

Iterated function system study group notes

ICMC USP

2023

Abstract

Notes written in the study group about the Iterated function system that happened in 2023 on the ICMC USP. These notes were written by Bella Figliaggi, Gabriel Lucidio, Kevin Kühn, Lucas Maekawa, Renan Wenzel, and Zheng Bian. In this group, we explored an Iterated function system (IFS) to construct fractals with self-similarities and define a Markov Chain using the IFS we studied stationary measure for the system.

Contents

1	The space where fractals live	1
2	The Hutchinson Operator	3
3	Contraction mappings on $H(M)$ and fixed point dependence on parameters	4
4	Space of the Choice	6
4.1	Motivations	6
4.2	The Usual Construction of the Cantor Set	6
4.2.1	Goal of the Part	6
4.2.2	Main Text	6
4.3	The Motivation Problem	7
4.3.1	Goal of the Part	7
4.3.2	Main Text	8
4.4	The Meaning of “Almost Every”	8
4.4.1	Goals of this Part	8
4.4.2	Main Text	9
4.5	The Topology of Cylinder Sets and Sequence Spaces	10
5	Coding map	10
6	Conditional Expectations	15
6.1	Application in L^2	17
7	Markov Chains	17
8	Stationary Measures	19
8.1	Stationary Measure	21
8.2	Weak* topology	21
8.3	Markov Operator	23
8.3.1	Letac Principle	26
9	Random Power Series	31

1 The space where fractals live

Welcome to the first session of this series of meetings on Iterated Function Systems. This is such an interesting subject because it is a unified way of constructing fractals (containing classical Cantor sets,

Sierpinski gaskets, the Barnsley fern, etc.) and also has quite nice applications, such as fractal image compressing.

Iterated function systems offer a new perspective for fractals, in opposition to the view that those sets resulted from successive microscopic refinement. We will now look at fractals as attractors of iterated function systems (a much more global perspective).

We'll start defining Iterated Function Systems as a complete metric space M together with a collection of maps \mathbf{f} on M , that is $\mathbf{f} = \{f_i : i = 1, \dots, k\}$ where $f_i : M \rightarrow M$



Figure 1: The Cantor ternary set can be produced using an IFS

Let us in this first session concentrate on constructing some important definitions and theorems about the space we'll be working on.

Let (M, d) be a metric space and define

1. For $a \in M, B \neq \emptyset, d_1(a, B) := \inf_{b \in B} d(a, b)$
2. For nonempty $A, B \subseteq M, d_2(A, B) := \sup_{a \in A} d_1(a, B)$
3. The Hausdorff distance $d_H(A, B) = \max\{d_2(A, B), d_2(B, A)\}$

Note that the need for d_H comes from the necessity of symmetry for the distance function (given that d_2 is not symmetric).

Now we define the central piece of this session, which is the collection of all nonempty compact subsets of M , denoted by $H(M)$ (called the space where fractals live).

Proposition 1.1. d_H a distance in $H(M)$.

Proof:

1. Symmetry: $\max\{d_2(A, B), d_2(B, A)\} = \max\{d_2(B, A), d_2(A, B)\}$.
2. $d_H(A, B) < \infty$
3. Non-negativity is immediate from definition (d is a distance).
4. If $A = B$ then $d_H(A, B) = 0$: $d_H(A, B) = d_2(A, A) = \sup_{a \in A} d_1(a, A) = \sup_{a \in A} \inf_{b \in A} d(a, b) = 0$.
5. If $d_H(A, B) = 0$ then $A = B$: $d_2(A, B) = 0 = d_2(B, A) \therefore d_1(a, B) = 0 \forall a \in A \therefore A \subseteq B$ (similarly we prove $B \subseteq A$).
6. Triangle inequality:

With A, B and C compact sets we write

$$\begin{aligned}
 d_1(a, C) &= \min_{c \in C} d(a, c) \leq \min_{c \in C} [d(a, b) + d(b, c)] \text{ for any } b \in B \\
 &= d(a, b) + \min_{c \in C} d(b, c) \text{ for any } b \in B \\
 d_1(a, C) &\leq \min_{b \in B} d(a, b) + \min_{b \in B} \min_{c \in C} d(b, c) \\
 &\leq \min_{b \in B} d(a, b) + \max_{b \in B} \min_{c \in C} d(b, c) \\
 &= d_1(a, B) + d_2(B, C)
 \end{aligned}$$

therefore $d_2(A, C) \leq d_2(C, B) + d_2(B, A)$. Now

$$\begin{aligned} d_H(A, C) &= \max\{d_2(A, C), d_2(C, A)\} \\ &\leq \max\{d_2(A, B) + d_2(B, C), d_2(C, B) + d_2(B, A)\} \\ &\leq \max\{d_2(A, B), d_2(B, A)\} + \max\{d_2(B, C), d_2(C, B)\} \\ &= d_H(A, B) + d_H(B, C) \end{aligned}$$

□

2 The Hutchinson Operator

Let us focus on contraction mappings on the space of fractals previously established. We specify $f_1, \dots, f_k : M \rightarrow M$ being contractions with factors $\lambda_1, \dots, \lambda_k$

$$\forall x, y \in M : d(f_i(x), f_i(y)) \leq \lambda_i d(x, y)$$

Also, denote $\lambda = \max\{\lambda_1, \dots, \lambda_k\}$.

Definition 2.1. The Barnsley-Hutchinson operator $B : H(M) \rightarrow H(M)$ is given by

$$B(A) = \cup_{i=1}^k f_i(A) \text{ for all } A \in H(M)$$

We'll be initially interested in four theorems regarding the Hutchinson operator, but before stating and proving them, we'll look at three interesting lemmas

Lemma 2.2. *Let $f : M \rightarrow M$ be a contraction mapping on the metric space (M, d) . Then f is continuous.*

Proof:

Let $\epsilon > 0$ be given. Let $\lambda > 0$ be the contraction factor for f . Then

$$d(f(x), f(y)) \leq \lambda d(x, y) < \epsilon$$

whenever $d(x, y) < \delta$ with $\delta = \frac{\epsilon}{\lambda}$.

□

Lemma 2.3. *Let $f : M \rightarrow M$ be a continuous mapping on the metric space (M, d) . Then f maps $H(M)$ into itself.*

Proof:

Let S be a nonempty compact subset of M . Then clearly $f(S) = \{f(x) : x \in S\}$ is nonempty.

We want to show that $f(S)$ is compact. Let $\{y_n = f(x_n)\}$ be an infinite sequence of points in S . Then $\{x_n\}$ is an infinite sequence of points in S . Since S is compact there is a subsequence $\{x_N\}$ that converges to a point $\hat{x} \in S$.

But as f is continuous it implies that $\{y_N = f(x_N)\}$ is a subsequence of $\{y_n\}$ that converges to $\hat{y} = f(\hat{x}) \in f(S)$.

□

Lemma 2.4. *Let $f : M \rightarrow M$ be a contraction mapping on the metric space (M, d) with contraction factor λ . Then $f : H(M) \rightarrow H(M)$ defined by*

$$f(B) = \{f(x) : x \in B\} \forall B \in H(M)$$

is a contraction mapping on $(H(M), d_H)$ with contraction factor λ .

Proof:

From lemma 2.1 $f : M \rightarrow M$ is continuous and from lemma 2.2 f maps $H(M)$ into itself.

Now let $B, C \in H(M)$, then

$$\begin{aligned} d_2(f(B), f(C)) &= \max\{\min\{d(f(x), f(y)) : y \in C\} : x \in B\} \\ &\leq \max\{\min\{\lambda d(x, y) : y \in C\} : x \in B\} = \lambda d_2(B, C) \end{aligned}$$

Similarly $d_2(f(C), f(B)) \leq \lambda d_2(C, B)$.

Finally, we write

$$\begin{aligned} d_H(f(B), f(C)) &= \max\{d_2(f(B), f(C)), d_2(f(C), f(B))\} \\ &\leq \lambda \max\{d_2(B, C), d_2(C, B)\} \\ &\leq \lambda d_H(B, C) \end{aligned}$$

□

After the above lemmas, let's now look at four theorems regarding the Hutchinson Operator over the space of fractals.

Theorem 2.5. *The operator B is a contraction in $(H(M), d_H)$ with factor λ .*

We first prove the following proposition

Proposition 2.6. *If $A_1, \dots, A_k, B_1, \dots, B_k \in H(M)$, then $d_H(\cup A_i, \cup B_i) \leq \max_i d_H(A_i, B_i)$*

Proof:

If $C \in H(M)$, then

$$d_2(\cup A_i, C) = \max_{x \in \cup A_i} d_1(x, C) = \max_i \max_{x \in A_i} d_1(x, C) = \max_i d_2(A_i, C)$$

Which yields

$$d_2(\cup A_i, \cup B_i) = \max_i d_2(A_i, \cup B_i) \leq \max_i d_2(A_i, B_i)$$

Where last inequality is true because $d_2(A_i, \cup B_i) \leq d_2(A_i, B_i)$ for each i . In analogous way, we can prove $d_2(\cup B_i, \cup A_i) \leq \max_i d_2(B_i, A_i)$.

Finally

$$\begin{aligned} d_H(\cup A_i, \cup B_i) &= \max\{d_2(\cup A_i, \cup B_i), d_2(\cup B_i, \cup A_i)\} \\ &\leq \max\{\max_i d_2(A_i, B_i), \max_i d_2(B_i, A_i)\} \\ &\leq \max_i d_H(A_i, B_i) \end{aligned}$$

□

3 Contraction mappings on $H(M)$ and fixed point dependence on parameters

We start by proving Theorem 2.5

Proof of Theorem 2.5:

For $K_1, K_2 \in H(M)$, we have

$$\begin{aligned} d_H(B(K_1), B(K_2)) &= d_H(\cup_i f_i(K_1), \cup_i f_i(K_2)) \\ &\leq \max_i d_H(f_i(K_1), f_i(K_2)) \\ &\leq \max_i \lambda_i d_H(K_1, K_2) \leq \lambda d_H(K_1, K_2) \end{aligned}$$

□

Theorem 3.1. *Let (M, d) be a complete metric space, and let $f : M \rightarrow M$ be a contraction with factor λ . Then, there is a unique point $p \in M$ such that $f(p) = p$. Moreover, p is an attracting fixed point (i.e., for any $y \in M$, the sequence $f^n(y)$ converges to p .)*

Proof:

Let $x \in M$, and consider the sequence $f^n(x) \in M$. For any $n \in \mathbf{N}$, we have

$$\begin{aligned} d(x, f^n(x)) &\leq d(x, f^1(x)) + d(f^1(x), f^2(x)) + \dots + d(f^{n-1}(x), f^n(x)) \\ &< d(x, f^1(x))(1 + \lambda + \dots + \lambda^{n-1}) \\ &< d(x, f^1(x))\sum_{n=0}^{\infty} \lambda^n < \infty \end{aligned}$$

Fixing $\epsilon > 0$ and choosing an integer N such that $\lambda^N d(x, f^1(x))\sum_{n=0}^{\infty} \lambda^n < \epsilon$. Then, for any $m \geq n \geq N$, we have

$$d(f^m(x), f^n(x)) = d(f^n \circ f^{m-n}, f^n(x)) \leq \lambda^n d(f^{m-n}(x), x) \leq \lambda^N d(x, f^1(x))\sum_{n=0}^{\infty} \lambda^n < \epsilon$$

This shows that $f^n(x)$ is indeed a Cauchy sequence in M . But since (M, d) is complete, $f^n(x)$ converges to some point p in M .

Also, p is a fixed point of f

$$f(p) = f(\lim_{n \rightarrow \infty} f^n(x)) = \lim_{n \rightarrow \infty} f \circ f^n(x) = \lim_{n \rightarrow \infty} f^{n+1}(x) = \lim_{n \rightarrow \infty} f^n(x) = p$$

To prove uniqueness, consider $q \in M$ also a fixed point of f . Then

$$d(p, q) = d(f(p), f(q)) \leq \lambda d(p, q) \therefore d(p, q) = 0 \text{ and } p = q$$

Finally, taking any $y \in M$

$$d(f^n(y), p) = d(f^n(y), f^n(p)) \leq \lambda^n d(y, p) \rightarrow 0$$

Which yields $f^n(y) \rightarrow p$.

□

The next theorem follows immediately from our previous results.

Theorem 3.2. (Collage) *If (M, d) is complete, then the Hutchinson operator B has a unique fixed point $K \in H(M)$. It is attractive and it is called Hutchinson attractor.*

Finally, we conclude this session with a view on the dependence of fixed points on the parameters of the contraction maps.

Theorem 3.3. *Let (Λ, ρ) be a metric space and (M, d) a complete metric space and let*

$$T : \Lambda \times M \rightarrow M$$

be a family of contractions with uniform contraction factor, that is

$$d(T(\lambda, x), T(\lambda, y)) \leq kd(x, y), \forall \lambda \in \Lambda \text{ e } \forall x, y \in M$$

Suppose that for $x \in M$ the map $\lambda \mapsto T(\lambda, x)$ is continuous from Λ to M . Then for each $\lambda \in \Lambda$, $T(\lambda, \cdot)$ has a unique fixed point $x(\lambda)$ and the map $\lambda \mapsto x(\lambda)$ is continuous.

Proof:

Given $\lambda \in \Lambda$, take $x \in M$ such that

$$T(\lambda, x) = x = x(\lambda)$$

The existence of x comes from the fact that T is a contraction for a fixed λ .

From the hypothesis, for a fixed $\epsilon_0 > 0$ there is a $\delta > 0$ such that $\beta \in \Lambda$ with $\rho(\beta, \lambda) \leq \delta$ implies in

$$d(T(\lambda, x), T(\beta, x)) \leq \epsilon_0$$

Consider then $\beta \in \Lambda$ in the neighborhood of λ and take $y \in M$ such that

$$T(\beta, y) = y = y(\beta)$$

We can then write

$$\begin{aligned} d(x(\lambda), y(\beta)) &= d(T(\lambda, x(\lambda)), T(\beta, y(\beta))) \\ &\leq \underbrace{d(T(\lambda, x(\lambda)), T(\lambda, y(\beta)))}_{kd(y(\beta), y(\beta))} + \underbrace{d(T(\lambda, y(\beta)), T(\beta, y(\beta)))}_{\leq \varepsilon_0} \end{aligned}$$

Which yields

$$d(x(\lambda), y(\beta)) \leq \frac{\varepsilon_0}{1-k}$$

Therefore, $\forall \varepsilon > 0$, taking $\delta > 0$ such that $\varepsilon_0 = \varepsilon(1-k)$, we have

$$\rho(\lambda, \beta) \leq \delta \Rightarrow d(x(\lambda), y(\beta)) \leq \varepsilon$$

□

To visualize the highlighted dependence, we head to <https://ifs-fractals.herokuapp.com/playground> for some examples.

4 Space of the Choice

4.1 Motivations

- Review Topology;
- Construct the σ -algebra that will be used;
- Construct the Middle-Third Cantor set via sequences of functions randomly picked.

4.2 The Usual Construction of the Cantor Set

4.2.1 Goal of the Part

Construction of the Cantor Set via nested intervals, analytical properties of the set, its topological properties and construction via functions.

4.2.2 Main Text

Before getting on the construction forementioned, it's important to know how it's usually done, even if just a bit. For that, we consider the interval $[0, 1]$. From it, we will delete the open middle third interval, i.e., $\left(\frac{1}{3}, \frac{2}{3}\right)$ which will leave us with two line segments, namely $\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] = C_1$. Now, repeat that for each one of these intervals and remove their open middle-third intervals, yielding

$$\begin{aligned} &\left[0, \frac{1}{9}\right], \quad \left[\frac{2}{9}, \frac{1}{3}\right] \\ &\left[\frac{2}{3}, \frac{7}{9}\right], \quad \left[\frac{8}{9}, 1\right]. \end{aligned}$$

Then, we call their union $C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$. We thus define inductively $C_n := \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right)$, and define the middle-third Cantor set by

$$\mathcal{C} := \lim_{n \rightarrow \infty} C_n = \bigcap_{n=0}^{\infty} C_n$$

Briefly speaking about this set's properties, it has the same cardinality as the real line. To prove it, we use Bernstein-Schröder theorem and compare the cardinality of \mathcal{C} with that of $[0, 1]$. Thus, $\aleph_1 \geq |\mathcal{C}|$ since $\mathcal{C} \subseteq [0, 1]$. On the other hand, to show the other side of the inequality, one constructs a surjection going from \mathcal{C} to $[0, 1]$. Jumping right to the definition, it is motivated by writing the points of the interval $[0, 1]$ in terms of base 3, and then removing on each n step, the numbers that have 1 as their n th digit in the ternary representation.

On the first step, digits of the form $0.1aaaaaaaa \dots$, where $aaaaaaaa \dots$ is between $00000 \dots_3$, and $22222 \dots_3$. leaving only the numbers whose first digit is either 0 or 2. The next step removes numbers with 1 on the second coordinate, leaving only those that assume the form

$$0.00aaaaa \dots, 0.02aaaaa \dots, \text{ or } 0.22aaaaa \dots.$$

We repeat it ad infinitum, and the numbers remaining on the Cantor set have representations consisting entirely of 0's and 2's.

Hence, the function from \mathcal{C} to $[0, 1]$ takes the ternary numerals constructed above and replaces every 2's by 1's, in a way that we interpret the sequence obtained as a binary form of a real number in $[0, 1]$, and reciprocally, any of them have a binary representation that can be translated to a ternary representation of a number in the Cantor set by replacing 1's by 2's:

$$f\left(\sum_{k \in \mathbb{N}} a_k 3^{-k}\right) = \sum_{k \in \mathbb{N}} \frac{a_k}{2} 2^{-k}, \quad \forall k \in \mathbb{N} : a_k \in \{0, 2\}$$

Therefore, f is surjective, so that $\aleph_1 \leq |\mathcal{C}| \leq \aleph_1$.

Some other interesting properties is that it is a set of measure 0, i.e., it is infinitesimally small, even though it has uncountably many elements. It is self-similar, a sort of primitive fractals, because it is equal to two copies of itself when translated and dilated by some factor. More specifically, dilating each copy by a factor of $\frac{1}{3}$ and translating it, and as a consequence, when you zoom in into a part of the set, it looks identical to the whole. This can be seen from the equation given before, i.e.,

$$C_n := \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right),$$

and from here, the self-similarity follow by showing that $\mathcal{C} = \frac{\mathcal{C}}{3} \cup \left(\frac{2}{3} + \frac{\mathcal{C}}{3}\right)$. The Cantor set is also perfect, meaning that any point in the Cantor Set has an infinite sequence of points also in the set converging to that point. The proof of this statement relies on taking an element x in \mathcal{C} , and arbitrary $\epsilon > 0$, an $N(\epsilon) \in \mathbb{N}$ and finding a good enough M so that $x \in [M, M+1] \subseteq (x - \epsilon, x + \epsilon)$. Then, we apply the process used on the construction of \mathcal{C} to the interval $[M, M+1]$, removing the interval $\left(\frac{3M+1}{3}, \frac{3M+2}{3}\right)$ from it, to see that if were to belong to the removed interval or not, it'd be a limit point in both cases. From arbitrariness of the choices, it follow that \mathcal{C} is indeed a perfect set.

Topologically speaking, the Cantor set has nice properties too. It's totally disconnected, compact, perfect, nonempty, and metrizable. Furthermore, any set with those properties are the Cantor Set up to homeomorphism.

4.3 The Motivation Problem

4.3.1 Goal of the Part

Review the Hutchinson Attractor, give the motivation problem behind construction the Cantor Set as a random application of functions that converge to the Attractor, set up the main problems to be studied.

4.3.2 Main Text

As Kevin talked about last week, we can construct the Cantor Middle-Third set through a function known as the Hutchinson Attractor, denoted by B , ranging from the set $\mathcal{H}(M)$ of nonempty compact subsets of a metric space M to $\mathcal{H}(M)$ itself and given by the union of the images of contractions in M . The way we construct it is, for instance, by applying B to a singleton set $A = \{x\}$ and using the previously defined metric d_H to translate the result to the Cantor Set. For each n iterations, we get an amount of parts of the CMT set equal to 2^n . Now consider two functions $f_i : [0, 1] \rightarrow [0, 1]$

$$f_i = \begin{cases} \frac{x}{3}, & i = 0 \\ \frac{x}{3} + \frac{2}{3}, & i = 1 \end{cases}$$

The infinite application of these functions yield the Cantor set, which shall be denoted by \mathcal{C} to make things easier. In other words, Consider now a representation of infinitely many combinations of 0's and 1's in any order, i.e., $\omega = (0, 1, 0, 0, 1, \dots)$ and let f_ω^n be the repeated use of f_i as defined before using the indices from ω , so that

$$f_\omega^n(x) = f_{\omega_{n-1}} \circ \dots \circ f_{\omega_1} \circ f_{\omega_0}(x)$$

indicates that we start using the n -th coordinate of ω on each f . Using the example of ω from before, f_ω^n would be

$$f_\omega^n(x) = f_{\omega_{n-1}} \circ \dots \circ f_0 \circ f_1 \circ f_0(x).$$

Using that notation, the thesis given at the start that the infinite use of f_i eventually covers all of \mathcal{C} is the same as

$$\lim_{l \rightarrow \infty} \overline{\{f_\omega^n(x) : n \geq l\}} = B$$

That begs the question: Will EVERY sequence of digits $\omega \in \{0, 1\}^{\mathbb{N}}$, when combined with f_i , give as its image the Cantor set? In fact, the answer is no, as the theorem is stated

Theorem 4.1. *Let ϕ be an independent and identically distributed random product of contractions $f_1, \dots, f_k : M \rightarrow M$ and K be the corresponding Hutchinson Attractor. Given $p \in M$, we have*

$$\lim_{l \rightarrow \infty} \overline{\{f_\omega^n : n \geq l\}} = K$$

for \mathbb{P} -almost all ω .

On that note, to grasp what we mean by almost all sequences work, consider a sequence of a single digit, i.e., $\omega = \{0, 0, 0, \dots\}$. Then, when we apply this to f_i , we get

$$\begin{aligned} f_\omega^n(x) &= f_0 \circ \dots \circ f_0 \circ f_0(x) \\ &= f_0 \circ \dots \circ f_0\left(\frac{x}{3}\right) \\ &= f_0 \circ \dots \circ f_0\left(\frac{x}{9}\right) \\ &\vdots \\ &= \frac{x}{3^n}. \end{aligned}$$

Because of this, as n goes to infinity, f_ω^n tends to 0, which is not the Cantor set. In other words,

$$\lim_{l \rightarrow \infty} \overline{\{f_\omega^n(x) : n \geq l\}} \neq B.$$

a new problem arises, that is, will the number of ω 's for which the initial thesis hold be way too small to be relevant? The answer is going to need quite a bit of measure theory, for the answer is, in fact, that the result holds for almost every ω .

4.4 The Meaning of ‘‘Almost Every’’

4.4.1 Goals of this Part

We want to define σ -algebras, the foundation of what is a measure. Then, define measures properly speaking, what is a property that happens almost everywhere, and some probabilistic concepts, such as independence and identical distributions.

4.4.2 Main Text

Now that the goal has been set, we can start to develop the theory needed to fully grasp it. Starting with a review of σ -algebra, they are the sets in which we define measures. They are sets that have closure under countable unions and complements. Formally speaking,

Definition: Given a set Ω , a σ -algebra \mathcal{F} is a nonempty collection of subsets of Ω that satisfy

- (i) If $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;
- (ii) If A_i is a countable sequence of sets, then $\bigcup_i A_i \in \mathcal{F}$;
- (iii) $\Omega \in \mathcal{F}$. \square

They are the defining ground that allow us to define measures and thus, find the meaning of “Almost Every”. Without further ado, we define a measure.

Definition 4.2. A measure is a nonnegative countably additive set function, i.e., a function $\mu : \mathcal{F} \rightarrow \mathbb{R}$ which satisfies

- i) $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$;
- ii) If $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i). \square$$

Now, we can formalize the meaning of almost every point satisfying something. The idea is that some property is going to hold for every point of a set except for an amount inside a set whose significance is naught. In other words, a set with a measure of zero, so that the set for which the property holds takes up nearly every single possibility.

Definition 4.3. Given a measure space $(\Omega, \mathcal{F}, \mu)$, in which Ω is a set, \mathcal{F} is a σ -algebra, and μ is a measure, a property P is said to hold almost everywhere if there exists a set $N \in \Sigma$ with $\mu(N) = 0$, and all $x \in \Omega/N$ satisfy that property.

It’s also important to understand what it means for a variable to be independent and identically distributed, since it allows for simplifications of results and ideas. Essentially, it means that the an element in the sequence of variables does not depend on the ones that came before it, and the value does not change with each subsequent application of it. The easiest example of such kinds of elements come up in the toss of coins, since the result of “head” or “tail” has no impact on the next tossing of coin, and neither does the probability change from 0.5 each time you flip it, meaning it is identically distributed. However, we need a mathematical definition to work better with it.

Definition 4.4. Suppose random variables X and Y are defined to assume values in $I \subseteq \mathbb{R}$. Let $F_X(x) = P(X \leq x)$, $F_Y(y) = P(Y \leq y)$ denote the cumulative distribution function of X and Y . Furthermore, denote their joint cumulative distributive function by

$$F_{X,Y}(x, y) = P(X \leq x \wedge Y \leq y).$$

Definition 4.5. Two random variables X and Y are independent if and only if

$$F_X(x) = F_Y(y) \quad \forall x, y \in I \subseteq \mathbb{R}.$$

Moreover, they are said to be identically distributed if and only if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \forall x, y \in I \subseteq \mathbb{R}.$$

When both happen, we say that they are i.i.d.

The thing is that we don’t work with only two variables (usually). Thankfully, we do have a definition for multiple variables.

Definition 4.6. We say that n random variables X_1, \dots, X_n are independent if and only if

$$F_{X_1}(x) = F_{X_k}(x) \quad \forall k \in \{1, \dots, n\} \text{ and } \forall x \in I$$

On the other hand, they are identically distributed if and only if

$$F_{X_1, \dots, X_n}(x) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \quad \forall x_1, \dots, x_n \in I$$

where $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1 \wedge \cdots \wedge X_n \leq x_n)$ denotes the joint cumulative distribution function of X_1, \dots, X_n .

4.5 The Topology of Cylinder Sets and Sequence Spaces

Regarding the topology of cylinder sets, **Goals of the Part**

Here, we shall define cylinders, working with coordinates of their elements, define a topology for the sequence space, and define bases for it.

Definition 4.7. Fix integers $n_1 < \cdots < n_k$ and $\alpha_1, \dots, \alpha_k \in \{0, 1, \dots, N-1\}$. Then, the cylinder of rank k is the set

$$C_{\alpha_1, \dots, \alpha_k}^{n_1, \dots, n_k} := \{\omega \in \Omega_N^R : \omega_{n_i} = \alpha_i\}$$

where i varies from 1 to k . Another notation that may be used is

$$[\alpha_0, \dots, \alpha_k] = C_{\alpha_1, \dots, \alpha_k}^{n_1, \dots, n_k}.$$

The elements of cylinders are exactly the k first coordinates of an element $\omega \in \{0, 1\}^{\mathbb{N}}$, and it should be noted that we may not know the other coordinates other than those k first ones. Cylinders are both open and closed sets simultaneously, as can be shown. The elements we are going to use as part of the definition of f will be members of the cylinders.

Furthermore, we endow the space $\{0, 1\}^{\mathbb{N}}$ with the smallest topology that makes projections continuous functions, the product topology, which has a subbasis formed exactly by the preimage of projections and a basis of cylinders.

Definition 4.8. Let $X = \prod_{\alpha \in J} X_\alpha$, where each X_α is a topological space. We define the product topology on X as the one given by the basis \mathcal{B} such that each basis element is defined as

$$B = \prod_{\alpha \in J} U_\alpha, \quad U_\alpha \neq X_\alpha \text{ for all but finitely many } \alpha$$

The product topology in question has as reference the discrete topology on $\{0, 1\}$

5 Coding map

Definition 5.1. Let $\Sigma = \{1, \dots, k\}^{\mathbb{N}}$ where we consider $\{1, \dots, k\}$ with the discrete topology and Σ with the product topology. Given a_0, \dots, a_l , we use the notation $[a_0, \dots, a_l]$ define the set

$$[a_0, \dots, a_l] = \{\omega \in \Sigma : \omega_0 = a_0, \dots, \omega_l = a_l\}.$$

Note that

$$[a_0, \dots, a_l] = \bigcap_{i=1}^l \pi_i^{-1}(\{a_i\})$$

where π_i is the projection on the i -th coordinate.

We can define a σ -algebra $\mathcal{B}(\Sigma)$ for Σ that is generated by the open sets and, in this particular case where $\Sigma = \{1, \dots, k\}^{\mathbb{N}}$, $\mathcal{B}(\Sigma)$ is also the σ -algebra generated by cylinder sets. Because the family of finite unions of cylinder sets is an algebra, we can define a measure \mathbb{P} where $\mathbb{P}([a_0, \dots, a_l]) = \prod_{i=0}^l p_{a_i}$ and $\sum_{i=0}^k p_i = 1$ (see theorem 1.14 in Real Analysis: Modern Techniques and Their Application, by Gerald B. Folland).

From now on, we will consider the probability space $(\Sigma, \mathcal{B}(\Sigma), \mathbb{P})$.

Theorem 5.2 (Coding map). *Let $f_1, \dots, f_k : M \rightarrow M$ be a family of contractions and let K be the corresponding Hutchinson attractor. For all sequence ω , the limit*

$$\pi(\omega) = \lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)$$

exists and is independent of p . Moreover, $\pi(\Sigma)$ is the Hutchinson attractor.

Proof. Let K be the Hutchinson attractor. For each $i = 1, \dots, k$, we have $f_i(K) \subset K$. In particular, for every sequence $\omega \in \Sigma$ and each $n \in \mathbb{N}$, we have

$$f_{\omega_0} \circ \dots \circ f_{\omega_{n+1}}(K) \subset f_{\omega_0} \circ \dots \circ f_{\omega_n}(K).$$

Because each f_i is a contraction, $\text{diam}(f_{\omega_0} \circ \dots \circ f_{\omega_n}(K)) \rightarrow 0$ as $n \rightarrow \infty$ and $f_i(K)$ is a closed set for $i = 1, \dots, k$, so that $\bigcap_n f_{\omega_0} \circ \dots \circ f_{\omega_n}(K)$ is just a single point. Therefore, the limit

$$\lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)$$

exists for all $p \in K$ and it depends only of ω .

For q outside K , as

$$d(f_{\omega_0} \circ \dots \circ f_{\omega_n}(q), f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

the sequence $\{f_{\omega_0} \circ \dots \circ f_{\omega_n}(q)\}_{n \in \mathbb{N}}$ converges to same point as the sequence $\{f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)\}_{n \in \mathbb{N}}$.

Now, to show that $\pi(\Sigma) = K$, we will show that $\pi(\Sigma)$ is compact and that it is the fixed point of the Hutchinson operator.

Claim. π is a continuous mapping.

Proof. To prove that π is continuous, we want to show that for each $\omega \in \Sigma$ and each $\epsilon > 0$, there exists a open set V containing ω such that for, $\nu \in V$,

$$d(\pi(\omega), \pi(\nu)) < \epsilon.$$

By definition of π , there exists $n \in \mathbb{N}$ such that $d(\pi(\omega), f_{\omega_0} \circ \dots \circ f_{\omega_n}(K)) < \epsilon$. Let $V = [\omega_0, \dots, \omega_n]$. For every $\nu \in V$, we have $\pi(\nu) \in f_{\omega_0} \circ \dots \circ f_{\omega_n}(K)$ and therefore $d(\pi(\omega), \pi(\nu)) < \epsilon$. \square

Claim. $f_i(\pi(\Sigma)) \subset \pi(\Sigma)$ for $i = 1, \dots, k$.

Proof. Because π is continuous, for each $\omega \in \Sigma$ we have

$$\begin{aligned} f_i(\pi(\omega)) &= f_i\left(\lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)\right) \\ &= \lim_{n \rightarrow \infty} f_i \circ f_{\omega_0} \circ \dots \circ f_{\omega_n}(p) = \pi(i\omega) \end{aligned}$$

where $i\omega$ is the sequence $i\omega_0\omega_1 \dots$. \square

Because π is continuous and Σ is compact, by Tychonoff's theorem, $\pi(\Sigma)$ is also compact, and by the last claim, $\mathcal{B}(\pi(\Sigma)) = \bigcup_{i=1}^k f_i(\pi(\Sigma)) = \bigcup_{i=1}^k \{\pi(i\omega) : \omega \in \Sigma\} = \pi(\Sigma)$ so $\pi(\Sigma)$ must be the Hutchinson attractor and this completes the proof. \square

We will need the following definitions to prove the next theorem.

Definition 5.3 (i.i.d. product of functions). Let $\Sigma = \{1, \dots, k\}^{\mathbb{N}}$ and $f_1, \dots, f_k : M \rightarrow M$ functions. The function $\varphi : \mathbb{N} \times \Sigma \times M \rightarrow M$ given by

$$\varphi(n, \omega, x) = f_{\omega_{n-1}} \circ \dots \circ f_{\omega_0}(x) = f_{\omega}^n(x)$$

is called the independent and identically distributed (i.i.d.) product of functions f_1, \dots, f_k .

Definition 5.4 (Density). Let X be a topological space and $f : X \rightarrow X$ be a continuous function. We say that a point $x \in X$ has a dense orbit if, for every non-empty open set $U \subset X$, there exists $n \in \mathbb{N}$ such that $f^n(x) \in U$.

Definition 5.5 (Shift operator). Let $\Sigma = \{1, \dots, k\}^{\mathbb{N}}$. The shift operator $\sigma : \Sigma \rightarrow \Sigma$ is the operator that takes a sequence and removes its first coordinate, i.e., $\sigma(\omega_0\omega_1\omega_2\cdots) = \omega_1\omega_2\omega_3\cdots$.

Observe that $\sigma^{-1}([a_0, \dots, a_l]) = \bigcup_{i=1}^k [i, a_0, \dots, a_l]$ so that σ is continuous. Moreover, $\mathbb{P}(\sigma^{-1}([a_0, \dots, a_l])) = \sum_{i=1}^k \mathbb{P}([i, a_0, \dots, a_l]) = \sum_{i=1}^k p_i \prod_{j=0}^l p_{a_j} = \prod_{j=0}^l p_{a_j}$, because $\sum_{i=1}^k p_i = 1$. This means that it does not matter which coordinates are fixed, the probabilities remain the same. We will prove an important fact about the shift operator that will be used in theorem 5.13.

Theorem 5.6. Let $A = [a_0, \dots, a_l]$ and $B = [b_0, \dots, b_m]$ be cylinder sets. For each $n \geq l$, $\mathbb{P}(A \cap \sigma^{-n}(B)) = \mathbb{P}(A)\mathbb{P}(B)$.

Proof. Observe that $[a_0, \dots, a_l] \cap \sigma^{-n}([b_0, \dots, b_m]) = \bigcup_{[c_1, \dots, c_{n-l}]} [a_0, \dots, a_l, c_1, \dots, c_{n-l}, b_0, \dots, b_m]$ where the union is over every cylinder set which fixes the first $n-l$ coordinates. So

$$\begin{aligned} \mathbb{P}(A \cap \sigma^{-n}(B)) &= \sum_{[c_1, \dots, c_{n-l}]} \mathbb{P}([a_0, \dots, a_l, c_1, \dots, c_{n-l}, b_0, \dots, b_m]) \\ &= \sum_{[c_1, \dots, c_{n-l}]} \left(\prod_{i=0}^l p_{a_i} \prod_{i=1}^{n-l} p_{c_i} \prod_{i=0}^m p_{b_i} \right) \\ &= \left(\prod_{i=0}^l p_{a_i} \prod_{i=0}^m p_{b_i} \right) \sum_{[c_1, \dots, c_{n-l}]} \prod_{i=1}^{n-l} p_{c_i} \\ &= \left(\prod_{i=0}^l p_{a_i} \prod_{i=0}^m p_{b_i} \right) \left(\sum_{i=1}^k p_i \right)^{n-l} \\ &= \left(\prod_{i=0}^l p_{a_i} \prod_{i=0}^m p_{b_i} \right) = \mathbb{P}(A)\mathbb{P}(B). \end{aligned}$$

□

Corollary 5.7. If C is a finite union of cylinder sets, there exists $n \in \mathbb{N}$ such that $\mathbb{P}(C \cap \sigma^{-n}(C)) = \mathbb{P}(C)^2$.

Proof. Let $C = \bigcup_{i=1}^l B_i$ be a finite union of disjoint cylinder sets. Take n big enough so that $\mathbb{P}(B_i \cap B_j) = \mathbb{P}(B_i)\mathbb{P}(B_j)$ for every i, j . Then

$$\begin{aligned} \mathbb{P}(C \cap \sigma^{-n}(C)) &= \sum_{i=1}^l \sum_{j=1}^l \mathbb{P}(B_i \cap \sigma^{-n}(B_j)) \\ &= \sum_{i=1}^l \sum_{j=1}^l \mathbb{P}(B_i)\mathbb{P}(\sigma^{-n}(B_j)) \\ &= \sum_{i=1}^l \mathbb{P}(B_i) \sum_{j=1}^l \mathbb{P}(B_j) \\ &= \sum_{i=1}^l \mathbb{P}(B_i)\mathbb{P}(C) = \mathbb{P}(C)^2 \end{aligned}$$

□

For the next theorem, we will use the fact that the set \mathcal{D} of sequences with dense orbit under the shift operator has measure 1. This will be proved later (theorem 5.13) and will use some notions of ergodic theory.

Theorem 5.8. *Let φ be an i.i.d. product of contractions $f_1, \dots, f_k : M \rightarrow M$ and let K be the corresponding Hutchinson attractor. Given $p \in M$, we have*

$$\lim_{t \rightarrow \infty} \overline{\{f_\omega^n(x) : n \geq l\}} = K$$

for \mathbb{P} -a.e. ω .

Proof. Fix $x \in M$, $\omega \in \mathcal{D}$ and define $A_l = \{f_\omega^n(x) : n \geq l\}$. To prove the limit, we need to show that for each $\epsilon > 0$ there exists $L \in \mathbb{N}$ such that $d_H(K, \overline{A_l}) = \max\{d(K, \overline{A_l}), d(\overline{A_l}, K)\} \leq \epsilon$ for all $l \geq L$.

Let's prove first that for each $\epsilon > 0$, $d(\overline{A_l}, K) \leq \epsilon$ for l big enough. Because $\mathcal{B}(\{x\}) \rightarrow K$, there exists l_0 such that

$$d(\mathcal{B}^l(\{x\}), K) \leq \epsilon \text{ for } l \geq l_0.$$

In particular, because $f_\omega^n(x) \in \mathcal{B}^n(\{x\})$, we have

$$d(f_\omega^n(x), K) \leq \sup_{x \in \mathcal{B}^n(\{x\})} d(x, K) = d(\mathcal{B}^l(\{x\}), K) \leq \epsilon$$

for all $n \geq l_0$, which implies $d(A_l, K) \leq \epsilon$. As $d(\overline{A_l}, K) = d(A_l, K)$, we have $d(\overline{A_l}, K) \leq \epsilon$ for all $l \geq l_0$.

Now we will show that $d(K, \overline{A_l})$ for every $l \in \mathbb{N}$. Note that this is equivalent to showing that for any $y \in K$, there exists $N \geq l$ such that $d(f_\omega^N(x), y) < \epsilon$. We will use the following claim for this:

Claim. Given $\omega \in \Sigma$ and $x \in M$, there exists $C \in \mathbb{R}$ such that $d(f_\omega^n(x), x) \leq C$.

Proof. Because K is compact in the metric space M , it is bounded and there exists a ball $B(x, r)$ centered around a point $z \in M$ such that $K \subset B(z, r)$. By the definition of convergence, there exists $N \in \mathbb{N}$ such that $d_{\mathcal{H}}(\mathcal{B}^n(\{x\}), K) < 1$ for $n \geq N$. Therefore, for each $f_\omega^n(x)$, $n \geq N$, there exists $y \in K$ such that $d(f_\omega^n(x), y) \leq 2$ and $d(f_\omega^n(x), z) \leq d(f_\omega^n(x), y) + d(y, z) \leq 2 + r$. Now, take

$$C = 2 \cdot \max\{r + 2, d(f_\omega^0(x), z), \dots, d(f_\omega^{N-1}(x), z)\}.$$

□

By the coding map (5.2), there exists a sequence $\nu \in \Sigma$ such that

$$y = \lim_{n \rightarrow \infty} f_{\nu_0} \circ \dots \circ f_{\nu_n}(x).$$

Let $\lambda = \max\{\lambda_1, \dots, \lambda_k\}$, where λ_i is the contraction factor of f_i . We can choose $m > l$ such that $\lambda^m C < \frac{\epsilon}{2}$ and

$$d(f_{\nu_0} \circ \dots \circ f_{\nu_{m-1}}(x), y) < \frac{\epsilon}{2}.$$

Because the orbit of ω is dense under the shift operator, there exists $N \in \mathbb{N}$ such that $\sigma^N(\omega) \in [\nu_m, \dots, \nu_0]$, i.e, $\omega_N = \nu_m, \dots, \omega_{N+1} = \nu_0$. So we have

$$\begin{aligned} d(f_\omega^{N+m+1}(x), y) &\leq d(f_\omega^{N+m+1}(x), f_{\nu_0} \circ \dots \circ f_{\nu_m}(x)) + d(f_{\nu_0} \circ \dots \circ f_{\nu_m}(x), y) \\ &\leq d(f_{\nu_0} \circ \dots \circ f_{\nu_m}(f_\omega^N(x)), f_{\nu_0} \circ \dots \circ f_{\nu_m}(x)) + \frac{\epsilon}{2} \\ &\leq \lambda^m d(f_\omega^N(x), x) + \frac{\epsilon}{2} \leq \lambda^m C + \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

As we chose $m \geq l$, taking $L = N + m + 1$, we have $L \geq l$ and $f_\omega^L(x) \in A_l$. □

Our next step is to show that $\mathbb{P}(\mathcal{D}) = 1$. To prove it, we will need to utilize tools from ergodic theory. The following definitions and theorems, as well as the proof of theorem 5.10, can be found in chapter 1 of the book *An Introduction to Ergodic Theory*, by Peter Walters.

Definition 5.9. Let (X, \mathcal{B}, μ) be a probability space.

- A transformation $T : X \rightarrow X$ is said to be measurable if $T^{-1}(B) \in \mathcal{B}$ for every $B \in \mathcal{B}$.
- A measurable transformation $T : X \rightarrow X$ is said to be measure-preserving if $\mu(B) = \mu(T^{-1}B)$ for every $B \in \mathcal{B}$.
- A measure-preserving transformation $T : X \rightarrow X$ is said to be ergodic if $T(B) = B$ implies $\mu(B) = 0$ or $\mu(B) = 1$.

Although these definitions seem a little arbitrary, there are intuitions and explanations behind them:

- The measure-preserving property appears in many important phenomena in nature, such as in Hamiltonian systems. An intuition behind it is that if you have a measure-preserving transformation, the probability of a point being in a set is the same as the probability of this point staying in this set.
- If a transformation is ergodic, we cannot divide this transformation into smaller ones. In fact, if $A \in \mathcal{B}$ is such that $T(A) = A$ and $0 < \mu(A) < 1$, then we can study the transformation as $T|_A : A \rightarrow A$ and $T|_{A^c} : A^c \rightarrow A^c$.

Theorem 5.10. Let (X, \mathcal{B}, μ) be a probability space, and let $T : X \rightarrow X$ be a measure-preserving transformation. The following statements are equivalent:

1. T is ergodic.
2. If $B \in \mathcal{B}$, then $\mu(T^{-1}(B) \Delta B) = 0 \Rightarrow \mu(B) = 0$ or $\mu(B) = 1$.
3. If $A \in \mathcal{B}$, then $\mu(A) > 0 \Rightarrow \bigcup_{n=1}^{\infty} T^{-n}A = 1$.
4. For every $A, B \in \mathcal{B}$, if $\mu(A) > 0$ and $\mu(B) > 0$, then there exists $n \in \mathbb{N}$ such that $\mu(T^{-n}A \cap B) > 0$.

Theorem 5.11. Let X be a compact metric space, $\mathcal{B}(X)$ the Borel σ -algebra, and let μ be a probability measure on $(X, \mathcal{B}(X))$. If $\mu(U) > 0$ for every non-empty open set U and $T : X \rightarrow X$ is a continuous ergodic transformation, then almost all points of X have a dense orbit under T i.e. $\mu(\{x \in X : x \text{ has a dense orbit under } T\}) = 1$.

Proof. Because X is a compact metric space, there exists a countable collection $\{U_n\}, U_n \neq \emptyset$ for each n , that is a base for the topology of X . Observe that $x \in X$ has a dense orbit under T if and only if $x \in \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{\infty} T^{-k}U_n$, so $\{x \in X : x \text{ has a dense orbit under } T\} = \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{\infty} T^{-k}U_n$. Since T is ergodic and $U_n \neq \emptyset$ is open, then $\mu(\bigcup_{k=0}^{\infty} T^{-k}U_n) = 1$ (theorem 5.10). Therefore,

$$\mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=0}^{\infty} T^{-k}U_n\right) = \lim_{m \rightarrow \infty} \mu\left(\bigcap_{n=1}^m \bigcup_{k=0}^{\infty} U_n\right) = 1.$$

□

The idea here is to use theorem 5.11 to show that $\mathbb{P}(\mathcal{D}) = 1$ but we need to show that our probability space $(\Sigma, \mathcal{B}(\Sigma), \mathbb{P})$ with our shift transformation σ satisfies our hypotheses. The fact that Σ is compact follows from Tychonoff theorem and we can define a metric $\rho(\omega, \nu) = \sum_{i=0}^{\infty} \frac{\min(1, |\omega_i - \nu_i|)}{2^i}$ which generates the product topology¹, so Σ is a compact metric space. Because our topology is generated by cylinder sets, $\mathbb{P}(U) > 0$ for all non-empty open set U , and because σ is continuous, it remains to show that σ is ergodic.

For this, we will need the following lemma:

Lemma 5.12. For each $\epsilon > 0$ and each $A \in \mathcal{B}(\Sigma)$ there exists a set B which is a finite union of cylinder sets and

$$\mathbb{P}(A \Delta B) < \epsilon.$$

¹Alternatively, you can show that Σ is regular and use Urysohn metrization theorem (see theorem 34.1 in Topology, by James Munkres).

Proof. Given $\epsilon > 0$, there exists a collection of cylinder sets $\{C_n\}$ such that $A \subset \bigcup_{n=1}^{\infty} C_n$ and $\mathbb{P}(\bigcup_{n=1}^{\infty} C_n) \leq \mathbb{P}(A) + \epsilon/2$. For N big enough, we have $\mathbb{P}(B) \geq \mathbb{P}(A)$, where $B = \bigcup_{n=1}^N C_n$, so that

$$\begin{aligned} \mathbb{P}(A\Delta B) &= \mathbb{P}(A\setminus B) + \mathbb{P}(B\setminus A) \\ &\leq \mathbb{P}\left(\bigcup_{n=1}^{\infty} C_n\setminus B\right) + \mathbb{P}\left(\bigcup_{n=1}^{\infty} C_n\setminus A\right) \\ &\leq \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

□

Theorem 5.13. *Let $\Sigma = \{1, \dots, k\}^{\mathbb{N}}$ be the space of sequences, $\mathcal{B}(\Sigma)$ the Borel σ -algebra, \mathbb{P} the probability measure and $\sigma : \Sigma \rightarrow \Sigma$ the shift operator. Then σ is ergodic.*

Proof. Let $A \in \mathcal{B}(\Sigma)$ be a measurable set such that $\sigma^{-1}(A) = A$. We will prove that $\mathbb{P}(A) = 1$ or $\mathbb{P}(A) = 0$ by showing that $\mathbb{P}(A)^2 = \mathbb{P}(A)$. By lemma 5.12, there exists a set B which is a finite union of cylinder sets and $\mathbb{P}(A\Delta B) < \epsilon/4$. So

$$\begin{aligned} |\mathbb{P}(A) - \mathbb{P}(B)| &= |[\mathbb{P}(A\setminus B) + \mathbb{P}(A \cap B)] - [\mathbb{P}(B\setminus A) + \mathbb{P}(A \cap B)]| \\ &= |\mathbb{P}(A\setminus B) - \mathbb{P}(B\setminus A)| \\ &\leq \mathbb{P}(A\setminus B) + \mathbb{P}(B\setminus A) \\ &= \mathbb{P}(A\Delta B) < \epsilon/4. \end{aligned}$$

Because B is a finite union of cylinder sets, there exists $n \in \mathbb{N}$ such that $\mathbb{P}(B \cap \sigma^{-n}(B)) = \mathbb{P}(B)^2$ (theorem 5.6). Then

$$\begin{aligned} \mathbb{P}(A\Delta\sigma^{-n}(B)) &= \mathbb{P}(\sigma^{-n}(A)\Delta\sigma^{-n}(B)) \\ &= \mathbb{P}(\sigma^{-n}(A\Delta B)) \\ &= \mathbb{P}(A\Delta B) \\ &< \epsilon/4. \end{aligned}$$

Observe now that $A\Delta(B \cap \sigma^{-n}(B)) \subset (A\Delta B) \cup (A\Delta\sigma^{-n}(B))$, so

$$\mathbb{P}(A\Delta(B \cap \sigma^{-n}(B))) \leq \mathbb{P}(A\Delta B) + \mathbb{P}(A\Delta\sigma^{-n}(B)) < \epsilon/4 + \epsilon/4 = \epsilon/2.$$

□

Therefore, we have

$$\begin{aligned} |\mathbb{P}(A) - \mathbb{P}(A)^2| &\leq |\mathbb{P}(A) - \mathbb{P}(B \cap \sigma^{-n}(B))| + |\mathbb{P}(B \cap \sigma^{-n}(B)) - \mathbb{P}(A)^2| \\ &\leq \mathbb{P}(A\Delta(B \cap \sigma^{-n}(B))) + |\mathbb{P}(B)^2 - \mathbb{P}(A)^2| \\ &\leq \epsilon/2 + \mathbb{P}(B)|\mathbb{P}(B) - \mathbb{P}(A)| + \mathbb{P}(A)|\mathbb{P}(B) - \mathbb{P}(A)| \\ &\leq \epsilon/2 + \epsilon/4 + \epsilon/4 = \epsilon \end{aligned}$$

where the last inequality holds because \mathbb{P} is a probability measure. Because ϵ is arbitrary, $|\mathbb{P}(A) - \mathbb{P}(A)^2| = 0$ and σ is ergodic.

6 Conditional Expectations

Let (X, \mathcal{B}, μ) be a probability space, and $\mathcal{F} \subset \mathcal{B}$ a sub- σ -algebra.

Definition 6.1. Let $f : X \rightarrow \mathbb{R}$ be a (\mathcal{B} -measurable) integrable random variable ($f \in L^1(X, \mathcal{B}, \mu)$). The *expectation* of f is defined to be:

$$\mathbb{E}(f) := \int_X f d\mu.$$

Definition 6.2. Let $f : X \rightarrow \mathbb{R}$ be a random variable and $\mathcal{F} \subset \mathcal{B}$ a σ -algebra. We say that a \mathcal{F} -measurable integrable random variable g is a *conditional expectation* of f given \mathcal{F} if:

$$\int_A f d\mu = \int_A g d\mu \quad \forall A \in \mathcal{F}.$$

In this case, we denote g by $\mathbb{E}(f|\mathcal{F})$.

The existence of conditional expectations comes from Radon-Nikodym Theorem, as it follows.

Definition 6.3. Let μ and ν be measures in a measurable space (X, \mathcal{B}) . We say that ν is *absolutely continuous* with respect to μ if $\mu(A) = 0$ implies $\nu(A) = 0$. In this case, we use the notation $\nu \ll \mu$.

Theorem 6.4. (*Radon-Nikodym Theorem*) Let μ and ν be finite measures in a measurable space (X, \mathcal{B}) , and suppose $\nu \ll \mu$. Then there exists an essentially unique² measurable function $f : X \rightarrow \mathbb{R}$ such that:

$$\nu(A) = \int_A f d\mu \quad \forall A \in \mathcal{B}. \quad (1)$$

The following notations are often used to represent (1): $d\nu = f d\mu$ or $f = \frac{d\nu}{d\mu}$. In this case, f is called the *Radon-Nikodym derivative* of ν with respect to μ .

Proof. The proof can be found in Theorem 3.2.2 of Bogachev. □

Proposition 6.5. Given a sub- σ -algebra $\mathcal{F} \subset \mathcal{B}$, the conditional expectation $\mathbb{E}(f|\mathcal{F})$ exists and is essentially unique.

Proof. For each $A \in \mathcal{F}$, define:

$$\nu(A) := \int_A f d\mu.$$

It follows that $\nu \ll \mu|_{\mathcal{F}}$. Then define $\mathbb{E}(f|\mathcal{F}) := \frac{d\nu}{d\mu|_{\mathcal{F}}}$, which exists and is essentially unique by the Radon-Nikodym Theorem. It is straightforward to check that $\mathbb{E}(f|\mathcal{F})$ so defined is in fact the conditional expectation of f given \mathcal{F} . □

Proposition 6.6. Here we present some basic properties about conditional expectations. Consider $\mathcal{F} \subset \mathcal{B}$ to be a sub- σ -algebra, and $f : X \rightarrow \mathbb{R}$ a random variable.

1. Given two random variables $f, g : X \rightarrow \mathbb{R}$ and $a, b \in \mathbb{R}$, we have:

$$\mathbb{E}(af + bg|\mathcal{F}) = a \mathbb{E}(f|\mathcal{F}) + b \mathbb{E}(g|\mathcal{F}).$$

2. $\mathbb{E}(\mathbb{E}(f|\mathcal{F})) = \mathbb{E}(f)$.

3. If f is \mathcal{F} -measurable, then $\mathbb{E}(f|\mathcal{F}) = f$ a.e.

4. If $f \leq g$, then $\mathbb{E}(f|\mathcal{F}) \leq \mathbb{E}(g|\mathcal{F})$.

5. If g is a \mathcal{F} -measurable bounded random variable, then $\mathbb{E}(gf|\mathcal{F}) = g \mathbb{E}(f|\mathcal{F})$.

6. If $\mathcal{G} \subset \mathcal{F}$, then $\mathbb{E}(\mathbb{E}(f|\mathcal{G})|\mathcal{F}) = \mathbb{E}(\mathbb{E}(f|\mathcal{F})|\mathcal{G}) = \mathbb{E}(f|\mathcal{G})$.

Proof. 1-3. These properties follow immediately from the definition of conditional expectation.

4. Since $g - f \geq 0$, then $\mathbb{E}((g - f)\chi_A) \geq 0$ for every $A \in \mathcal{F}$, i.e., $\mathbb{E}(f\chi_A) \leq \mathbb{E}(g\chi_A)$. Therefore, by definition of conditional expectation, we conclude that $\mathbb{E}(f|\mathcal{F}) \leq \mathbb{E}(g|\mathcal{F})$ a.e.

5. Since g is bounded, there exists $M > 0$ such that $|g| \leq M$; so $\mathbb{E}(|gf|) \leq M \mathbb{E}(|f|) < \infty$, proving that gf is integrable. Moreover, since g is \mathcal{F} -measurable, we obtain:

$$g \mathbb{E}(f|\mathcal{F}) = \mathbb{E}(gf|\mathcal{F}) \stackrel{(3)}{=} \mathbb{E}(g|\mathcal{F}) \mathbb{E}(f|\mathcal{F}) = \mathbb{E}(gf|\mathcal{F}).$$

²To be essentially unique means that two functions satisfying (1) are equal a.e.

6. Given $A \in \mathcal{F}$, we have:

$$\int_A \mathbb{E}(\mathbb{E}(f|\mathcal{G})|\mathcal{F}) d\mu = \int_A \mathbb{E}(f|\mathcal{G}) d\mu \stackrel{(3)}{=} \int_A \mathbb{E}(\mathbb{E}(f|\mathcal{F})|\mathcal{G}). \quad \square$$

Example 6.7. Let \mathcal{F} be the σ -algebra generated by a finite partition $\{A_1, \dots, A_n\}$ of X , i.e. the elements of the partition are nonempty, disjoint and $\bigcup_{i=1}^n A_i = X$. Suppose that $\mu(A_i) > 0$ for every $i \in \{1, \dots, n\}$. Then:

$$\mathbb{E}[f|\mathcal{F}] = \sum_{i=1}^n \frac{\mathbb{E}[f\chi_{A_i}]}{\mu(A_i)} \chi_{A_i}.$$

In fact, given $A_j \in \{A_1, \dots, A_n\}$, we have:

$$\int_{A_j} \sum_{i=1}^n \frac{\mathbb{E}[f\chi_{A_i}]}{\mu(A_i)} \chi_{A_i} d\mu = \sum_{i=1}^n \frac{\mathbb{E}[f\chi_{A_i}]}{\mu(A_i)} \int_{A_j} \chi_{A_i} d\mu = \int_{A_j} f d\mu.$$

Moreover, since $X \setminus A_j = \bigcup_{i \neq j} A_i$, then:

$$\int_{X \setminus A_j} \sum_{i=1}^n \frac{\mathbb{E}[f\chi_{A_i}]}{\mu(A_i)} \chi_{A_i} d\mu = \sum_{\substack{i=1 \\ i \neq j}}^n \frac{\mathbb{E}[f\chi_{A_i}]}{\mu(A_i)} \int_{X \setminus A_j} \chi_{A_i} d\mu = \sum_{\substack{i=1 \\ i \neq j}}^n \mathbb{E}[f\chi_{A_i}] = \int_{X \setminus A_j} f d\mu.$$

Finally, given $A_j, A_k \in \{A_1, \dots, A_n\}$, we conclude that:

$$\begin{aligned} \int_{A_j \cup A_k} \sum_{i=1}^n \frac{\mathbb{E}[f\chi_{A_i}]}{\mu(A_i)} \chi_{A_i} d\mu &= \sum_{i=1}^n \frac{\mathbb{E}[f\chi_{A_i}]}{\mu(A_i)} \int_{A_j \cup A_k} \chi_{A_i} d\mu = \mathbb{E}[f\chi_{A_j}] + \mathbb{E}[f\chi_{A_k}] \\ &= \int_{A_j \cup A_k} f d\mu. \end{aligned}$$

In particular, if \mathcal{F} is the σ -algebra generated by $A \subset X$, i.e. $\mathcal{F} = \{\emptyset, X, A, A^c\}$, we obtain:

$$\mathbb{E}[f|\mathcal{F}] = \frac{\mathbb{E}[f\chi_A]}{\mu(A)} \chi_A + \frac{\mathbb{E}[f\chi_{A^c}]}{\mu(A^c)} \chi_{A^c}.$$

6.1 Application in L^2

Suppose $f \in L^2(X, \mathcal{B}, \mu)$, that is, $\mathbb{E}f^2 < \infty$. Recall that L^2 is a Hilbert space with the inner product:

$$\langle f, g \rangle = \int_X fg d\mu.$$

Given a sub- σ -algebra $\mathcal{F} \subset \mathcal{B}$, the set $L^2(\mathcal{F})$ is a closed subspace. It is quite useful to look for the projection of f onto $L^2(\mathcal{F})$, which is the unique point in the subspace closest to f .

Theorem 6.8. *The projection of f onto the closed subspace $L^2(\mathcal{F})$ is the conditional expectation $\mathbb{E}(f|\mathcal{F})$.*

Proof. By the Hilbert Projection Theorem, there exists a unique random variable $g \in L^2(\mathcal{F})$ that minimizes the “mean square error” $\mathbb{E}(f - g)^2$, which is precisely the unique element in $L^2(\mathcal{F})$ such that $f - g$ is orthogonal to $L^2(\mathcal{F})$. Hence for any $A \in \mathcal{F}$, we have:

$$\langle \chi_A, f - g \rangle = 0 \implies \int \chi_A(f - g) d\mu = 0 \implies \int_A f d\mu = \int_A g d\mu. \quad \square$$

7 Markov Chains

Definition 7.1. Let $h : X \rightarrow Y$ be a function from a set X to a measurable space (Y, \mathcal{G}) . We define on X the σ -algebra generated by h to be the smallest σ -algebra with respect to which h is measurable:

$$\sigma(h) := \{h^{-1}(A); A \in \mathcal{G}\}.$$

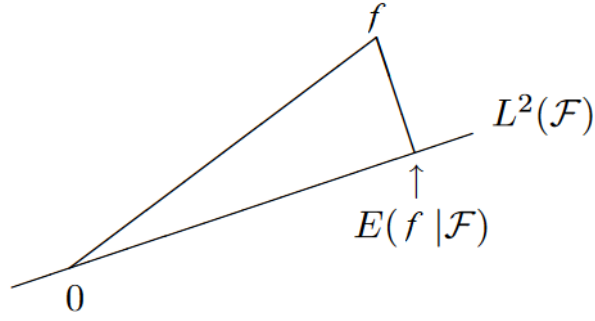


Figure 2: Conditional expectation as projection in L^2 .

If $\{(Y_\alpha, \mathcal{G}_\alpha), \alpha \in \Gamma\}$ is a family of measurable spaces and $\mathcal{C} := \{h_\alpha; \alpha \in \Gamma\}$ is a family of functions $h_\alpha : X \rightarrow Y_\alpha$, we then define the σ -algebra generated by \mathcal{C} to be:

$$\sigma(h_\alpha; \alpha \in \Gamma) := \sigma \left(\bigcup_{\alpha \in \Gamma} \sigma(h_\alpha) \right).$$

Definition 7.2. Let (X, \mathcal{F}, μ) be a probability space.

- Two sets $A, B \in \mathcal{F}$ are *independent* if $\mu(A \cap B) = \mu(A)\mu(B)$.
- Two σ -algebras \mathcal{F}, \mathcal{G} are *independent* if, for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$, A and B are independent.
- A random variable $h : X \rightarrow \mathbb{R}$ is *independent* of the σ -algebra \mathcal{G} if $\sigma(h)$ is independent of \mathcal{G} .
- Two random variables $h_1, h_2 : X \rightarrow \mathbb{R}$ are *independent* if their generated σ -algebras $\sigma(h_1)$ and $\sigma(h_2)$ are independent.
- A sequence $(h_n)_{n \in \mathbb{N}_0}$ of random variables is *independent* if any finite sub-collection of them is independent, in the following sense:

$$\mu \left(\bigcap_{j=1}^k h_{i_j}^{-1}(A_{i_j}) \right) = \prod_{j=1}^k \mu \left(h_{i_j}^{-1}(A_{i_j}) \right),$$

for any finite subset $\{i_1, \dots, i_k\} \subset \mathbb{N}_0$ and $A_{i_j} \in \mathcal{B}(\mathbb{R})$.

Proposition 7.3. If h_1 and h_2 are independent random variables, then $\mathbb{E}(h_1 h_2) = \mathbb{E}(h_1)\mathbb{E}(h_2)$. Also, if $h : X \rightarrow \mathbb{R}$ is independent of a σ -algebra \mathcal{G} , then $\mathbb{E}(h|\mathcal{G}) = \mathbb{E}(h)$ almost everywhere.

Proof. First consider the case of indicator functions, $h_1 = \chi_A$ and $h_2 = \chi_B$. Then:

$$\mathbb{E}(\chi_A \chi_B) = \int \chi_A \chi_B d\mu = \int_{A \cap B} d\mu = \mu(A \cap B) = \mu(A)\mu(B) = \int \chi_A d\mu \int \chi_B d\mu = \mathbb{E}(\chi_A)\mathbb{E}(\chi_B).$$

By linearity, the result follows for any elementary function and, by convergence theorems on Measure Theory, it follows also for any positive integrable function. Finally, summing up the positive and negative parts, the result is valid for any integrable function.

For the second statement, let $A \in \mathcal{G}$ and define $h_2 := \chi_A$. Then h and h_2 are independent random variables. From the first part we have:

$$\int_A (E|_{\mathcal{G}}) d\mu = \int_A h d\mu = \mathbb{E}(h\chi_A) = \mathbb{E}(h)\mathbb{E}(\chi_A) = \mathbb{E}(h)\mu(A) = \int_A \mathbb{E}(h) d\mu. \quad \square$$

Let M be a metric space with the Borel σ -algebra $\mathcal{B}(M)$, and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Definition 7.4. A function $p : M \times \mathcal{B}(M) \rightarrow \mathbb{R}$ is said to be a *transition probability* if:

1. For each $x \in M$, the function $A \mapsto p(x, A)$ is a probability measure on $(M, \mathcal{B}(M))$;
2. For each $A \in \mathcal{B}(M)$, the function $x \mapsto p(x, A)$ is measurable.

Definition 7.5. Let I be a subset of \mathbb{R} . A *stochastic process* is a family $Z = \{Z_t; t \in I\}$ of random variables on Ω with values in M .

Remark 7.6. In this work we will consider only the case where $I = \mathbb{N}_0$.

Definition 7.7. Let $(Z_n)_{n \in \mathbb{N}_0}$ be a stochastic process on Ω with values in M . We say that $(Z_n)_{n \in \mathbb{N}_0}$ is a *Markov chain with transition probability* p if, for every $A \in \mathcal{B}(M)$:

$$\mathbb{P}(Z_{n+1} \in A \mid \mathcal{F}_n) = p(Z_n, A). \quad (2)$$

Here, $\mathcal{F}_n := \sigma(Z_0, \dots, Z_n)$, and the notation $\mathbb{P}(Z_{n+1} \in A \mid \mathcal{F}_n)$ means the conditional expectation (with respect to the measure \mathbb{P}) $\mathbb{E}(\chi_{Z_{n+1}^{-1}(A)} \mid \mathcal{F}_n)$. So (2) means that, for \mathbb{P} -almost every $\omega \in \Omega$:

$$\mathbb{P}(Z_{n+1} \in A \mid \mathcal{F}_n)(\omega) = \mathbb{E}(\chi_{Z_{n+1}^{-1}(A)} \mid \mathcal{F}_n)(\omega) = p(Z_n(\omega), A).$$

Remark 7.8. In the case where M is countable, the Markov property is equivalent to say that, for all $z_0, \dots, z_n, z \in M$, we have:

$$\mathbb{P}(Z_{n+1} = z \mid Z_0 = z_0, \dots, Z_n = z_n) = \mathbb{P}(Z_{n+1} = z \mid Z_n = z_n).$$

Definition 7.9. Let (X, \mathcal{F}, μ) be a probability space and (Y, \mathcal{G}) a measurable space. If $h : X \rightarrow Y$ is a measurable function, we can define the *push-forward measure* $h_*\mu$ on Y by:

$$h_*\mu(A) := \mu(h^{-1}(A)) \quad \forall A \in \mathcal{G}.$$

We call $h_*\mu$ the *distribution* of h .

Remark 7.10. We say that a stochastic process $(X_n)_{n \in \mathbb{N}_0}$ is *independent and identically distributed (i.i.d.)* if it is independent, in the sense of Definition 7.2, and each X_n has the same distribution $(X_n)_*\mu = \nu$.

Example 7.11. Given a measurable space (Y, \mathcal{G}) and a point $y \in Y$, we can define the *Dirac Delta measure* on Y :

$$\delta_y(A) := \begin{cases} 1 & \text{if } y \in A, \\ 0 & \text{if } y \notin A \end{cases} \quad \forall A \in \mathcal{G}.$$

8 Stationary Measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X = (X_n)_{n \in \mathbb{N}_0}$ an independent stochastic process on Ω with values on a finite set $\{1, \dots, k\}$, such that each X_n has the same distribution $\nu = p_1\delta_1 + \dots + p_k\delta_k$. Let Z_0 be a random variable on Ω with values in a metric space M independent of X .

Consider $f_1, \dots, f_k : M \rightarrow M$ to be continuous functions, and define the sequence of random variables:

$$Z_n(\omega) := f_{X_{n-1}(\omega)} \circ \dots \circ f_{X_0(\omega)}(Z_0(\omega)) = f_{X_{n-1}}(Z_{n-1}).$$

Proposition 8.1. *The sequence $(Z_n)_{n \in \mathbb{N}_0}$ is a Markov chain with transition probability given by:*

$$p(x, A) = \int \chi_A(f_i(x)) d\nu(i) = \sum_{i=1}^k p_i \chi_A(f_i(x)).$$

Proof. We need to prove that, for each $A \in \mathcal{B}(M)$ and \mathbb{P} -almost every $\omega \in \Omega$:

$$\mathbb{P}(Z_{n+1} \in A \mid \sigma(Z_0, \dots, Z_n))(\omega) = p(Z_n(\omega), A).$$

Remember: to prove that some conditional expectation $\mathbb{E}(f \mid \mathcal{G})$ is essentially equal to a function g is the same as proving that g is \mathcal{G} -measurable and that, for each $B \in \mathcal{G}$, we have $\int_B g = \int_B f$.

Step 1. Given $B \in \sigma(Z_0, \dots, Z_n)$, we must prove that:

$$\int_B \chi_A(Z_{n+1}(\omega)) d\mathbb{P}(\omega) = \int_B p(Z_n(\omega), A) d\mathbb{P}(\omega). \quad (3)$$

Observe that $\sigma(Z_0, \dots, Z_n) \subset \sigma(Z_0, X_0, \dots, X_{n-1})$. In fact, we have the following compositions:

$$\begin{array}{ccc} \Omega & \xrightarrow{(X_0, \dots, X_{n-1}, Z_0)} & \{1, \dots, k\}^n \times M & \xrightarrow{f} & M \\ \omega & \longmapsto & (X_0(\omega), \dots, X_{n-1}(\omega), Z_0(\omega)) & \longmapsto & Z_n(\omega). \end{array}$$

Now take a basis set $B \in \sigma(Z_0, \dots, Z_n)$, $B \in \sigma(Z_0) \cup \dots \cup \sigma(Z_n)$. If $B \in \sigma(Z_0)$, then of course it belongs to $\sigma(Z_0, X_0, \dots, X_{n-1})$. Suppose that $B \in \sigma(Z_j)$, $j \neq 0$. Then $B = Z_j^{-1}(C)$ for some $C \in \mathcal{B}(M)$. By the above composition, we also have $B = Z_j^{-1}(C) = (X_0, \dots, X_{n-1}, Z_0)^{-1}(f^{-1}(C)) \in \sigma(X_0, \dots, X_{n-1}, Z_0)$; this concludes the claim.

We also know, as above, that $\sigma(Z_0, X_0, \dots, X_{n-1})$ is the σ -algebra generated by the following map $\iota: \Omega \rightarrow \{1, \dots, k\}^n \times M$:

$$\iota(\omega) := (X_0(\omega), \dots, X_{n-1}(\omega), Z_0(\omega)).$$

So B can be chosen as a set of the form $\iota^{-1}(B')$ for some measurable set B' on $\{1, \dots, k\}^n \times M$. We then have:

$$\begin{aligned} \int_B \chi_A(Z_{n+1}) d\mathbb{P} &= \int \chi_A(Z_{n+1}) \chi_B d\mathbb{P} \\ &= \int \chi_A(f_{X_n} \circ \dots \circ f_{X_0}(Z_0)) \chi_{B'}(X_0, \dots, X_{n-1}, Z_0) d\mathbb{P}. \end{aligned}$$

Since X_0, \dots, X_n, Z_0 are independent, the distribution of the vector (X_0, \dots, X_n, Z_0) is $\nu^{n+1} \times \mu$. Denoting by $(\alpha_0, \dots, \alpha_n, x)$ a typical vector on $\{1, \dots, k\}^{n+1} \times M$, we then have:

$$\int_B \chi_A(Z_{n+1}) d\mathbb{P} = \iint \chi_A(f_{\alpha_n} \circ \dots \circ f_{\alpha_0}(x)) \chi_{B'}(\alpha_0, \dots, \alpha_{n-1}, x) d\nu^{n+1} d\mu. \quad (4)$$

Moreover, by definition of p , we know that:

$$\int \chi_A(f_{\alpha_n} \circ \dots \circ f_{\alpha_0}(x)) d\nu(\alpha_n) = p(f_{\alpha_{n-1}} \circ \dots \circ f_{\alpha_0}(x), A).$$

From (4) we obtain:

$$\begin{aligned} \int_B \chi_A(Z_{n+1}) d\mathbb{P} &= \int p(f_{\alpha_{n-1}} \circ \dots \circ f_{\alpha_0}(x), A) \chi_{B'}(\alpha_0, \dots, \alpha_{n-1}, x) d\nu^n d\mu \\ &= \int_B p(f_{X_{n-1}} \circ \dots \circ f_{X_0}(Z_0), A) d\mathbb{P}. \end{aligned}$$

This proves formula (3), concluding the first step.

Step 2. We now prove that the function $p(Z_n, A): \Omega \rightarrow \mathbb{R}$ is $\sigma(Z_0, \dots, Z_n)$ -measurable. Since $A \in \mathcal{B}(M)$ is fixed, we have the following compositions.

$$\begin{array}{ccccc} \Omega & \longrightarrow & M & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & Z_n(\omega) & \longmapsto & p(Z_n(\omega), A). \end{array}$$

By definition of transition probability, the second arrow is a measurable function. The first arrow is $\sigma(Z_n)$ -measurable by definition of generated σ -algebra; since $\sigma(Z_n) \subset \sigma(Z_0, \dots, Z_n)$, we know that it is also $\sigma(Z_0, \dots, Z_n)$ -measurable. Finally, the composition of measurable functions is also measurable, and we conclude the proof. \square

8.1 Stationary Measure

In this part our goal is to prove the previously Markov Chain has a stationary measure. For that, let (M, d) be a metric space and $\mathcal{M}_1(M)$ be the set of all probability measures on M .

Suppose Z_0 has distribution μ , then the distribution of Z_n is $\mathcal{T}^n\mu$, with

$$\begin{aligned}\mathcal{T} : \mathcal{M}_1(M) &\rightarrow \mathcal{M}_1(M) \\ \mu &\mapsto \mathcal{T}\mu(A) = \int p(x, A)d\mu(x)\end{aligned}$$

being the Markov Operator.

In fact, suppose Z_n has distribution ν so the distribution of $Z_{n+1} : \Omega \rightarrow M$ can be define by $Z_{n+1} * \mathbb{P}$ such that

$$Z_{n+1} * \mathbb{P}(A) = \mathbb{P}(Z_{n+1}^{-1}(A)) = \{w \in \Omega : Z_1(\omega) \in A\} = \mathbb{P}(Z_{n+1} \in A).$$

We already saw that

$$Z_{n+1} * \mathbb{P}(A) = \mathbb{P}(Z_{n+1} \in A) = \int \chi_A(Z_{n+1}(\omega))d\mathbb{P}(\omega) = \int p(Z_n(\omega, A))d\mathbb{P}(\omega) = \int p(x, A)dZ_n * \mathbb{P}(x) = \int p(x, A)d\nu(x).$$

A measure μ is said to be stationary measure if Z_0 has distribution μ then so does Z_n for all $n \geq 1$. In other words, μ is a fixed point for the Markov Operator.

So to be able to determine the stationary measure we need first define a topology for $\mathcal{M}_1(M)$ called by Weak* Topology.

8.2 Weak* topology

To define the Weak* topology in the probability measures set, we need to see the following definitions.

Definition 8.2. Given $B \subset M$, we call the δ -neighborhood of B the set

$$B^\delta = \{x \in M : d_1(x, B) < \delta\},$$

with $d_1(x, B) = \inf\{d(x, b) : b \in B\}$.

Definition 8.3. Given $\mu \in \mathcal{M}_1(M)$, $\epsilon > 0$ and $\Phi = \{\phi_i : i = 1, \dots, N\}$ such that

$$\phi_i : M \rightarrow \mathbb{R}, \forall i = 1, \dots, N$$

are bounded continuous functions. Define the set

$$V(\mu, \Phi, \epsilon) = \{\nu \in \mathcal{M}_1(M) : \left| \int \phi_i d\nu - \int \phi_i d\mu \right| < \epsilon, \forall i = 1, \dots, N\}.$$

Proposition 8.4. The family $\mathcal{B} = \{V(\mu, \Phi, \epsilon)\}$ may be taken as a base for a topology in $\mathcal{M}_1(M)$.

Proof. To prove that it's sufficient to show the following two conditions:

- each $\nu \in \mathcal{M}_1(M)$ is contained in some subset in \mathcal{B} ;
- if $V_1, V_2 \in \mathcal{B}$ and $\nu \in V_1 \cap V_2$, then exists $W \in \mathcal{B}$ with $\nu \in W \subset V_1 \cap V_2$.

In fact, by definition $\nu \in V(\nu, \Phi, \epsilon)$ for every Φ and $\epsilon > 0$.

Moreover, consider $\nu \in V(\mu_1, \Phi_1, \epsilon_1) \cap V(\mu_2, \Phi_2, \epsilon_2)$ with $\Phi_1 = \{\phi_{1i} : i = 1, \dots, N_1\}$, $\Phi_2 = \{\phi_{2i} : i = 1, \dots, N_2\}$ and $\epsilon_1, \epsilon_2 > 0$.

Then

$$\left| \int \phi_{1i} d\nu - \int \phi_{1i} d\mu_1 \right| < \epsilon, \forall i = 1, \dots, N_1$$

and

$$\left| \int \phi_{2i} d\nu - \int \phi_{2i} d\mu_2 \right| < \epsilon, \forall i = 1, \dots, N_2.$$

Consider

$$r_1 = \max \left\{ \left| \int \phi_{1i} d\nu - \int \phi_{1i} d\mu_1 \right| : i = 1, \dots, N_1 \right\}$$

and

$$r_2 = \max \left\{ \left| \int \phi_{2i} d\nu - \int \phi_{2i} d\mu_2 \right| : i = 1, \dots, N_2 \right\}$$

Taking $\epsilon = \min\{\frac{\epsilon_1 - r_1}{2}, \frac{\epsilon_2 - r_2}{2}\}$ and $\Phi = \Phi_1 \cup \Phi_2$, define the set $V(\nu, \Phi, \epsilon)$.

If $z \in V(\nu, \Phi, \epsilon)$ then

$$\left| \int \phi_{1i} dz - \int \phi_{1i} d\nu \right| < \epsilon, \quad \forall i = 1, \dots, N_1$$

therefore

$$\begin{aligned} \left| \int \phi_{1i} dz - \int \phi_{1i} d\mu_1 \right| &\leq \left| \int \phi_{1i} dz - \int \phi_{1i} d\nu \right| + \left| \int \phi_{1i} d\nu - \int \phi_{1i} d\mu_1 \right| \\ &< \epsilon + r_1 \\ &< \epsilon_1 \end{aligned}$$

that means $z \in V(\mu_1, \Phi_1, \epsilon_1)$.

With a similar process we get $z \in V(\mu_2, \Phi_2, \epsilon_2)$.

Definition 8.5 (Weak* topology). The weak* topology it's the topology generated by \mathcal{B} .

Lemma 8.6 (Convergence in $\mathcal{M}_1(M)$). A sequence $(\mu_n)_{n \in \mathbb{N}}$ of $\mathcal{M}_1(M)$ elements converges to a measure $\mu \in \mathcal{M}_1(M)$ in the weak* topology if and only if

$$\int \phi d\mu_n \rightarrow \int \phi d\mu \quad \forall \text{ bounded continuous functions } \phi : M \rightarrow \mathbb{R}.$$

Proof. (\implies) Suppose $\mu_n \rightarrow \mu$ in the weak* topology, consider any $\phi : M \rightarrow \mathbb{R}$ bounded continuous function and the set $\Phi = \{\phi\}$.

So for any $\epsilon > 0$ there is a $n_0 \in \mathbb{N}$ that

$$\mu_n \in V(\mu, \Phi, \epsilon), \quad \forall n \geq n_0.$$

In other way:

$$\left| \int \phi d\mu_n - \int \phi d\mu \right| < \epsilon, \quad \forall n_0 \leq n,$$

and that implies

$$\int \phi d\mu_n \rightarrow \int \phi d\mu$$

in \mathbb{R} .

(\impliedby) Suppose that $\int \phi d\mu_n \rightarrow \int \phi d\mu$ in \mathbb{R} for all bounded continuous function $\phi : M \rightarrow \mathbb{R}$.

Given any $\Phi = \{\phi_i : i = 1, \dots, N\}$ and $\epsilon > 0$, so for every $i = 1, \dots, N$ exists $n_0^i \in \mathbb{N}$ such that

$$n > n_0^i \implies \left| \int \phi_i d\mu_n - \int \phi_i d\mu \right| < \epsilon.$$

Choosing $n_0 = \max\{n_0^i : i = 1, \dots, N\}$ we have

$$n > n_0 \implies \left| \int \phi_i d\mu_n - \int \phi_i d\mu \right| < \epsilon, \quad \forall i = 1, \dots, N$$

and that implies

$$\mu_n \in V(\mu, \Phi, \epsilon), \quad \forall n > n_0.$$

Therefore $\mu_n \xrightarrow{*} \mu$ in weak* topology.

8.3 Markov Operator

The Markov Operator is define by:

$$\begin{aligned}\mathcal{T} : \mathcal{M}_1(M) &\rightarrow \mathcal{M}_1(M) \\ \mu &\mapsto \mathcal{T}\mu(A) = \int p(x, A)d\mu(x).\end{aligned}$$

Consider the Markov Chain (Z_n) previously defined, if Z_0 has distribution μ then the distribution of Z_n is given by $\mathcal{T}^n\mu$.

Definition 8.7. The distribution μ on M is called stationary measure if it is a fixed point of the Markov Operator.

Proposition 8.8. If $\phi : M \rightarrow \mathbb{R}$ is a bounded continuous function, then

$$\int \phi(y)p(x, dy) = \sum_{i=1}^k p_i \phi \circ f_i(x).$$

And denote $P\phi(x) = \int \phi(y)p(x, dy)$.

Proof. Remember that

$$p(x, A) = \int \chi_A(f_i(x))d\nu(i) = \sum_{i=1}^k p_i \chi_A(f_i(x)).$$

We will divide this proof in four steps.

Step 1. Suppose that ϕ is a characteristic function, that is, $\phi = \chi_B$, for some $B \subset M$. So

$$\begin{aligned}\int \phi(y)p(x, dy) &= \int \chi_B(y)p(x, dy) \\ &= p(x, B) \\ &= \int \chi_B(f_i(x))d\nu(i) \\ &= \sum_{i=1}^k p_i \chi_B(f_i(x)) \\ &= \sum_{i=1}^k p_i \phi(f_i(x))\end{aligned}$$

Step 2. Suppose that ϕ is a simple function, that is, $\phi = \sum_{j=1}^m \alpha_j \chi_{A_j}$, with $M = \cup_{j=1}^m A_j$ and $A_j \cap A_K = \emptyset$. So

$$\begin{aligned}\int \phi(y)p(x, dy) &= \int \sum_{j=1}^m \alpha_j \chi_{A_j} p(x, dy) \\ &= \sum_{j=1}^m \alpha_j \int \chi_{A_j} p(x, dy) \\ &= \sum_{j=1}^m \alpha_j \sum_{i=1}^k p_i \chi_{A_j}(f_i(x)) \\ &= \sum_{i=1}^k p_i \sum_{j=1}^m \alpha_j \chi_{A_j}(f_i(x)) \\ &= \sum_{i=1}^k p_i \phi(f_i(x)).\end{aligned}$$

Step 3. Suppose that ϕ is a positive function. So there is a sequence of samples functions $(s_n)_n$ that $s_1 \leq s_2 \leq \dots \leq \phi$.

We know that

$$\int \phi(y)p(x, dy) = \sup \left\{ \int s_n(y)p(x, dy) : n \in \mathbb{N} \right\}.$$

For the step 2 we have:

$$\begin{aligned} \int \phi(y)p(x, dy) &= \sup \left\{ \sum_{i=1}^k p_i s_n \circ f_i(x) : n \in \mathbb{N} \right\} \quad (\sup s_n = \phi) \\ &= \sum_{i=1}^k p_i \phi \circ f_i(x) \end{aligned}$$

Step 4. Let ϕ be any function, so $\phi = \phi^+ - \phi^-$ where ϕ^+ and ϕ^- are positive functions, so

$$\begin{aligned} \int \phi(y)p(x, dy) &= \int \phi^+(y)p(x, dy) - \int \phi^-(y)p(x, dy) \quad (\text{Step 3}) \\ &= \sum_{i=1}^k p_i \phi^+(f_i(x)) - \sum_{i=1}^k p_i \phi^-(f_i(x)) \\ &= \sum_{i=1}^k p_i (\phi^+(f_i(x)) - \phi^-(f_i(x))) \\ &= \sum_{i=1}^k p_i \phi \circ f_i(x). \end{aligned}$$

Affirmation. If $\Phi : M \rightarrow \mathbb{R}$ it's a integral function so $\int \phi d\mathcal{T}\mu = \int \phi(y) \int p(x, dy) d\mu_n(x) = \int \int \phi(y)p(x, dy) d\mu(x)$.

In fact, we will to show this affirmation in four steps.

Step 1. Suppose $\phi = \chi_B$ for some $B \subset M$ measurable, then

$$\begin{aligned} \int \phi d\mathcal{T}\mu &= \int \chi_B(y) d\mathcal{T}\mu = \mathcal{T}\mu(B) \\ &= \int p(x, B) d\mu(x) \\ &= \int \int \chi_B(y) p(x, dy) d\mu(x) \\ &= \int \int \phi(y) p(x, dy) d\mu(x). \end{aligned}$$

Step 2. Suppose ϕ is a simple function, that is, $\phi = \sum_{j=1}^m \alpha_j \chi_{A_j}$, with $M = \cup_{j=1}^m A_j$ and $A_j \cap A_K = \emptyset$, then

$$\begin{aligned} \int \phi d\mathcal{T}\mu &= \int \sum_{j=1}^m \alpha_j \chi_{A_j} d\mathcal{T}\mu = \sum_{j=1}^m \alpha_j \mathcal{T}\mu(A_j) \\ &= \sum_{j=1}^m \alpha_j \int p(x, A_j) d\mu(x) \\ &= \int \sum_{j=1}^m \alpha_j p(x, A_j) d\mu(x) \\ &= \int \int \phi(y) p(x, dy) d\mu(x). \end{aligned}$$

Step 3. Suppose that ϕ is a positive function. So there is a sequence of samples functions $(s_n)_n$ that $s_1 \leq s_2 \leq \dots \leq \phi$.

$$\begin{aligned} \int \phi d\mathcal{T}\mu &= \sup \left\{ \int s_n(y) d\mathcal{T}\mu : n \in \mathbb{N} \right\} \\ &= \sup \left\{ \int \int s_n(y) p(x, dy) d\mu(x) : n \in \mathbb{N} \right\} \\ &= \int \int \phi(y) p(x, dy) d\mu(x). \end{aligned}$$

Step 4. If ϕ be any function, so $\phi = \phi^+ - \phi^-$ where ϕ^+ and ϕ^- are positive functions and the prove is immediatly from the previously step.

Proposition 8.9. If $\mu_n \xrightarrow{*} \mu$ (that is μ_n converges to μ in the weak* topology), then $\mathcal{T}\mu_n \xrightarrow{*} \mathcal{T}\mu$.

Proof. For the hypothesis $\mu_n \xrightarrow{*} \mu$ implies that for any continuous bounded function $\phi : M \rightarrow \mathbb{R}$ we have

$$\lim_{n \rightarrow \infty} \int \phi d\mu_n = \int \phi d\mu.$$

By the other side

$$\int \phi d\mathcal{T}\mu_n = \int \phi(y) \int p(x, dy) d\mu_n(x) = \int \int \phi(y) p(x, dy) d\mu_n(x).$$

Afirmation. We already showed that

$$P\phi(x) = \int \phi(y) p(x, dy) = \sum_{i=1}^k p_i \phi \circ f_i(x)$$

because f_i ($\forall i = 1, \dots, k$), ϕ are continuous and ϕ is bounded we have $P\phi(y)$ also is a bounded and continuous function.

Therefore,

$$\begin{aligned} \lim \int \phi(y) d\mathcal{T}\mu_n(y) &= \lim \int \phi(y) \int p(x, dy) d\mu_n(x) \\ &= \lim \int \int \phi(y) p(x, dy) d\mu_n(x) \\ &= \lim \int P\phi(x) d\mu_n(x) \\ &= \int P\phi(x) d\mu(x) \\ &= \int \int \phi(y) p(x, dy) d\mu(x) \\ &= \int \phi(y) \int p(x, dy) d\mu(x) \\ &= \int \phi(y) d\mathcal{T}\mu(y). \end{aligned}$$

That means $\mathcal{T}\mu_n \xrightarrow{*} \mathcal{T}\mu$.

Definition 8.10. Let $\pi : \Omega \rightarrow M$ be defined by

$$\pi(\omega) = \lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(p).$$

The pushforward measure $\pi_*\mathbb{P}$ in M is defined by

$$\pi_*\mathbb{P}(A) = \mathbb{P}(\pi \in A) = (\{\omega : \pi(\omega) \in A\}).$$

8.3.1 Letac Principle

Proposition 8.11 (Letac Principle). *Consider an independent and identically distributed random product of continuous functions $f_1, \dots, f_k : M \rightarrow M$. Suppose the limit*

$$\pi(\omega) = \lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)$$

exists for \mathbb{P} -almost every point $\omega \in \Omega$ and it is independent of $p \in M$.

Then for every $\nu \in \mathcal{M}_1(M)$ we have

$$\mathcal{T}^n \nu \xrightarrow{*} \pi_* \mathbb{P}.$$

In particular, $\pi_ \mathbb{P}$ it's a stationary measure.*

Proof. We want to show that for every bounded continuous function $\phi : M \rightarrow \mathbb{R}$ is true

$$\lim_{n \rightarrow \infty} \int \phi d\mathcal{T}^n \nu = \int \phi d\pi_* \mathbb{P}.$$

Because $\pi_* \mathbb{P}$ is a pushforward measure we know that for every $B \in \mathcal{B}(M)$

$$\begin{aligned} \int_B \phi d\pi_* \mathbb{P} &= \int_{\pi^{-1}(B)} \phi(\pi(\omega)) d\mathbb{P}(\omega) \\ &= \int_{\pi^{-1}(B)} \phi\left(\lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)\right) d\mathbb{P}(\omega) \end{aligned}$$

by the hypothesis ϕ is continuous

$$\int_B \phi d\pi_* \mathbb{P} = \lim_{n \rightarrow \infty} \int_{\pi^{-1}(B)} \phi(f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)) d\mathbb{P}(\omega)$$

and

$$\int \phi d\pi_* \mathbb{P} = \lim_{n \rightarrow \infty} \int \phi(f_{\omega_0} \circ \dots \circ f_{\omega_n}(p)) d\mathbb{P}(\omega).$$

By the other side

$$\int \phi(f_{\omega_0}(x)) d\mathbb{P}(\omega) = \sum_{i=1}^k p_i \phi(f_i(x)).$$

Furthermore $P\phi(x) = \int \phi(y) p(x, dy) = \sum_{i=1}^k p_i \phi \circ f_i(x)$, that means

$$P\phi(x) = \int \phi(f_{\omega_0}(x)) d\mathbb{P}(\omega).$$

In addition,

$$\begin{aligned} P^2\phi(x) &= P(P\phi(x)) \\ &= \sum_{i=1}^k p_i P(\phi \circ f_i(x)) \\ &= \sum_{i=1}^k p_i \sum_{j=1}^k p_j \phi \circ f_i \circ f_j(x) \\ &= \sum_{i=1}^k \sum_{j=1}^k p_i p_j \phi(f_i \circ f_j(x)) \\ &= \int \phi(f_{\omega_0} \circ f_{\omega_1}(x)) d\mathbb{P}(\omega) \end{aligned}$$

And by induction

$$P^n \phi(x) = \int \phi(f_{\omega_0} \circ \dots \circ f_{\omega_n}(x)) d\mathbb{P}(\omega)$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} P^n \phi(x) &= \lim_{n \rightarrow \infty} \int \phi(f_{\omega_0} \circ \dots \circ f_{\omega_n}(x)) d\mathbb{P}(\omega) \\ &= \int \phi(\lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(x)) d\mathbb{P}(\omega) \\ &= \int \phi d\pi_* \mathbb{P}. \end{aligned}$$

Soon

$$\int \phi(x) d\mathcal{T}^n \nu(x) = \int P^n \phi(x) d\nu(x)$$

that implies

$$\begin{aligned} \lim_{n \rightarrow \infty} \int \phi(x) d\mathcal{T}^n \nu(x) &= \lim_{n \rightarrow \infty} \int P^n \phi(x) d\nu(x) \\ &= \int \lim_{n \rightarrow \infty} P^n \phi(x) d\nu(x) \quad (P^n \phi \text{ continuous and bounded}) \\ &= \int \int \phi d\pi_* \mathbb{P} d\nu(x) \quad (\int \phi d\pi_* \mathbb{P} \text{ don't depend on } x) \\ &= \int \phi d\pi_* \mathbb{P} \nu(M) \\ &= \int \phi d\pi_* \mathbb{P}. \end{aligned}$$

Finally, $\pi_* \mathbb{P}$ it's a fixed point for \mathcal{T} . In fact, by the proposition 8.9 we have that

$$\mathcal{T}^n(\pi_* \mathbb{P}) \xrightarrow{*} \pi_* \mathbb{P}$$

implies

$$\mathcal{T}^{n+1}(\pi_* \mathbb{P}) \xrightarrow{*} \mathcal{T}(\pi_* \mathbb{P}).$$

And by the uniqueness of the limit $\mathcal{T}(\pi_* \mathbb{P}) = \pi_* \mathbb{P}$.

Now suppose there are other fixed point ν , that is $\mathcal{T}(\nu) = \nu$ which implies $\mathcal{T}^n(\nu) = \nu$. For the other side

$$\nu = \lim_{n \rightarrow \infty} \mathcal{T}^n(\nu) = \pi_* \mathbb{P}.$$

Corollary Letac Principle

Let

- (M, d) be a metric space;
- $\mathcal{B}(M)$ be the Borel σ -algebra of M ;
- $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space
- $Z = \{Z_n : \Omega \rightarrow M; n \in \mathbb{N}_0\}$ be a stochastic process;
- $M^{\mathbb{N}}$ be the space of sequence of M with the σ -algebra $\mathcal{B}^{\mathbb{N}}$

Definition 8.12 (Stationary process). A stochastic process $Z = \{Z_n; n \in \mathbb{N}_0\}$ is stationary if

$$\mathbb{P}(\{\omega : (Z_0, Z_1, \dots, Z_n) \in C\}) = \mathbb{P}(\{\omega : (Z_k, Z_{1+k}, \dots, Z_{n+k}) \in C\})$$

for any $k \geq 1$, $n \geq 0$ and any $C \in \mathcal{B}^{\mathbb{N}}$.

Definition 8.13 (Invariant event). Let $Z = \{Z_n; n \in \mathbb{N}_0\}$ be a stationary process.

- $A \in \mathcal{F}$ it's **invariant** if exists $C \in \mathcal{B}^{\mathbb{N}}$ such that

$$A = \{\omega : (Z_k(\omega), Z_{k+1}(\omega), \dots) \in C\} \quad \forall k \geq 0.$$

- $A \in \mathcal{F}$ it's **almost invariant** if exists B invariant such that

$$\mathbb{P}(A \Delta B) = 0.$$

Definition 8.14 (Ergodicity). A stationary sequence $Z = \{Z_n; n \in \mathbb{N}_0\}$ is ergodic if every invariant event has probability one or zero.

Theorem 8.15. Suppose $\{X_n : n \geq 0\}$ is a stationary Markov process on a state space $(M, \mathbf{B}(M))$, with a transition probability p and a invariant initial distribution m .

The process is ergodic if and only if there does not exist an invariant distribution $\nu \neq m$ such that $\nu \ll m$.

Proof. The proof can be founded in the book "Stationary Processes and Discrete Parameter Markov Processes, Bhattacharya, R., Waymire, E. C. (2022)", Theorem 16.4 page 218.

Lemma 8.16. Let $Z = \{Z_n; n \in \mathbb{N}_0\}$ be a Markov Chain with transition probability $p : M \times \mathcal{B}(M) \rightarrow \mathbb{R}$, and a unique stationary measure m . If Z_0 has distribution m , then Z it's a stationary and ergodic process.

Proof. (Stationary Process)

Affirmation.

$$\mathbb{P}(Z_i \in B_i, 0 \leq i \leq n) = \int_{B_0} dZ_0 * \mathbb{P} \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_{n-1}, dx_n).$$

In fact,

- Let's see for $\mathbb{P}(Z_0 \in B_0, Z_1 \in B_1)$.

$$\begin{aligned} \mathbb{P}(Z_0 \in B_0, Z_1 \in B_1) &= \int \chi_{Z_1^{-1}(B_1)} \chi_{Z_0^{-1}(B_0)} d\mathbb{P} \\ &= \int_{Z_0^{-1}(B_0)} \chi_{Z_1^{-1}(B_1)} d\mathbb{P} \\ &= \int_{Z_0^{-1}(B_0)} \chi_{Z_1^{-1}(B_1)} d\mathbb{P} \\ &= \int_{Z_0^{-1}(B_0)} p(Z_0, B_1) d\mathbb{P} \quad (\text{page 15}) \\ &= \int_{B_0} p(x_0, B_1) d\mathbb{P} * Z_0 \quad (\text{change of variables}) \\ &= \int_{B_0} \int_{B_1} p(x_0, dx_1) d\mathbb{P} * Z_0. \end{aligned}$$

Using the affirmation we can show that $(Z_n)_n$ it's stationary.

$$\begin{aligned} \mathbb{P}(Z_i \in B_i, 0 \leq i \leq n) &= \int_{B_0} dZ_0 * \mathbb{P} \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_{n-1}, dx_n) \\ &= \int_{B_0} \int_{B_1} \cdots \int_{B_n} p(x_{n-1}, dx_n) \cdots p(x_0, dx_1) dZ_0 * \mathbb{P} \\ &= \int_{B_0} \int_{B_1} \cdots \int_{B_n} p(x_{n-1}, dx_n) \cdots p(x_0, dx_1) dm \\ &= \int_{B_0} \int_{B_1} \cdots \int_{B_n} p(x_{n-1}, dx_n) \cdots p(x_0, dx_1) dZ_k * \mathbb{P} \\ &= \mathbb{P}(Z_{i+k} \in B_i, 0 \leq i \leq n) \end{aligned}$$

(Ergodic Process) Immediate of the theorem 8.15.

Theorem 8.17 (Birkhoff's Ergodic Theorem). *Let $\{X_n : \Omega \rightarrow M, n \geq 0\}$ be a stationary sequence on the state space M and $f : M \rightarrow \mathbb{R}$ an integrable function, then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} f(X_m(\omega))$$

exists for almost every $\omega \in \Omega$.

If $\{X_n\}$ is ergodic then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} f(X_m(\omega)) = \int f dm$$

for almost every $\omega \in \Omega$.

Corollary 8.18 (Law of the Large Numbers). *Consider a compact metric space M and an i.i.d random product of contractions*

$$f_i : M \rightarrow M \quad \forall i = 1, \dots, k$$

Then the stationary measure m satisfies:

1. *the support of m it's the Hutchinson attractor $\text{supp}(m) = K$.*
2. *For every ϕ continuous and bounded exists some subset C of $\{1, \dots, k\}^{\mathbb{N}}$ with full-bernoulli measure such that*

$$\lim_{n \rightarrow \infty} \phi(f_{s_i} \circ \dots \circ f_{s_0}(x)) = \int \phi dm$$

for all $s = (s_n)_n \in C$ and $x \in M$.

Proof. Consider the Markov Chain $Z_n : \Omega \rightarrow M$

$$Z_n(\omega) := f_{X_{n-1}(\omega)} \circ \dots \circ f_{X_0(\omega)}(Z_0(\omega)) = f_{X_{n-1}}(Z_{n-1}).$$

where $X = (X_n)_{n \in \mathbb{N}_0}$ an independent stochastic process on Ω with values on a finite set $\{1, \dots, k\}$, such that each X_n has the same distribution $\nu = p_1 \delta_1 + \dots + p_k \delta_k$.

Follows from Letac Principle that exists only one stationary measure $m = \pi * \mathbb{P}$, where

$$\pi(\omega) = f_{\omega_0} \circ \dots \circ f_{\omega_{n-1}}(p).$$

For this Markov Chain we have

- $\text{supp}(\pi * \mathbb{P}) = K$.

In fact, remember that

$$K = \lim_{l \rightarrow \infty} \overline{\{f_{\omega}^n(x) : n \geq l\}},$$

with $f_{\omega}^n(x) = f_{\omega_{n-1}} \circ \dots \circ f_{\omega_0}(x)$.

For every ω we have $\pi(\omega) \in K$, and by definition $y \in \text{supp}(m)$ if $\mu(V) > 0$ for any neighborhood V of y , and this happen only if $y \in K$.

Suppose the opposite, that is exists $V, y \in V$ such that $V \cap K = \emptyset$ and $m(V) > 0$.

We also have

$$m(K) = \mathbb{P}(\{\omega \in \Omega : \pi(\omega) \in K\}) = 1,$$

so that implies

$$m(K \cap V) = m(K) + m(V) > 1,$$

but $m(A) < 1$ for all $A \in \mathcal{B}(M)$.

On the other way, take $y \in K$, and a neighborhood V of y , then by the π continuity $\pi^{-1}(V)$ is a open subset of Ω , that means exists a cylinder $C_{i, a_0, \dots, a_k} \subset \pi^{-1}(V)$.

Because $\mathbb{P}(C_{i, a_0, \dots, a_k}) > 0$, then $\mathbb{P}(\pi^{-1}(V)) = m(V) > 0$.

Hence $\text{supp}(\pi * \mathbb{P}) = K$.

- Suppose the Markov Chain $\{Z_n : n \geq 0\}$ the element Z_0 has distribution $m = \pi * \mathbb{P}$ (the stationary measure by the Letac Principle), then this Markov Chain it's a ergodic and stationary process.

Let $g : M \rightarrow \mathbb{R}$ be an integrable function, so follows from Birkhoff's Ergodic Theorem

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(Z_i(\omega)) = \int g dm,$$

for \mathbb{P} -almost every $\omega \in \Omega$.

That is, exists $\Omega_g \subset \Omega$ with full-measure such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(Z_i(\omega)) = \int g dm,$$

for all $\omega \in \Omega_g$.

Now consider the bounded continuous functions $\phi : M \rightarrow \mathbb{R}$, $|\phi(x)| \leq L$ for all $x \in M$.

We want to prove that:

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} : n > n_0 \implies \left| \frac{1}{n} \sum_{i=0}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \int \phi dm \right| < \varepsilon.$$

For that, let $C_\phi = (X_0, X_1, \dots)(\Omega_\phi)$ because Ω_ϕ has full measure then C_ϕ has full measure. Moreover, if $c \in C_\phi$ that means $c = (X_0, X_1, \dots)(\omega)$ for some $\omega \in \Omega_\phi$ and

$$\frac{1}{n} \sum_{i=0}^{n-1} \phi(Z_i(\omega)) \rightarrow \int \phi dm$$

Furthermore, take $c \in C_\phi$ we get

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=0}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \int \phi dm \right| &\leq \left| \frac{1}{n} \sum_{i=0}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \frac{1}{n} \sum_{i=0}^{n-1} \phi(Z_i(\omega)) \right| \\ &\quad + \left| \frac{1}{n} \sum_{i=0}^{n-1} \phi(Z_i(\omega)) - \int \phi dm \right|. \end{aligned}$$

We already know the last part converges, that is

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} \phi(Z_i(\omega)) - \int \phi dm \right| \rightarrow 0 \quad (n \rightarrow \infty)$$

On the other side

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=0}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \frac{1}{n} \sum_{i=0}^{n-1} \phi(Z_i(\omega)) \right| &= \left| \frac{1}{n} \sum_{i=0}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \phi(Z_i(\omega)) \right| \\ &\leq \left| \frac{1}{n} \sum_{i=0}^{k-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \phi(Z_i(\omega)) \right| \\ &\quad + \left| \frac{1}{n} \sum_{i=k}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \phi(Z_i(\omega)) \right| \\ &\leq \left| \frac{2Lk}{n} \right| + \left| \frac{1}{n} \sum_{i=k}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \phi(Z_i(\omega)) \right| \end{aligned}$$

For a sufficiently large n we have

$$\left| \frac{2Lk}{n} \right| \rightarrow 0 \quad (n \rightarrow \infty).$$

Moreover

$$\left| \frac{1}{n} \sum_{i=k}^{n-1} \phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \phi(Z_i(\omega)) \right| \leq \frac{1}{n} \sum_{i=k}^{n-1} |\phi(f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)) - \phi(Z_i(\omega))|.$$

Because ϕ it's uniformly continuous in M , and that implies for every $\varepsilon > 0$ exists $\delta > 0$:

$$d(x, y) < \delta \implies |\phi(x) - \phi(y)| < \varepsilon.$$

On the other side

$$d(f_c^i(x), Z_i(\omega)) < \lambda^i d(x, Z_0(\omega)),$$

with $\lambda = \max\{\lambda_j : j = 1, \dots, k\}$, and λ_j the factors of the contractions f_j and $f_c^i(x) = f_{c_{i-1}} \circ \dots \circ f_{c_0}(x)$.

Because $\lambda \in (0, 1)$ taking i sufficiently large we get

$$\lambda^i < \frac{\delta}{d(x, Z_0(\omega))} \implies d(f_c^i(x), Z_i(\omega)) < \delta \implies |\phi(f_c^i(x)) - \phi(Z_i(\omega))| < \varepsilon.$$

Then

$$\left| \frac{1}{n} \sum_{i=k}^{n-1} \phi(f_c^i(x)) - \phi(Z_i(\omega)) \right| \leq \frac{1}{n} \sum_{i=k}^{n-1} |\phi(f_c^i(x)) - \phi(Z_i(\omega))| \rightarrow 0 \quad (n \rightarrow \infty).$$

Finally, because (M, d) is compact then the space of continuous and bounded function ($C_0(M)$) is separable, that is, exists $\{\phi_m : m \in \mathbb{N}\}$ such that $\overline{\{\phi_m\}} = C_0(M)$. So it's enough that $C = \cap C_{\phi_n}$ and approximate every continuous and bounded function for some function in $\{\phi_m : m \in \mathbb{N}\}$.

9 Random Power Series

The main result for this part uses all the theory developed until this point to give us a way to know how frequently an IFS goes through a set. The proof of this is done specifically for intervals I , and it can be mathematically seen as

$$\mathbb{P}\left(\sum_{i=0}^{\infty} \pm \lambda^n \in I\right) = \text{frequency with which } f_{\omega}^n(x) \text{ goes through } I,$$

where $\lambda \in (0, 1)$. Meanwhile, the more general result can be stated in a way similar to von Mises' definition of probability - the one using an event's repetition:

Theorem 9.1. For all “well-behaved” sets A , it follows that

$$\mathbb{P}(\pi \in A) = \lim_{n \rightarrow \infty} \frac{1}{n} \# \left\{ i : f_{\omega}^i(x) \in A, 0 \leq i \leq n-1 \right\}.$$

Remark 9.2 (von Mises' Definition of Probability). If n_A is the number of times event A occurs in n independent repetitions of a given “experiment”, then

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}. \quad \square$$

Using the remark, we may give the following interpretation for the theorem - if A is the event “ $f_{\omega}^i(x)$ passes through A ”, then n_A would be the number of times that the coding map's terms have passed through A after n repetitions, and thus

$$\mathbb{P}(\pi \in A) = \lim_{n \rightarrow \infty} \frac{1}{n} \# \left\{ i : f_{\omega}^i(x) \in A, 0 \leq i \leq n-1 \right\} = \lim_{n \rightarrow \infty} \frac{n_A}{n}.$$

Leaving that aside, let's start the proof of the theorem's special case of intervals, stated as

Theorem 9.3. For every interval I , we have

$$\mathbb{P}\left(\sum_{n=0}^{\infty} \pm \lambda^n \in I\right) = \text{frequency with which } f_{\omega}^n(x) \text{ goes through } I,$$

where $f_{\omega}^n(x)$ is the **random orbit** of the contraction systems $f_i(x) = \lambda x + i, i \in \{-1, 1\}$.

Remark 9.4. The **random orbit** consists of all the many ways to combine values which $\omega_i, 0 \leq i \leq n$, can assume in an iteration $f_{\omega}^n(x) = f_{\omega_n} \circ f_{\omega_{n-1}} \circ \dots \circ f_{\omega_0}(x), \omega_i \in \{-1, 1\}$.

Proof. Given the data in the theorem's statement, our system of functions is of the form $\{f_1, f_{-1}\}$, i.e., $\omega \in \{-1, 1\}^{\mathbb{N}}$. Consider the coding map

$$\pi(\omega) = \lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(p),$$

which is known to be independent of p to us. We start by proving that, given $\lambda \in (0, 1)$,

$$\pi(\omega) = \sum_{n=0}^{\infty} \omega_n \lambda^n.$$

In other words, we are going to show that the coding map can be seen as a random power series with respect to the λ , and what makes this power series a *random* power series is the fact that each ω_n is random.

In fact, given that $f_i(x) = \lambda x + i, i \in \{-1, 1\}$, the following is true for $n \in \mathbb{N}$:

- $f_{\omega_n}(x) = \lambda x + \omega_n$
- $f_{\omega_{n-1}} \circ f_{\omega_n}(x) = \lambda(f_{\omega_n}(x)) + \omega_{n-1} = \lambda(\lambda x + \omega_n) + \omega_{n-1} = \lambda^2 x + \lambda \omega_n + \omega_{n-1}$
- $f_{\omega_{n-2}} \circ f_{\omega_{n-1}} \circ f_{\omega_n}(x) = \lambda(\lambda^2 x + \lambda \omega_n + \omega_{n-1}) + \omega_{n-2} = \lambda^3 x + \lambda^2 \omega_n + \lambda \omega_{n-1} + \omega_{n-2}$
- \vdots
- $f_{\omega_0} \circ \dots \circ f_{\omega_n}(x) = \lambda^{n+1} x + \lambda^n \omega_n + \dots + \omega_0 = \lambda^{n+1} x + \sum_{i=0}^n \lambda^i \omega_i.$

Taking the limit as $n \rightarrow \infty$, we find out that

$$\pi(\omega) = \lim_{n \rightarrow \infty} f_{\omega_0} \circ \dots \circ f_{\omega_n}(x) = \lim_{n \rightarrow \infty} \left(\lambda^{n+1} x + \sum_{i=0}^n \lambda^i \omega_i \right) = \lim_{n \rightarrow \infty} \lambda^{n+1} x + \lim_{n \rightarrow \infty} \sum_{i=0}^n \lambda^i \omega_i.$$

Because $\lambda \in (0, 1)$, λ^{n+1} goes to 0, so that

$$\pi(\omega) = \lim_{n \rightarrow \infty} \lambda^{n+1} x + \lim_{n \rightarrow \infty} \sum_{i=0}^n \lambda^i \omega_i = 0 + \lim_{n \rightarrow \infty} \sum_{i=0}^n \lambda^i \omega_i = \sum_{n=0}^{\infty} \lambda^n \omega_n.$$

Thus,

$$\pi(\omega) = \sum_{n=0}^{\infty} \lambda^n \omega_n.$$

According to the Letac Principle, the pushforward of the coding map is the unique stationary measure, given by

$$\pi_* \mathbb{P} = \mathbb{P}(\pi \in A) = \mathbb{P}\left(\left\{\omega \in \{-1, 1\}^{\mathbb{N}} : \sum \omega_n \lambda^n \in I\right\}\right)$$

Next, our goal is to prove that a stationary measure has no atoms. This is related to the “well-behaved” part of the statement since, as long as we have shown that it indeed has no atoms, then the boundary will have null measure, which is exactly what we wanted.

Before moving to that, recall that an atom is a singleton set $\{x\}$ such that $\mu(\{x\}) > 0$. To prove that a stationary measure has no atoms, thus, it suffices to show that any singleton has null measure. On that note, to make notation easier, let μ be the stationary measure. Put K as the Hutchinson attractor and

define the function $M : K \rightarrow \mathbb{R}_{\geq 0}$, $M(a) = m(\{a\})$. Taking a closer look at M, because of the way it is defined, it follows that given any $a \in K$,

$$M(a) = m(\{a\}) = \pi_*\mathbb{P}(\{a\}) = \mathbb{P}(\pi \in \{a\}) = \mathbb{P}(\{\omega : \pi(\omega) \in \{a\}\}) < \infty.$$

Moreover, the continuity from above of m is inherited by M. Hence, given an increasing sequence $a_n \uparrow a \in K$, it follows that $M(a_n) \uparrow M(a)$. In particular, continuity from above also implies semi-continuity from above. This is important to prove that M attains its maximum in K. As a matter of fact, using the semi-continuity from above and the fact that K is a compact set, it follows readily that M indeed reaches its maximum in K, i.e., there exists a $q \geq 0$ such that $q \geq M(x)$ for all $x \in K$ and $M(x_0) = q$ for some $x_0 \in K$.

Now let q be the maximum of M in K and consider the set of all singletons of elements of K whose measures are equal to q :

$$B = \{a \in K : m(\{a\}) = q\}.$$

If $q > 0$, then because we are in a probability space, $m(\{a\}) = \mathbb{P}(\pi \in \{a\}) \leq 1$, so that only finitely many a 's in K can be mapped to q , meaning B is a finite set. This will be important to find a contradiction and conclude that $q = 0$.

Firstly, notice that

$$f_i(x) = \lambda x + i \Rightarrow f_i(x) - i = \lambda x \Rightarrow x = \frac{f_i(x) - i}{\lambda},$$

so that $f_i^{-1}(x) = \frac{x-i}{\lambda}$. Hence,

$$q = m\{a\} = \sum p_j m(\{f_i^{-1}(a)\}) \leq \sum p_i q = q \sum p_j = q,$$

meaning that $m(\{f_i^{-1}(a)\}) = q$ for all i , contradicting the finiteness of B. Therefore, $q = 0$, singletons are all zero measured and there are no atoms in this measure space.

In conclusion, for any interval I formed by real number a and b , its boundary is $\partial I = \{a, b\}$. However, notice that

$$\{a, b\} = \{a\} \cup \{b\} \quad \& \quad \{a\} \cap \{b\} = \emptyset.$$

Consequently, $\pi_*\mathbb{P}(\partial I) = \pi_*\mathbb{P}(\{a\} \cup \{b\}) = \pi_*\mathbb{P}(\{a\}) + \pi_*\mathbb{P}(\{b\}) = 0 + 0 = 0$. □