaining a potential that upper-bounds the (total) conditional probability of failing any currently relevant normality constraint. In Subsection 3.1 we define, for each output length n, a finite family F_n of aligned block-frequency constraints for the first t_n shufflers. In Subsection 3.2 we prove a uniform tail bound showing that each individual constraint has tiny failure probability even after conditioning on short fixed prefixes. In Subsection 3.3 we aggregate these conditional failure probabilities into a bad-mass functional $B_n(u,v)$ and then into a rolling potential $\Phi_L(u,v)$ over a sparse checkpoint sequence N_j. Subsection 3.4 gives the greedy algorithm (Algorithm 1) and proves correctness via an integrality argument at checkpoints (Proposition 1) and an interpolation step that yields normality for every shuffler output. Finally, Subsection 3.5 shows that the algorithm runs in time $N^{O(1)}$ by a dynamic program for computing each conditional probability term.

3.1 Aligned block counts and the F_n constraints

We now set up the finite families of frequency constraints that our construction will enforce at a sparse sequence of output lengths. These constraints are stated in terms of aligned block counts and are imposed simultaneously for an initial segment of the shuffler enumeration. Fix a concrete enumeration $(S_i)_{i \geq 1}$ of all shufflers, chosen so that the map $i \mapsto S_i$ is efficiently computable and the size of S_i is polynomially bounded in i. To this end, fix a canonical explicit binary encoding $\text{enc}(S)$ of a shuffler $S = (Q, q_0, \delta, \tau)$ in which the state set is identified with $\{1, \ldots, |Q|\}$ and the transition table $\delta : Q \times \Sigma \to Q$ together with the tape-choice function $\tau : Q \to \{1, 2\}$ are written out in full. For fixed alphabet size $k = |\Sigma|$, such an encoding has length $|\text{enc}(S)| = \Theta(|Q| \, k \log |Q|)$.

Let $(b_i)_{i \geq 1}$ be the length-lexicographic enumeration of all binary strings. Define S_i to be the shuffler obtained by decoding b_i if b_i is a valid encoding of some shuffler, and otherwise let S_i be a fixed trivial shuffler (e.g. the one that always reads tape 1). With this convention, given i one can compute b_i and check validity/perform decoding in time $\text{poly}(|b_i|) = \text{poly}(\log i)$, so the enumeration is effective with explicit resource bounds. Moreover, since the encoding is explicit, any valid b_i can describe only shufflers whose number of states satisfies $|Q_i| \leq \text{poly}(|b_i|) = \text{poly}(\log i) \leq \text{poly}(i)$, a bound that will be used in the runtime analysis.

Next we specify, for each output length n, the range of shufflers, block lengths, and error tolerance that will constitute the finite constraint family F_n.

For $n \geq 3$ define:

$$t_n := n, \qquad \ell_n := \left\lfloor \tfrac{1}{3} \log_k n \right\rfloor, \qquad \varepsilon_n := 2\sqrt{\frac{\log n \cdot \log_k n}{n}}.$$

For each $r \leq \ell_n$, set $m_{n,r} := \lfloor n/r \rfloor$.

Using these parameters, we now formalize what it means for one specific shuffler output to have the correct aligned frequency for one specific word.

Definition 5. *For a shuffler S, $n \geq 3$, $1 \leq r \leq \ell_n$, and $w \in \Sigma^r$, define*

$$E_S(n, r, w) := \left\{ (x, y) : \left| \text{occ}^{\text{al}}(S(x, y){\restriction}n, w) - m_{n,r}/k^r \right| < \varepsilon_n \, m_{n,r} \right\}.$$

Finally, F_n is the intersection of all such good events over the first t_n shufflers, all block lengths up to ℓ_n, and all words of the corresponding length.

Definition 6 (F_n). *Define*

$$F_n := \bigcap_{i=1}^{t_n} \bigcap_{r=1}^{\ell_n} \bigcap_{w \in \Sigma^r} E_{S_i}(n, r, w).$$

5

3.2 Conditional tail bounds under short prefix conditioning

We will repeatedly use a standard multiplicative Chernoff bound; we record a convenient form.

Lemma 1. *Let $X \sim \mathrm{Bin}(M,p)$ and assume $0 < \delta \leq 1$. Then*

$$\Pr\big[\, |X - Mp| \geq \delta Mp \,\big] \;\leq\; 2\exp\left(-\frac{\delta^2}{3}\, Mp\right).$$

The next lemma is the key uniform estimate: it bounds the conditional failure probability of a *single* constraint even after fixing short prefixes on both input tapes.

Lemma 2. *Fix $n \geq 3$, $1 \leq r \leq \ell_n$, and $w \in \Sigma^r$. Let S be any shuffler. Let $L \geq 0$ and $u, v \in \Sigma^L$. Write $m = m_{n,r} = \lfloor n/r \rfloor$. Assume $4L \leq \varepsilon_n\, m$ and $\varepsilon_n \leq k^{-r}$. Then*

$$\mu\big(E_S(n,r,w)^c \mid [u] \times [v]\big) \;\leq\; 2\exp\left(-\frac{\varepsilon_n^2}{12}\, m\, k^r\right).$$

Proof. Work under the conditional distribution $\mu(\cdot \mid [u] \times [v])$ and write $Z := S(X,Y){\upharpoonright}n$. Partition Z into the $m = \lfloor n/r \rfloor$ aligned length-r blocks.

During the production of the first n output symbols, every time the shuffler reads an input symbol from tape 1 at a position $\leq L$ it outputs the corresponding fixed letter of u, and similarly for tape 2 with v. Let $T \subseteq \{1, 2, \ldots, n\}$ be the (random) set of output times at which the read input position lies in a fixed prefix (i.e. at which the output letter is one of the $2L$ fixed prefix symbols). Since each of the first L symbols on each tape can be read at most once, we have the deterministic bound

$$|T| \;\leq\; 2L.$$

Now construct an auxiliary length-n word $U \in \Sigma^n$ as follows. For each $t \in \{1, \ldots, n\}$, set

$$U[t] \;:=\; \begin{cases} Z[t], & t \notin T, \\ R_t, & t \in T, \end{cases}$$

where $(R_t)_{t \in T}$ are fresh independent uniform letters from Σ, independent of (X, Y).

Claim 1: U is a uniformly random word in Σ^n (equivalently, its letters are i.i.d. uniform).

Let $\mathcal{F}_{t-1}$ be the σ-field generated by all symbols of X and Y that have been read by S during the first $t-1$ output steps (together with all auxiliary letters $R_{t'}$ for $t' < t$). At time t, the shuffler's state, tape choice $\tau(q)$, and head positions are $\mathcal{F}_{t-1}$-measurable, since they are deterministic functions of the previously read input symbols. If $t \in T$, then $U[t] = R_t$ is uniform on Σ and independent of $\mathcal{F}_{t-1}$ by construction. If $t \notin T$, then the shuffler reads the next unused symbol from one tape at some position $j > L$. Under the product measure μ, the tail symbols $\{X_{L+1}, X_{L+2}, \ldots\}$ and $\{Y_{L+1}, Y_{L+2}, \ldots\}$ are independent i.i.d. uniform and are independent of the fixed prefixes u, v; moreover, the particular index j to be read at time t is $\mathcal{F}_{t-1}$-measurable. Hence the next unread symbol that is read at time t is uniform on Σ and independent of $\mathcal{F}_{t-1}$, and therefore $U[t] = Z[t]$ is uniform and independent of the past. In all cases, conditional on $\mathcal{F}_{t-1}$, the random variable $U[t]$ is uniform on Σ and independent of $\mathcal{F}_{t-1}$; thus $(U[1], \ldots, U[n])$ are i.i.d. uniform and U is uniform on Σ^n.

Claim 2: Let $C := \mathrm{occ}^{\mathrm{al}}(Z, w)$ and $C' := \mathrm{occ}^{\mathrm{al}}(U, w)$. Then $|C - C'| \leq 2L$.

Changing one symbol of a word can affect $\mathrm{occ}^{\mathrm{al}}$ by at most 1, because each position lies in exactly one aligned length-r block. The words Z and U differ only at positions in T, hence in at most $|T| \leq 2L$ positions. Therefore $|C - C'| \leq 2L$.

Now, if $|C - m/k^r| \geq \varepsilon_n m$, then by Claim 2,

$$\left|C' - m/k^r\right| \ \geq\ \varepsilon_n m - 2L \ \geq\ \frac{\varepsilon_n m}{2},$$

using the hypothesis $4L \leq \varepsilon_n m$.

By Claim 1, the m aligned blocks of U are i.i.d. uniform over Σ^r, so

$$C' \sim \mathrm{Bin}\big(m,\, k^{-r}\big), \qquad \mathbb{E}[C'] = m/k^r.$$

Let $\mu_{C'} := m/k^r$ and $t := \varepsilon_n m/2$, so the relative deviation is

$$\delta \ :=\ \frac{t}{\mu_{C'}} \ =\ \frac{\varepsilon_n m/2}{m/k^r} \ =\ \frac{\varepsilon_n k^r}{2}.$$

By the hypothesis $\varepsilon_n \leq k^{-r}$ we have $\delta \leq 1/2 \leq 1$ and we may apply Lemma 1:

$$\Pr\big[|C' - \mu_{C'}| \geq t\big] \ \leq\ 2\exp\left(-\frac{\delta^2}{3}\mu_{C'}\right) \ =\ 2\exp\left(-\frac{1}{3}\cdot\frac{\varepsilon_n^2 k^{2r}}{4}\cdot\frac{m}{k^r}\right) \ =\ 2\exp\left(-\frac{\varepsilon_n^2}{12}\, m\, k^r\right).$$

Since $\mu(E_S(n,r,w)^c \mid [u] \times [v]) = \Pr[|C - m/k^r| \geq \varepsilon_n m]$, the implication above yields the lemma. $\qquad\square$

3.3   Aggregating conditional failure probabilities via a rolling potential function

We now aggregate the conditional failure probabilities of all constraints at a fixed length n.

Definition 7. *For $n \geq 3$ and prefixes $u, v \in \Sigma^L$, define*

$$B_n(u,v) \ :=\ \sum_{i=1}^{t_n}\sum_{r=1}^{\ell_n}\sum_{w\in\Sigma^r} \mu\big(E_{S_i}(n,r,w)^c \mid [u]\times[v]\big).$$

The next corollary shows that once the prefix length is at most $\sqrt{n}$, the total conditional bad mass at length n is already extremely small.

Corollary 1. *There exists n_0 such that for all $n \geq n_0$, all $L \leq \sqrt{n}$, and all $u, v \in \Sigma^L$,*

$$B_n(u,v) \ \leq\ \frac{1}{n^2}.$$

Proof. Fix $n \geq n_0$ and $L \leq \sqrt{n}$. For every $r \leq \ell_n$ and all $n \geq 2r$ we have

$$m_{n,r} = \lfloor n/r \rfloor \ \geq\ \frac{n}{r} - 1.$$

Therefore

$$\varepsilon_n\, m_{n,r} \ \geq\ 2\sqrt{\frac{\log n \log_k n}{n}}\left(\frac{n}{r} - 1\right) \ =\ \frac{2\sqrt{n\log n \log_k n}}{r} - 2\sqrt{\frac{\log n \log_k n}{n}}.$$

Using $r \leq \ell_n \leq \frac{1}{3}\log_k n$ gives

$$\frac{2\sqrt{n\log n \log_k n}}{r} \ \geq\ \frac{2\sqrt{n\log n \log_k n}}{(1/3)\log_k n} \ =\ 6\sqrt{\log k}\,\sqrt{n}.$$

7

Hence, for all sufficiently large n (increase n_0 if needed), the subtracted term $2\sqrt{\frac{\log n \log_k n}{n}} = o(\sqrt{n})$ is negligible and we obtain the uniform bound

$$\varepsilon_n \, m_{n,r} \ \geq \ 4.9\sqrt{n}$$

for all $r \leq \ell_n$. In particular, for all $L \leq \sqrt{n}$ we have

$$4L \ \leq \ 4\sqrt{n} \ \leq \ \varepsilon_n \, m_{n,r}$$

uniformly over all $r \leq \ell_n$. Also, for $r \leq \ell_n$ we have $k^r \leq k^{\ell_n} \leq n^{1/3}$, hence

$$\varepsilon_n k^r \ \leq \ 2\, n^{1/3}\sqrt{\frac{\log n \log_k n}{n}} \ = \ 2\, n^{-1/6}\sqrt{\log n \log_k n},$$

which is < 1 for all sufficiently large n; increase n_0 so that $\varepsilon_n \leq k^{-r}$ holds for all $n \geq n_0$ and all $r \leq \ell_n$. Thus Lemma 2 applies to every triple (S_i, r, w).

Therefore

$$B_n(u,v) \ \leq \ \sum_{i \leq n} \sum_{r \leq \ell_n} \sum_{w \in \Sigma^r} 2\exp\left(-\frac{\varepsilon_n^2}{12} \, m_{n,r} k^r\right).$$

Using $m_{n,r} \geq n/(2r)$ and $\varepsilon_n^2 = 4\frac{\log n \log_k n}{n}$ gives

$$\frac{\varepsilon_n^2}{12} \, m_{n,r} k^r \ \geq \ \frac{4}{12}\cdot\frac{\log n \log_k n}{n}\cdot\frac{n}{2r}\cdot k^r \ = \ \frac{1}{6}\cdot\frac{k^r}{r}\, \log n \, \log_k n.$$

In particular, for every $r \geq 1$ we have $\frac{k^r}{r} \geq k$, so the exponent is at least

$$\frac{k}{6}\, \log n \, \log_k n = \frac{k}{6\log k}\, (\log n)^2.$$

Hence there is a constant $c > 0$ with

$$2\exp\left(-\frac{\varepsilon_n^2}{12} \, m_{n,r} k^r\right) \leq 2e^{-c(\log n)^2} \leq n^{-10}$$

for all sufficiently large n (increase n_0 so this holds). Now count terms:

$$\sum_{i \leq n} \sum_{r \leq \ell_n} \sum_{w \in \Sigma^r} 1 \leq n\cdot\ell_n\cdot\sum_{r=1}^{\ell_n} k^r \leq n\cdot\ell_n\cdot k^{\ell_n+1} \leq n\cdot\ell_n\cdot k\cdot n^{1/3} = O(n^{4/3}\log n).$$

Therefore

$$B_n(u,v) \ \leq \ O(n^{4/3}\log n)\cdot n^{-10} \ \leq \ \frac{1}{n^2}$$

for all large enough n. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We will enforce all constraints at a sparse set of checkpoint lengths N_j. Our construction chooses the next symbols of (x,y) greedily, at each prefix length L, so as to keep a suitable potential function small. To make this greedy choice work uniformly for every L, we use a rolling potential that only includes a checkpoint N_j once the current prefix length is at least its activation length $A_j = \sqrt{N_j}$.

Define checkpoint lengths

$$N_j := (j + m_0)^4, \qquad j \geq 1,$$

where m_0 is a fixed constant chosen so that $\sum_{j \geq 1} 1/N_j^2 < 1/4$ and $N_1 \geq n_0$ (e.g. any m_0 large enough).

Define activation lengths

$$A_j := \sqrt{N_j} = (j + m_0)^2.$$

Definition 8. *For each $L \geq 0$, define the active index set*

$$J(L) := \{\, j \geq 1 : \; A_j \leq L \leq N_j \,\}.$$

For prefixes $u, v \in \Sigma^L$, define

$$\Phi_L(u, v) \; := \; \sum_{j \in J(L)} B_{N_j}(u, v).$$

The next lemma gives the basic averaging identity for one-step extensions of the current prefixes; it will be used to justify the greedy choice of the next symbols.

Lemma 3. *Fix $n \geq 3$ and prefixes $u, v \in \Sigma^L$. Then*

$$B_n(u, v) \; = \; \frac{1}{k^2} \sum_{a \in \Sigma} \sum_{b \in \Sigma} B_n(ua, vb).$$

Proof. Each term $\mu(E^c \mid [u] \times [v])$ is a conditional probability under the product measure. Conditioning on the next symbols (X_{L+1}, Y_{L+1}) yields the law of total probability: the conditional probability under $[u] \times [v]$ is the average of the conditional probabilities under the refined cylinders $[ua] \times [vb]$. Summing over all constraints preserves equality. $\square$

The next lemma formalizes the averaging step that guarantees the existence of a one-symbol extension with small potential.

Lemma 4. *For every $L \geq 0$ and every prefixes $u, v \in \Sigma^L$,*

$$\min_{a, b \in \Sigma} \Phi_{L+1}(ua, vb) \; \leq \; \sum_{j \in J(L+1)} B_{N_j}(u, v).$$

Proof. By Lemma 3, for each fixed $j \in J(L+1)$, $B_{N_j}(u, v)$ equals the average of $B_{N_j}(ua, vb)$ over $(a, b) \in \Sigma \times \Sigma$. Summing over $j \in J(L+1)$ gives

$$\sum_{j \in J(L+1)} B_{N_j}(u, v) \; = \; \frac{1}{k^2} \sum_{a, b \in \Sigma} \Phi_{L+1}(ua, vb),$$

so the minimum is at most the average. $\square$

The next lemma packages the invariant needed for correctness: the rolling potential stays strictly below 1.

Lemma 5. *For every $L \geq 0$ there exist prefixes $u, v \in \Sigma^L$ such that*

$$\Phi_L(u, v) \; < \; 1.$$

Proof. We prove by induction on L the stronger bound

$$\Phi_L(u, v) \; \leq \; \sum_{j:\; A_j \leq L} \frac{1}{N_j^2}.$$

Base $L = 0$: $J(0) = \emptyset$, so $\Phi_0(\lambda, \lambda) = 0$.

Inductive step: let $(u, v) \in \Sigma^L \times \Sigma^L$ be the produced prefixes. For any L, the algorithm chooses $(a^\star, b^\star)$ minimizing $\Phi_{L+1}(ua, vb)$, hence by Lemma 4

$$\Phi_{L+1}(ua^\star, vb^\star) \leq \sum_{j \in J(L+1)} B_{N_j}(u, v).$$

Split $J(L + 1)$ into old-active indices and newly activated ones:

$$J(L + 1) = \big(J(L) \cap J(L + 1)\big) \cup \{j : A_j = L + 1\}.$$

The first part is a subset of $J(L)$, so

$$\sum_{j \in J(L) \cap J(L+1)} B_{N_j}(u, v) \leq \Phi_L(u, v).$$

For any newly activated j with $A_j = L + 1$, we have $L \leq A_j \leq \sqrt{N_j}$, so by Corollary 1 (applied with $n = N_j$ and prefix length $L \leq \sqrt{N_j}$),

$$B_{N_j}(u, v) \leq \frac{1}{N_j^2}.$$

Therefore

$$\Phi_{L+1}(ua^\star, vb^\star) \leq \Phi_L(u, v) + \sum_{j:\; A_j = L+1} \frac{1}{N_j^2} \leq \sum_{j:\; A_j \leq L+1} \frac{1}{N_j^2}.$$

Finally, by choice of m_0, $\sum_{j \geq 1} 1/N_j^2 < 1/4 < 1$, so $\Phi_L < 1$ for all L along the run. $\qquad \square$

3.4   The algorithm and correctness

We now give the greedy construction of the prefixes $(x{\upharpoonright}N, y{\upharpoonright}N)$. At each length L we consider all one-step extensions of the current prefixes and choose the pair of next symbols that minimizes the next-step potential Φ_{L+1}. The preceding subsection shows that this choice keeps the rolling potential strictly below 1 throughout the run; in particular, it will force every checkpoint family F_{N_j} to hold.

Algorithm 1 Compute $(x{\upharpoonright}N, y{\upharpoonright}N)$

Require: $N \in \mathbb{N}$

1: $u \leftarrow \lambda$, $v \leftarrow \lambda$, $L \leftarrow 0$
2: **while** $L < N$ **do**
3:     Compute $\Phi_{L+1}(ua, vb)$ for all $(a, b) \in \Sigma \times \Sigma$
4:     Choose $(a^\star, b^\star)$ minimizing $\Phi_{L+1}(ua, vb)$
5:     $u \leftarrow ua^\star$, $v \leftarrow vb^\star$, $L \leftarrow L + 1$
6: **end while**
7: **return** (u, v)

We next show that the invariant $\Phi_L < 1$ forces every checkpoint constraint family F_{N_j} to hold.

To connect the potential bound to satisfaction of constraints at a fixed length n, we use the following integrality observation once the conditioning prefixes determine the first n symbols on both tapes.

Lemma 6. *If $L \geq n$, then $B_n(u, v)$ is an integer (hence either 0 or at least 1).*

Proof. When $L \geq n$, the first n symbols on each tape are fixed under $[u] \times [v]$. Thus each event $E_{S_i}(n, r, w)$ (and its complement) is determined, so its conditional probability is 0 or 1. Summing finitely many 0–1 values gives an integer. $\qquad\square$

The next proposition applies Lemma 6 at each checkpoint length N_j to turn the strict inequality $\Phi_{N_j} < 1$ into the exact satisfaction of all constraints in F_{N_j}.

Proposition 1. *Let (x, y) be the infinite output of Algorithm 1. Then for every $j \geq 1$, $(x, y) \in F_{N_j}$.*

Proof. Fix j and let $L := N_j$. Then $j \in J(L)$ (because $A_j \leq N_j \leq N_j$), so by Lemma 5,

$$0 \leq B_{N_j}(x{\restriction}L, y{\restriction}L) \ \leq \ \Phi_L(x{\restriction}L, y{\restriction}L) \ < \ 1.$$

By Lemma 6 (with $n = L = N_j$), $B_{N_j}(x{\restriction}L, y{\restriction}L)$ is an integer, hence it must equal 0. Since B_{N_j} is a sum of non-negative conditional probabilities, the sum is 0 iff each summand is 0, i.e. every constraint in F_{N_j} holds. Thus $(x, y) \in F_{N_j}$. $\qquad\square$

We now derive normality for every shuffler output from the fact that all checkpoint constraints are met.

Proposition 2. *For every shuffler S, the sequence $S(x, y)$ is normal. In particular, x and y are normal.*

Proof. Fix $S = S_i$. Since $t_{N_j} = N_j$, for all sufficiently large j we have $i \leq t_{N_j}$, hence $(x, y) \in F_{N_j}$ implies that for every $r \leq \ell_{N_j}$ and every $w \in \Sigma^r$,

$$\left| \mathrm{occ}^{\mathrm{al}}(S(x, y){\restriction}N_j, w) - \frac{m_{N_j, r}}{k^r} \right| < \varepsilon_{N_j} \, m_{N_j, r}.$$

Dividing by $m_{N_j, r}$ gives aligned frequency error $< \varepsilon_{N_j}$ at the subsequence N_j.

Now fix any block length $r \geq 1$ and any $w \in \Sigma^r$. For large j, we have $r \leq \ell_{N_j}$. Let n be large and choose j with $N_j \leq n < N_{j+1}$. Let $z = S(x, y)$ and write $m_n = \lfloor n/r \rfloor$ and $m_j = \lfloor N_j/r \rfloor$. Since each additional aligned block can increase $\mathrm{occ}^{\mathrm{al}}$ by at most 1, we have

$$0 \leq \mathrm{occ}^{\mathrm{al}}(z{\restriction}n, w) - \mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w) \leq m_n - m_j.$$

Hence

$$\left| \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}n, w)}{m_n} - \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w)}{m_n} \right| \leq \frac{m_n - m_j}{m_n}.$$

Also, since $0 \leq \mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w) \leq m_j$, we have

$$\left| \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w)}{m_n} - \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w)}{m_j} \right| = \mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w) \left| \frac{1}{m_n} - \frac{1}{m_j} \right| \leq \frac{m_n - m_j}{m_n}.$$

By the triangle inequality,

$$\left| \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}n, w)}{m_n} - \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w)}{m_j} \right| \leq 2 \cdot \frac{m_n - m_j}{m_n}.$$

Moreover,

$$\frac{m_n - m_j}{m_n} \leq \frac{\lfloor (n - N_j)/r \rfloor + 1}{\lfloor n/r \rfloor} \leq \frac{(n - N_j)/r + 1}{n/r - 1} = \frac{n - N_j + r}{n - r}.$$

For $n > 2r$ this implies

$$\frac{m_n - m_j}{m_n} \le 3\frac{n - N_j}{n} + 3\frac{r}{n} \le 3\frac{N_{j+1} - N_j}{N_j} + 3\frac{r}{N_j}.$$

Finally, since $(x, y) \in F_{N_j}$, for all sufficiently large j (so that $i \le t_{N_j}$ and $r \le \ell_{N_j}$) we have

$$\left| \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}N_j, w)}{m_j} - \frac{1}{k^r} \right| < \varepsilon_{N_j}.$$

Combining the last two displays gives

$$\left| \frac{\mathrm{occ}^{\mathrm{al}}(z{\restriction}n, w)}{m_n} - \frac{1}{k^r} \right| \le 2 \cdot \frac{m_n - m_j}{m_n} + \varepsilon_{N_j} \le 6\frac{N_{j+1} - N_j}{N_j} + 6\frac{r}{N_j} + \varepsilon_{N_j}.$$

Since $N_j = (j + m_0)^4$, we have $(N_{j+1} - N_j)/N_j \to 0$ and $r/N_j \to 0$, and also $\varepsilon_{N_j} \to 0$. Therefore the aligned block frequency at length n converges to $1/k^r$.

By the equivalence between aligned-count normality and non-aligned count normality (see the remark in the preliminaries), z is normal.

Finally, taking the shuffler that always reads only tape 1 (resp. tape 2) gives x and y normal. $\quad\square$

3.5   Polynomial-time implementation

We now verify that Algorithm 1 can be implemented within polynomial time (in bit-complexity) when the alphabet size $k = |\Sigma|$ is fixed.

Lemma 7. *For fixed alphabet size $k = |\Sigma|$, Algorithm 1 runs in time $N^{O(1)}$ (bit-complexity) on input N.*

Proof. Fix $N \in \mathbb{N}$. At iteration L (with current prefixes $u, v \in \Sigma^L$), the algorithm evaluates $\Phi_{L+1}(ua, vb)$ for all $(a, b) \in \Sigma \times \Sigma$ and picks the minimizer. Since k is fixed, it suffices to show that for each fixed (a, b), $\Phi_{L+1}(ua, vb)$ can be computed in time polynomial in N, uniformly over all $L \le N$.

Let $L' := L + 1$ and write $u' = ua$ and $v' = vb$. We compute

$$\Phi_{L'}(u', v') = \sum_{j \in J(L')} B_{N_j}(u', v').$$

If $j \in J(L')$, then $A_j \le L'$ implies $j + m_0 \le \sqrt{L'} + m_0$, hence

$$N_j = (j + m_0)^4 \ \le \ (\sqrt{L'} + m_0)^4 \ = \ O\big((L')^2\big) \ \le \ O(N^2).$$

Also $A_j \le L'$ implies $j \le \sqrt{L'}$, so $|J(L')| \le \lfloor \sqrt{L'} \rfloor = O(\sqrt{N})$. Thus, for each $L' \le N$, computing $\Phi_{L'}(u', v')$ reduces to computing $B_n(u', v')$ for $O(\sqrt{N})$ values n with $n \le cN^2$ for a constant c.

We next explain how to compute one term appearing in $B_n(u', v')$ by dynamic programming. Fix parameters

$$n, \ i \le t_n = n, \ 1 \le r \le \ell_n, \ w \in \Sigma^r, \ u', v' \in \Sigma^{L'} \ (L' \le n),$$

and write $S_i = (Q_i, q_0, \delta, \tau)$. We compute

$$p := \mu\big(E_{S_i}(n, r, w)^c \mid [u'] \times [v']\big).$$

Let $m = \lfloor n/r \rfloor$. We run a forward DP over the production of the first n output symbols. A DP state consists of

$$(q, a, b, s, \sigma, c),$$

where:

- $q \in Q_i$ is the current shuffler state,

- $a \in \{0, \ldots, n\}$ and $b \in \{0, \ldots, n\}$ are the numbers of symbols consumed from tapes 1 and 2 (so the output time is $t = a + b$),

- $s \in \{0, \ldots, r-1\}$ is the position inside the current aligned length-r block ($s = t \bmod r$),

- $\sigma \in \{0, 1\}$ indicates whether the current block built so far still matches the prefix of w (i.e. $\sigma = 1$ iff the last s symbols equal $w[1..s]$),

- $c \in \{0, \ldots, m\}$ is the number of completed aligned blocks equal to w among the first $\lfloor t/r \rfloor$ completed blocks.

Initialize $\mathrm{DP}[q_0, 0, 0, 0, 1, 0] = 1$ and all other entries to 0. For each state with $t = a + b < n$, we update as follows. Let $\ell = \tau(q) \in \{1, 2\}$ be the chosen tape. If $\ell = 1$, the next tape symbol is at position $a + 1$; if $a + 1 \leq L'$ it equals the fixed letter $u'[a+1]$, otherwise it is uniform over Σ. Similarly for $\ell = 2$ with v' and $b + 1$.

Thus each transition from (q, a, b, s, σ, c) branches over $\alpha \in \Sigma$ with probability

$$
\Pr[\alpha] = \begin{cases} 1, & \text{if the next position is fixed and equals } \alpha, \\ 0, & \text{if the next position is fixed and differs from } \alpha, \\ 1/k, & \text{if the next position is unfixed.} \end{cases}
$$

For each α with $\Pr[\alpha] > 0$, set $q' = \delta(q, \alpha)$ and increment the appropriate head counter ($a' = a + 1, b' = b$ if $\ell = 1$, else $a' = a, b' = b + 1$). Update the within-block variables:

$$
\sigma' = \begin{cases} 1, & \sigma = 1 \text{ and } \alpha = w[s+1], \\ 0, & \text{otherwise}, \end{cases} \qquad s' = s + 1.
$$

If $s' < r$ (block not finished), keep count $c' = c$ and store $(q', a', b', s', \sigma', c')$. If $s' = r$ (block finished), then set

$$
c' = \begin{cases} c + 1, & \sigma' = 1, \\ c, & \sigma' = 0, \end{cases} \qquad s' = 0, \ \sigma' = 1
$$

(start the next block fresh), and store $(q', a', b', 0, 1, c')$. All updates add probability mass accordingly.

After processing all states with $a + b = n$, we obtain the distribution of the random variable $C = \mathrm{occ}^{\mathrm{al}}(S_i(X, Y) {\restriction} n, w)$ under the conditioning:

$$
\Pr[C = c] = \sum_{q \in Q_i} \sum_{a+b=n} \sum_{s, \sigma} \mathrm{DP}[q, a, b, s, \sigma, c].
$$

We then compute

$$
p = \Pr\Big[\big|C - m/k^r\big| \geq \varepsilon_n m\Big]
$$

by summing $\Pr[C = c]$ over all integers c outside the allowed interval.

The DP table has at most

$$
|Q_i| \cdot (n+1)^2 \cdot r \cdot 2 \cdot (m+1) \;=\; O\big(|Q_i|\, n^3 r\big) \;=\; O\big(|Q_i|\, n^3 \log n\big)
$$

states, since $r \leq \ell_n = O(\log n)$ and $m \leq n$. Each state performs $O(k) = O(1)$ arithmetic updates. Probabilities are rationals with denominator at most k^n, so numerators/denominators have $O(n)$

bits; all additions/multiplications therefore cost polynomial time in n (with fixed k). Hence this computation runs in time $n^{O(1)} \cdot |Q_i|$.

We now compute $B_n(u', v')$ itself. By definition,

$$B_n(u', v') = \sum_{i=1}^{n} \sum_{r=1}^{\ell_n} \sum_{w \in \Sigma^r} \mu\big(E_{S_i}(n, r, w)^c \mid [u'] \times [v']\big).$$

There are n choices of i, $O(\log n)$ choices of r, and

$$\sum_{r=1}^{\ell_n} |\Sigma^r| = \sum_{r=1}^{\ell_n} k^r \leq k^{\ell_n + 1} \leq k \cdot n^{1/3}$$

choices of w. Therefore $B_n(u', v')$ is a sum of at most $O(n^{4/3} \log n)$ terms. Computing each term as above and summing them yields a total time polynomial in n.

Finally, for fixed $L' \leq N$ and fixed (a, b), we compute

$$\Phi_{L'}(u', v') = \sum_{j \in J(L')} B_{N_j}(u', v').$$

As shown at the start, $|J(L')| = O(\sqrt{N})$ and every $N_j \leq cN^2$. Each $B_{N_j}(u', v')$ is computable in time polynomial in N_j, hence polynomial in N, so $\Phi_{L'}(u', v')$ is computable in time polynomial in N. We do this for $k^2 = O(1)$ candidate pairs (a, b) at each of N iterations, so the full run time is $N^{O(1)}$. $\square$

Proposition 2 establishes that the limit sequences x, y are normal and that $S(x, y)$ is normal for every shuffler S, and Lemma 7 shows that the construction runs in time $N^{O(1)}$ on input N. This completes the proof of Theorem 2. And finally due to Theorem 1, the pair (x, y) is finite-state independent.

4 A computable finite-state independent word for a fixed computable normal input

Throughout this subsection, fix a finite alphabet Σ of size $k \geq 2$ and a computable normal word $x \in \Sigma^\infty$. We construct a computable $y \in \Sigma^\infty$ such that y is normal and $S(x, y)$ is normal for every shuffler S. By the shuffler characterization theorem stated earlier in the paper, this implies that x and y are finite-state independent.

Proof outline. The construction has two ingredients.

1. We first show an almost-sure statement: for each fixed shuffler S, if $Y \sim \text{Unif}(\Sigma^\infty)$ then $S(x, Y)$ is normal with probability 1.

2. We then effectivize this: from a uniformly computable family of clopen sets of high measure (one for each finite collection of shuffler tests), we compute a single computable word y that lies in all of them simultaneously.

4.1 Finite test sets and a computable extraction lemma

For parameters $m \geq 1$ and $\varepsilon > 0$, write

$$\text{Test}(u; m, \varepsilon) \; :\Longleftrightarrow \; \forall w \in \Sigma^m : \; \left| \frac{\text{occ}(u, w)}{|u|} - k^{-m} \right| \leq \varepsilon.$$

For a shuffler S and inputs (x, y) define

$$E_S(n; m, \varepsilon) := \{(x, y) \in \Sigma^\infty \times \Sigma^\infty : \; \text{Test}(S(x, y)[1..n]; m, \varepsilon)\}.$$

The next lemma records two basic facts needed for the effective intersection step: the relevant test sets are clopen, and their measures (and cylinder intersections) are uniformly computable when x is computable.

Lemma 8. *For fixed S, n, m, ε, the set $E_S(n; m, \varepsilon)$ is clopen in $\Sigma^\infty \times \Sigma^\infty$. Moreover, for each fixed $x \in \Sigma^\infty$, the slice*

$$E_S(n; m, \varepsilon)(x) := \{y : \; (x, y) \in E_S(n; m, \varepsilon)\}$$

is clopen in Σ^∞.

If x is computable then for every cylinder $[v] \subseteq \Sigma^\infty$ the quantity $\mu(E_S(n; m, \varepsilon)(x) \cap [v])$ is computable uniformly in $(S, n, m, \varepsilon, v)$. In particular, $\mu(E_S(n; m, \varepsilon)(x))$ is computable uniformly in (S, n, m, ε).

Proof. To determine membership in $E_S(n; m, \varepsilon)$ it suffices to know the output prefix $S(x, y)[1..n]$. During the first n output steps, the shuffler reads at most n symbols from each tape, hence $S(x, y)[1..n]$ is determined by $(x[1..n], y[1..n])$. Therefore $E_S(n; m, \varepsilon)$ is a finite union of rectangles $[u] \times [v]$ with $u, v \in \Sigma^n$, and is clopen; each slice is a finite union of cylinders in Σ^∞.

Assume x is computable. Fix a cylinder $[v]$ with $|v| = r$. If $r \geq n$, then on $[v]$ the prefix $y[1..n]$ is fixed to be $v[1..n]$, hence either $[v] \subseteq E_S(n; m, \varepsilon)(x)$ or $[v] \cap E_S(n; m, \varepsilon)(x) = \emptyset$; in either case $\mu(E_S(n; m, \varepsilon)(x) \cap [v]) \in \{0, k^{-r}\}$ is computable.

If $r < n$, enumerate all $u \in \Sigma^n$ extending v. For each such u, simulate S for n output steps on inputs $(x[1..n], u)$, decide whether the resulting output prefix passes $\text{Test}(\cdot; m, \varepsilon)$, and count the number of passing u. Then $\mu(E_S(n; m, \varepsilon)(x) \cap [v])$ equals this count times k^{-n}, hence is a computable rational, uniformly in $(S, n, m, \varepsilon, v)$. □

The next lemma is a standard effective intersection principle: from a uniformly computable sequence of high-measure clopen sets with rapidly summable error, one can produce a single computable point in their intersection.

Lemma 9. *Let $G_1, G_2, \cdots \subseteq \Sigma^\infty$ be clopen sets such that for each t,*

$$1 - \mu(G_t) \leq 2^{-2t}.$$

Assume that for every L and cylinder $[v]$, the number

$$\mu\left(\left(\bigcap_{t \leq L} G_t \right) \cap [v] \right)$$

is computable as an exact rational uniformly in (L, v) (hence comparisons with rational thresholds are decidable). Then there exists a computable $y \in \bigcap_{t \geq 1} G_t$.

Proof. Write $\varepsilon_t := 2^{-2t}$ and let

$$\mathrm{Tail}(L) := \sum_{t > L} \varepsilon_t.$$

For $L \geq 0$ let $F_L := \bigcap_{t \leq L} G_t$ (with $F_0 = \Sigma^\infty$).

We build prefixes $v_\ell \in \Sigma^\ell$ such that

$$\mu\big([v_\ell] \cap \bigcap_{t \geq 1} G_t\big) > 0 \qquad (\ell \geq 0). \tag{1}$$

This holds at $\ell = 0$ because

$$\mu\big(\bigcap_{t \geq 1} G_t\big) \geq 1 - \sum_{t \geq 1} \mu(G_t^c) \geq 1 - \sum_{t \geq 1} \varepsilon_t = 1 - \frac{1}{3} = \frac{2}{3} > 0.$$

Inductive step. Assume 1 holds for some ℓ and set

$$\alpha := \mu\big([v_\ell] \cap \bigcap_{t \geq 1} G_t\big) > 0.$$

Since $F_L \downarrow \bigcap_{t \geq 1} G_t$ as $L \to \infty$ and measures are continuous from above,

$$\mu([v_\ell] \cap F_L) \downarrow \alpha \quad (L \to \infty).$$

Also $\mathrm{Tail}(L) \to 0$. Hence there exists $L \geq 0$ such that

$$\mu([v_\ell] \cap F_L) > k \, \mathrm{Tail}(L). \tag{2}$$

By the hypothesis (exact rational computation), we can find the least such L by brute force search.

Fix this least L. Since

$$\mu([v_\ell] \cap F_L) = \sum_{a \in \Sigma} \mu([v_\ell a] \cap F_L),$$

(2) implies that there exists at least one $a \in \Sigma$ with

$$\mu([v_\ell a] \cap F_L) > \mathrm{Tail}(L),$$

otherwise the sum would be at most $k \, \mathrm{Tail}(L)$. Let a be the least such symbol and set $v_{\ell+1} := v_\ell a$.

Now we show 1. Using the union bound and $\mu([v_{\ell+1}] \cap G_t^c) \leq \mu(G_t^c) \leq \varepsilon_t$,

$$\mu\big([v_{\ell+1}] \cap \bigcap_{t \geq 1} G_t\big) \geq \mu([v_{\ell+1}] \cap F_L) - \sum_{t > L} \mu([v_{\ell+1}] \cap G_t^c) \geq \mu([v_{\ell+1}] \cap F_L) - \mathrm{Tail}(L) > 0,$$

so 1 holds.

Let $y := \lim_\ell v_\ell \in \Sigma^\infty$. Since the cylinders $[v_\ell]$ are nested, y exists and lies in every $[v_\ell]$. Fix t. Since G_t is clopen, membership in G_t is decided by some prefix length r. Choose $\ell \geq r$. By 1, $[v_\ell]$ intersects $\bigcap_{s \geq 1} G_s \subseteq G_t$, hence $[v_\ell] \cap G_t \neq \emptyset$. Because $|v_\ell| \geq r$ and G_t is clopen, either $[v_\ell] \subseteq G_t$ or $[v_\ell] \cap G_t = \emptyset$. The latter is impossible, so $[v_\ell] \subseteq G_t$ and hence $y \in G_t$. As t was arbitrary, $y \in \bigcap_{t \geq 1} G_t$.

Computability: at stage ℓ we search effectively for the least L satisfying 2, then compute $\mu([v_\ell a] \cap F_L)$ for each $a \in \Sigma$ and choose the least a with value $> \mathrm{Tail}(L)$. Thus $v_{\ell+1}$ is computable from v_ℓ, and $y = \lim_\ell v_\ell$ is computable. $\qquad\square$

4.2 Almost-sure normality for a fixed shuffler

Fix a shuffler $S = (Q, q_0, \delta, \tau)$ over Σ. We prove: if x is normal and $Y \sim \mathrm{Unif}(\Sigma^\infty)$, then $S(x, Y)$ is normal almost surely.

We decompose the run into segments according to how many symbols have been consumed from tape X. Let $Z := S(x, Y)$. Let i_t be the number of x-symbols consumed after producing t output symbols. Fix an integer $N \geq 1$ and define stopping times

$$T_0 := 0, \qquad T_\ell := \min\{t : i_t = \ell N\} \in \mathbb{N} \cup \{\infty\}.$$

For each $\ell \geq 1$, define the ℓth segment by

$$W_\ell := \begin{cases} Z[T_{\ell-1} + 1..T_\ell] \in \Sigma^* & \text{if } T_\ell < \infty, \\ \lambda & \text{if } T_\ell = \infty, \end{cases}$$

where λ denotes the empty word (so $|W_\ell| = 0$ when $T_\ell = \infty$). Note that $|W_\ell| \geq N$ whenever $T_\ell < \infty$, since in that case W_ℓ contains exactly N symbols copied from tape X.

We will condition on the information revealed up to segment boundaries. Let $(\mathcal{G}_\ell)_{\ell \geq 0}$ be the natural filtration generated by the run up to time T_ℓ, equivalently by the output prefix $Z[1..T_\ell]$ (and hence by the Y-symbols revealed up to time T_ℓ). All conditional probabilities/expectations below are with respect to this filtration.

We next record the block-frequency deviation statistic used within a segment. For a finite word v and $m \geq 1$, write

$$\mathrm{occ}^{\mathrm{in}}(v, w) := \#\{1 \leq j \leq |v| - m + 1 : v[j..j+m-1] = w\}, \qquad \Delta_m(v) := \max_{w \in \Sigma^m} \left| \frac{\mathrm{occ}^{\mathrm{in}}(v, w)}{|v|} - k^{-m} \right|.$$

We also isolate the part of the state space that is relevant on the event of infinitely many X-reads. Let $Q_X := \tau^{-1}(1)$ and $Q_Y := \tau^{-1}(2)$. Consider the directed graph on Q with edges $q \to \delta(q, a)$ for $a \in \Sigma$. Let $R \subseteq Q$ be the union of all *sink strongly connected components* (i.e. strongly connected components with no outgoing edge) that intersect Q_X.

For fixed m, set

$$\mathcal{C}_R := R \times \Sigma^{\leq m-1}.$$

Lemma 10. *Define the segment-level hitting time*

$$\ell_R := \min\{\ell \geq 1 : \text{the shuffler state at time } T_{\ell-1} \text{ lies in } R\} \in \mathbb{N} \cup \{\infty\}.$$

On the event that tape X is read infinitely often (equivalently, $T_\ell < \infty$ for all ℓ), we have $\ell_R < \infty$ and for all $\ell \geq \ell_R$ the segment-start state at time $T_{\ell-1}$ lies in R. Moreover, for each L, the event $\{\ell_R \leq L\}$ is determined by the run up to time T_{L-1} (i.e. it is $\mathcal{G}_{L-1}$-measurable).

Proof. For any infinite run of a finite-state automaton, the set of states visited infinitely often is contained in a sink strongly connected component (SCC) of the underlying transition graph. On the event that tape X is read infinitely often, some state in Q_X is visited infinitely often, so the sink SCC visited infinitely often intersects Q_X and hence is contained in R. Therefore the run enters R and never leaves it. Since $T_\ell \to \infty$ on this event, some segment start time $T_{\ell-1}$ occurs after the entry time into R, hence $\ell_R < \infty$ and all later segment-start states lie in R.

The measurability claim holds because the shuffler state at time T_j is determined by the run up to time T_j, so whether some T_j with $j \leq L - 1$ has state in R is determined by the run up to T_{L-1}. $\qquad \square$

Finally, we define the segment experiment from a boundary context. For fixed m, let $\mathcal{C}_Q :=$ $Q \times \Sigma^{\leq m-1}$ be the set of *boundary contexts*, recording the current shuffler state together with the last $\leq m - 1$ output symbols. For $u \in \Sigma^N$ and $\sigma \in \mathcal{C}_Q$, define $W(u, \sigma)$ as the random segment produced when:

- the next N symbols read from tape X are fixed to be u,

- the segment starts from boundary context σ,

- and every Y-symbol used during this segment is fresh i.i.d. uniform in Σ.

(If the shuffler fails to consume N symbols from tape X under this experiment, we declare $W(u, \sigma)$ undefined.)

We now identify those X-blocks whose induced segments typically have near-uniform internal block statistics, uniformly over the relevant boundary contexts.

Definition 9. *Fix $m \geq 1$ and $\epsilon \in (0, 1/10)$. A word $u \in \Sigma^N$ is (m, ϵ)-balanced if for all $\sigma \in \mathcal{C}_R$,*

$$\mathbb{P}\big(W(u, \sigma) \text{ is defined and } \Delta_m(W(u, \sigma)) \leq \epsilon\big) \geq 1 - \epsilon,$$

where the probability is over the fresh Y-symbols in the segment experiment.

The next lemma shows that for sufficiently large segment length N, a uniformly random X-block is balanced with high probability.

Lemma 11. *For each fixed m and ϵ, there exists $N_0 = N_0(S, m, \epsilon)$ such that for all $N \geq N_0$, if $U \sim \mathrm{Unif}(\Sigma^N)$ then*
$$\mathbb{P}\big(U \text{ is } (m, \epsilon)\text{-balanced}\big) \geq 1 - \epsilon.$$

Proof. Fix $\sigma \in \mathcal{C}_R$ and consider the fully random model $(X, Y) \sim \mathrm{Unif}(\Sigma^\infty)^2$, started from boundary context σ. In this model, conditional on the past, the next output symbol is the next unused symbol from either tape, hence uniform and independent of the past; therefore the output process is i.i.d. uniform on Σ.

Let $U = X[1..N]$, and let W_1 be the first segment produced while consuming these N symbols from tape X. Then W_1 has the same distribution as $W(U, \sigma)$ (when defined). Since $\sigma \in \mathcal{C}_R$ and R is a union of sink strongly connected components intersecting Q_X, the state process stays inside the corresponding sink SCC; since this SCC is strongly connected, the induced finite Markov chain is irreducible, so every state in the SCC (in particular some state of Q_X) is visited infinitely often almost surely. In particular, W_1 is defined almost surely for every N.

Because $|W_1| \geq N$, we have $|W_1| \to \infty$ as $N \to \infty$. Since the output is i.i.d. uniform, $\Delta_m(Z[1..L]) \to 0$ almost surely as $L \to \infty$ (apply the strong law to each $w \in \Sigma^m$ and union bound). Therefore $\mathbb{P}(\Delta_m(W_1) \leq \epsilon) \to 1$ as $N \to \infty$.

Now define $f_\sigma(u) := \mathbb{P}(W(u, \sigma) \text{ defined and } \Delta_m(W(u, \sigma)) \leq \epsilon)$. Then

$$\mathbb{E}_{U \sim \mathrm{Unif}(\Sigma^N)}[f_\sigma(U)] = \mathbb{P}\big(\Delta_m(W_1) \leq \epsilon\big).$$

Thus for all sufficiently large N we have $\mathbb{E}[f_\sigma(U)] \geq 1 - \epsilon^2$. Markov's inequality on $1 - f_\sigma(U)$ yields $\mathbb{P}(f_\sigma(U) \leq 1 - \epsilon) \leq \epsilon$. Finally, union bound over the finite set $\mathcal{C}_R$ completes the proof. $\square$

The next lemma quantifies how rare unbalanced blocks are, both in the uniform distribution on Σ^N and along the aligned N-block decomposition of a normal word.

Lemma 12. *Let $B \subseteq \Sigma^N$ be the set of (m, ϵ)-unbalanced words. Then $|B| \leq \epsilon k^N$. Moreover, if $x = u_1 u_2 \cdots$ is the aligned N-block decomposition of a normal word, then*

$$\limsup_{L \to \infty} \frac{\#\{1 \leq \ell \leq L : u_\ell \in B\}}{L} \leq 2\epsilon.$$

Proof. The size bound is immediate from Lemma 11. For the density bound, normality implies that aligned N-blocks have limiting frequencies k^{-N} for each $u \in \Sigma^N$; summing over $u \in B$ gives limiting density $|B|/k^N \leq \epsilon$, and the displayed bound allows slack. $\square$

We will use the standard bounded-difference Azuma–Hoeffding inequality: if $(D_j)_{j \geq 1}$ is a martingale difference sequence with $|D_j| \leq 1$ almost surely, then for all $t > 0$,

$$\mathbb{P}\Big(\sum_{j=1}^L D_j \leq -t\Big) \leq \exp\Big(-\frac{t^2}{2L}\Big) \qquad \text{and} \qquad \mathbb{P}\Big(\sum_{j=1}^L D_j \geq t\Big) \leq \exp\Big(-\frac{t^2}{2L}\Big).$$

The next lemma shows that, along the segment decomposition, almost all segments have small internal m-block deviation (conditioned on tape X being read infinitely often).

Lemma 13. *Fix m, ϵ and $N \geq N_0(S, m, \epsilon)$. Let B_ℓ be the event that $T_\ell < \infty$ and $\Delta_m(W_\ell) \leq \epsilon$. Then on the event that $T_L < \infty$ for all L (i.e. tape X is read infinitely often), we have*

$$\mathbb{P}\left(\liminf_{L \to \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbf{1}_{B_\ell} \geq 1 - 5\epsilon\right) = 1.$$

Proof. Write $x = u_1 u_2 \cdots$ in aligned N-blocks and call ℓ *balanced* if u_ℓ is (m, ϵ)-balanced. By Lemma 12, for all sufficiently large L, at least $(1 - 3\epsilon)L$ indices in $\{1, \ldots, L\}$ are balanced. Let $I_L := \{\ell \leq L : u_\ell \text{ is balanced}\}$, so $|I_L| \geq (1 - 3\epsilon)L$ for all large L.

Work on the event $\mathcal{I} := \{T_\ell < \infty \; \forall \ell\}$. For each ℓ, let $p_\ell := \mathbb{P}(B_\ell \mid \mathcal{G}_{\ell-1})$. Conditioned on $\mathcal{G}_{\ell-1}$, the next N symbols to be consumed from tape X are the fixed block u_ℓ, and all as-yet-unread Y-symbols that may be used during segment ℓ are fresh i.i.d. uniform on Σ; therefore, the conditional law of W_ℓ agrees with the segment experiment $W(u_\ell, \sigma_\ell)$ where σ_ℓ is the realized boundary context at time $T_{\ell-1}$.

If $\ell \in I_L$ and $\ell \geq \ell_R$, then $\sigma_\ell \in \mathcal{C}_R$ and by balance of u_ℓ we have $p_\ell \geq 1 - \epsilon$.

Define, for each fixed L, the process

$$M_j^{(L)} := \sum_{\ell \leq j, \; \ell \in I_L} \big(\mathbf{1}_{B_\ell} - p_\ell\big), \qquad j = 0, 1, \ldots, L,$$

with $M_0^{(L)} = 0$. Since B_ℓ is determined by the run up to time T_ℓ, the indicator $\mathbf{1}_{B_\ell}$ is $\mathcal{G}_\ell$-measurable; by definition $p_\ell = \mathbb{P}(B_\ell \mid \mathcal{G}_{\ell-1}) = \mathbb{E}[\mathbf{1}_{B_\ell} \mid \mathcal{G}_{\ell-1}]$. Therefore for each $\ell \leq L$,

$$\mathbb{E}\big[M_\ell^{(L)} - M_{\ell-1}^{(L)} \mid \mathcal{G}_{\ell-1}\big] = \mathbf{1}_{\{\ell \in I_L\}} \mathbb{E}\big[\mathbf{1}_{B_\ell} - p_\ell \mid \mathcal{G}_{\ell-1}\big] = 0,$$

so $(M_j^{(L)})_{j=0}^L$ is a martingale with respect to $(\mathcal{G}_j)_{j \geq 0}$. Moreover, its increments are bounded: for $\ell \in I_L$, $\mathbf{1}_{B_\ell} - p_\ell \in [-1, 1]$ since $\mathbf{1}_{B_\ell} \in \{0, 1\}$ and $p_\ell \in [0, 1]$, and for $\ell \notin I_L$ the increment is 0. Finally, write $M_L := M_L^{(L)} = \sum_{\ell \in I_L}(\mathbf{1}_{B_\ell} - p_\ell)$. The Azuma–Hoeffding inequality for martingales with bounded differences gives, for each L,

$$\mathbb{P}\big(M_L \leq -\epsilon |I_L|\big) \leq \exp(-\epsilon^2 |I_L|/2),$$

19

and the right-hand side is summable in L because $|I_L| \geq cL$ eventually. By Borel–Cantelli, almost surely for all sufficiently large L,

$$\sum_{\ell \in I_L} \mathbf{1}_{B_\ell} \geq \sum_{\ell \in I_L} p_\ell - \epsilon |I_L|.$$

On $\mathcal{I}$, we have $\ell_R < \infty$ by Lemma 10, hence for all large L, at most ℓ_R indices in I_L are $< \ell_R$. Therefore,

$$\sum_{\ell \in I_L} p_\ell \geq (1 - \epsilon)\big(|I_L| - \ell_R\big).$$

Combining,

$$\sum_{\ell \in I_L} \mathbf{1}_{B_\ell} \geq (1 - 2\epsilon)|I_L| - (1 - \epsilon)\ell_R \qquad \text{for all large } L \text{ on } \mathcal{I}.$$

Since $\sum_{\ell=1}^{L} \mathbf{1}_{B_\ell} \geq \sum_{\ell \in I_L} \mathbf{1}_{B_\ell}$ and $|I_L| \geq (1 - 3\epsilon)L$ eventually, dividing by L and letting $L \to \infty$ yields

$$\liminf_{L \to \infty} \frac{1}{L} \sum_{\ell=1}^{L} \mathbf{1}_{B_\ell} \geq 1 - 5\epsilon \qquad \text{almost surely on } \mathcal{I},$$

as required. $\qquad\square$

We also need a uniform tail bound on segment lengths to rule out the possibility that a few exceptionally long segments dominate the statistics.

Lemma 14. *There exist constants $\theta > 0$ and $C \in (0, \infty)$, depending only on S, such that for every $N \geq 1$, every segment index ℓ, and every history $\mathcal{G}_{\ell-1}$ on which segment ℓ starts in R,*

$$\mathbb{E}\big[e^{\theta |W_\ell|} \mid \mathcal{G}_{\ell-1}\big] \leq e^{CN}.$$

Proof. Since R is finite and each sink strongly connected component in R intersects Q_X, there exists L_0 such that from every $q \in R \cap Q_Y$ there is a word $a_1 \cdots a_{L_0}$ whose induced state trajectory stays in Q_Y until it first hits Q_X at step $\leq L_0$. Let $p := k^{-L_0}$. Then, starting from any $q \in R \cap Q_Y$, the probability to hit Q_X within L_0 Y-steps is at least p, so the Y-waiting time until the next visit to Q_X is stochastically dominated by $L_0 \cdot \mathrm{Geom}(p)$ and hence has a uniform exponential moment for some $\theta > 0$.

Within segment ℓ, exactly N X-symbols are consumed. Between consecutive consumptions from X, the number of output steps spent reading from Y has a conditional exponential moment bounded uniformly over segment-start histories with start state in R by the preceding domination argument. Iterating N times and applying the tower property yields $\mathbb{E}[e^{\theta |W_\ell|} \mid \mathcal{G}_{\ell-1}] \leq e^{CN}$ for some C depending only on S and θ. $\qquad\square$

As a direct consequence, we obtain an exponential tail bound for the segment length via Markov's inequality.

Lemma 15. *Assume segment ℓ starts in R. With θ, C as in Lemma 14, for every $t \geq 0$,*

$$\mathbb{P}\big(|W_\ell| > t \mid \mathcal{G}_{\ell-1}\big) \leq \exp(CN - \theta t).$$

Proof. Markov's inequality and Lemma 14 give

$$\mathbb{P}(|W_\ell| > t \mid \mathcal{G}_{\ell-1}) \leq e^{-\theta t} \mathbb{E}[e^{\theta |W_\ell|} \mid \mathcal{G}_{\ell-1}] \leq \exp(CN - \theta t).$$

$\qquad\square$

We next show that segments longer than a fixed multiple of N contribute only a vanishing fraction of the total output length.

Lemma 16. *Fix $N \geq 1$ and $D \geq 1$. On the event that tape X is read infinitely often, almost surely*

$$\limsup_{L \to \infty} \frac{1}{LN} \sum_{\ell=1}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}} \leq \frac{8}{\theta N} \exp\left(\left(C - \frac{\theta D}{2}\right) N\right),$$

where θ, C are as in Lemma 14.

Proof. Work on the event $\mathcal{I} := \{T_\ell < \infty \; \forall \ell\}$, so $\ell_R < \infty$ by Lemma 10.

Fix $D \geq 1$. For $j \geq 0$ define the length class events

$$A_{\ell,j} := \{2^j DN < |W_\ell| \leq 2^{j+1} DN\}, \qquad N_{L,j} := \sum_{\ell=1}^{L} \mathbf{1}_{\{\ell_R \leq \ell\}} \mathbf{1}_{A_{\ell,j}}.$$

By Lemma 15, on the event $\{\ell_R \leq \ell\}$ we have

$$\mathbb{P}(A_{\ell,j} \mid \mathcal{G}_{\ell-1}) \leq \mathbb{P}(|W_\ell| > 2^j DN \mid \mathcal{G}_{\ell-1}) \leq \exp(CN - \theta \cdot 2^j DN) =: p_j.$$

Define the martingale (with respect to $(\mathcal{G}_\ell)_{\ell \geq 0}$)

$$M_{L,j} := \sum_{\ell=1}^{L} \left(\mathbf{1}_{\{\ell_R \leq \ell\}} \mathbf{1}_{A_{\ell,j}} - \mathbb{E}[\mathbf{1}_{\{\ell_R \leq \ell\}} \mathbf{1}_{A_{\ell,j}} \mid \mathcal{G}_{\ell-1}]\right),$$

whose increments are bounded by 1. Since $\mathbb{E}[\mathbf{1}_{\{\ell_R \leq \ell\}} \mathbf{1}_{A_{\ell,j}} \mid \mathcal{G}_{\ell-1}] \leq p_j \mathbf{1}_{\{\ell_R \leq \ell\}}$, we have

$$\sum_{\ell=1}^{L} \mathbb{E}[\mathbf{1}_{\{\ell_R \leq \ell\}} \mathbf{1}_{A_{\ell,j}} \mid \mathcal{G}_{\ell-1}] \leq p_j L.$$

Hence the event $N_{L,j} \geq 2p_j L$ implies $M_{L,j} \geq p_j L$, and the Azuma–Hoeffding inequality yields

$$\mathbb{P}(N_{L,j} \geq 2p_j L) \leq \mathbb{P}(M_{L,j} \geq p_j L) \leq \exp(-p_j^2 L/2),$$

which is summable in L. By Borel–Cantelli, almost surely for each fixed j, for all sufficiently large L,

$$N_{L,j} \leq 2p_j L,$$

and hence $\limsup_{L \to \infty} N_{L,j}/L \leq 2p_j$.

Now,

$$\sum_{\ell=1}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}} = \sum_{\ell < \ell_R} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}} + \sum_{\ell=\ell_R}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}}.$$

The first term is $O(1)$ on $\mathcal{I}$ and vanishes after dividing by LN. For the second term,

$$\sum_{\ell=\ell_R}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}} \leq \sum_{j \geq 0} 2^{j+1} DN \cdot N_{L,j}.$$

For each J,

$$\limsup_{L \to \infty} \frac{1}{LN} \sum_{\ell=\ell_R}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}} \leq \sum_{j=0}^{J} 2^{j+2} D \cdot \limsup_{L \to \infty} \frac{N_{L,j}}{L},$$

and letting $J \to \infty$ (monotone in J) gives

$$\limsup_{L \to \infty} \frac{1}{LN} \sum_{\ell=\ell_R}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}} \leq \sum_{j \geq 0} 2^{j+2} D \cdot 2 p_j = 4D\, e^{CN} \sum_{j \geq 0} 2^j e^{-\theta DN 2^j}.$$

We use the elementary bound: for all $a > 0$,

$$\sum_{j \geq 0} 2^j e^{-a2^j} \leq \frac{2}{a} e^{-a/2}.$$

Indeed, for each $j \geq 0$ and all $u \in [2^{j-1}, 2^j]$ we have $e^{-au} \geq e^{-a2^j}$, hence

$$\int_{2^{j-1}}^{2^j} e^{-au}\, du \geq 2^{j-1} e^{-a2^j} \quad \Rightarrow \quad 2^j e^{-a2^j} \leq 2 \int_{2^{j-1}}^{2^j} e^{-au}\, du.$$

Summing over $j \geq 0$ yields

$$\sum_{j \geq 0} 2^j e^{-a2^j} \leq 2 \int_{1/2}^{\infty} e^{-au}\, du = \frac{2}{a} e^{-a/2}.$$

With $a := \theta DN$, the right-hand side is at most

$$\frac{8}{\theta N} \exp\left(\left(C - \frac{\theta D}{2}\right) N\right).$$

$\square$

Finally, we record that the next segment becomes negligible compared to the total output produced so far.

Lemma 17. *On the event that tape X is read infinitely often, we have*

$$\frac{|W_{L+1}|}{T_L} \to 0 \qquad (L \to \infty)$$

almost surely.

Proof. Work on the event $\mathcal{I} := \{T_\ell < \infty \ \forall \ell\}$, so $\ell_R < \infty$ by Lemma 10. Let

$$b_L := \frac{C}{\theta} N + \frac{2}{\theta} \log(L+2).$$

By Lemma 15, on the event $\{\ell_R \leq L\}$ we have

$$\mathbb{P}(|W_L| > b_L \mid \mathcal{G}_{L-1}) \leq \exp(CN - \theta b_L) \leq (L+2)^{-2}.$$

Hence

$$\mathbb{P}(|W_L| > b_L \text{ and } \ell_R \leq L) \leq (L+2)^{-2},$$

and the right-hand side is summable. By Borel–Cantelli, almost surely only finitely many L satisfy $|W_L| > b_L$ and $\ell_R \leq L$. On $\mathcal{I}$ we have $\ell_R < \infty$, so for all sufficiently large L, $\ell_R \leq L$ and thus $|W_L| \leq b_L$.

Also $T_L = \sum_{\ell \leq L} |W_\ell| \geq LN$, hence for large L,

$$\frac{|W_{L+1}|}{T_L} \leq \frac{b_{L+1}}{LN} \to 0.$$

$\square$

We now pass from segment-level control to prefix frequencies. For a concatenation $W_1 \cdots W_L$ and any $w \in \Sigma^m$, at most $(m-1)L$ occurrences of w start in the last $(m-1)$ symbols of some W_ℓ, hence cross a segment boundary. Apart from these boundary-crossing occurrences, the count of w in $Z[1..T_L] = W_1 \cdots W_L$ is the sum of the within-segment counts. The next lemma formalizes this bookkeeping: if most segments have small internal deviation and the total length contributed by very long segments is small, then the prefix $Z[1..T_L]$ has near-uniform m-block frequencies.

Lemma 18. *Fix $m \geq 1$, $\epsilon \in (0, 1/10)$, and $N \geq N_0(S, m, \epsilon)$. Let B_ℓ be the event that $T_\ell < \infty$ and $\Delta_m(W_\ell) \leq \epsilon$. Fix $D \geq 1$ and $\eta > 0$. On the event that $T_L < \infty$ for all L, for all sufficiently large L, if*

$$\frac{1}{L} \sum_{\ell=1}^{L} \mathbf{1}_{B_\ell} \geq 1 - 6\epsilon \qquad and \qquad \frac{1}{LN} \sum_{\ell=1}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell| > DN\}} \leq \eta,$$

then for every $w \in \Sigma^m$,

$$\left| \frac{\mathrm{occ}(Z[1..T_L], w)}{T_L} - k^{-m} \right| \leq \epsilon + 6\epsilon D + \eta + \frac{m-1}{N}.$$

Proof. Write $Z[1..T_L] = W_1 \cdots W_L$. Boundary-crossing occurrences contribute at most $(m-1)L$, hence at most $(m-1)L/T_L \leq (m-1)/N$ since $T_L = \sum_{\ell \leq L} |W_\ell| \geq LN$.

For internal occurrences, on B_ℓ we have $\mathrm{occ}^{\mathrm{in}}(W_\ell, w) = k^{-m} |W_\ell| \pm \epsilon |W_\ell|$, so

$$\sum_{\ell=1}^{L} \mathrm{occ}^{\mathrm{in}}(W_\ell, w) = k^{-m} T_L \pm \epsilon T_L + \sum_{\ell \notin B} \mathrm{occ}^{\mathrm{in}}(W_\ell, w),$$

where $B := \{\ell \leq L : B_\ell \text{ holds}\}$. For $\ell \notin B$, use $\mathrm{occ}^{\mathrm{in}}(W_\ell, w) \leq |W_\ell|$ to get

$$\left| \sum_{\ell=1}^{L} \mathrm{occ}^{\mathrm{in}}(W_\ell, w) - k^{-m} T_L \right| \leq \epsilon T_L + \sum_{\ell \notin B} |W_\ell|.$$

Split the bad-length sum into short and long segments:

$$\sum_{\ell \notin B} |W_\ell| \leq \sum_{\ell \notin B, |W_\ell| \leq DN} |W_\ell| + \sum_{\ell : |W_\ell| > DN} |W_\ell|.$$

The first term is at most $DN \cdot \#\{\ell \notin B\} \leq DN \cdot 6\epsilon L$. The second term is at most ηLN by hypothesis. Dividing by $T_L \geq LN$ yields

$$\frac{1}{T_L} \sum_{\ell \notin B} |W_\ell| \leq 6\epsilon D + \eta.$$

Combine with the boundary term to obtain the stated inequality. $\qquad \square$

We can now combine the segment analysis to obtain the desired almost-sure normality statement for a fixed shuffler.

Theorem 3. *Let S be any shuffler. If x is normal and $Y \sim \mathrm{Unif}(\Sigma^\infty)$, then $S(x, Y)$ is normal almost surely.*

Proof. If along the run the shuffler eventually reads only from tape X, then the output is a suffix of x and hence normal. If it eventually reads only from tape Y, then the output is a suffix of Y and hence normal almost surely.

Let $\mathcal{I}$ be the remaining event that tape X is read infinitely often (equivalently, $T_L < \infty$ for all L). It suffices to prove that on $\mathcal{I}$ the output Z is normal almost surely.

Fix $m \geq 1$ and $\delta > 0$. Run the argument below with target accuracy $\delta/2$ and then extend to all n.

Choose $D \geq 1$ so that $\theta D/2 > C$, where θ, C are from Lemma 14. Choose $\epsilon \in (0, 1/10)$ so that $\epsilon(1 + 6D) \leq \delta/8$. Choose $N \geq N_0(S, m, \epsilon)$ so large that $(m-1)/N \leq \delta/8$ and

$$\frac{8}{\theta N} \exp\left(\left(C - \frac{\theta D}{2}\right) N\right) \leq \delta/16.$$

(Note that $(m-1)/N < 1$ implies $N \geq m$.)

By Lemma 16, on $\mathcal{I}$ almost surely,

$$\limsup_{L\to\infty} \frac{1}{LN} \sum_{\ell=1}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell|>DN\}} \leq \delta/16.$$

Hence on $\mathcal{I}$ almost surely, for all sufficiently large L,

$$\frac{1}{LN} \sum_{\ell=1}^{L} |W_\ell| \mathbf{1}_{\{|W_\ell|>DN\}} \leq \delta/8.$$

By Lemma 13, on $\mathcal{I}$ almost surely,

$$\liminf_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L} \mathbf{1}_{B_\ell} \geq 1 - 5\epsilon.$$

Hence on $\mathcal{I}$ almost surely, for all sufficiently large L,

$$\frac{1}{L} \sum_{\ell=1}^{L} \mathbf{1}_{B_\ell} \geq 1 - 6\epsilon.$$

Therefore, on $\mathcal{I}$ almost surely, for all sufficiently large L the hypotheses of Lemma 18 hold with $\eta = \delta/8$, and hence

$$\max_{w \in \Sigma^m} \left| \frac{\operatorname{occ}(Z[1..T_L], w)}{T_L} - k^{-m} \right| \leq \delta/2$$

for all sufficiently large L.

By Lemma 17, $|W_{L+1}|/T_L \to 0$ on $\mathcal{I}$ almost surely. Fix such an ω. Then for all sufficiently large L, for every $n \in [T_L, T_{L+1}]$ and every $w \in \Sigma^m$, writing $r := n - T_L \leq |W_{L+1}|$, the number of new occurrences of w created by extending from T_L to n is at most $r + (m-1)$, so

$$|\operatorname{occ}(Z[1..n], w) - \operatorname{occ}(Z[1..T_L], w)| \leq r + (m-1) \leq |W_{L+1}| + (m-1) \leq 2|W_{L+1}|,$$

using $|W_{L+1}| \geq N \geq m$. Therefore

$$\left| \frac{\operatorname{occ}(Z[1..n], w)}{n} - \frac{\operatorname{occ}(Z[1..T_L], w)}{T_L} \right| \leq \frac{2|W_{L+1}|}{T_L} + \frac{|W_{L+1}|}{T_L} = 3\frac{|W_{L+1}|}{T_L}.$$

For all sufficiently large L, the right-hand side is $\leq \delta/2$, uniformly over $n \in [T_L, T_{L+1}]$ and $w \in \Sigma^m$. Combining with the $\delta/2$ bound at time T_L gives the desired δ bound for all sufficiently large n.

Since m and δ were arbitrary, $Z = S(x, Y)$ is normal almost surely. $\qquad\square$

4.3 From almost-sure normality to a computable y

We now compute a single computable y that simultaneously forces $S(x, y)$ to be normal for every shuffler S.

Fix a computable enumeration $S_1, S_2, \ldots$ of all shufflers over Σ. For $t \geq 1$ and $N \geq 2$, define the *quadratic window slice*

$$H_t(N)(x) := \bigcap_{i=1}^{t} \bigcap_{\ell=1}^{t} \bigcap_{n=N}^{N^2} E_{S_i}\big(n; \ell, 2^{-t}\big)(x) \subseteq \Sigma^\infty.$$

This is clopen (finite intersection of clopen sets). Since x is computable, each measure $\mu(H_t(N)(x) \cap [v])$ is computable uniformly in (t, N, v) by Lemma 8 and closure under finite intersections. The next lemma records that, for a fixed normal source x, the finite test family defining $H_t(N)(x)$ succeeds with probability tending to 1 as the cutoff N goes to infinity.

Lemma 19. *Fix $t \geq 1$ and a normal x. Then for $Y \sim \mathrm{Unif}(\Sigma^\infty)$,*

$$\mathbb{P}\big(Y \in H_t(N)(x)\big) \to 1 \qquad (N \to \infty).$$

Proof. Fix $i \leq t$ and $\ell \leq t$. By Theorem 3, $S_i(x, Y)$ is normal almost surely. For the fixed tolerance $2^{-t} > 0$, almost surely there exists $N_{i,\ell}$ such that for all $n \geq N_{i,\ell}$, $\mathrm{Test}(S_i(x, Y)[1..n]; \ell, 2^{-t})$ holds. Let $N_* := \max_{i \leq t, \ell \leq t} N_{i,\ell}$. Then almost surely, for all $N \geq N_*$, all tests in the definition of $H_t(N)(x)$ hold (since $[N, N^2] \subseteq [N_*, \infty)$), and hence $\mathbf{1}_{H_t(N)(x)}(Y) \to 1$ almost surely. By dominated convergence, $\mathbb{P}(Y \in H_t(N)(x)) \to 1$. $\square$

We next choose concrete cutoffs N_t so that each slice $H_t(N_t)(x)$ has very high measure, while keeping N_t growing in a controlled way.

Lemma 20. *Assume x is computable and normal. There exists a computable increasing sequence $2 \leq N_1 < N_2 < N_3 < \cdots$ such that for each $t \geq 1$,*

$$\mu\big(H_t(N_t)(x)\big) \geq 1 - 2^{-2t} \qquad \text{and} \qquad N_{t+1} \leq N_t^2.$$

Proof. We construct (N_t) inductively by search with one-step lookahead. Here the searches are effective because each $H_t(N)(x)$ is clopen, so $\mu(H_t(N)(x))$ is a rational that can be computed exactly (by Lemma 8 and closure under finite intersections), hence the threshold tests are decidable.

For $t = 1$, search $N \geq 2$ until $\mu(H_1(N)(x)) \geq 1 - 2^{-2}$ and there exists some M with $N < M \leq N^2$ such that $\mu(H_2(M)(x)) \geq 1 - 2^{-4}$ and there exists some M' with $M < M' \leq M^2$ such that $\mu(H_3(M')(x)) \geq 1 - 2^{-6}$. Set $N_1 := N$ and N_2 to be the least such M.

Now assume $t \geq 2$ and N_t has been chosen. Search for the least M with $N_t < M \leq N_t^2$ such that

$$\mu(H_{t+1}(M)(x)) \geq 1 - 2^{-2(t+1)}$$

and there exists some M' with $M < M' \leq M^2$ such that

$$\mu(H_{t+2}(M')(x)) \geq 1 - 2^{-2(t+2)}.$$

Set $N_{t+1} := M$.

Existence follows from Lemma 19 (applied to the finitely many indices involved at each stage): since $\mu(H_s(N)(x)) \to 1$ as $N \to \infty$ for each fixed s, one can take M sufficiently large so that $\mu(H_{t+1}(M)(x))$ meets the stage threshold and also M^2 exceeds some stage-$(t + 2)$ witness. By construction, (N_t) is strictly increasing and satisfies $N_{t+1} \leq N_t^2$ and the stated measure bounds. $\square$

Now, we prove the main result of this section.

Theorem 4. *If $x \in \Sigma^\infty$ is computable and normal, then there exists a computable $y \in \Sigma^\infty$ such that:*

(a) y is normal;

(b) for every shuffler S, the output $S(x,y)$ is normal.

In particular, x and y are finite-state independent by the shuffler characterization theorem.

Proof. Assume x is computable and normal, and let $(N_t)_{t \geq 1}$ be the computable increasing sequence given by Lemma 20. Define $G_t := H_t(N_t)(x)$. Then each G_t is clopen with $\mu(G_t) \geq 1 - 2^{-2t}$, so

$$1 - \mu(G_t) \leq 2^{-2t} \qquad \text{and} \qquad \sum_{t \geq 1} 2^{-2t} = 1/3 < 1.$$

Apply Lemma 9 to obtain a computable $y \in \bigcap_{t \geq 1} G_t$.

Fix $i \geq 1$ and a block length $\ell \geq 1$, and let $z := S_i(x,y)$. Let $\delta > 0$ be arbitrary and choose $t \geq \max\{i, \ell\}$ with $2^{-t} \leq \delta$.

We claim that for every $n \geq N_t$,

$$\text{Test}\big(z[1..n]; \ell, \delta\big) \text{ holds.}$$

Indeed, let $s \geq t$ be the maximal index such that $N_s \leq n$ (well-defined since $N_s \to \infty$). Then $n < N_{s+1} \leq N_s^2$ by Lemma 20, hence $n \in [N_s, N_s^2]$. Since $y \in G_s = H_s(N_s)(x)$, we have

$$\text{Test}\big(S_i(x,y)[1..n]; \ell, 2^{-s}\big)$$

and since $2^{-s} \leq 2^{-t} \leq \delta$, this implies $\text{Test}(z[1..n]; \ell, \delta)$ as claimed.

Because $\delta > 0$ was arbitrary, it follows that for each fixed ℓ, the ℓ-block frequencies of z converge to $k^{-\ell}$, hence z is normal. Since i was arbitrary, this proves (b).

For (a), let S^{allY} be the shuffler that always reads tape Y. Then $S^{\text{allY}}(x,y) = y$, so applying (b) to S^{allY} shows that y is normal. $\square$

References

[1] Verónica Becher, Olivier Carton, and Pablo Ariel Heiber. Finite-state independence. *Theory of Computing Systems*, 62(7):1555–1572, 2018.

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

[3] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.

[4] Jack H. Lutz and Elvira Mayordomo. Computing absolutely normal numbers in nearly linear time. *Information and Computation*, 281:104746, 2021.

[5] Alan Turing. A note on normal numbers. *Collected Works of AM Turing: Pure Mathematics*, pages 117–119, 1992.

[6] Nicolás Álvarez, Verónica Becher, and Olivier Carton. Finite-state independence and normal sequences. *Journal of Computer and System Sciences*, 103:1–17, 2019.

[7] M. Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 27(1):247–271, December 1909.