Ergodic Theory and the Logistic Map

Universidad de Zaragoza
Facultad de Ciencias, *Grado en Matemáticas*

Autor:
Juan Manuel López Jarque

Directores:
Luciano Abadías Ullod
Carlos Gómez Ambrosi

Noviembre 2024

Resumen

El principal objetivo del presente trabajo es probar que la aplicación logística de Ulam T_4, dada por

$$x \mapsto 4x(1-x),\ \ x \in [0,1],$$

es ergódica.

Comenzaremos dando cuenta del comportamiento asintótico de las trayectorias (u órbitas) para el sistema dinámico discreto definido por la aplicación logística $T_r\colon [0,1] \to [0,1]$, dada por

$$T_r(x) = rx(1-x),$$

en función del valor del parámetro $r \in [0,4]$. Para ello, primeramente realizaremos una serie de simulaciones numéricas, tales como representar de manera gráfica el comportamiento asintótico de algunas órbitas, así como el cálculo del diagrama de órbita (o diagrama de Feigenbaum) para la familia logística $\{T_r\}_{r\in[0,4]}$. A través de estos experimentos, clasificaremos las funciones de dicha familia principalmente en dos grupos: hablaremos de aplicaciones regulares y de aplicaciones estocásticas. Este hecho se formaliza en el teorema regular o estocástico de Lyubich (ver [6]).

En aras de cumplir con nuestro objetivo, en el Capítulo 1 realizamos otro experimento en el que calculamos un histograma que refleja la distribución de casi todas las órbitas para la aplicación de Ulam T_4.

El razonamiento llevado a cabo en este experimento se formalizará en el marco de la teoría de la medida, y en particular dentro del marco de la teoría ergódica y el teorema ergódico de Birkhoff, que nos ayudarán a estudiar y entender el comportamiento a largo plazo y en promedio de casi todas las órbitas para la aplicación de Ulam.

Muy relacionado con lo anterior es el hecho de que la aplicación de Ulam T_4 es topológicamente transitiva y de que casi todas sus trayectorias son densas en el intervalo unidad $[0,1]$. Para probar dichas afirmaciones tendremos que recurrir a otro sistema dinámico discreto que está estrechamente relacionado con la aplicación T_4: la función de duplicación $S\colon [0,1] \to [0,1]$, dada por

$$S(x) = \mathrm{Frac}(2x),$$

donde el operador $\mathrm{Frac}(x)$ nos proporciona la parte decimal de un número $x \in \mathbb{R}$. Probaremos que S es topológicamente transitiva y que casi todas sus órbitas son densas en $[0,1]$. Para ello, dotaremos a S de una nueva interpretación: pasando a la representación binaria de los números reales $x \in [0,1]$. La relación entre T_4 y S se formalizará con la noción de semiconjugación topológica.

Hemos hablado acerca de la teoría ergódica, la cual puede verse como una aplicación de la teoría de la medida al estudio del comportamiento promedio a largo plazo de los sistemas dinámicos.

Las principales nociones serán las de medida invariante con respecto a una aplicación, y la de aplicación ergódica con respecto a una medida invariante. Además, la principal idea de la teoría ergódica será el teorema ergódico de Birkhoff, según el cual, con probabilidad uno, la media de una función integrable a lo largo de una órbita de una aplicación ergódica es igual a la integral de dicha función. Desarrollaremos todas las nociones y resultados necesarios para probar dicho teorema en el Capítulo 2 del presente trabajo.

Por último, en el Capítulo 3, probaremos que la aplicación de Ulam T_4 es ergódica con respecto a una medida de probabilidad μ. Para ello, nuevamente, nos serviremos del hecho de que T_4 será topológicamente semiconjugada con S, y de que S será ergódica con respecto a la medida de Lebesgue. Además, probaremos que la función de densidad ρ, asociada a la medida de probabilidad μ, estará dada por la siguiente expresión analítica:

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \ x \in (0,1).$$

Casi todas las órbitas de T_4 se distribuirán con arreglo a esta función de densidad. Finalmente, explicaremos como hacer un bosquejo de la gráfica de ρ utilizando el teorema ergódico de Birkhoff.

La principal fuente utilizada para la realización del presente trabajo ha sido el libro "Computational Ergodic Theory" de Choe (ver [3]).

Contents

Chapter 1

Introduction to the Logistic Family

The present work is framed within the theory of *discrete dynamical systems,* which studies the iteration of a function of a set into itself. In our case, the set will be the unit interval $[0, 1]$, which we will denote by I, and the function will be the so-called *logistic map* $T_r \colon I \to I$, given by

$$T_r(x) = rx(1 - x),$$

where the values of the parameter r are restricted to the interval $[0, 4]$, so that $T_r(x) \in I$ for every $x \in I$.

When we iterate the logistic map starting at an initial value $x_0 \in I$, we obtain what is called the *orbit* of x_0, i.e. the set of points $x_0, x_1, x_2, \ldots, x_n, \ldots$, where $x_n = T_r(x_{n-1})$ for $n \geq 1$.

We will adopt a *measure-theoretic* viewpoint, in the sense that we are interested in describing the *asymptotic behavior* of *almost all* orbits of T_r for *almost any* parameter value $r \in [0, 4]$.
The terms "almost all" and "almost any" mean that the corresponding assertions are true except on a set of Lebesgue measure zero.[1] We will denote by $\mathcal{B}$ the Borel σ-algebra over I and by λ the Lebesgue measure on $(I, \mathcal{B})$.

1.1 Orbit Diagram

We will say that the orbit of $x_0 \in I$ is *periodic* if there exists $n \geq 1$ such that $x_n = x_0$. The least positive integer n for which $x_n = x_0$ is called the *period* of the orbit. In this case, the orbit of x_0 is just the set of points $x_0, x_1, x_2, \ldots, x_{n-1}$, and we will say that the *lenght* of the orbit is n, and that the orbit is a *n-cycle*.

Let's start by doing some experiments. In Figures 1.1 and 1.2 we show graphically the asymptotic behavior of four orbits for different r-values and different initial conditions $x_0 \in I$.
On the horizontal axis, the number of iterations n is marked, while on the vertical axis, the iterates x_n are given for each n. The points (n, x_n) are connected by line segments to enhance visibility.

In the first picture, we observe that, after a *transient phase,* the orbit tends to a 1-cycle (or *fixed point*), while in the second and third, also after a transient, the orbit leads to a 2-cycle and a 3-cycle, respectively (*periodic points*). In contrast, the last image exhibits an irregular pattern all the time (*chaotic behavior*).

[1] Both the parameter interval $[0, 4]$ and the unit interval I are endowed with the Lebesgue measure.

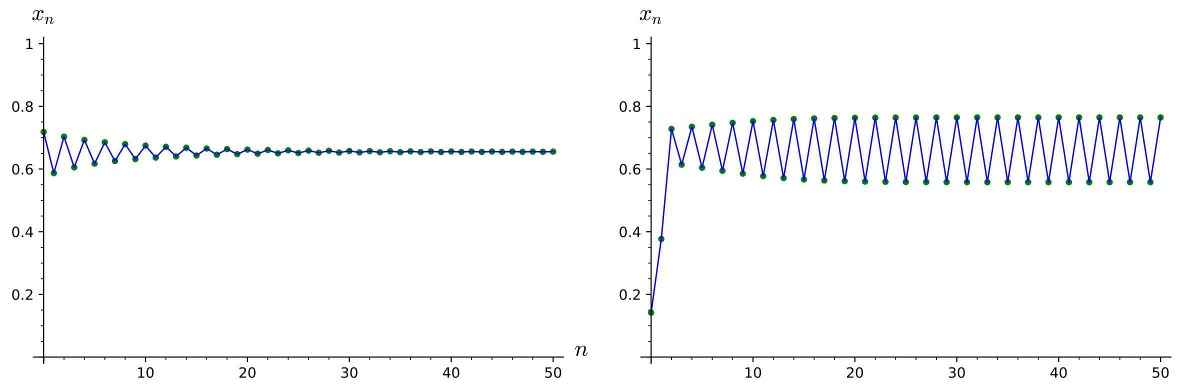

Figure 1.1: Orbits for $r = 2.9$ and $x_0 = e - 2$ (left) and $r = 3.1$ and $x_0 = \pi - 3$ (right).

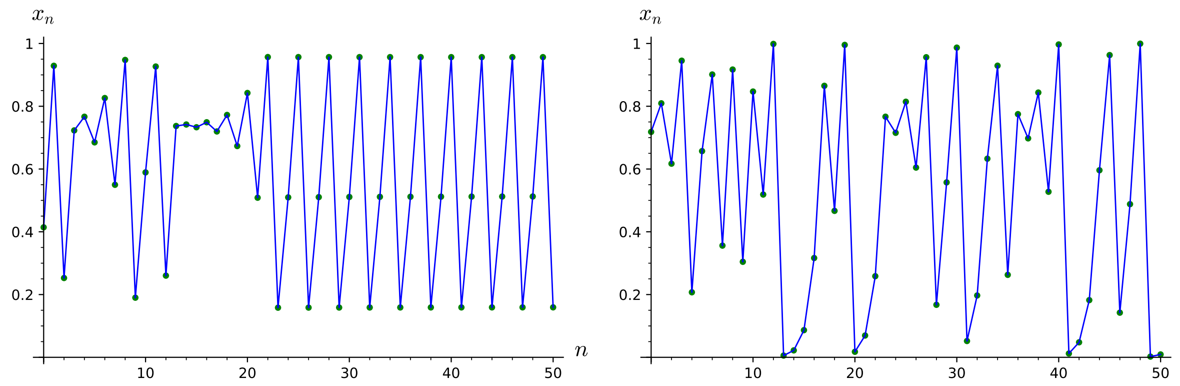

Figure 1.2: Orbits for $r = 1 + \sqrt{8}$ and $x_0 = \sqrt{2} - 1$ (left) and $r = 4$ and $x_0 = e - 2$ (right).

Experiments of this type show that the asymptotic behavior of the orbits depends on the parameter value $r \in [0, 4]$, but it does not seem to depend substantially on the initial condition $x_0 \in I$. Further, we observe that there are r-values for which the orbits seem to *stabilize* in a cycle, while there are others for which the orbits show an apparently chaotic behavior.

Hence, we carry out another experiment which consists in ploting, for each parameter value $r \in [0, 4]$ and starting at a randomly chosen initial value $x_0 \in I$, the part of the orbit which we obtain after disregarding a transient phase, thus getting the so-called *orbit diagram* (or *bifurcation diagram*, or *Feigenbaum diagram*) which we show in Figure 1.3.

More precisely, the horizontal axis denotes the r-values, and the vertical axis represents the unit interval I. For each of 1001 equally spaced r-values[2] the iterations $x_1, x_2, \ldots, x_N$, say $N = 1000$, are computed, but we drop the first 801 iterations $x_0, x_1, \ldots, x_{800}$. This eliminates the transient, so that the dependence on the initial choice x_0 is diluted to almost zero. Only the remaining 200 iterations $x_{801}, x_{802}, \ldots, x_N$ are plotted in the image.

Therefore, given a parameter value $r \in [0, 4]$, if there exists a set of points to which almost all orbits of T_r converge, what we will informally call the *attractor*, this experiment will be able to locate it.

[2] In the orbit diagrams of Figure 1.3, we have considered the parameter values $r_k = 4\frac{k}{1000}$ (left) and $r_k = 3 + \frac{k}{1000}$ (right) for $k = 0, 1, \ldots, 1000$.

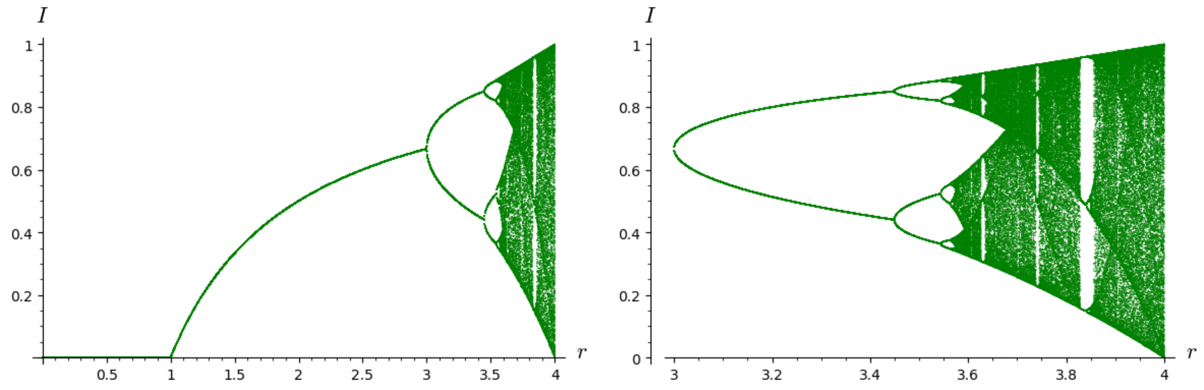

Figure 1.3: Orbit diagrams of T_r for $0 \leq r \leq 4$ (left) and $3 \leq r \leq 4$ (right).

1.1.1 Regularity

Let's analyze the orbit diagram a bit more in depth. If we look at the left-hand side of the orbit diagrams in Figure 1.3, we can see that there is a lot of *regularity*, in the sense that the asymptotic behavior of the orbits is quite simple. We will say that a parameter value $r \in [0, 4]$ is *regular* (see [6]) if there exists an *attracting cycle* to which almost all orbits of T_r converge. This attracting cycle is unique. Thus, for regular values, the attractor is the attracting cycle.

As examples of regularity, we have that $x = 0$ is an attracting fixed point for $r \in [0, 1]$ and that $x = 1 - \frac{1}{r}$ is an attracting fixed point for $r \in (1, b_1]$, where $b_1 = 3$. In Figure 1.4 we can also observe an attracting 2-cycle for $r \in (b_1, b_2]$, where $b_2 = 1 + \sqrt{6} \approx 3.449489$, an attracting 4-cycle for $r \in (b_2, b_3]$, where $b_3 \approx 3.544090$, and an attracting 8-cycle for $r \in (b_3, b_4]$, where $b_4 \approx 3.564407$. In general, at $r = b_k$ a new attracting 2^k-cycle appears. These r-values are called *bifurcation points*. Hence, we have an increasing sequence of parameter values b_k, which tends to the so-called *Feigenbaum point* $b_\infty \approx 3.569945$[3] (see [7]). The region between $r = 0$ and $r = b_\infty$ is called the *period-doubling tree* of the orbit diagram.

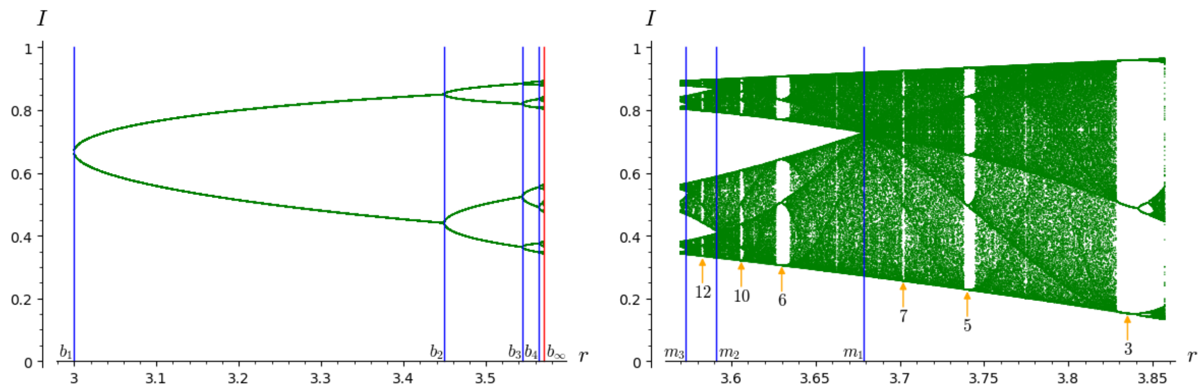

Figure 1.4: Orbit diagrams of T_r for $3 \leq r \leq b_\infty$ (left) and $b_\infty \leq r \leq 3.857$ (right) along with some notable r-values and some periodic windows.

At the right-hand side of the Feigenbaum point not everything is chaotic. For example, the

[3]The Feigenbaum point is neither regular nor stochastic (see Section 1.1.2). Furthermore, the attractor for this r-value is a Cantor set: we may have some intuition about this as the branches of the period-doubling tree fork at each b_k, but understanding the dynamics at this point is beyond the scope of the present work.

parameter value $r = 1 + \sqrt{8} \approx 3.828427$ is a regular value at which an attracting 3-cycle is born, in the sense that there is no 3-cycle (attractor or not) for any smaller value of r. In Figure 1.4, we oberve an attracting 5-cycle, 7-cycle, 6-cycle, 10-cycle and 12-cycle. The regions of the orbit diagram where chaos seems to be interrupted by attracting cycles are called *periodic windows*. Further, we observe that at $r = m_1 \approx 3.678573$ the orbit diagram splits into 2 parts (from right to left), at $r = m_2$ it splits into 4 parts, and at $r = m_3$ it splits into 8 parts. In general, at $r = m_k$ the orbit diagram splits into 2^k parts. These r-values are called *band-merging points*. Hence, we have a decreasing sequence of parameter values m_k, which also leads to the Feigenbaum point $m_\infty = b_\infty$ (see [7]).

1.1.2 Stochasticity

Now, we ask what happens to the orbits for apparently chaotic r-values such as $r = 4$. For this parameter value, T_4 is the so-called *Ulam logistic map* (see Figure 1.7). Since we are interested in knowing the asymptotic behavior of the orbits, we wonder whether we can give a satisfactory answer to questions such as what is the probability that the orbit $x_0, x_1, x_2, \dots, x_n = T_4(x_{n-1}), \dots,$ (*in the long term* and *on average*) lies in the interval $[0.68, 0.69]$.

To answer this question, we do another experiment which consists of seeing how the points of the orbit of x_0 are distributed in the unit interval I. Thus, we compute a histogram reflecting which parts of the unit interval are visited by the orbit and how often (see Figure 1.5).

To this end, we pick an initial point $x_0 \in I$ at random and iterate the Ulam map, say $N = 10^6$ times. Now, we divide the unit interval I into a large number of small subintervals, say $M = 1000$, given by

$$I_k = \left(\frac{k-1}{M}, \frac{k}{M}\right], \ k = 1, 2, \dots, M.$$

Then, we count how many of the iterates $x_0, x_1, x_2, \dots, x_N$ fall into each interval I_k. Let this number be n_k. In other words, n_k is the number of events in I_k, i.e. the absolute frequency for the interval I_k. Thus, noting that the lenght of the truncated orbit of x_0 is $N + 1$, we may consider the associated relative frequencies, given by

$$f_k = \frac{n_k}{N+1}, \ k = 1, 2, \dots, M.$$

Moreover, the relative frequency f_k can be interpreted as a probability: it is the probability that we guess correctly (whitout calculations) the interval I_k into which a point falls, randomly chosen from the $N + 1$ points of the truncated orbit of x_0.

Therefore, the area of the k^{th} column of the histogram will be f_k, and since the width of each column is $\frac{1}{M}$ (the lenght of I_k), we conclude that the height of the k^{th} column will be given by

$$\rho_k = M f_k = M \frac{n_k}{N+1}.$$

In Figure 1.5 we see a distribution for the ρ_k which is symmetric with respect to $x = \frac{1}{2}$, and which is rather flat in the center while having steep boundary spikes at $x = 0$ and $x = 1$. This means that, during the course of the iteration, the probability that we see a point of the truncated orbit near $x = 0$ or $x = 1$ is comparatively much higher than that of seeing it in the center of the unit interval I.

Running the same experiment again for different initial values $x_0 \in I$, results in histograms which are indistinguishable from the one above. As we increase the number M of subintervals and the number N of iterations, the effect is a smoothing of the shape of the columns of the histogram.

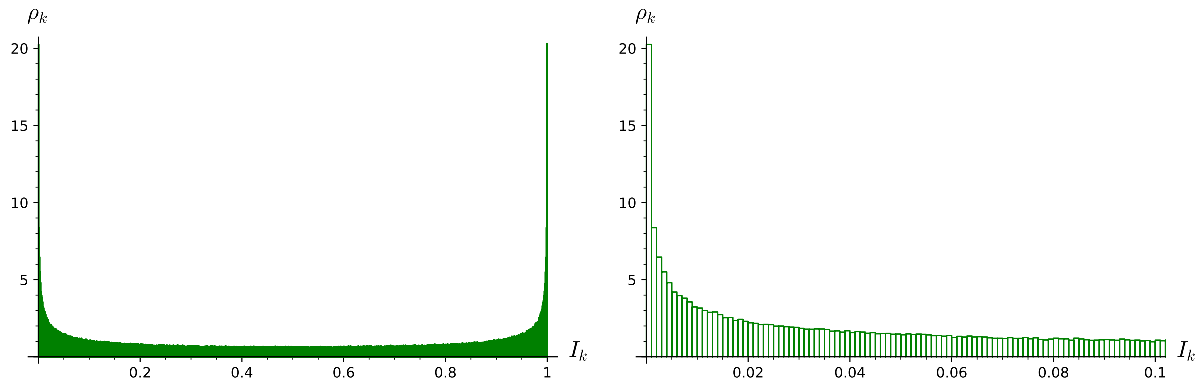

Figure 1.5: Histogram for the distribution of the truncated orbit $x_0, x_1, x_2, \ldots, x_N$ in the unit interval I.

In the limit, we would approximate a well-known curve (see Figure 1.6): the probability density function[4] of the *arcsine distribution*, given by

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \quad x \in (0,1).$$

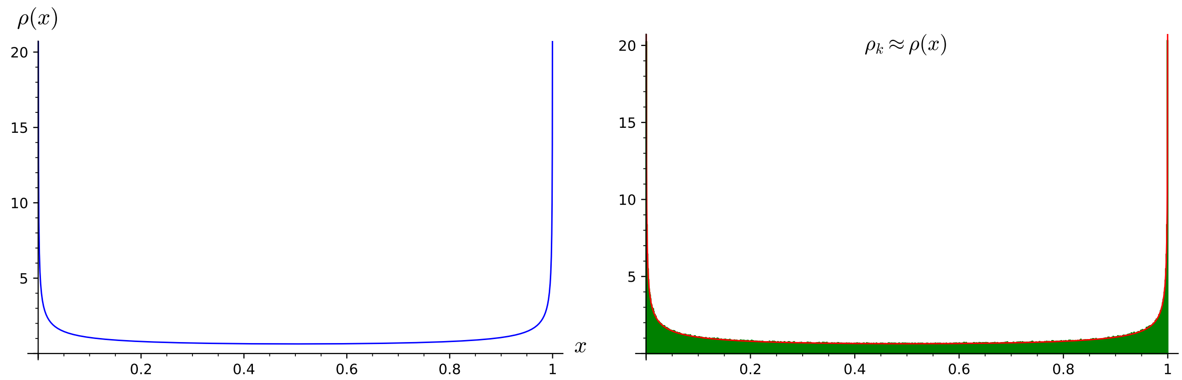

Figure 1.6: $y = \rho(x)$ (left) and the histogram approximation to the graph of ρ (right).

Let's now calculate the arcsine distribution function F:

$$F(x) = \int_0^x \rho(t)\,\mathrm{d}t = \int_0^x \frac{\mathrm{d}t}{\pi\sqrt{t(1-t)}} = \frac{2}{\pi}\int_0^{\sqrt{x}} \frac{\mathrm{d}y}{\sqrt{1-y^2}} = \frac{2}{\pi}\arcsin(\sqrt{x}), \quad x \in [0,1],$$

where we have used the change of variables $y = \sqrt{t}$, $0 \le t \le x$.

Therefore, the probability that the orbit $x_0, x_1, x_2, \ldots, x_n, \ldots$ (in the long-term and on average) lies in the interval $[0.68, 0.69]$ is exactly

$$F(0.69) - F(0.68) = \frac{2}{\pi}\left(\arcsin(\sqrt{0.69}) - \arcsin(\sqrt{0.68})\right) \approx 0.0068527.$$

[4]Only for $r = 4$ an analytical expression for the probability density function is known, albeit for other values such as $r = m_1$ there also seems to be one.

On the other hand, we can also estimate this probability summing the corresponding relative frequencies f_k, i.e. since

$$[0.68, 0.69] = \bigcup_{k=681}^{690} I_k,^{5}$$

we conclude that the probability that a randomly chosen point of the truncated orbit $x_0, x_1, x_2, \ldots, x_N$ belongs to the interval $[0.68, 0.69]$ is

$$\sum_{k=681}^{690} f_k \approx 0.0069059,^{6}$$

which roughly agrees with the exact value.

The reasoning carried out in the experiments above is formalized in the framework of *ergodic theory* and the Birkhoff ergodic theorem, which will be discussed in Chapter 2. The main concepts are: *invariant measure* with respect to a map, and *ergodic map* with respect to an invariant measure.

In the previous experiments, the probability measure μ defined over the measurable space $(I, \mathcal{B})$, whose density function is ρ, is invariant with respect to the Ulam map T_4, and T_4 is ergodic with respect to μ. In this situation, the Birkhoff ergodic theorem states that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T_4^k(x)) = \int_I f \, \mathrm{d}\mu \tag{1.1}$$

for almost all $x \in I$ with respect to μ, and for any integrable function f. If we take $f = \chi_{[0.68,0.69]}$, we obtain the probability calculated above.

Moreover, since μ is a probability measure with density function ρ, μ is absolutely continuous with respect to λ, and since ρ is positive in $(0, 1)$, λ is also absolutely continuous with respect to μ. Hence, μ and λ are *equivalent,* so an assertion is true μ-a.e. if and only if it is true λ-a.e. Thus, Equality 1.1 holds also for almost all $x \in I$ with respect to λ.
The details of all this will be discussed in Chapter 3.

At this point, we have built sufficient background to address the notion of *stochasticity.* As opposed to the notion of regularity, a parameter value $r \in [0, 4]$ is said to be *stochastic* (see [6]) if there exists a T_r-invariant probability measure μ_r which is absolutely continuous with respect to the Lebesgue measure λ. In that case, T_r is μ_r-ergodic, and we can apply the Birkhoff ergodic theorem. The probability measure μ_r then explains the asymptotic behavior of almost all orbits of T_r.

It is known that T_r cannot be simultaneously *regular* and *stochastic,* and that both, the set of r-values for which T_r is regular and the set of r-values for which T_r is stochastic, have positive Lebesgue measure λ.

Furthermore, Lyubich Regular or Stochastic theorem states that for almost every $r \in [0, 4]$, the logistic map T_r is either regular or stochastic (see [6]). This theorem, which was a milestone in the understanding of *one-dimensional dynamics,* provides a complete *qualitative* description of the asymptotic behavior of almost all orbits for almost any parameter value r in the logistic family T_r.

[5] Equality modulo Lebesgue measure zero sets (see Chapter 3).
[6] See A.3 to verify this value.

1.2 Chaos and Dense Orbits

Closely related to the above is the fact that the Ulam map T_4 is *chaotic* on the unit interval I and that almost all of its orbits are *dense* in I. Given a topological space X and a map $T\colon X \to X$, the orbit of $x_0 \in X$ is **dense** in X if for any open non-empty subset $U \subseteq X$ there exists $n > 0$ such that $T^n(x_0) \in U$. In our case, I is a metric space, so the orbit of $x_0 \in I$ is dense in I if and only if for any $x \in I$ and any $\varepsilon > 0$ there exists $n > 0$ such that $|x - T^n(x_0)| < \varepsilon$.

Definition (Devaney's Definition of Chaos). Let X be a metric space. A continuous map $T\colon X \to X$ is said to be **chaotic** on X if

(i) T is **topologically transitive,** i.e. for any pair of open non-empty subsets $U, V \subseteq X$ there exists $k > 0$ such that $T^k(U) \cap V \neq \emptyset$,

(ii) the set of **periodic points** of T is **dense** in X, i.e. for any $x \in X$ and any $\varepsilon > 0$ there exists a periodic point $y \in X$ of T such that $|x - y| < \varepsilon$, and

(iii) T has **sensitive dependence on initial conditions,** i.e. there exists $\delta > 0$ such that, for any $x \in X$ and any neighborhood N of x, there exist $y \in N$ and $n \geq 0$ such that $|T^n(x) - T^n(y)| > \delta$. In this case, δ is called the **sensitivity constant.**

A few years after the publication of this denition (see [4]), a group of five mathematicians (see [2]) showed that transitivity plus density of periodic points implied sensitivity. And two years later, other two mathematicians (see [8]) proved that, for the case of continuous maps defined on an interval, transitivity implied chaos.

Therefore, to prove that the Ulam map T_4 is chaotic, the only condition that has to be checked is transitivity. We will accomplish this through the introduction of another map.

1.2.1 The Doubling Map

The goal of this section is to prove transitivity, as well as density of almost all orbits, for the so-called *doubling map* $S\colon I \to I$, given by

$$S(x) = \mathrm{Frac}(2x).$$

Remembering that the *fractional part* (or *decimal part*) of a number $x \in \mathbb{R}$ is defined by

$$\mathrm{Frac}(x) = x - k, \text{ if } k \leq x < k + 1, \ k \in \mathbb{Z},$$

we may develop the term on the right to obtain an explicit formula for the doubling map (see Figure 1.7). Then

$$\mathrm{Frac}(2x) = 2x - k, \text{ if } k \leq 2x < k + 1, \ k \in \mathbb{Z}.$$

Since $0 \leq 2x \leq 2$, we conclude that the only possible integer values are $k = 0, 1$ and 2. Hence,

$$S(x) = \mathrm{Frac}(2x) = \begin{cases} 2x, & \text{if } 0 \leq x < \frac{1}{2}, \\ 2x - 1, & \text{if } \frac{1}{2} \leq x < 1, \\ 0, & \text{if } x = 1. \end{cases}$$

Let's start by proving some basic properties of the Frac-operator.

Proposition. *Let $x \in \mathbb{R}$ and $m \in \mathbb{Z}$. Then*

(i) $\mathit{Frac}(x + m) = \mathit{Frac}(x)$.

(ii) $\mathit{Frac}(m\mathit{Frac}(x)) = \mathit{Frac}(mx)$.

Proof. (i) Let $k \in \mathbb{Z}$ such that $k \le x < k+1$. Thus $k+m \le x+m < k+m+1$, and

$$\mathrm{Frac}(x+m) = x+m-(k+m) = x-k = \mathrm{Frac}(x).$$

(ii) Let $k \in \mathbb{Z}$ such that $k \le x < k+1$. Then

$$\mathrm{Frac}(m\mathrm{Frac}(x)) = \mathrm{Frac}(m(x-k)) = \mathrm{Frac}(mx-mk) = \mathrm{Frac}(mx),$$

where the last equality follows from the first property. □

From this proposition we can derive a closed-form expression for the iteration of the doubling map S.

Proposition. *Let $x_0 \in I$. Then*

$$x_n = S^n(x_0) = \mathit{Frac}(2^n x_0), \ \ n \ge 1.$$

Proof. We prove the statement by induction on n. The first iterate, $x_1 = S(x_0) = \mathrm{Frac}(2x_0)$, is already in the closed form by definition. For the induction step from n to $n+1$, let us assume the hypothesis $x_n = \mathrm{Frac}(2^n x_0)$ for some $n \ge 1$. Then, we compute

$$x_{n+1} = S(x_n) = \mathrm{Frac}(2x_n) = \mathrm{Frac}(2\mathrm{Frac}(2^n x_0)) = \mathrm{Frac}(2^{n+1}x_0),$$

where the last equality follows from the second property of the Frac-operator. This concludes the proof by induction. □

Now, we reveal a new interpretation for the doubling map S by passing to *binary representations* of the numbers $x \in I$. Recall that any $x \in I$ can be written as $x = 0.a_1a_2a_3\ldots$, where the a_i are binary digits, i.e. $a_i \in \{0,1\}$, and

$$x = a_1 2^{-1} + a_2 2^{-2} + a_3 2^{-3} + \cdots .$$

Proposition. *Let $x, y \in I$ with binary expansions $x = 0.a_1a_2a_3\ldots$ and $y = 0.b_1b_2b_3\ldots$ Then*

$$|x-y| \le 2^{-k},$$

provided that $a_i = b_i$ for $i = 1,\ldots,k$.

Proof. Without loss of generality, we can assume that $x \ge y$. Then

$$|x-y| = |0.a_1a_2a_3\ldots - 0.b_1b_2b_3\ldots| \le |0.a_1\ldots a_k\widehat{1} - 0.b_1\ldots b_k\widehat{0}| = |0.0\ldots0\widehat{1}| = 2^{-k}.$$

□

If we apply the doubling map S to a binary representation $x = 0.a_1a_2a_3\ldots$ then

$$S(x) = \mathrm{Frac}(2x) = \mathrm{Frac}(a_1.a_2a_3a_4\ldots) = 0.a_2a_3a_4\ldots,$$

i.e. one application of S consists in first shifting all binary digits one place to the left, and then erasing the digit that is moved in front of the decimal point. Due to the type of this almost mechanical procedure, S is also called the *binary shift operator*.

Note. There is a technicality which we must address here, namely, the ambiguity of the binary representations. For example, $x = \frac{1}{2}$ has two possible binary versions: 0.1 and $0.0\widehat{1}$. If we apply the shift operator to them we obtain 0 and $0.\widehat{1} = 1$, respectively. On the other hand, $S(\frac{1}{2}) = \mathrm{Frac}(1) = 0$. Hence, to avoid the ambiguity and to be consistent with the initial definition of S, we discard all binaries ending with repeating digits 1.[7] Thus, we represent $x = \frac{1}{2} = 0.1$ and $x = \frac{1}{4} = 0.01$, but not as $0.0\widehat{1}$ or $0.00\widehat{1}$.

[7] Our convention implies that we cannot represent $x = 1$ in the form $0.\widehat{1}$. This point is the only one that will not have a binary representation in our analysis. But this is not significant for the dynamics of the iteration of S, since $x = 1$ has not preimages (no point of I is mapped into $x = 1$ through S), and this point is mapped into the fixed point $x = 0$.

Note. The notion of transitivity does not require a map to be continuous. Thus, we may try transitivity for the doubling map S.

Now, we are ready to prove the following key theorems.

Theorem. *The doubling map is topologically transitive.*

Proof. Let U and V be two open non-empty subsets in I. Then, there exist two open non-empty subintervals $\tilde{U} \subseteq U$ and $\tilde{V} \subseteq V$. Now, let $k > 0$ such that $\tilde{U}$ has a length greater than $2^{-(k-1)}$. Further, let $x = 0.a_1a_2a_3\dots$ be the binary representation of the midpoint of $\tilde{U}$ and let $y = 0.b_1b_2b_3\dots$ be the binary representation of a point of $\tilde{V}$.

We construct an initial point $x_0 \in \tilde{U}$ which, after exactly k iterations of the shift operator, will be equal to y, thus providing the required point in the target interval $\tilde{V}$. To define x_0 we copy the first k digits of the center of $\tilde{U}$ and then append all digits of the target point y:

$$x_0 = 0.a_1a_2a_3\dots a_kb_1b_2b_3\dots$$

Now, we check that x_0 belongs to $\tilde{U}$: by the previous proposition $|x - x_0| \leq 2^{-k}$, i.e. x_0 differs from the center of $\tilde{U}$ by at most 2^{-k}. Since the width of $\tilde{U}$ is greater than twice this distance, necessarily $x_0 \in \tilde{U}$. Secondly, after k iterations we have

$$x_k = S^k(x_0) = 0.b_1b_2b_3\dots = y \in \tilde{V}.$$

Therefore,

$$S^k(U) \cap V \supseteq S^k(\tilde{U}) \cap \tilde{V} \neq \emptyset.$$

□

Remark. In the proof above we have even over-fulfilled the requirement, since in the case of the shift operator we can hit any target point $y \in V$.

Theorem. *Almost all orbits of the doubling map are dense in I.*

Proof. We have to prove the following: if we pick an initial condition $x_0 \in I$ at random, then for any $x \in I$ and any $\varepsilon > 0$ almost surely there exists $n > 0$ such that $|x - S^n(x_0)| < \varepsilon$.

So, let $x_0 \in I$ for which we choose its binary digits at random and let $x = 0.a_1a_2a_3\dots$ be the binary representation of a point of I. Given $\varepsilon > 0$, there exists $k > 0$ such that $2^{-k} < \varepsilon$.

Now, the Infinite monkey theorem (see [1]) implies that the string of digits $a_1a_2a_3\dots a_k$ (the first k digits of x) almost surely appears in the binary expansion of x_0 at some place, and therefore sufficiently many shifts will bring this string to the leading digits, i.e. almost surely there exists $n > 0$ such that $x_n = S^n(x_0) = 0.a_1a_2a_3\dots a_k\dots$, and thus,

$$|x - x_n| \leq 2^{-k} < \varepsilon.$$

□

1.2.2 Topological Conjugacy and Semi-conjugacy

The goal of this section is to relate the dynamics of the doubling map S to the dynamics of the Ulam map T_4, in order to transfer to the latter what we have just proved for the former.

Definition. Let X and Y be two topological spaces and $f\colon X \to X$ and $g\colon Y \to Y$ two maps.

(i) f and g are said to be **topologically conjugate** provided f and g are continuous and there exists a homeomorphism $h\colon X \to Y$ such that $h \circ f = g \circ h$. In this case, h is called a **topological conjugacy.**

(ii) g is said to be **topologically semi-conjugate** to f provided there exists a continuous and onto map $h\colon X \to Y$ such that $h \circ f = g \circ h$. In this case, h is called a **topological semi-conjugacy.**

To express that the *conjugacy equation* $h \circ f = g \circ h$ holds, we say that the following diagram is commutative:

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\ Y & \xrightarrow[g]{} & Y \end{array}$$

Proposition. *Let g be topologically semi-conjugate to f via h. Then*

(i) The functional equation $h \circ f^n = g^n \circ h$ holds for every $n \geq 1$.

(ii) If f is topologically transitive, then also g is topologically transitive.

(iii) If the orbit of x_0 is dense in X, then the orbit of $y_0 = h(x_0)$ is dense in Y.

Proof. (i) We prove the statement by induction on n. For $n = 1$, the functional equation $h \circ f = g \circ h$ holds since h is a topological semi-conjugacy. For the induction step from n to $n + 1$, let us assume that $h \circ f^n = g^n \circ h$. Then,

$$h \circ f^{n+1} = (h \circ f^n) \circ f = (g^n \circ h) \circ f = g^n \circ (h \circ f) = g^n \circ (g \circ h) = g^{n+1} \circ h.$$

(ii) Let U and V be two open non-empty subsets in Y. We have to find $y \in U$ and $k > 0$ such that $g^k(y) \in V$. To do this, we take the preimages $A = h^{-1}(U)$ and $B = h^{-1}(V)$, which are non-empty subsets since h is onto, and also, they are open in X since h is continuous. Thus, there exists $x \in A$ and $k > 0$ such that $f^k(x) \in B$, since f is topologically transitive. Now, we take $y = h(x)$, and using the functional equation $h \circ f^k = g^k \circ h$ we obtain

$$g^k(y) = g^k(h(x)) = h(f^k(x)) \in h(B) = V,$$

since h is onto.

(iii) Let U an open non-empty subset in Y. We have to find $n > 0$ such that $g^n(y_0) \in U$. To do this, we take the preimage $A = h^{-1}(U)$, which is a non-empty subset since h is onto, and also, it is open in X since h is continuous. Thus, there exists $n > 0$ such that $f^n(x_0) \in A$, since the orbit of x_0 is dense. Now, using the functional equation $h \circ f^n = g^n \circ h$ we obtain

$$g^n(y_0) = g^n(h(x_0)) = h(f^n(x_0)) \in h(A) = U,$$

since h is onto. □

Proposition. *The Ulam map is topologically semi-conjugate to the doubling map via the map $\psi\colon I \to I$, given by*

$$\psi(x) = \sin^2(\pi x).$$

Proof. It is clear that ψ is continuous and onto, but not one-to-one, since $y = \psi(x)$ if and only if $x = z$ or $x = 1 - z$, where

$$z = \frac{1}{\pi} \arcsin(\sqrt{y}) \in \left[0, \frac{1}{2}\right].$$

On the one hand,

$$(\psi \circ S)(x) = \left\{ \begin{array}{ll} \psi(2x), & \text{if } 0 \leq x < \frac{1}{2} \\ \psi(2x-1), & \text{if } \frac{1}{2} \leq x < 1 \\ \psi(0), & \text{if } x = 1 \end{array} \right\} = \sin^2(2\pi x),$$

and, on the other hand,

$$(T_4 \circ \psi)(x) = 4\sin^2(\pi x)(1 - \sin^2(\pi x)) = (2\sin(\pi x)\cos(\pi x))^2 = \sin^2(2\pi x), \ x \in I.$$

□

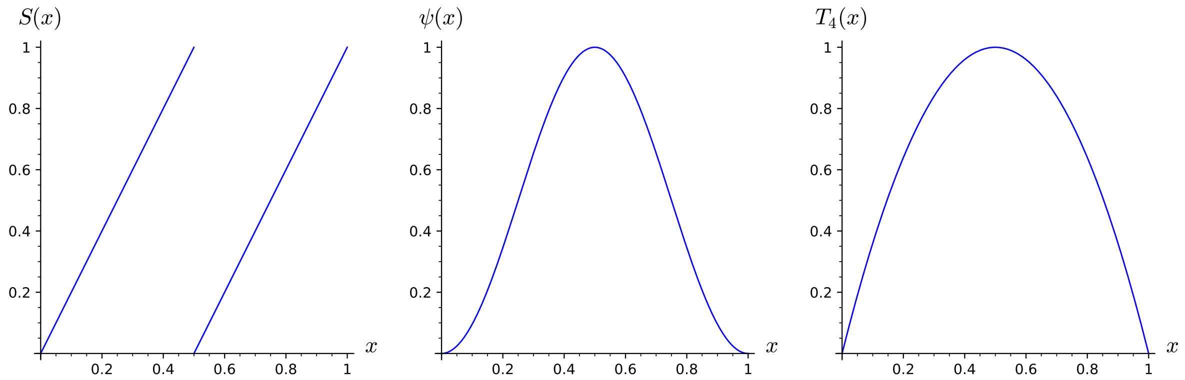

Figure 1.7: $y = S(x)$ (left), $y = \psi(x)$ (center) and $y = 4x(1-x)$ (right).

Corollary. *We conclude the following two facts:*

(i) The Ulam map is chaotic on I.

(ii) Almost all orbits of the Ulam map are dense in I.

Proof. (i) This follows from the fact that the Ulam map T_4 is continuous and that it inherits from the doubling map S, via ψ, the property of topological transitivity.
(ii) This also follows from the fact that T_4 inherits from S, via ψ, the property of density of almost all orbits. □

The fact that almost all orbits of T_4 are dense in I suggests that the attractor for the Ulam map is the whole interval I.
In Chapter 3, we will follow the same proof strategy as above to show that the Ulam map is ergodic.

1.2.3 Dense Orbit implies Topological Transitivity

We start by proving the following result:

Proposition. *Let X be a topological space without isolated points, D a dense subset in X and $x_1, x_2, \ldots, x_n \in D$. Then, the set $D_n = D \setminus \{x_1, x_2, \ldots, x_n\}$ is dense in X.*

Proof. We must prove the following: $D_n \cap U \neq \emptyset$, for any open non-empty subset $U \subseteq X$.

So, let U an open non-empty subset in X. We prove the statement by induction on n. For $n = 1$,

$$D_1 \cap U = (D \setminus \{x_1\}) \cap U = (D \cap U) \setminus \{x_1\} \neq \emptyset,$$

since x_1 is not isolated. Hence, D_1 is dense in X. For the induction step from n to $n+1$, let us assume that $D_n = D \setminus \{x_1, x_2, \dots, x_n\}$ is dense in X. Then,

$$D_{n+1} \cap U = (D \setminus \{x_1, x_2, \dots, x_n, x_{n+1}\}) \cap U = (D_n \setminus \{x_{n+1}\}) \cap U = (D_n \cap U) \setminus \{x_{n+1}\} \neq \emptyset,$$

since x_{n+1} is not isolated. Therefore, D_{n+1} is dense in X. □

Theorem. *Let X be a topological space without isolated points and $T\colon X \to X$ a continuous map. If T has a dense orbit in X, then T is topologically transitive.*

Proof. Suppose that the orbit of an initial value $x_0 \in X$ is dense in X. Let U and V be two open non-empty subsets in X. Then, there exists $n > 0$ such that $x_n = T^n(x_0) \in U$. Now, the orbit of $x_n \in X$ is also dense in X, since, by the previous proposition, the set

$$\{x_n, x_{n+1}, \dots\} = \{x_0, x_1, \dots\} \setminus \{x_0, x_1, \dots, x_{n-1}\},$$

is dense in X. Thus, there exists $k > 0$ such that $T^k(x_n) \in V$. □

In separable complete metric spaces, the converse is also true.

Theorem (**Birkhoff Transitivity Theorem**). *Let X be a separable complete metric space without isolated points and $T\colon X \to X$ a continuous map. The following statements are equivalent:*

(i) T is topologically transitive.

(ii) T has a dense orbit in X.

For a proof of this theorem see [5].

Chapter 2

Ergodic Theory

The most fundamental idea in ergodic theory is the Birkhoff ergodic theorem, which states that, with probability one, the average of a function along an orbit of an ergodic map is equal to the integral of the given function.

2.1 Image Measure

Let's start by remembering the notion of an image measure. Given a measurable space $(X, \mathcal{A})$ and a map $T\colon X \to X$, we know that the set

$$T(\mathcal{A}) = \{E \subseteq X : T^{-1}(E) \in \mathcal{A}\}$$

is a σ-algebra over X, the so-called **image σ-algebra** of $\mathcal{A}$ under T. If a measure μ is chosen for $(X, \mathcal{A})$ we may consider the **image measure** of μ under T, given by

$$\begin{array}{rccl} \mu \circ T^{-1}\colon & T(\mathcal{A}) & \longrightarrow & [0, +\infty] \\ & E & \longmapsto & \mu(T^{-1}(E)). \end{array}$$

We will denote the image measure $\mu \circ T^{-1}$ by $T_*(\mu)$.
Further, if $\mathcal{A} \subseteq T(\mathcal{A})$, we may consider the restriction measure of $T_*(\mu)$ to $\mathcal{A}$, so that $(X, \mathcal{A}, T_*(\mu)_{|\mathcal{A}})$ is a measure space. For ease of notation, we will write simply $T_*(\mu)_{|\mathcal{A}} = T_*(\mu)$ in that case.
Note that $\mathcal{A} \subseteq T(\mathcal{A})$ is equivalent to $T^{-1}(E) \in \mathcal{A}$ for every $E \in \mathcal{A}$.

2.2 Invariant Measure

Definition. Let $(X, \mathcal{A}, \mu)$ be a measure space. A map $T\colon X \to X$ is said to be **measure preserving** with respect to μ (or **μ-preserving**) if

$$\mathcal{A} \subseteq T(\mathcal{A}) \quad \text{and} \quad T_*(\mu) = \mu.$$

In this case, μ is said to be **invariant** under T (or **T-invariant**).

Definition. Let $(X, \mathcal{A})$ be a measurable space. A function $f\colon X \to \mathbb{R}$ is said to be **measurable** (or **$\mathcal{A}$-measurbale**) if

$$f^{-1}(E) \in \mathcal{A} \text{ for every } E \in \mathcal{B}(\mathbb{R}),$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra. If f is also a complex function, we say that f is **measurable** if its real and imaginary parts are measurable functions.

Definition. Let $(X, \mathcal{A}, \mu)$ be a measure space. A measurable function $f\colon X \to \mathbb{R}$ is said to be **integrable** with respect to μ (or **μ-integrable**) if

$$\int_X |f| \, \mathrm{d}\mu < +\infty.$$

In this case, we denote by $L^1(X,\mu)$ the set of integrable functions on X. If f is also a complex function, we say that f is **integrable** if its real and imaginary parts are integrable functions.

Theorem. *Let $(X,\mathcal{A},\mu)$ be a finite measure space and $T\colon X\to X$ a map such that $\mathcal{A}\subseteq T(\mathcal{A})$. The following statements are equivalent:*

(i) T is μ-preserving.

(ii) For any $f\in L^1(X,\mu)$ we have

$$\int_X f\,d\mu = \int_X (f\circ T)\,d\mu\,.$$

Proof. (ii)⇒(i). Let $E\in\mathcal{A}$ and take $f=\chi_E$. Since $\mu(X)<+\infty$, then $f\in L^1(X,\mu)$. Hence,

$$\mu(E)=\int_X \chi_E(x)\,\mathrm{d}\mu=\int_X \chi_E(T(x))\,\mathrm{d}\mu=\int_X \chi_{T^{-1}(E)}(x)\,\mathrm{d}\mu=\mu(T^{-1}(E))=T_*(\mu)(E)\,.$$

(i)⇒(ii). First observe that a complex-valued measurable function f can be written as a sum

$$f=f_1-f_2+i(f_3-f_4),$$

where $i=\sqrt{-1}$, and each function f_j is real, nonnegative and measurable. Thus, we may assume that f is real-valued and $f\geq 0$. Let $E\in\mathcal{A}$ and take $f=\chi_E$. Then,

$$\int_X f(x)\,\mathrm{d}\mu=\mu(E)=T_*(\mu)(E)=\int_X \chi_{T^{-1}(E)}(x)\,\mathrm{d}\mu=\int_X f(T(x))\,\mathrm{d}\mu\,.$$

By linearity, the same relation holds for a simple measurable function f. Now, for a general nonnegative function $f\in L^1(X,\mu)$, choose an increasing sequence of simple measurable nonnegative functions $\{s_n\}_{n\geq1}$ converging to f pointwise. Then, $\{s_n\circ T\}_{n\geq1}$ is an increasing sequence and it converges to $f\circ T$ pointwise. Finally, the Monotone Convergence Theorem implies that

$$\int_X f(T(x))\,\mathrm{d}\mu=\lim_{n\to\infty}\int_X s_n(T(x))\,\mathrm{d}\mu=\lim_{n\to\infty}\int_X s_n(x)\,\mathrm{d}\mu=\int_X f(x)\,\mathrm{d}\mu\,.$$

□

2.3 Ergodic Map

Definition. Let X be a set and $T\colon X\to X$ a map.

(i) A set $E\subseteq X$ is said to be **invariant** under T (or T-**invariant**) if $T^{-1}(E)=E$.

(ii) A function $f\colon X\to\mathbb{C}$ is said to be **invariant** under T (or T-**invariant**) if $f\circ T=f$.

Definition. Let $(X,\mathcal{A},\mu)$ be a measure space and $E,F\in\mathcal{A}$. We say that $E=F$ **modulo measure zero** if

$$\mu((E\setminus F)\cup(F\setminus E))=0.$$

In this case, we write $E\mathrel{\mathring{=}}F$.

Definition. Let $(X,\mathcal{A},\mu)$ be a probability space. A μ-preserving map $T\colon X\to X$ is said to be **ergodic** with respect to μ (or μ-**ergodic**) if for any $E\in\mathcal{A}$

$$E \text{ is } T\text{-invariant modulo measure zero if and only if } \mu(E)=0 \text{ or } \mu(E)=1.$$

Definition. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P(x)$ a property whose validity depends on $x \in X$. We say that $P(x)$ is true **almost everywhere** with respect to μ (or μ-**a.e.**) if

there exists $N \in \mathcal{A}$ with $\mu(N) = 0$ such that $P(x)$ holds true on $X \setminus N$.

Example. Let X be a set and $T\colon X \to X$ a map. Given $f\colon X \to \mathbb{C}$, we say that f is T-invariant μ-a.e. if there exists $N \in \mathcal{A}$ with $\mu(N) = 0$ such that $f(T(x)) = f(x)$ at least for any $x \in X \setminus N$.

Theorem. *Let $(X, \mathcal{A}, \mu)$ be a probability space and $T\colon X \to X$ a μ-preserving map. The following statements are equivalent:*

(i) T is μ-ergodic.

(ii) If a measurable function $f\colon X \to \mathbb{C}$ is T-invariant μ-a.e., then f is constant μ-a.e.

(iii) If an integrable function $f\colon X \to \mathbb{C}$ is T-invariant μ-a.e., then f is constant μ-a.e.

(iv) If a square-integrable function $f\colon X \to \mathbb{C}$ is T-invariant μ-a.e., then f is constant μ-a.e.

Proof. (i)$\Rightarrow$(ii). Let f be measurable and T-invariant μ-a.e. By considering real and imaginary parts, we may assume that f is real-valued. Put

$$E_{n,k} = \left\{x \in X : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\right\}, \ n \geq 1, \ k \in \mathbb{Z}.$$

Then, $\{E_{n,k}\}_{k\in\mathbb{Z}}$ is a partition of X for every $n \geq 1$. Note that

$$T^{-1}(E_{n,k}) = \left\{x \in X : \frac{k}{2^n} \leq f(T(x)) < \frac{k+1}{2^n}\right\} \doteq \left\{x \in X : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\right\} = E_{n,k},$$

since f is T-invariant μ-a.e. Since T is μ-ergodic, $\mu(E_{n,k}) = 0$ or 1; more precisely, for each $n \geq 1$ there exists a unique $k \in \mathbb{Z}$, say k_n, such that $\mu(E_{n,k_n}) = 1$ and $\mu(E_{n,k}) = 0$ for $k \neq k_n$, since $\{E_{n,k}\}_{k\in\mathbb{Z}}$ is a partition of X and $\mu(X) = 1$. Let

$$X_0 = \bigcap_{n=1}^{\infty} E_{n,k_n}.$$

By taking complements, we have

$$\mu(X_0^c) = \mu\left(\bigcup_{n=1}^{\infty} E_{n,k_n}^c\right) \leq \sum_{n=1}^{\infty} \mu(E_{n,k_n}^c) = 0,$$

and so $\mu(X_0) = 1$. We will prove by contradiction that f is constant on X_0. Suppose that there exist $x, y \in X_0$ such that $f(x) \neq f(y)$, i.e. $\varepsilon = |f(x) - f(y)| > 0$.
On the one hand, there exists $n_0 \geq 1$ sufficiently large such that

$$\frac{1}{2^n} < \varepsilon$$

for any $n \geq n_0$.
On the other hand, $x, y \in E_{n,k_n}$ for every $n \geq 1$. In particular, x, y satisfy

$$\frac{k_{n_0}}{2^{n_0}} \leq f(x), f(y) < \frac{k_{n_0}+1}{2^{n_0}}.$$

Hence,

$$\varepsilon = |f(x) - f(y)| < \frac{1}{2^{n_0}},$$

which is a contradiction.
(ii)⇒(i) Let $E \in \mathcal{A}$ and take $f = \chi_E$. Then χ_E is measurable. Suppose that $T^{-1}(E) \mathrel{\mathring{=}} E$. Thus χ_E is T-invariant μ-a.e., and hence constant μ-a.e. Since the possible values of χ_E are 0 and 1, we conclude that either $\chi_E = 0$ μ-a.e. or $\chi_E = 1$ μ-a.e., equivalently, $\mu(E) = 0$ or $\mu(E) = 1$. Conversely, let $E \in \mathcal{A}$ such that $\mu(E) = 0$ or $\mu(E) = 1$. Since T is μ-preserving, $\mu(T^{-1}(E)) = \mu(E) = 0$ or $\mu(T^{-1}(E)) = \mu(E) = 1$, equivalently, $T^{-1}(E) \mathrel{\mathring{=}} E$.
(ii)⇒(iii). This is immediate since every integrable function f is measurable.
(ii)⇒(iv). Again, every square-integrable function f is measurable.
The implications (iii)⇒(i) and (iv)⇒(i) are identical to (ii)⇒(i), since $\mu(X) < +\infty$. □

2.4 The Birkhoff Ergodic Theorem

The proof of the Birkhoff ergodic theorem is rather technical. We need the following two previous results.

Lemma. *Let $(X, \mathcal{A}, \mu)$ be a measure space, $T\colon X \to X$ a μ-preserving map and $f\colon X \to \mathbb{R}$ an integrable function. Define $f_0 = 0$,*

$$f_n = \sum_{k=0}^{n-1} (f \circ T^k), \;\; n \geq 1,$$

and

$$F_N = \max_{0 \leq n \leq N} f_n, \;\; N \geq 0.$$

Put $A_N = \{x \in X : F_N(x) > 0\}$, $N \geq 0$. Then,

$$\int_{A_N} f \, d\mu \geq 0.$$

Proof. Note that $f_n, F_N \in L^1(X, \mu)$ and $F_N \geq 0$. For $N = 0$ we have

$$\int_{A_0} f \,\mathrm{d}\mu = 0,$$

since $A_0 = \emptyset$. Now, for $0 \leq n \leq N$ we have $F_N \geq f_n$, and so $F_N \circ T \geq f_n \circ T$. Hence,

$$F_N \circ T + f \geq f_n \circ T + f = \sum_{k=1}^{n} (f \circ T^k) + f = f_{n+1}.$$

Thus,

$$F_N \circ T + f \geq \max_{1 \leq n \leq N+1} f_n.$$

If $F_{N+1} > 0$, then the right-hand side of the inequality is equal to $\max\limits_{0 \leq n \leq N+1} f_n = F_{N+1}$, and hence $f \geq F_{N+1} - F_N \circ T$ on A_{N+1}. Now, we have

$$\int_{A_{N+1}} f \,\mathrm{d}\mu \geq \int_{A_{N+1}} F_{N+1} \,\mathrm{d}\mu - \int_{A_{N+1}} F_N \circ T \,\mathrm{d}\mu = \int_X F_{N+1} \,\mathrm{d}\mu - \int_{A_{N+1}} F_N \circ T \,\mathrm{d}\mu \,,$$

since $F_{N+1} = 0$ on $X \setminus A_{N+1}$. Due to $F_N \circ T \geq 0$ and $F_N \leq F_{N+1}$, we conclude that

$$\int_{A_{N+1}} f \,\mathrm{d}\mu \geq \int_X F_{N+1} \,\mathrm{d}\mu - \int_X F_N \circ T \,\mathrm{d}\mu \geq \int_X F_{N+1} \,\mathrm{d}\mu - \int_X F_{N+1} \circ T \,\mathrm{d}\mu = 0,$$

where the last equality follows from the fact that T is μ-preserving. □

Theorem (Maximal Ergodic Theorem). *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, $T: X \to X$ a μ-preserving map, and $g: X \to \mathbb{R}$ an integrable function. Define*

$$B_\alpha = \left\{ x \in X : \sup_{n \geq 1} \left(\frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) \right) > \alpha \right\}, \ \alpha \in \mathbb{R}.$$

Then,

$$\int_{B_\alpha} g \, d\mu \geq \alpha \mu(B_\alpha).$$

Furthermore, if $E \in \mathcal{A}$ is T-invariant then

$$\int_{E \cap B_\alpha} g \, d\mu \geq \alpha \mu(E \cap B_\alpha).$$

Proof. Let $f = g - \alpha$. Since $f \in L^1(X, \mu)$, we may consider f_n, F_N and A_N as in the previous lemma. Now, we want to prove the following equality:

$$B_\alpha = \bigcup_{N=0}^{\infty} A_N.$$

Indeed,

$$x \in \bigcup_{N=0}^{\infty} A_N \Leftrightarrow \exists N \geq 1, \ \max_{1 \leq n \leq N} \left(\sum_{k=0}^{n-1} (g(T^k(x)) - \alpha) \right) > 0.$$

Let

$$S_n = \sum_{k=0}^{n-1} (g(T^k(x)) - \alpha), \ 1 \leq n \leq N.$$

Observe that the following equivalence holds:

$$\exists N \geq 1, \ \max_{1 \leq n \leq N} S_n > 0 \Leftrightarrow \exists N \geq 1, \ \max_{1 \leq n \leq N} \frac{S_n}{n} > 0.$$

Therefore, the last inequality is equivalent to:

$$\sup_{n \geq 1} \left(\frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) \right) > \alpha \Leftrightarrow x \in B_\alpha.$$

Since $\{F_N\}_{N \geq 1}$ is an increasing sequence of functions, we have the increasing sequence of sets

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_N \subseteq A_{N+1} \subseteq \cdots \subseteq B_\alpha,$$

and, equivalently, we have the increasing sequence of characteristic functions

$$0 \leq \chi_{A_1} \leq \chi_{A_2} \leq \cdots \leq \chi_{A_N} \leq \chi_{A_{N+1}} \leq \cdots \leq \chi_{B_\alpha}.$$

It is clear that the sequence $\{\chi_{A_N}\}_{N \geq 1}$ converges to χ_{B_α} pointwise. Thus, the sequence $\{\chi_{A_N} f\}_{N \geq 1}$ also converges to $\chi_{B_\alpha} f$ pointwise. Observe that $|\chi_{A_N} f| \leq |f|$ for any $N \geq 1$. The Dominated Convergence Theorem implies that

$$\int_{B_\alpha} f \, \mathrm{d}\mu = \int_X \chi_{B_\alpha} f \, \mathrm{d}\mu = \lim_{N \to \infty} \int_X \chi_{A_N} f \, \mathrm{d}\mu = \lim_{N \to \infty} \int_{A_N} f \, \mathrm{d}\mu \geq 0,$$

where the last inequality is by the previous lemma. Hence,

$$\int_{B_\alpha} g \, \mathrm{d}\mu - \alpha \mu(B_\alpha) \geq 0.$$

For the second part we consider the restriction map of T to E. Since E is T-invariant, then $T(E) \subseteq E$. In this situation, the subset E plays the role of X in the first case. □

Theorem (**Birkhoff Ergodic Theorem**). *Let $(X, \mathcal{A}, \mu)$ be a probability space, $T \colon X \to X$ a μ-preserving map and $f \colon X \to \mathbb{C}$ an integrable function. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = f^*(x) \quad \mu\text{-a.e.},$$

and

$$\int_X f^* \, d\mu = \int_X f \, d\mu \,,$$

for some T-invariant $f^* \in L^1(X, \mu)$. Moreover, if T is μ-ergodic, then f^* is constant μ-a.e. and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \int_X f \, d\mu \quad \mu\text{-a.e.}$$

Proof. By considering real and imaginary parts, we may prove the statement for real-valued functions f. Let

$$(A_n f)(x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)), \; n \geq 1.$$

Put

$$f^*(x) = \limsup_{n\to\infty} (A_n f)(x) \quad \text{and} \quad f_*(x) = \liminf_{n\to\infty} (A_n f)(x).$$

Then $f_*(x) \leq f^*(x)$. Now, let us see that both functions are T-invariant. Indeed,

$$(A_n f)(T(x)) = \frac{1}{n} \sum_{k=1}^{n} f(T^k(x)) = \frac{n+1}{n} \frac{1}{n+1} \sum_{k=1}^{n} f(T^k(x)) =$$

$$\frac{n+1}{n} \left(\frac{1}{n+1} \sum_{k=0}^{n} f(T^k(x)) - \frac{1}{n+1} f(x) \right) = \frac{n+1}{n} (A_{n+1} f)(x) - \frac{1}{n} f(x).$$

Thus,

$$f^*(T(x)) = \limsup_{n\to\infty} (A_n f)(T(x)) = \limsup_{n\to\infty} (A_{n+1} f)(x) = f^*(x) \,.$$

The same holds for f_*. Now, we will show that $f_* = f^*$ and that both are integrable.
Put

$$E_{\alpha,\beta} = \{x \in X : f_*(x) < \beta \text{ and } \alpha < f^*(x)\}, \; \alpha, \beta \in \mathbb{Q}.$$

Note that

$$\{x \in X : f_*(x) < f^*(x)\} = \bigcup_{\beta < \alpha} E_{\alpha,\beta},$$

where the right-hand side is a countable union of sets. Moreover, $T^{-1}(E_{\alpha,\beta}) = E_{\alpha,\beta}$, since f_* and f^* are T-invariant. Put

$$B_\alpha = \left\{ x \in X : \sup_{n \geq 1} (A_n f)(x) > \alpha \right\}.$$

Then, $E_{\alpha,\beta} \subseteq B_\alpha$, since if $x \in E_{\alpha,\beta}$ then

$$\alpha < f^*(x) = \limsup_{n\to\infty} (A_n f)(x) \leq \sup_{n \geq 1} (A_n f)(x).$$

From the Maximal Ergodic Theorem we have

$$\int_{E_{\alpha,\beta}} f \, \mathrm{d}\mu = \int_{E_{\alpha,\beta} \cap B_\alpha} f \, \mathrm{d}\mu \geq \alpha \mu(E_{\alpha,\beta} \cap B_\alpha) = \alpha \mu(E_{\alpha,\beta}).$$

Note that $(-f)^* = -f_*$ and $(-f)_* = -f^*$, so that

$$E_{\alpha,\beta} = \{x \in X : (-f)^*(x) > -\beta \text{ and } -\alpha > (-f)_*(x)\}\,.$$

If we replace f, α and β by $-f, -\beta$ and $-\alpha$ respectively in the previous inequality, then we have

$$\int_{E_{\alpha,\beta}} (-f)\,\mathrm{d}\mu \geq -\beta\mu(E_{\alpha,\beta}), \text{ i.e. } \int_{E_{\alpha,\beta}} f\,\mathrm{d}\mu \leq \beta\mu(E_{\alpha,\beta}).$$

Thus, we obtain $\alpha\mu(E_{\alpha,\beta}) \leq \beta\mu(E_{\alpha,\beta})$, which implies that if $\beta < \alpha$ then $\mu(E_{\alpha,\beta}) = 0$.
Therefore, $f_* = f^*$ μ-a.e. and

$$\lim_{n\to\infty} (A_n f)(x) = f^*(x) \quad \mu\text{-a.e.}$$

Now, we show that f^* is integrable. Let

$$g_n(x) = |(A_n f)(x)|,\ n \geq 1.$$

Then, $\lim_{n\to\infty} g_n(x) = |f^*(x)|$ μ-a.e. and

$$\int_X g_n\,\mathrm{d}\mu \leq \frac{1}{n}\sum_{k=0}^{n-1}\int_X |f(T^k(x))|\,\mathrm{d}\mu = \int_X |f|\,\mathrm{d}\mu\,,$$

where the last equality follows from the fact that T is μ-preserving.
Fatou's lemma implies that

$$\int_X |f^*|\,\mathrm{d}\mu = \int_X \liminf_{n\to\infty} g_n\,\mathrm{d}\mu \leq \liminf_{n\to\infty}\int_X g_n\,\mathrm{d}\mu \leq \int_X |f|\,\mathrm{d}\mu < +\infty.$$

It remains to show that

$$\int_X f^*\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu\,.$$

Put

$$D_{n,k} = \left\{x \in X : \frac{k}{n} \leq f^*(x) < \frac{k+1}{n}\right\},\ n \geq 1,\ k \in \mathbb{Z}.$$

Note that $\{D_{n,k}\}_{k\in\mathbb{Z}}$ is a partition of X for every $n \geq 1$. Further, $T^{-1}(D_{n,k}) = D_{n,k}$, since f^* is T-invariant. For sufficiently small $\varepsilon > 0$ we have

$$D_{n,k} \subseteq B_{\frac{k}{n}-\varepsilon}.$$

The Maximal Ergodic Theorem implies that

$$\int_{D_{n,k}} f\,\mathrm{d}\mu \geq \left(\frac{k}{n} - \varepsilon\right)\mu(D_{n,k}),$$

for every sufficiently small $\varepsilon > 0$, and hence,

$$\int_{D_{n,k}} f\,\mathrm{d}\mu \geq \frac{k}{n}\mu(D_{n,k}).$$

By the definition of $D_{n,k}$,

$$\int_{D_{n,k}} f^*\,\mathrm{d}\mu \leq \frac{k+1}{n}\mu(D_{n,k}) \leq \frac{1}{n}\mu(D_{n,k}) + \int_{D_{n,k}} f\,\mathrm{d}\mu\,.$$

Summing over $k \in \mathbb{Z}$ we obtain

$$\int_X f^* \,\mathrm{d}\mu \le \frac{1}{n} + \int_X f \,\mathrm{d}\mu\,,$$

since $\mu(X) = 1$. This holds true for every $n \ge 1$. By letting $n \to \infty$ we have

$$\int_X f^* \,\mathrm{d}\mu \le \int_X f \,\mathrm{d}\mu\,.$$

Applying the same procedure to $-f$ we obtain

$$\int_X (-f)^* \,\mathrm{d}\mu \le \int_X (-f) \,\mathrm{d}\mu\,, \text{ i.e. } \int_X f_* \,\mathrm{d}\mu \ge \int_X f \,\mathrm{d}\mu\,.$$

Since $f_* = f^*$ μ-a.e., we conclude that

$$\int_X f^* \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\mu\,.$$

Finally, if T is μ-ergodic, then f^* is constant μ-a.e. and

$$\lim_{n\to\infty} (A_n f)(x) = f^*(x) = \int_X f^* \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\mu \quad \mu\text{-a.e.},$$

where the second equality follows from the fact that $\mu(X) = 1$. □

Corollary. *Let $(X, \mathcal{B}(X), \mu)$ be a probability space, where X is a topological space and $\mathcal{B}(X)$ denotes the Borel σ-algebra generated by the topology of X, and $T\colon X \to X$ a map. Assume also that the topology of X is generated by a countable basis $\mathcal{U} = \{U_1, U_2, \ldots\}$. If T is μ-ergodic, then almost all orbits of T are dense in X.*

Proof. Suppose that the orbit of an initial value $x_0 \in X$ is not dense in X. Thus, there exists an open set U_i which does not intersect the orbit of x_0. Now, if we take $f = \chi_{U_i}$, then $f \in L^1(X, \mu)$, since $\mu(X) < +\infty$. On the one hand,

$$\int_X \chi_{U_i} \,\mathrm{d}\mu = \mu(U_i) > 0,$$

since $U_i \neq \emptyset$, and, on the other hand,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{U_i}(T^k(x_0)) = 0,$$

since the orbit of x_0 does not meet U_i. Hence $x_0 \in E_i$, where

$$E_i = \left\{ x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{U_i}(T^k(x)) \neq \int_X \chi_{U_i} \,\mathrm{d}\mu \right\}.$$

Therefore, the set of points which do not have a dense orbit is included in the union

$$\bigcup_{n=1}^{\infty} E_i.$$

But the Birkhoff ergodic theorem implies that

$$\mu\left(\bigcup_{n=1}^{\infty} E_i \right) = 0,$$

which proves the claim. □

Chapter 3

Ergodicity of the Ulam Logistic Map

In this chapter we will focus on maps $T\colon I \to I$ continuous λ-a.e., i.e. $T^{-1}(E) \in \mathcal{B}$ for every $E \in \mathcal{B}$, which is equivalent to

$$\mathcal{B} \subseteq T(\mathcal{B}) = \{E \subseteq I : T^{-1}(E) \in \mathcal{B}\}.$$

3.1 Ergodicity of the Doubling Map

Proposition. *The doubling map preserves Lebesgue measure.*

Proof. It is clear that S is continuous λ-a.e., so $\mathcal{B} \subseteq S(\mathcal{B})$. Let $E \in \mathcal{B}$. Then,

$$S_*(\lambda)(E) = \lambda(S^{-1}(E)) = \int_I \chi_{S^{-1}(E)}(x)\,\mathrm{d}x = \int_I \chi_E(S(x))\,\mathrm{d}x\,.$$

On the one hand,

$$\int_0^{\frac{1}{2}} \chi_E(2x)\,\mathrm{d}x = \frac{1}{2}\int_0^1 \chi_E(x)\,\mathrm{d}x\,,$$

and, on the other hand,

$$\int_{\frac{1}{2}}^1 \chi_E(2x-1)\,\mathrm{d}x = \frac{1}{2}\int_0^1 \chi_E(x)\,\mathrm{d}x\,.$$

Adding up we obtain

$$S_*(\lambda)(E) = \int_I \chi_E(S(x))\,\mathrm{d}x = \int_I \chi_E(x)\,\mathrm{d}x = \lambda(E).$$

□

Theorem. *The doubling map is ergodic with respect to Lebesgue measure.*

Proof. It is clear that λ is a probability measure. Let

$$f(x) = \sum_{n\in\mathbb{Z}} c_n e^{2\pi i n x},\ x \in I,$$

be the Fourier series expansion of an S-invariant λ-a.e. function $f \in L^2(I,\lambda)$. Then,

$$f(S(x)) = \sum_{n\in\mathbb{Z}} c_n e^{2\pi i 2 n x},\ x \in I.$$

Comparing the Fourier coefficients of $f(x)$ and $f(S(x))$, we conclude that

$$c_n = \begin{cases} c_{\frac{n}{2}}, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

Therefore, $c_n = 0$ if $n \neq 0$, so $f = c_0$ λ-a.e. □

3.2 Ergodicity of the Ulam Logistic Map

Proposition. *The Ulam map preserves the image measure of* λ *under the topological semi-conjugacy* $\psi(x) = \sin^2(\pi x)$*, i.e. the measure* $\mu = \psi_*(\lambda)$.

Proof. Since T_4 and ψ are continuous, then $\mathcal{B} \subseteq T_4(\mathcal{B})$ and $\mathcal{B} \subseteq \psi(\mathcal{B})$. Thus, we can consider the image measure

$$\mu = \psi_*(\lambda)\colon \quad \begin{array}{ccc} \mathcal{B} & \longrightarrow & [0,+\infty] \\ E & \longmapsto & \lambda(\psi^{-1}(E)). \end{array}$$

Then,

$$(T_4)_*(\mu) = (T_4)_*(\psi_*(\lambda)) = (T_4 \circ \psi)_*(\lambda) = (\psi \circ S)_*(\lambda) = \psi_*(S_*(\lambda)) = \psi_*(\lambda) = \mu,$$

where we have used the fact that T_4 is topologically semi-conjugate to S via ψ and that S is λ-preserving. □

Theorem. *The Ulam map is ergodic with respect to the probability measure* $\mu = \psi_*(\lambda)$.

Proof. It is clear that $\mu = \psi_*(\lambda)$ is a probability measure. Let $E \in \mathcal{B}$ such that $T_4^{-1}(E) = E$. Then,

$$\mu(E) = \psi_*(\lambda)(E) = \lambda(\psi^{-1}(E))$$

and

$$S^{-1}(\psi^{-1}(E)) = \psi^{-1}(T_4^{-1}(E)) = \psi^{-1}(E),$$

where we have used the fact that T_4 is topologically semi-conjugate to S via ψ. Therefore, since S is λ-ergodic, we obtain

$$\mu(E) = \lambda(\psi^{-1}(E)) = 0 \text{ or } 1.$$

□

3.3 The Probability Density Function ρ

In the following, λ is not necessarily the Lebesgue measure. Given a measure space $(X, \mathcal{A}, \lambda)$ and a measurable function $\varphi\colon X \to [0, +\infty)$, we know that the function

$$\mu\colon \quad \begin{array}{ccc} \mathcal{A} & \longrightarrow & [0,+\infty] \\ E & \longmapsto & \int_E \varphi \, \mathrm{d}\lambda \end{array}$$

is a measure over $(X, \mathcal{A})$. In this case, φ is called the **density function** of μ with respect to λ, and we write $\mathrm{d}\mu = \varphi \, \mathrm{d}\lambda$. If $(X, \mathcal{A}, \mu)$ is also a probability space, φ is called the **probability density function.**

Definition. Let μ and λ be two measures over a measurable space $(X, \mathcal{A})$. The measure μ is said to be **absolutely continuous** with respect to λ if

$$\mu(E) = 0 \text{ for any } E \in \mathcal{A} \text{ such that } \lambda(E) = 0.$$

In this case, we write $\mu \ll \lambda$. If also $\lambda \ll \mu$, μ and λ are said to be **equivalent,** and we write $\mu \sim \lambda$.

Remark. In view of the above, we have the following two immediate facts:

(i) If $\mathrm{d}\mu = \varphi \, \mathrm{d}\lambda$ then $\mu \ll \lambda$.

(ii) If $\mu \sim \lambda$, then a property $P(x)$ is true μ-a.e. if and only if $P(x)$ is true λ-a.e.

Theorem. *The probability measure $\mu = \psi_*(\lambda)$ is absolutely continuous with respect to Lebesgue measure λ, and its probability density function is given by*

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \ x \in (0,1).$$

Moreover, μ and λ are equivalent.

Proof. Let $E \in \mathcal{B}$. Then,

$$\mu(E) = \lambda(\psi^{-1}(E)) = \int_I \chi_{\psi^{-1}(E)}(x)\,\mathrm{d}x = \int_I \chi_E(\psi(x))\,\mathrm{d}x\,.$$

Thus,

$$\int_0^1 \chi_E(\psi(x))\,\mathrm{d}x = 2\int_0^{\frac{1}{2}} \chi_E(\psi(x))\,\mathrm{d}x\,,$$

where we have used the fact that the function ψ is symmetric with respect to $x = \frac{1}{2}$. Using the change of variables

$$\left\{ \begin{array}{l} y = \psi(x) = \sin^2(\pi x), \ 0 \le x \le \frac{1}{2} \\ \mathrm{d}x = \frac{\mathrm{d}y}{2\pi\sqrt{y(1-y)}} \end{array} \right\},$$

we conclude that

$$2\int_0^{\frac{1}{2}} \chi_E(\psi(x))\,\mathrm{d}x = 2\int_0^1 \chi_E(y)\frac{\mathrm{d}y}{2\pi\sqrt{y(1-y)}} = \int_I \chi_E(y)\rho(y)\,\mathrm{d}y = \int_E \rho\,\mathrm{d}\lambda\,,$$

where

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \ x \in (0,1).$$

Therefore,

$$\mu(E) = \int_E \rho\,\mathrm{d}\lambda\,,$$

and $\mu \ll \lambda$. Finally, since $\mathrm{d}\mu = \rho\,\mathrm{d}\lambda$ and ρ is positive in $(0,1)$, then

$$\mathrm{d}\lambda = \frac{1}{\rho}\,\mathrm{d}\mu\,,$$

and thus $\mu \sim \lambda$. □

3.3.1 Sketch of the Graph of ρ

In this section we explain how to sketch the graph of ρ using the Birkhoff ergodic theorem. In our case, the probability space is $(I, \mathcal{B}, \mu)$ and the μ-ergodic map is the Ulam map $T_4\colon I \to I$, given by

$$T_4(x) = 4x(1-x).$$

First, we choose an initial point $x_0 \in I$ at random and iterate the Ulam map $N = 10^6$ times. Again, we divide the unit interval I into $M = 1000$ subintervals, given by

$$I_k = \left(\frac{k-1}{M}, \frac{k}{M}\right], \ k = 1, 2, \ldots, M.$$

Hence, Birkhoff ergodic theorem, applied to the integrable function $f = \chi_{I_k}$, implies that

$$\frac{1}{N+1}\sum_{j=0}^{N} \chi_{I_k}(T_4^j(x_0)) \approx \int_I \chi_{I_k}\,\mathrm{d}\mu = \int_{I_k} \rho\,\mathrm{d}\lambda \approx \rho(\overline{x}_k)\lambda(I_k) = \rho(\overline{x}_k)\frac{1}{M},$$

where the second approximation is due to the mean value theorem for definite integrals, and $\overline{x}_k$ is the midpoint of I_k, i.e

$$\overline{x}_k = \frac{k - \frac{1}{2}}{M}.$$

Therefore, we approximate the graph of ρ by

$$\rho(\overline{x}_k) \approx \frac{M}{N+1} \sum_{j=0}^{N} \chi_{I_k}(T_4^j(x_0)),\ \ k = 1, 2, \ldots, M.$$

Furthermore, note that the previous sum counts how many of the points $T_4^j(x_0)$ for $j = 0, 1, \ldots, N$, visit the interval I_k, then

$$\sum_{j=0}^{N} \chi_{I_k}(T_4^j(x_0)) = n_k,$$

where n_k was the absolute frequency for the interval I_k, and thus,

$$\rho(\overline{x}_k) \approx \frac{M}{N+1} n_k = \rho_k,\ \ k = 1, 2, \ldots, M,$$

which agrees with what we had seen in Chapter 1.

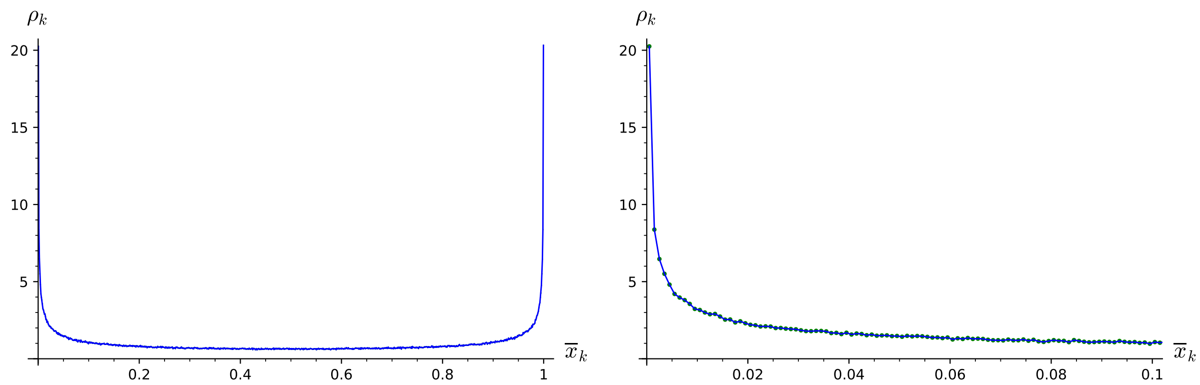

Figure 3.1: Sketch of the graph of ρ using the Birkhoff ergodic theorem.

In Figure 3.1 we have connected the points $(\overline{x}_k, \rho_k)$ by line segments to smooth the shape of the graph of ρ.

Bibliography

[1] A. Abdesselam, *The weakly dependent strong law of large numbers revisited,* Graduate J. Math. 3, 2018. Pages 94-97.

[2] J. Banks, J. Brooks, G. Cairns, G. Davis, P. Stacey, *On Devaney's Definition of Chaos,* American Mathematical Monthly 99.4, 1992. Pages 332-334.

[3] G. H. Choe, *Computational Ergodic Theory,* Algorithms and Computation in Mathematics, Volume 13, Springer, 2005. Chapters 1-3.

[4] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems, 3rd Edition,* CRC Press, 2022. Chapters 1-15.

[5] K-G. Grosse-Erdmann, A. P. Manguillot, *Linear Chaos,* Springer, 2011. Chapter 1.

[6] M. Lyubich, *The Quadratic Family as a Qualitatively Solvable Model of Chaos,* American Mathematical Society 47.9, 2000. Pages 1042-1052.

[7] H.-O. Peitgen, H. Jürgens, D. Saupe *Chaos and Fractals, 2nd Edition,* New Frontiers of Science, Springer, 2004. Chapters 10-11.

[8] M. Vellekoop, R. Berglund, *On Intervals, Transitivity=Chaos,* American Mathematical Monthly 101.4, 1994. Pages 353-355.

Appendix A

Python Programs

In this section we show the python code that we have used to obtain the images of the present work, together with some comments that will almost surely be very useful to understand the code.

A.1 Orbits

```
# Parameter value [2.9, 3.1, (1+sqrt(8)).n(), 4]
r = 2.9

# Initial condition [(e-2).n(), (pi-3).n(), (sqrt(2)-1).n()]
x_0 = (e-2).n()

# Logistic map
T(x) = r*x*(1-x)

# Number of iterations
N = 50

# Orbit of x_0: {x_0, x_1, x_2,..., x_N}
seed = [0.0 for i in range(N+1)]
seed[0] = x_0

# Iteration of T starting at x_0
for i in range(1,N+1):
    seed[i] = T(seed[i-1])

# We plot the points (n,x_n) and connect them by line segments
g = plot(points([[i,seed[i]] for i in range(N+1)], color='green', pointsize=20))
g += plot(line([[i,seed[i]] for i in range(N+1)], color='blue'))

# Labels for the axes
g.axes_labels(['$n$', '$x_n$'])

# We show the orbit in the range from 0 to 1
show(g, ymin=0, ymax=1)

# We save the image
g.save('1-CYCLE.pdf')
```

Listing A.1: Code for Figures 1.1 and 1.2.

A.2 Orbit Diagrams

```
# Initial condition
x_0 = (pi-3).n()

# Number of iterations
N = 1000

# Number of parameter values
M = 1000

# For each parameter value: {r_0, r_1, r_2,..., r_M}
# we have the orbit of x_0: {x_0, x_1, x_2,..., x_N}
seed = [[0.0 for i in range(N+1)] for k in range(M+1)]

# Iteration of T_r for each parameter value between 0 and 4
# starting at x_0
for k in range(M+1):
    seed[k][0] = x_0
    r = 4*k/M        # [3+k/M]
    T(x) = r*x*(1-x)
    for i in range(1,N+1):
        seed[k][i] = T(seed[k][i-1])

# For each r we plot the remaining 200 iterations: {x_801, x_802,..., x_N}
g = plot(points([[4*k/M,seed[k][i]] for i in range(801,N+1)
                 for k in range(M+1)], color='green', pointsize=1))

g.axes_labels(['$r$', '$I$'])

show(g)

g.save('ORBIT-DIAGRAM.png')
```

Listing A.2: Code for Figure 1.3.

```
# Initial condition
x_0 = (pi-3).n()

# Number of iterations
N = 1000

# Number of parameter values
M = 1000

# For each parameter value: {r_0, r_1, r_2,..., r_M}
# we have the orbit of x_0: {x_0, x_1, x_2,..., x_N}
seed = [[0.0 for i in range(N+1)] for k in range(M+1)]

# Iteration of T_r for each parameter value between 3 and 3.56994567
# starting at x_0
for k in range(M+1):
    seed[k][0] = x_0
    r = 3+(3.56994567-3)*k/M
    T(x) = r*x*(1-x)
    for i in range(1,N+1):
        seed[k][i] = T(seed[k][i-1])

# For each r we plot the remaining 200 iterations: {x_801, x_802,..., x_N}
g = plot(points([[3+(3.56994567-3)*k/M,seed[k][i]] for i in range(801,N+1)
                 for k in range(M+1)], color='green', pointsize=1))

# We plot a vertical line segment for some bifurcation points
g += plot(line([[3,0], [3,1]], color='blue'))
g += text('$b_1$', [3-0.01,0.03], horizontal_alignment='center',
          color='black', fontsize=12)
```

```
g += plot(line([[1+sqrt(6),0], [1+sqrt(6),1]], color='blue'))
g += text('$b_2$', [1+sqrt(6)-0.01,0.03], horizontal_alignment='center',
          color='black', fontsize=12)

g += plot(line([[3.544090,0], [3.544090,1]], color='blue'))
g += text('$b_3$', [3.544090-0.01,0.03], horizontal_alignment='center',
          color='black', fontsize=12)

g += plot(line([[3.564407,0], [3.564407,1]], color='blue'))
g += text('$b_4$', [3.564407-0.01,0.03], horizontal_alignment='center',
          color='black', fontsize=12)

g += plot(line([[3.56994567,0], [3.56994567,1]], color='red'))
g += text('$b_\infty$', [3.56994567+0.015,0.03], horizontal_alignment='center',
          color='black', fontsize=12)

g.axes_labels(['$r$', '$I$'])

show(g)

g.save('ORBIT-DIAGRAM2.png')
```

Listing A.3: Code for Figure 1.4 (left).

```
# Initial condition
x_0 = (pi-3).n()

# Number of iterations
N = 1000

# Number of parameter values
M = 1000

# For each parameter value: {r_0, r_1, r_2,..., r_M}
# we have the orbit of x_0: {x_0, x_1, x_2,..., x_N}
seed = [[0.0 for i in range(N+1)] for k in range(M+1)]

# Iteration of T_r for each parameter value between 3 and 3.56994567
# starting at x_0
for k in range(M+1):
    seed[k][0] = x_0
    r = 3.56994567+(3.857-3.56994567)*k/M
    T(x) = r*x*(1-x)
    for i in range(1,N+1):
        seed[k][i] = T(seed[k][i-1])

# For each r we plot the remaining 200 iterations: {x_801, x_802,..., x_N}
g = plot(points([[3.56994567+(3.857-3.56994567)*k/M,seed[k][i]]
                 for i in range(801,N+1) for k in range(M+1)], color='green',
                 pointsize=1))

# We plot a vertical arrow for some periodic windows
g += text('$3$', [3.835,0.04], horizontal_alignment='center', color='black',
          fontsize=12)
g += arrow([3.835,0.05], [3.835,0.165], arrowshorten=8, arrowsize=2, width=1,
           color='orange')

g += text('$5$', [3.740,0.115], horizontal_alignment='center', color='black',
          fontsize=12)
g += arrow([3.740,0.125], [3.740,0.24], arrowshorten=8, arrowsize=2, width=1,
           color='orange')

g += text('$7$', [3.702,0.145], horizontal_alignment='center', color='black',
```

```
         fontsize=12)
g += arrow([3.702,0.155], [3.702,0.27], arrowshorten=8, arrowsize=2, width=1,
          color='orange')

g += text('$6$', [3.630,0.195], horizontal_alignment='center', color='black',
         fontsize=12)
g += arrow([3.630,0.205], [3.630,0.32], arrowshorten=8, arrowsize=2, width=1,
          color='orange')

g += text('$10$', [3.606,0.21], horizontal_alignment='center', color='black',
         fontsize=12)
g += arrow([3.606,0.22], [3.606,0.335], arrowshorten=8, arrowsize=2, width=1,
          color='orange')

g += text('$12$', [3.583,0.2218], horizontal_alignment='center', color='black',
         fontsize=12)
g += arrow([3.583,0.2318], [3.583,0.3468], arrowshorten=8, arrowsize=2, width=1,
          color='orange')

# We plot a vertical line segment for some band-merging points
g += text('$m_1$', [3.67857351-0.006,0.025], horizontal_alignment='center',
         color='black', fontsize=12)
g += plot(line([[3.67857351,0], [3.67857351,1]], color='blue'))

g += text('$m_2$', [3.591+0.008,0.025], horizontal_alignment='center',
         color='black', fontsize=12)
g += plot(line([[3.591,0], [3.591,1]], color='blue'))

g += text('$m_3$', [3.573-0.0065,0.025], horizontal_alignment='center',
         color='black', fontsize=12)
g += plot(line([[3.573,0], [3.573,1]], color='blue'))

g.axes_labels(['$r$', '$I$'])

show(g)

g.save('ORBIT-DIAGRAM3.png')
```

Listing A.4: Code for Figure 1.4 (right).

A.3 Histograms

```
# The probability density function of the arcsine distribution
rho(x) = 1/(pi*sqrt(x*(1-x)))

# The Ulam map
T(x) = 4*x*(1-x)

# Initial condition
x_0 = (pi-3).n()

# Number of iterations
N = 1000000

# Orbit of x_0: {x_0, x_1, x_2,..., x_N}
seed = [0.0 for i in range(N+1)]
seed[0] = x_0

# Iteration of T starting at x_0
for i in range(1,N+1):
    seed[i] = T(seed[i-1])
```

```
# Number of subintervals
M = 1000

# Absolut frequencies: {n_0, n_1, n_2,..., n_M}
# (freq[0] is not used in calculations)
freq = [0.0 for k in range(M+1)]

# We count the number of iterates n_k which fall into each interval I_k
for i in range(N+1):
    slot = ceil(M*seed[i])
    freq[slot] = freq[slot]+1

# Coordinates for the four vertices of each column of the histogram
# (P[0] is not used in calculations)
P = [[0.0,0.0] for k in range(4*M+1)]

# We calculate the coordinates for the four vertices of each column
# based on the number of hits n_k
for k in range(1,M+1):
    P[4*k-3] = [(k-1)/M,0]
    P[4*k-2] = [(k-1)/M,freq[k]*M/(N+1)]
    P[4*k-1] = [k/M,freq[k]*M/(N+1)]
    P[4*k] = [k/M,0]

# We plot line segments between the four vertices of each column
g = plot(line([P[k] for k in range(1,4*M+1)], color='green'))

# Graph of rho in the unit interval
g += plot(rho,0,1, color='red')

g += text('$\\rho_k\\approx\\rho$', (0.5,20), horizontal_alignment='center',
          color='black', fontsize=15)

g.axes_labels([' ', ' '])

# We adjust the graph to our convenience (ymax=freq[M]*M/(N+1))
show(g, ymax=20.3179796820203)

g.save('DENSITY-MAP2.pdf')
```

Listing A.5: Code for Figure 1.6 (right).

```
suma=0.0

# Sum of the absolute frequencies from the first interval to the last
for k in range(681,690+1):
    suma=freq[k]+suma

# Sum of the corresponding relative frequencies
suma/(N+1)
```

Listing A.6: Code for numerical estimations.

A.4 Graphs of our Maps

```
# The probability density function of the arcsine distribution
rho(x) = 1/(pi*sqrt(x*(1-x)))

# Graph of rho in the unit interval
g = plot(rho,0,1, color='blue')

g.axes_labels(['$x$', '$\\rho(x)$'])
```

```
# We adjust the graph to our convenience (ymax=freq[M]*M/(N+1))
show(g, ymax=20.3179796820203, ymin=0)

g.save('DENSITY-MAP.pdf')
```
Listing A.7: Code for Figure 1.7 (left).

```
# The doubling map [psi(x)=sin(pi*x)*sin(pi*x), T(x)=4*x*(1-x)]
S(x) = frac(2*x)

# Discontinuity points of S
excl = [1/n for n in range(1,2+1)]

# We exclude the discontinuity points of S
g = plot(S,0,1, exclude=excl)

g.axes_labels(['$x$', '$S(x)$'])

# We put the same scale in both axes
show(g, aspect_ratio=1)

g.save('DOUBLING-MAP.pdf', aspect_ratio=1)
```
Listing A.8: Code for Figure 1.7.

A.5 Birkhoff Ergodic Theorem in Practice

```
# The Ulam map
T(x) = 4*x*(1-x)

# Initial condition
x_0 = (pi-3).n()

# Number of iterations
N = 1000000

# Orbit of x_0: {x_0, x_1, x_2,..., x_N}
seed = [0.0 for i in range(N+1)]
seed[0] = x_0

# Iteration of T starting at x_0
for i in range(1,N+1):
    seed[i] = T(seed[i-1])

# Number of subintervals
M = 1000

# Absolut frequencies: {n_0, n_1, n_2,..., n_M}
# (freq[0] is not used in calculations)
freq = [0.0 for k in range(M+1)]

# We count the number of iterates n_k which fall into each interval I_k
for i in range(N+1):
    slot = ceil(M*seed[i])
    freq[slot] = freq[slot]+1

# We plot the points (x_k,rho_k)) and connect them by line segments
g = plot(points([[(k-0.5)/M,freq[k]*M/N] for k in range(1,M+1)], color='green',
                pointsize=1))
g += plot(line([[(k-0.5)/M,freq[k]*M/N] for k in range(1,M+1)], color='blue'))
```

```
g.axes_labels(['$\\overline{x}_k$', '$\\rho_k$'])

# We adjust the graph to our convenience (ymax=freq[M]*M/(N+1))
show(g, ymax=20.3179796820203)

g.save('DENSITY-MAP3.pdf')
```

Listing A.9: Code for Figure 3.1.