Lecture Notes
Introduction to Ergodic Theory*

Tiago Pereira
Department of Mathematics
Imperial College London

Our course consists of five introductory lectures on probabilistic aspects of dynamical systems, known as ergodic theory. In simple terms, ergodic theory studies dynamics systems that preserve a probability measure. Let us first discuss some definitions and a motivation for the study.

Dynamical Systems: There are various definitions for a dynamical system, some quite general. Loosely speaking, a dynamical system is a rule for time evolution on a state space. Throughout these lecture we will focus on two models. Most of our time we will be concerned with discrete time dynamical systems, That is, transformations $f : M \to M$ on a metric or topological space.

Heuristically, we think of f as mapping a state $x \in M$ to another state $f(x)$. Then we can follow the iterates or in other words the evolution of the state

$$M \ni x \mapsto f(x) \mapsto f(f(x)) = f^2(x).$$

We say that the sequence $\{f^n(x)\}$ is the trajectory of x. Our goal is to describe the behaviour of the trajectory as $n \to \infty$.

Another model are flows, which continuous time dynamical systems. A flow in M is a family of diffeomorphisms $f^t : M \to M$ with $t \in \mathbb{R}$ of transformation satisfying

$$f^0 = \text{identity and} \quad f^t \circ f^s = f^{t+s} \quad \text{for every } t, s \in \mathbb{R}$$

Flow appear in the context of differential equations with complete flows. Take as f^t the transformation that associate to each point x the value of the solution of the equation at the time t.

*These notes are based on some Lectures given by Marcelo Viana in 2003

Why Invariant Measures? Many natural phenomena are model as dynamical systems that preserve an invariant measure. Historically, the most important example is Hamiltonian systems that describe the evolution of conservative systems in Newtonian mechanics. These systems preserve the Liouville measure. It is very difficult to understand and predict the behavior of orbits of a dynamical systems Surprisingly, the study of invariant measures can give detailed and non-trivial information about the statistical behavior of the system.

Invariant Measure

Measure theory is a mature discipline and lies at the heart of ergodic theory. Instead of providing a review on measure theory, we will discuss the necessary results as we need them. Consider the space M endowed with a $\sigma-$algebra $\mathcal{B}$. Lets also consider a measure $\mu : \mathcal{B} \to \mathbb{R}$. The new concept we want to introduce here is the invariant measure. Assume that our transformation $f : M \to M$ is measurable. The centerpiece of this lecture is the following

Definition 1. *We say that f preserves μ or, equivalently, μ is said to be f-invariant, if*

$$\mu(f^{-1}(B)) = \mu(B)$$

for any $B \in \mathcal{B}$.

At first it may seem strange to have in the definition the pre-image f^{-1} instead of f. There are deep reasons for this definition. And the theory only works with this definition. Lets first discuss a simple reasoning for this choice. First because if the transformation is measurable then for all $B \in \mathcal{B}$ its pre-image is also measurable $f^{-1}(B) \in \mathcal{B}$, hence the definition is well defined. The same is not true if we consider f. It may happen that measurable sets are mapped into non-measurable sets.

Lets now discuss an example which will give other hints of why this definition is appropriate. Consider

$$f : [0,1] \to [0,1], \quad x \mapsto 2x \text{ mod } 1$$

For this transformation we have

Proposition 1. *The Lebesgue measure on $[0,1]$ is invariant under f*

To prove this proposition we need to consider all measurable sets $B \in \mathcal{B}$ and check that definition applies. There is only one problem. There are too many measurable sets! Moreover, we don't have a nice formula for measurable sets, so even writing them explicitly is a problem. But we can check the invariant for nice sets, that is, for intervals $(a,b) \subset [0,1]$. For intervals checking that invariance is quite easy as $f^{-1}(a,b)$ consists of two intervals of length $|b-a|/2$. So, we can easily prove the claim for intervals. Now, notice that had we defined the notion of invariance with respect to f, the Lebesgue measure would not be invariant. The image of the interval (a, b) is an interval two as large.

So, the invariance works for intervals. But the proof for intervals actually will imply that the invariance follows for any measurable sets. To see this consider the case where B is a finite union of disjoint intervals

$$B = B_1 \cup B_2 \cup \cdots \cup B_k$$

Now $\mu(B) = \sum_i \mu(B_i)$ and $f^{-1}(B) = \cup_i f^{-1}(B_i)$. So, we also verify the invariance for finite union of pairwise disjoint sets. Lets introduce the set

$$\mathcal{A} := \{ \text{ all finite union of intervals } \}$$

Now we need the following observations: i) $\mathcal{A}$ is an algebra, and ii) $\mathcal{A}$ generates the σ-algebra $\mathcal{B}$. The next lemma is very useful as it will spare us quite a bit of bureaucratic work.

Lemma 1. *Assume that $\mu(M) < \infty$. If $\mu(B) = \mu(f^{-1}(B))$ for any set B in the generating algebra, then μ is invariant under f.*

Using this Lemma, we can then prove the Proposition with the remarks we have up to now. To this end, we just need to notice that any union of elements of $\mathcal{A}$ can be written as a disjoint union elements of $\mathcal{A}$. For example, given $A_1, A_2 \in \mathcal{A}$ we can write

$$A_1 \cup A_2 = A_1 \cup (A_2 \backslash A_1),$$

now define $B_1 = A_1$ and $B_2 = A_2 \backslash A_1$, so the union is write as a disjoint union.

Invariant measure in terms of functions

From the notion of invariance in terms of measures $\mu(f^{-1}(B)) = \mu(B)$, we can construct a dictionary in terms of functions. First notice that

$$\mu(B) = \int \chi_B d\mu \tag{1}$$

where χ_B is the characteristic function of the set B

$$\chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise} . \end{cases}$$

Moreover, we notice the following

Exercise 1. *Show that*

$$\mu(f^{-1}(B)) = \int \chi_{f^{-1}(B)} d\mu = \int \chi_B \circ f d\mu \tag{2}$$

Therefore, from the definition of invariant together with Eqs. (1) and (2)

$$\int \chi_B d\mu = \int \chi_B \circ f d\mu$$

Now using the linearity of the integral we can immediately extend the previous properties to a simple function

$$\psi = \sum_i c_i \chi_{B_i}.$$

Then by linearity

$$\int \psi d\mu = \int \psi \circ f d\mu.$$

Next, consider a positive measurable function $\psi : M \to [0, +\infty)$. Then, there exists a sequence of simple functions ψ_n converging monotonously to ψ. So by the Lebesgue monotone convergence theorem

$$\int \psi d\mu = \int \lim_n \psi_n d\mu = \lim_n \int \psi_n \circ f d\mu = \int \psi \circ f d\mu.$$

Next, let $\psi : M \to \mathbb{R}$ be any measurable function, then it can be represented as a difference of two positive functions

$$\psi = \psi_+ - \psi_-$$

where $\psi_+ = \max(\psi, 0)$ and $\psi_- = \max(-\psi, 0)$. Then by linearity

$$\int \psi \circ f d\mu = \int \psi d\mu,$$

whenever the integrals make sense. That is, when the functions are integrable. The space of integrable functions is defined to be

$$L^1(M, \mathcal{B}, \mu) = \{\psi : M \to \mathbb{R} | \psi \text{ is measurable and } \int |\psi| d\mu < \infty\}$$

The dictionary for invariance in terms of functions. Bringing together the results we have proved the following

Lemma 2. *The following are equivalent*

(i) μ *is* $f-$*invariant;*

(ii) for each $\psi \in L^1(M, \mathcal{B}, \mu)$*, we have*

$$\int \psi d\mu = \int \psi \circ f d\mu$$

More examples

Measures supported on periodic points: Suppose that x is a periodic point for the map f, that is, there exists $n \geq 1$ such that $x = f^n(x)$. Then consider the probability measure

$$\mu = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{f^j(x)}$$

We claim that this measure is f-invariant. Indeed, using our last Lemma it suffices to check (ii) of Lemma 2.

$$\begin{aligned}
\int \psi \circ f d\mu &= \frac{1}{n} \left(\psi(f(x)) + \cdots + \psi(f^n(x))\right) \\
&= \frac{1}{n} \big(\psi(x) + \cdots + \psi(f^{n-1}(x))\big) \\
&= \int \psi d\mu
\end{aligned}$$

where we used that $f^n(x) = x$.

This has a deep consequence for dynamics. A typical dynamical systems has many period orbits. Our previous example $f(x) = 2x$ mod1 has infinitely many periodic orbits. This means that typical dynamical systems will preserve many invariant measures. Therefore, typically one looks for some restriction on the measure in order to capture interesting behavior. For example, we may only look for invariant measure that are absolutely continuous with respect to the Lebesgue measure. To fix idea lets us discuss some concepts.

Suppose that μ_1 and μ_2 are two measures on $(M, \mathcal{B})$. We say that μ_1 is absolutely continuous with respect to μ_2, and we write $\mu_1 \ll \mu_2$, if

$$\mu_2(B) = 0 \Rightarrow \mu_1(B) = 0.$$

We say that the measures are equivalent if $\mu_1 \ll \mu_2$ and $\mu_2 \ll \mu_1$. That is, these measure have the same zero measure sets.

Another approach is to consider the *natural measure*. Let $\nu(C, x_0, T)$ be the fraction of time that the orbit $\{f^n(x_0)\}_{n=0}^{T}$ spends in the set C and consider the limit

$$\nu(C, x_0) = \lim_{T \to \infty} \nu(C, x_0, T),$$

if it exists. This measure is the "histogram", we will show in the following lectures that depending on f this limit exists and is independent of the point (ν almost surely).

Degree k map: Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be the map $f(x) = kx$, where $k \in \mathbb{N}$. Proof that the Lebesgue measure is f-invariant. (The proof is the same as in the case $k = 2$.) Show that this map has k^n periodic points for period n. Construct the invariant measure for the periodic points.

Gauss Transformation: Consider the map $f : (0, 1] \to [0, 1]$ given map

$$f(x) = \frac{1}{x} - [1/x],$$

where $[1/x]$ is the integer part of 1/x. This map preserves the Gauss measure

$$\mu(B) = \frac{1}{\log 2} \int_B \frac{1}{1+x} dx.$$

Notice that if $x \in (1/(k+1), 1/k)$ for some $k \in N$ then the integer part of $1/x$ equals k so

$$f(x) = 1/x - k$$

Notice that $f(1/k) = 0$ hence $f^2(1/k)$ is not defined (the third iterate is not defined on its pre-image and so on). This means that rigorously f is not a dynamical systems in the sense we defined earlier. But this imposes no problem, because all iterates of f are well defined on the set of irrational numbers. For us it is enough to treat properties that are defined almost everywhere.

Notice that

$$m(E)/2 \le \mu(E) \le m(E)$$

so μ is equivalent to the Lebesgue measure.

There are many ways to prove that μ is f-invariant we will use the following

Exercise 2. *Let $f : U \to U$ be a local C^1 diffeomorphism, and let ρ be a continuous function. Show that f preserves the measure $\mu = \rho m$ if and only if*

$$\sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|det Df(x)|} = \rho(y)$$

Lets use this exercise to prove the invariant. Hence, we have to show that

$$\sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|f'(x)|} = \rho(y) \text{ where } \rho(x) = c/(1+x)$$

Lets start by observing that each y has exactly one pre-image in each interval $(1/(1+k), 1/k]$ given by

$$f(x_k) = \frac{1}{x_k} - k = y \Leftrightarrow x_k = \frac{1}{y+k}$$

Moreover, notice that $f'(x) = -1/x^2$. Hence, using the exercise

$$\sum_{k=1}^{\infty} \frac{cx_k^2}{1+x_k} \Leftrightarrow \sum_{k=1}^{\infty} \frac{1}{(y+k)(y+k+1)} = \frac{c}{1+y}$$

To check this we observe

$$\frac{1}{(y+k)(y+k+1)} = \frac{1}{y+k} - \frac{1}{y+k+1}$$

So the sum can be written as a telescopic sum: all terms, except for the first, will appear twice but with different signs, which concludes the proof.

Rotations: Let $M = \mathbb{R}\backslash\mathbb{Z}$, and consider the rigid rotation of the circle $f : M \to M$ with

$$f_\theta(x) = x + \theta.$$

The Lebesgue measure is $f_\theta-$invariant. To see this, let $\psi : M \to \mathbb{R}$.

$$\int_0^1 (\psi \circ f_\theta)(x)dx = \int_0^1 \psi(x+\theta)dx = \int_0^1 \psi(x)dx$$

Exercise 3. *Prove that if $f : M \to M$ preserves a probability μ, then for any $k \geq 2$ f^k preserves μ. Is the reciprocal true?*

Exercise 4. *Let $f : U \to U$ be a diffeomorphism and $U \subset \mathbb{R}^d$ an open set. Show that the Lebesgue measure m is $f-$invariant if and only if $|detDf| = 1$*

Exercise 5. *Let $f, g : M \to M$ be two transformations. We say that f is conjugated to g if there exists a continuous one-to-one transformation (change of coordinates) h such that*

$$h \circ f = g \circ h.$$

i) Show that $h \circ f^n = g^n \circ h$, for every $n \geq 1$.

Next, consider the tent map

$$g(x) = \begin{cases} 2x & \text{for } x < 1/2 \\ 2 - 2x & \text{for } x \geq 1/2 \end{cases}$$

ii) Show that the Lebesgue measure is $g-$invariant.

The logistic map $f(x) = 4x(1-x)$ and the tent map are conjugated by

$$h(x) = (1 - \cos \pi x)/2.$$

iii) Use this fact to show that the measure $\mu = \varphi m$ is $f-$invariant with

$$\varphi(x) = \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$$

Recurrence

Now we will study the Poincaré recurrence theorem. The theorem says that given any finite $f-$invariant measure $\mu-$almost every point of any measurable set E will return to E infinitely often. This results has profound implies for mechanics, in particular, for statistical physics.

Theorem 1. *Let $f : M \to M$ be a measurable transformation and μ a $f-$invariant measure satisfying $\mu(M) < \infty$. Let $E \subset M$ be a measurable set with $\mu(E) > 0$. Then, $\mu-$almost every point $x \in E$ there exists $n \geq 1$ such that $f^n(x) \in E$. Moreover, there are infinitely many values of n such that $f^n(x)$ belongs to E*

Proof. Let E_0 be the set of points $x \in E$ that never return to E. We wish to show that E_0 has zero measure. First, lets notice that the pre-images $f^{-n}(E_0)$ are disjoint

$$f^{-n}(E_0) \bigcap f^{-m}(E_0) = \emptyset \text{ for all } m \neq n \geq 1$$

Suppose that there are $m > n \geq 1$ such that the $f^{-m}(E_0)$ intersections $f^{-n}(E_0)$. Let x be a point in the intersection. Let $y = f^n(x)$, then clearly

$$y \in E_0 \text{ and } f^{m-n}(y) = f^m(x) \in E_0$$

this means that y returns to E_0 contradicting the definition of E_0. This proves that the pre-images are pairwise disjoint.

Now recall that the measure is invariant $\mu(f^{-n}(E_0)) = \mu(E_0)$ for all n, hence we conclude that

$$\mu\left(\bigcup_{i=1}^{\infty} f^{-n}(E_0)\right) = \sum_{i=1}^{\infty} \mu(f^{-n}(E_0)) = \sum_{i=1}^{\infty} \mu(E_0)$$

But we assumed that the measure is finite, hence the expression in the left side is finite. In the other hand, the in the right side we have infinitely many terms all equal. The only way this sum is finite is that $\mu(E_0) = 0$ as we promised.

Now let F be the set of points $x \in E$ that return to E only a finite number of times. By direct consequence of the definition, every point $x \in F$ has some iterate $f^k(x) \in E_0$. That is,

$$F \subset \bigcup_{k=0}^{\infty} f^{-k}(E_0)$$

Then

$$\mu(F) \leq \mu\left(\bigcup_{k=0}^{\infty} f^{-k}(E_0)\right) \leq \sum_{k=0}^{\infty} \mu\left(f^{-k}(E_0)\right) = \sum_{k=0}^{\infty}(E_0) = 0.$$

Hence, $\mu(F) = 0$. □

Example 1. *Consider $f : \mathbb{R} \to \mathbb{R}$ be the translation by one*

$$f(x) = x + 1.$$

Then the Lebesgue measure is invariant. Notice however, that the measure in this case is not finite. Clearly, there is no recurrent point under f. On the other hand, by the recurrence theorem, f does preserve any finite measure.

Kac Lemma

Assume that the system of many interacting particles (molecules in a room) has a fixed (finite) total energy. Then the dynamics takes place in bounded subsets of the phase space. Roughly speaking, the second law of thermodynamics claims that the system will evolve that the mess increases, that is, the system tries to occupy the maximum number of states. The recurrence theorem shows that the system will eventually return arbitrarily close to its initial state. In statistical mechanics this is the so-called recurrence paradox in statistical mechanics. Kac Lemma gives the average return time of almost every point to the set.

Let again $f : M \to M$ be a measurable transformation and μ a $f-$invariant finite measure. Let $E \subset M$ be any measurable set with $\mu(E)$. Consider the function called *first return time* $\rho_E : E \to \mathbb{N} \cup \{\infty\}$ defined by

$$\rho_E(x) = \min\{n \geq 1 : f^n(x) \in E\}$$

whenever the set in the right-hand side is non-empty, otherwise $\rho_E(x) = \infty$

Now we will show that this function ρ_E is integrable. To this end, we introduce

$$E_0 = \{x \in E : f^n(x) \notin E \text{ for all } n \geq 1\} \quad \text{and} \tag{3}$$

$$E_0^* = \{x \in M : f^n(x) \notin E \text{ for all } n \geq 0\} \tag{4}$$

that is, E_0 is the set of points of E that never return to E, and E_0^* is the set of points of M that never enter in E. Note that $\mu(E_0) = 0$, by the Poincaré recurrence theorem.

Theorem 2 (Kac). *Let $f : M \to M$ be a measurable transformation, μ a finite $f-$invariant measure and $E \subset M$ a subset of positive measure. Then, the function ρ_E is integrable and*

$$\int_E \rho_E d\mu = \mu(M) - \mu(E_0^*).$$

Proof. For each $n \geq 1$ let us define

$$\begin{aligned} E_n &= \{x \in E : f(x) \notin E, \dots, f^{n-1}(x) \notin E \text{ but } f^n(x) \in E\} \quad \text{and} \quad (5) \\ E_n^* &= \{x \in M : x \notin E \dots, f^{n-1}(x) \notin E \text{ but } f^n(x) \in E\} \quad (6) \end{aligned}$$

This means that E_n is the set of points of E that return to E for the first time precisely at the moment n,

$$E_n = \{x \in E : \rho_E(x) = n\},$$

and E_n^* is the set of points that is not in E and enter in E for the first time at the moment n. These sets are measurable and so the function ρ_E is measurable. Moreover, for $n \geq 1$ the sets E_n and E_n^* are pairwise disjoint and the union is the whole space M. Hence,

$$\mu(M) = \sum_{n=0}^{\infty}[\mu(E_n) + \mu(E_n^*)] = \mu(E_0^*) + \sum_{n=1}^{\infty}[\mu(E_n) + \mu(E_n^*)] \tag{7}$$

Notice that

$$f^{-1}(E_n^*) = E_{n+1}^* \bigcup E_{n+1} \text{ for all } n. \tag{8}$$

In fact, $f(y) \in E_n^*$ means that the first iterate of $f(y)$ to land in E is $f^n(f(y)) = f^{n+1}(y)$, but this happens if and only if $y \in E_{n+1}^*$ or $y \in E_{n+1}$. This proves (8). Hence, using that the measure is invariant

$$\mu(E_n^*) = \mu(f^{-1}(E_n^*)) = \mu(E_{n+1}^*) + \mu(E_{n+1}) \text{ for all } n.$$

Applying this relation multiple times, we obtain

$$\mu(E_n^*) = \mu(E_m^*) + \sum_{i=n+1}^{m} \mu(E_i) \text{ for all } m > n. \tag{9}$$

Equation (7) implies that $\mu(E_m^*) \to 0$ as $m \to \infty$. Therefore, taking the limit $m \to \infty$ in (9) we obtain

$$\mu(E_n^*) = \sum_{i=n+1}^{\infty} \mu(E_i) \tag{10}$$

Replacing (10) in (7) we obtain

$$\mu(M) - \mu(E_0^*) = \sum_{n=1}^{\infty}\left(\sum_{i=n}^{\infty}\mu(E_i)\right) = \sum_{n=1}^{\infty} n\mu(E_n) = \int_E \rho_E d\mu$$

concluding the proof. □

If the system (f, μ) is ergodic (we will study this property later) the set E_0^* has zero measure. Moreover, if measures is actually a probability then $\mu(M) = 1$. Hence the conclusion of Kac Lemma

$$\frac{1}{\mu(E)} \int_E \rho_E d\mu = \frac{1}{\mu(E)}$$

for every measurable set E. In the left-hand side we have the mean return time to E. Hence, the equality means that the mean return time is inversely proportional to the measure of E.

Topological Flavours

Assume that M is a topological space endowed with the Borel $\sigma-$algebra.

Definition 2. *We say that a point $x \in M$ is recurrent for the transformation $f : M \to M$ if there is a sequence $n_j \to \infty$ such that $f^{n_j}(x) \to x$*

Our next goal is to prove the following

Theorem 3. *Let $f : M \to M$ be a continuous transformation in a compact metric space M. Then, there exists some point $x \in M$ recurrent for f.*

Proof. Consider the family $\mathcal{I}$ of all closed non-empty sets $X \subset M$ that are invariant $f(X) \subset X$. This family is non-empty since $M \in \mathcal{I}$. We say that an element $X \in \mathcal{I}$ is minimal for the inclusion relation if and only if the orbit of the point $x \in X$ is dense em X.

Indeed, since X is closed and invariant then X contains the closure of the orbits. Hence, X is minimal if it coincides with any of the orbits closures. Likewise, if X coincides with its closure of the orbit of any of its points then it coincides with any close invariant subset, that is, X is minimal. This proves our claim. In particular, any point x in a minimal set is recurrent. Hence, to prove to theorem it suffices to show that there exists a minimal set.

Now we claim that a ordered set $\{X_\alpha\} \subset \mathcal{I}$ admits a lower bound. Indeed, consider $X = \bigcap X_\alpha$. Notice that X is non-empty since X_α are compact and the family is ordered. Clearly X is closed and invariant under f and it is also a *lower bound* for the set $\{X_\alpha\}$. This proves our claim. Now we apply Zorn Lemma to conclude that $\mathcal{I}$ really contains minimal elements. $\square$

Exercise 6 (Numerics)**.** *Estimate the mean return time to the set $E = [0.2, 0.3]$ for $f(x) = 10x$ mod 1.*

Exercise 7 (Numerics)**.** *Consider the transformation $f(x) = 3x$ mod1. Consider the set $A = [0.1, 0.11]$, and a point $x \in A$. What is the typical distribution of first return times?*

Exercise 8. *Consider the map* $f : [0,1] \to [0,1]$ *given by* $x \mapsto 10x$ *mod1. Show that almost every number* $x \in [0,1]$ *whose decimal expansion starts with the digit* 7 *will have infinitely many digits equal to* 7.

Exercise 9. *Let* f *be the Gauss map. Show that a number* $x \in (0,1)$ *is rational if and only if, there is* $n \geq 1$ *such that* $f^n(x) = 0$.

Ergodic Theorems

In the past lecture we introduce the concept of natural measure which concerns the time an orbit spends in a set. In many cases, the time the orbit spends in the set equals to measure of the set. That is, the orbit spends as much time as the importance the invariant measure attributes to that set. This idea is called *ergodicity*.

Statement and Discussion

Lets recall the definition of the time an orbit spend in a set. Let $x \in M$ and $E \subset M$ be a measurable set, and consider

$$\tau_n(E,x) = \frac{1}{n}\#\{j = 0, 1, \cdots, n-1 : f^j(x) \in E\},$$

and notice that

$$\tau_n(E,x) = \frac{1}{n}\sum_{j=0}^{n-1} \chi_E(f^j(x)),$$

where again χ_E is the characteristic function of the set E. Taking $n \to \infty$, if the limit exists we introduce E

$$\tau(E,x) = \lim_n \tau_n(E,x),$$

the time of orbit spends in the set E. This in principle depends on the starting point x. However, along the orbit the function $\tau(E,x)$ is constant. Indeed, first lets us notice that

$$\tau(E,x) = \tau(E,f(x))$$

Hence, if the limit exist for $x \in M$, it will also exist for its iterates. To see this notice that

$$\begin{aligned}
\tau(E,f(x)) &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \chi_E(f(x)) \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \chi_E(f(x)) + \frac{1}{n}[\chi_E(f(x)) - \chi_E(x)] \\
&= \tau(E,x) + \lim_{n\to\infty} \frac{1}{n}[\chi_E(f(x)) - \chi_E(x)]
\end{aligned}$$

It is not obvious that this limit exist as the following example shows

Exercise 10. *Consider the transformation $f : M \to M$, with $f(x) = 10$ mod1. And take*

$$x = 0.3355333355555555333333333333333335\ldots,$$

that is, x consists of blocks of 3's and 5's with the block size twice as large as the previous (except the second). Consider $E = [0.3, 0.4]$. Show that

$$\tau_2(E,x) = 1, \quad \tau_8(E,x) = \frac{3}{4}, \cdots \quad \tau_{2^{2k-1}}(E,x) \to \frac{2}{3}$$

while

$$\tau_4(E,x) = \frac{1}{2}, \quad \tau_{16}(E,x) = \frac{3}{8}, \cdots \quad \tau_{2^{2k}}(E,x) \to \frac{1}{3}$$

Hence, the time the orbit of x spends in E does not exist.

However, the ergodic theorem states that the above case is atypical, and the limit exits for almost every point.

Theorem 4 (Birkhoff)**.** *Let $f : M \to M$ be a measurable transformation and μ a $f-$invariant probability. Given any measurable set $E \subset M$ the mean visit time*

$$\tau(E,x) = \lim_n \frac{1}{n} \#\{j = 0, 1, \cdots, n-1 : f^j(x) \in E\}$$

exists for μ almost every point $x \in M$. Moreover,

$$\int \tau(E,x) d\mu(x) = \mu(E).$$

This is a particular case of the ergodic theorem, but bears precisely the idea we had just discussed. Soon we will show that $\tau(E,x)$ is constant almost everywhere point with respect to μ. And hence,

$$\tau(E) = \mu(E) \quad \text{for } \mu - \text{almost every point.}$$

A more general claim is the following

Theorem 5. *Let $f : M \to M$ be a measurable transformation and μ a $f-$invariant probability. Then given a integrable function $\varphi : M \to \mathbb{R}$ the limit*

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x))$$

exists for μ almost every point $x \in M$. Moreover, the function $\tilde{\varphi}$ is integrable and

$$\int \tilde{\varphi}(x) d\mu(x) = \int \varphi(x) d\mu(x)$$

The proof of the last statement can be obtained from the first. So, we will focus on the first statement

Proof. Theorem 4. Let $E \subset M$ be a measurable set. For each $x \in M$ we introduce

$$\bar{\tau}(E,x) = \limsup \tau_n(E,x) \quad \text{and} \quad \underline{\tau}(E,x) = \liminf \tau_n(E,x).$$

The main idea is the show that

$$\bar{\tau}(E,x) = \underline{\tau}(E,x) \quad \text{for almost every } x \in M,$$

this of course implies that the limit exists. The strategy is the following: By definition of $\limsup$ and $\liminf$ we have $\underline{\tau}(E,x)$ is always smaller than $\bar{\tau}(E,x)$, so if we show that

$$\int \bar{\tau}(E,x)d\mu(x) \leq \mu(E) \leq \int \underline{\tau}(E,x)d\mu(x)$$

the statement will follow.

We start with the first inequality. Let $\varepsilon > 0$ be given, then we definition of $\limsup$ for any $x \in M$ there is a sequence of n_k such that

$$\frac{1}{n_k}\#\{j \in \{0,\ldots,n_k-1\}\} \geq \bar{\tau}(E,x) - \varepsilon$$

Let $t(x)$ be the smallest integer with this property.

We first consider a particular case where

$$x \mapsto t(x) \text{ is bounded} \quad \exists T \in \mathbb{N} \text{ such that} \quad t(x) < T$$

Given $x \in M$ and $n > 0$ we define a sequence $x_0, x_1, \ldots, x_s$ of points of M and a sequence $t_0, t_1, \ldots, t_s$ of natural numbers. The sequence is defined iteratively as follows

1. First, take $x_0 = x$.
2. Assume that x_i was defined, we take $t_i = t(x_i)$ and $x_{i+1} = f^{t_i}(x_i)$.
3. We finish the process when we find x_s such that $t_0 + \cdots + t_s > n$.

Hence, every x_i is an iterate of x: $x_i = f^{t_0+\cdots+t_{i-1}}(x)$. Notice that because our definition of t_i, from the first t_i iterates of x_i at least

$$t_i(\bar{\tau}(E,x_i) - \varepsilon) = t_i(\bar{\tau}(E,x) - \varepsilon)$$

are in the set E. This observation holds for every $i = 0,\ldots,s-1$, so at least

$$(t_0 + \cdots t_{s-1})(\bar{\tau}(E,x) - \varepsilon)$$

are in E. Moreover, using the last rule of the iterative process that defines t_i we have

$$t_0 + \cdots t_{s-1} \geq n - t_s \geq n - T.$$

So we just showed that at least $(n-T)(\bar{\tau}(E,x) - \varepsilon)$ out of the first n iterates of x are in E. That is,

$$\sum_{j=0}^{n-1} \chi_E(f^j(x)) \geq (n-T)(\bar{\tau}(E,x) - \varepsilon)$$

for all $x \in M$ and $n \geq 1$. Integrating we obtain

$$\sum_{j=0}^{n-1} \int (\chi_E \circ f^j)(x) d\mu(x) \geq (n-T) \int (\bar{\tau}(E,x) - \varepsilon) d\mu(x)$$

Next, we use that the measure is invariant, so every term in the sum in the left-hand side equals $\mu(E)$. So,

$$n\mu(E) \geq (n-T) \int \bar{\tau}(E,x) d\mu(x) - (n-T)\varepsilon.$$

Now dividing by n and taking the limit $n \to \infty$ we obtain

$$\mu(E) \geq \int \bar{\tau}(E,x) d\mu(x) - \varepsilon.$$

Notice that since ε is arbitrary we have just obtained the first inequality.

The second inequality can be obtained from this one by noticing that

$$\mu(E) = 1 - \mu(E^c) \quad \text{and} \quad \underline{\tau}(E,x) = 1 - \bar{\tau}(E^c, x)$$

where E^c is the complement of the set E.

The general Case: Given $\varepsilon > 0$ we fix $T > 1$ such that the measure of the set

$$B := \{y \in M : t(y) > T\}$$

is smaller than ε.

Exercise 11. *Why the above statement on $\mu(B) \leq \varepsilon$ is true?*

Then, we update the rule 2 by

2a If $t(x_i) < T$, we take $t_i = t(x_i)$ and $x_{i+1} = f^{t_i}(x_i)$,

2b If $t(x_i) > T$, we take $t_i = 1$ and $x_{i+1} = f(x_i)$.

Then we obtain

$$\sum_{j=0}^{t_i-1} \chi_E(f^j(x)) \geq t_i(\bar{\tau}(E,x) - \varepsilon)$$

This implies that

$$\sum_{j=0}^{t_i-1} \chi_E(f^j(x)) \geq t_i(\bar{\tau}(E,x) - \varepsilon) - \sum_{j=0}^{t_i-1} \chi_B(f^j(x))$$

Now this last inequality is valid for all values of i. This yields

$$\sum_{j=0}^{t_i-1} \chi_E(f^j(x)) \geq (n-T)(\bar{\tau}(E,x) - \varepsilon) - \sum_{j=0}^{t_i-1} \chi_B(f^j(x))$$

where we have used the in the complement of B the function t_i is bounded. Integrating we obtain

$$n\mu(E) \geq (n-T) \int \bar{\tau}(E,x) d\mu(x) - (n-T)\varepsilon - n\mu(B).$$

Dividing by n and taking the limit (also recalling that $\mu(B) \leq \varepsilon$) we obtain

$$\mu(E) \geq \int \bar{\tau}(E,x) d\mu(x) - 2\varepsilon.$$

since ε is arbitrary the claim follows. $\square$

Ergodicity

We say that the system (f, μ) is ergodic if given any measurable set E if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \int \varphi d\mu, \tag{11}$$

for every $\varphi \in L^1(M, \mathcal{B}, \mu)$. That is, if the temporal averages coincides almost everywhere with the spacial averages. In particular, if

$$\tau(E, x) = \mu(E) \text{ for } \mu \text{ almost every point } x \in M$$

We want to show that this implies that the system is dynamically indivisible, that is, any invariant has either full measure or zero measure. We say that a set $A \subset M$ is invariant if $f^{-1}(A) = A$.

Definition 3. *A measurable function $g : M \to \mathbb{R}$ is called invariant with respect to f if*

$$g \circ f = g \textit{ almost everywhere}$$

The next theorem rephrases the ergodic theorem in terms of the invariant sets and invariant functions

Proposition 2. *Let $f : M \to M$ be a measurable transformation and μ be a $f-$invariant probability. The following statements are equivalent.*

1. *The system (f, μ) is ergodic.*
2. *For every invariant set $A \subset M$ we have that either $\mu(A) = 1$ or $\mu(A) = 0$.*
3. *Every invariant function is constant almost everywhere, that is, in a set of full measure.*

Proof. First we show that (1) implies (2): Consider $\varphi = \chi_A$ from (11) we obtain

$$\tilde{\varphi} = \int \varphi d\mu = \mu(A)$$

for almost every $x \in M$. Since A is invariant $x \in A$ if and only if $f(x) \in A$. Hence,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(f^j(x)) = \chi_A(x) = \mu(A).$$

Therefore, either $\mu(A) = 1$ if $x \in A$ or $\mu(A) = 0$ if $\chi_A(x) = 0$.

Next, we show that (2) implies (3): Let ψ be any invariant function $\psi \circ f = \psi$, and define

$$B_t = \{x \in M : \psi(x) > t\} \quad \text{for all} \quad t \in \mathbb{R}$$

Now notice that all sets B_t are invariant

$$\begin{aligned} f^{-1}(B_t) &= \{x \in M : f(x) \in B_t\} \\ &= \{x \in M : \psi(f(x)) > t\} \\ &= \{x \in M : \psi(x) > t\} = B_t \end{aligned}$$

where in the last line we used the invariant of ψ. Hence, (2) shows that either $\mu(B_t) = 1$ or $\mu(B_t) = 0$. If ψ is not constant almost everywhere then we can find t_0 such $\mu(B_{t_0}) \in (0, 1)$ which contradicts (2). Therefore, the function ψ must to constant almost everywhere.

(3) implies (1): Let ψ be any integrable function. Notice that the time average $\tilde{\psi}$ is an invariant function. Hypothesis (3) implies that this function is constant almost everywhere, so applying the ergodic theorem

$$\tilde{\psi}(x) = \int \tilde{\psi} d\mu = \int \psi d\mu \quad \text{almost everywhere.}$$

$\square$

Exercise 12. *Consider the rotation by a angle α*

$$R_\alpha : S^1 \to S^1, \qquad R_\alpha(z) = e^{\alpha i} z$$

Show that

1. *The Lebesgue measure m is invariant for every value of α.*

   Hint: Show that $\int \psi \circ R_\alpha dm = \int \psi dm$.

2. *R_α is rational (every point is periodic) if and only if $e^{\alpha i}$ is a root of the unit.*

3. *If R_α is rational then R_α is not ergodic for m.*

4. *If R_α is irrational then R_α is ergodic for m.*

   Hint: Show that invariant functions are constant. You can write any measurable function as $\varphi = \sum c_k z^k$. Conclude that if α is irrational and φ invariant then $\varphi(z) = c_0$