

Bernoulli Schemes with a finite alphabet

! Bernoulli Schemes

We are concerned with sequences of *very long* words (x_0, x_1, x_2, \dots) , where each x_i is taken from a finite alphabet $\mathcal{A} := \{a, b, \dots\}$. We will consider infinitely-long words, which is a standard approach if the words length is much larger than the number of iterations of the shift map (defined below). These structures are a basic building block for Statistical Mechanics, Dynamical Systems and Finite Automata. Here we investigate the simplest example, the case where we sample each letter independently.

This page contains interactive content, feel free to explore and modify it. For more involved modifications, you are encouraged to run and edit the interactive [Jupyter Notebook](#) instead. You can either:

- Run it on a computer with a Python/Jupyter-lab installation (requires ipywidgets, numpy, matplotlib).
- Run it using an online service. The official [Jupyter website](#) provides a free and open service. If you need more computing power, you can import the notebook in [Google Colab](#) (requires a Google account).

Definitions

Let \mathcal{A} be a finite or countable set and let $M: \mathcal{A} \times \mathcal{A} \rightarrow \{0, 1\}$. \mathcal{A} represents the alphabet, and M encodes which pairs of letters can be consecutive (more complex rules for the allowed sequences of letters fit within this framework just by modifying the alphabet space).

The space $\Sigma \equiv \Sigma(\mathcal{A}, M)$ of M -compatible words is the space of sequences

$$\Sigma := \left\{ x \in \mathcal{A}^{\mathbb{N}} : M_{x_i, x_{i+1}} = 1, \forall i \in \mathbb{N} \right\}$$

In other words, M tells us which letters can be consecutive, and Σ is the space of infinitely long words that only contain consecutive pairs of allowed letters. Σ is a measurable space, being a measurable subset of $\mathcal{A}^{\mathbb{N}}$ (equipped with the product σ -algebra). Elements of Σ are usually denoted by x, y, z , and we write $x = (x_0, x_1 \dots)$ and so on.

Definition 0.1 (Shift map). The map

$$\begin{aligned} T: \Sigma &\rightarrow \Sigma \\ (Tx)_i &= x_{i+1} \end{aligned}$$

is called the **shift** map.

Usually the pair (\mathcal{A}, M) (and thus the ensuing space Σ equipped with the shift map T) is called a **symbolic dynamical system**.

Notice that T is not invertible (unless \mathcal{A} has just one element), since the first letter of x disappears in Tx , the second letter becomes the first, and so on.

Definition 0.2 (Invariant measure). A probability measure \mathbb{P} on Σ is called **invariant** if

$$\mathbb{P}(E) = \mathbb{P}(T^{-1}(E)) \quad (1)$$

Let $\mu \in \mathcal{P}(\mathcal{A})$ be a probability measure on \mathcal{A} . If $M_{a,b} = 1$ for all $a, b \in \mathcal{A}$, then the product measure $\mathbb{P} = \mu^{\otimes \mathbb{N}}$ is invariant. This statement is obvious, it is just a difficult way of saying that

$$\mathbb{P}(X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}) = \mathbb{P}(X_1 \in A_0, \dots, X_n \in A_{n-1})$$

if the X_i are i.i.d.. Indeed, by the definition of the product σ -algebra, it is enough to check Equation 1 for cylindrical sets $E = A_0 \times \dots \times A_{n-1} \times \mathcal{A} \times \dots \mathcal{A}$.

Definition 0.3 (Bernoulli Scheme). A **Bernoulli Scheme** is a symbolic dynamical system where \mathcal{A} is finite or countable, $M_{a,b} = 1$ for all $a, b \in \mathcal{A}$, and $\mathbb{P} = \mu^{\otimes \mathbb{N}}$ is a product probability measure on $\Sigma = \mathcal{A}^{\mathbb{N}}$, for some probability μ on \mathcal{A} . In other words, it is just a sequence of i.i.d. random variables over a finite or countable space \mathcal{A} , each with distribution μ . If \mathcal{A} is finite, then the scheme is called **finite**.

Example 0.1. Consider a Bernoulli scheme over the alphabet $\mathcal{A} = \{0, 1\}$. This is identified by a parameter $p \in [0, 1]$ given by $p = \mu(\{1\})$. To avoid trivial

situations, we choose $p \in (0, 1)$. Up to a countable set (and thus a set of \mathbb{P} -measure 0), $\Sigma = \{0, 1\}^{\mathbb{N}}$ is in bijection with the interval $[0, 1]$ via the map

$$\Phi(x) := \sum_{i \geq 0} 2^{-i-1} x_i$$

The point $\Phi(x)$ will be in the interval $[0, 1/2]$ if $x_0 = 0$, and in the interval $[1/2, 1]$ if $x_0 = 1$. Similarly $\Phi(x) \in [0, 1/4]$ if $x_0 = 0$ and $x_1 = 0$, $\Phi(x) \in [1/4, 1/2]$ if $x_0 = 0$ and $x_1 = 1$, and so on. This is not entirely exact, since dyadic numbers admit two binary representations, e.g. $\Phi(01111111 \dots) = 1/2 = \Phi(10000000 \dots)$. But this is just a countable set that we will neglect since it has measure 0 (it is possible to make Φ bijective and measurable changing its definition on a countable set but we do not need it here).

So we can look at this Bernoulli scheme on $[0, 1]$ instead. What does the shift map $x \mapsto Tx$ correspond to? In other words, can we identify a map $T': [0, 1] \rightarrow [0, 1]$ such that $T'(\Phi(x)) = \Phi(Tx)$? This is just a way of saying: I first change the space to represent my points, from $\Sigma \ni x$ to $[0, 1] \ni \Phi(x)$. Now x changes to Tx , can I read this change *autonomously* on $[0, 1]$, just a function of the image $\Phi(x)$? Of course we can, since Φ is invertible a.e.. So $T'y = \Phi(T\Phi^{-1}y)$. We claim that $T'y = 2y(\text{mod } 1)$; which it is easy to check. Indeed

$$\Phi(Tx) = \sum_{i \geq 0} 2^{-i-1} x_{i+1} = \begin{cases} 2\Phi(x) & \text{if } x_0 = 0 \\ 2\Phi(x) - 1 & \text{if } x_0 = 1 \end{cases}$$

[../media/bs_doubling.webm](#)

Finite Schemes

Let us focus on finite Bernoulli schemes. So let $\mathcal{A} = \{0, 1, \dots, n-1\}$, $\Sigma = \mathcal{A}^{\mathbb{N}}$ and $p_k := \mu(\{k\})$ so that $\sum_i p_i = 1$. We assume $p_i > 0$ (otherwise just neglect the letters of the alphabet with 0 probability). As in Example 0.1 consider the n -ary representation

$$\begin{aligned} \Phi: \Sigma &\rightarrow [0, 1] \\ \Phi(x) &:= \sum_{i \geq 0} n^{-i-1} x_i \end{aligned}$$

Exercise 0.1. Prove that in this case, the n -ary representation pushes the shift map on Σ to the map $[0, 1] \ni y \mapsto ny(\text{mod } 1)$ on $[0, 1]$.

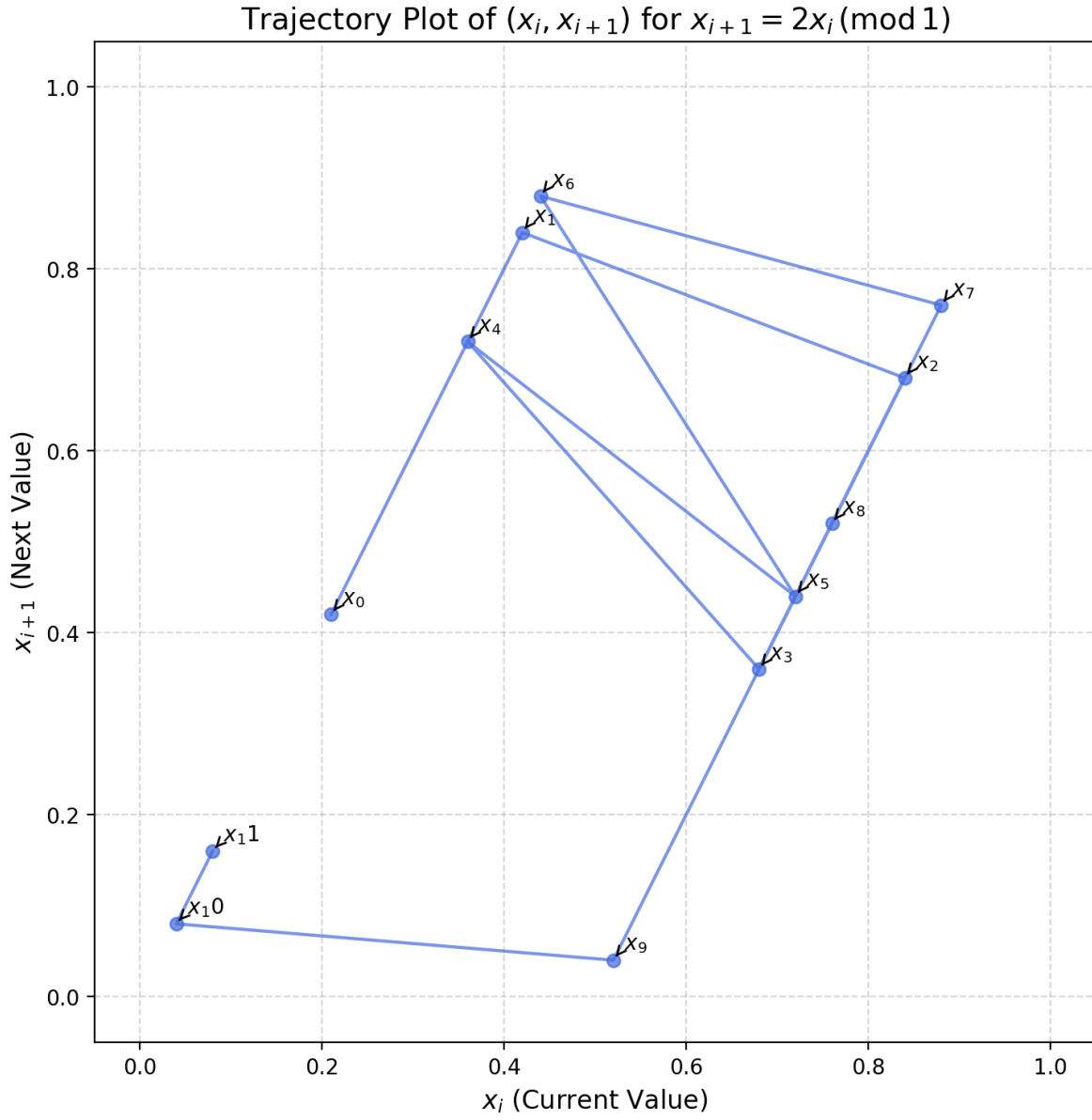


Figure 1: Graph of the pair (x_i, x_{i+1}) , where $x_i = T'x_{i-1}$. The plot fills-in the graph of the map T .

We want to answer the following question. Fix μ , that is fix p_0, \dots, p_{n-1} . What is the pushforward of the measure $\mathbb{P} = \mu^{\otimes \mathbb{N}}$ to $[0, 1]$. In other words, if we take $(X_i)_{i \geq 0}$ i.i.d. with distribution μ (so $\mathbb{P}(X_i = k) = p_k$) what will be the distribution of the random variable

$$Y = \sum_{i \geq 0} n^{-i-1} X_i$$

The sum on the r.h.s. is well-defined, so Y is well-defined and we want to investigate its distribution. Let us denote $\mathbb{Q}_\mu = \mu^{\otimes \mathbb{N}} \circ \Phi^{-1}$ the law of Y when the X_i are i.i.d. with distribution μ .

Proposition 0.1. *If $p_k \equiv p = 1/n$, namely if the X_i are uniformly on the finite set \mathcal{A} , then \mathbb{Q}_μ is the Lebesgue measure (but defined on the Borel σ -algebra) on $[0, 1]$.*

Proposition 0.1. Indeed, we fix an interval of type $I = [kn^{-j}, (k+1)n^{-j}]$, for some $j \geq 1$ and $0 \leq k \leq n^j - 1$. We will have that $Y = \Phi(X) \in I$ iff $\sum_{i=0}^{j-1} n^{-i-1} X_i = kn^{-j}$ (up to a countable set of measure 0). In turn, this holds for a unique value of (X_0, \dots, X_{j-1}) . Therefore

$$\mathbb{Q}_\mu(I) = \mathbb{P}(Y \in I) = \mathbb{P}\left(\sum_{i=0}^{j-1} n^{-i-1} X_i = kn^{-j}\right) = p^j = n^{-j} = \text{Leb}(I)$$

Since intervals of this type determine the measure entirely, we have $\mathbb{Q}_\mu = \text{Leb}$. \square

In other words, we can just obtain the Lebesgue measure by flipping a coin countably many times, a rather intuitive construction for an object that was ill-defined until the beginning of the 20th century. By the law of large numbers, Proposition 0.1 means that Lebesgue-almost all real numbers have the same frequency of each digit. Say, if we fix $n = 10$, then define $\eta_{k,\ell}(x)$ the number of times a digit $k = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ occurs in the representation of x in base 10, in the first ℓ digits of x . Then

$$\text{Leb}\left(\left\{x \in [0, 1] : \limsup_{\ell} \frac{\eta_{k,\ell}}{\ell} = \liminf_{\ell} \frac{\eta_{k,\ell}}{\ell} = \frac{1}{10}\right\}\right) = 1$$

Indeed, once we are back to the $n = 10$ -ary representation, $\eta_{k,\ell}$ is nothing but the sum $\sum_{i=0}^{\ell-1} \mathbf{1}_{X_i=k}$, which is a sum of i.i.d. random variables with expected value $1/n$. So the strong law of large numbers implies that the limit exists and equals $1/10$ with probability 1.

Theorem 0.1. For a probability $\mu \in \mathcal{P}(\mathcal{A})$, $\mu(\{k\}) := p_k$ over the finite alphabet $\mathcal{A} = \{0, \dots, n-1\}$, define

$$E_\mu := \left\{ x \in [0, 1] : \limsup_{\ell} \frac{\eta_{k,\ell}}{\ell} = \liminf_{\ell} \frac{\eta_{k,\ell}}{\ell} = p_k, \quad \forall k \in \mathcal{A} \right\}$$

Then the E_μ are measurable, disjoint sets and $\mathbb{Q}_\mu(E_\mu) = 1$ (and thus $\mathbb{Q}_\mu(E_{\mu'}) = 0$ for $\mu \neq \mu'$). Thus

A. If μ is uniform, $p_k = 1/n$, \mathbb{Q}_μ is the Lebesgue measure. B. If there exists $k \in \mathcal{A}$ such that $p_k = 1$, then \mathbb{Q}_μ is Dirac mass (it is concentrated on one point). C. If $p_k < 1$ for all k , but the measure is not uniform, \mathbb{Q}_μ is of Cantor type, namely it gives measure 0 to every point, but is singular w.r.t. the Lebesgue measure.

The proof is left as a guided exercise.

Exercise 0.2.

- Notice that, by their very definition, the E_μ are Borel-measurable and disjoint.
- Reasoning as in the text after Proposition 0.1, prove in detail that $\mathbb{Q}_\mu(E_\mu) = 1$.
- Recall Proposition 0.1 and solve the trivial case where $p_k = 1$ for some k , to prove the statements A) and B) in the theorem.
- In the case C), it remains to prove that \mathbb{Q}_μ gives measure 0 to each point.

Testing convergence

We can look at the trajectory of points of the map $x \mapsto nx(\text{mod } 1)$.

We can check that if we start with N points sampled with a given distribution μ , we can see numerically that in a few step they will spread out along the interval. The idea is that, if we start with a distribution μ absolutely continuous w.r.t. an invariant measure, then after many iterations of the shift map T , the distribution of points will get close to the invariant measure (Lebesgue).

However, Lebesgue is not the only invariant measure, and as we iterate the map many times, we see the problems. Numerical precision (the fact that we cannot exactly sample a.c. w.r.t. Lebesgue) will have an impact when we iterate the map many times, e.g. periodic orbits (that have 0 probability w.r.t. Lebesgue) have positive probability when we sample on a computer. The effect is visible in this video.

[../media/bs_simulation.webm](#)

Feel free to experiment how different distributions converge to Lebesgue. Can you imagine how to mathematically state a convergence result?

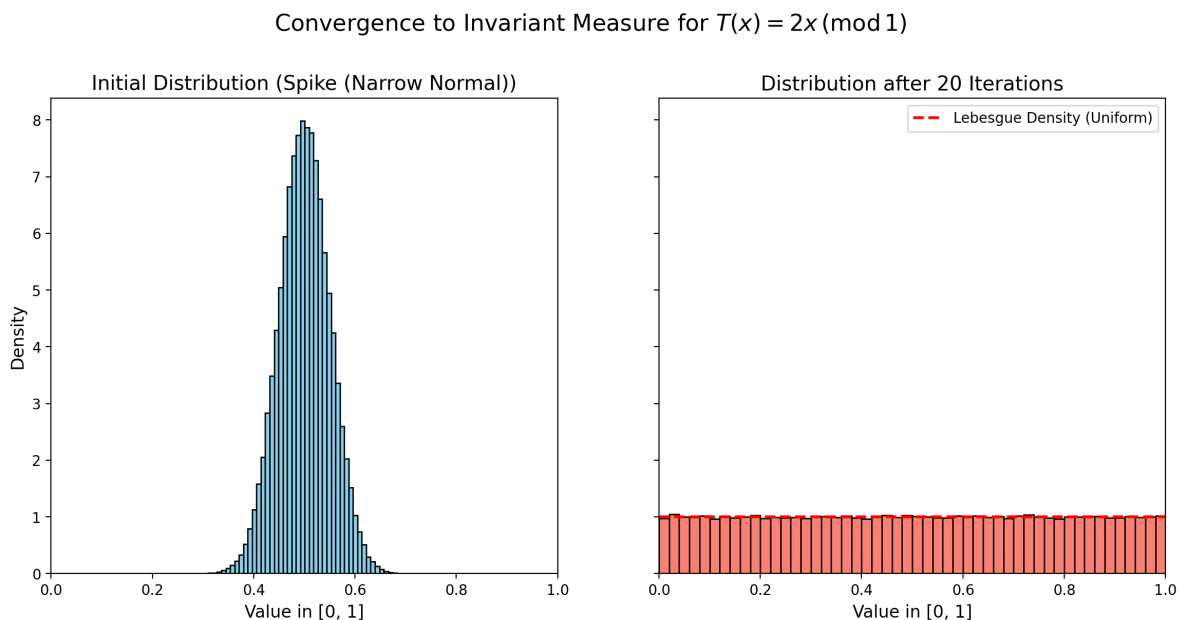


Figure 2: If at time 0 we sample our points randomly, with a measure absolutely continuous w.r.t. Lebesgue, their distribution will converge exponentially fast to the Lebesgue measure. This also holds for the other invariant measures.